ASIA-PACIFIC APPLIED
ECONOMICS ASSOCIATION

Email:
contact@a-paea.org
contact.apaea@gmail.com

Asia-Pacific Applied Economics Association Conference Proceedings

ISSN 2208-6765

4th Applied Financial Modelling Conference

Deakin University, Melbourne, Australia, 1-2 February 2018

Editorial Note

Editor-in-Chief: ***Paresh Kumar Narayan, Alfred Deakin Professor, Deakin University***

Series Editor: ***Seema Narayan, Associate Professor, RMIT University***

Asia-Pacific Applied Economics Association (APAEA) conference proceedings publishes high quality papers selected out of papers presented at APAEA's conferences. Each APAEA conference is affiliated with either SCOPUS indexed or social science citation indexed journals. The APAEA conferences encourage presentation of papers broadly in the fields of economics and finance that make use of advanced econometric techniques and new datasets to test economic models and hypotheses related to finance and economics. Common topics of importance to conference participants are those that test economic models and hypothesis using new datasets and/or methods, forecasting financial time-series data, financial market performance, macroeconomic stability issues, panel data models, energy finance, economic growth and productivity, and econometrics methods including financial econometrics. These are the types of papers that are ultimately published in the *APAEA conference proceedings*.

The *APAEA conference proceedings* follow a single blind review procedure. All papers submitted to the conference go through a single blind review procedure such that those papers that are ultimately published in the *Proceedings* have undergone a review process. The conference and, therefore, the *Proceedings* rejection rate stands at 50%. The low quality papers, which in the view of the conference scientific committee and the Editor of the *Proceedings* have low chances of advancing knowledge and contributing to the literature are desk rejected without sending the papers for a formal review.

All APAEA publications, including the *Proceedings*, follow the publication ethics and malpractice statements developed for editors and authors by Wagner & Kleinert (2011). See https://publicationethics.org/node/11184 for details and full bibliographical information on Wagner & Kleinert (2011).

Table of Contents

Deep Neural Networks Understand Investors Better

Nader Mahmoudi, Department of Finance, Newcastle Business School, the University of Newcastle

Paul Docherty. Department of Banking and Finance, Monash Business School, Monash University

Pablo Moscato, Department of Computer Science and Software Engineering, School of Electrical Engineering and Computing, The University of Newcastle

ABSTRACT

The presence of noise traders in financial markets may result in investor sentiment affecting stock returns. However, studies that seek to examine the impact of investor sentiment have been affected by inaccurate sentiment measurement and the use of inappropriate data. This study presents a superior measure of investor sentiment by applying advanced sentiment classification techniques to data from StockTwits, an investment-focused social media platform. The inclusion of emojis is also shown to result in significantly better sentiment classification in traditional algorithms. Moreover, deep neural networks with domain-specific word embeddings outperform the traditional approaches for the classification of investor sentiment. The approach to sentiment classification outlined in this paper can be applied in future empirical tests that examine the impact of investor sentiment on financial markets.

Keywords: Investor Sentiment, Domain-specific, Emojis, Deep Neural Network (DNN), Word Embeddings, StockTwits

1. Introduction

Although the neoclassical finance paradigm of efficient markets provides the proposition that stock returns are unpredictable (Fama, 1970), a large body of contradictory empirical evidence has brought this theory into question (Baker & Wurgler, 2000; Cochrane, 2000). In light of this evidence, *behavioral finance* has been proposed as an alternative theoretical paradigm to explain stock returns. The key implication of behavioral finance is that the emotions and moods of investors play an important role in financial decisions (Nofsinger, 2005). Moreover, the presence of irrationality and the emotive basis of decisions made by noise-traders, who comprise a relatively large proportion of stock market participants (Black, 1986), has resulted in investor sentiment being considered to influence investor decision-making, and hence stock returns. This new paradigm of stock market behavior has resulted in the need to develop accurate measures of investor sentiment (Chan & Chong, 2017). This study examines the incorporation of non-text features, such as emojis, the development of domain-specific word embeddings and the use of deep learning to classify investor sentiment.

Despite a large number of studies proposing a relationship between investor sentiment and stock market returns, there is limited empirical evidence supporting this proposition. Proponents of behavioral finance argue that this lack of empirical evidence can be attributed to problems with the measurement of investor sentiment in existing studies of financial markets. These problems include: the absence of an accurate approach for measuring investor sentiment (Bollen, Mao, & Zeng, 2011; Oh & Sheng, 2011); use of datasets from platforms that do not accurately represent investors (Bollen et al., 2011; Ranco, Aleksovski, Caldarelli, Grčar, & Mozetič, 2015); and the use of short sample periods (Bollen et al., 2011; Oh & Sheng, 2011). Our study seeks to provide a resolution for these problems by applying sentiment classificiation (SC), across a dataset that can appropriately reflect investors' beliefs.

The application of sentiment analysis is widely employed across the social sciences. For example, in the marketing discipline measures of customer sentiment are applied as a proxy for preferences for a particular product or service (Pang & Lee, 2008). The evolution in general-purpose SC techniques can be divided into two separate paradigms: traditional machine learning and deep learning. Traditional methods depend on feature extraction, the process of transforming the raw texts into features from which classification algorithms can learn. Various manually engineered and complex feature types have been proposed to capture this information, including *n*-grams, parts of speech, negation, and emojis/emoticons (Aggarwal & Zhai, 2012; Pang & Lee, 2008). Moreover, different domain-general and domain-specific sentiment lexicon resources have been constructed for sentiment classification where polarity indices such as positivity, negativity, or objectivity are assigned to every word (Baccianella, Esuli, & Sebastiani, 2010; Deng, Sinha, & Zhao, 2017).

Deep neural networks (DNNs) reach state-of-the-art performance in most of the NLP problems without any need for enhanced pre-engineered features (Kalchbrenner, Grefenstette, & Blunsom, 2014). These models can capture deep local features by convolution kernels or capture long-distance dependencies by memory units over the input sentences (X. Wang, Jiang, & Luo, 2016). The success of word embedding construction algorithms, which take a large corpus as input and produce a high-dimensional vector space (Mikolov, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014), has led to an increase in the implementation of DNNs on NLP problems. The word embeddings play a strongly significant role in solving the NLP problems as they succeed in representing semantic and syntactic relationships between words in a context. Meanwhile, various intrinsic or extrinsic methods have been proposed in order to evaluate the word embeddings: similarity (relatedness), analogy, POS tagging, and sentiment classification (Schnabel, Labutov, Mimno, & Joachims, 2015). Despite the above advancements in NLP, extant studies that classify investor sentiment have only applied simple structures with basic feature types and shallow classification techniques. To date, DNNs have not been implemented for the classification of investor sentiment.

The collection of investor sentiment data from Internet-based microblogs overcomes issues that have been identified from the use of questionnaires, such as errors due to impaired questionnaire design (Brace, 2008) and careless or untruthful participant responses (Singer, 2002). While previous studies have sought to measure investor sentiment using other microblogs, such as Twitter (Bollen et al., 2011; Ranco et al., 2015), StockTwits should provide a more relevant source of information to measure investor sentiment, given its focus on stock-related information (Oliveira, Cortez, & Areal, 2013). However, it is difficult to use basic classification approaches to classify investor sentiment in StockTwits given the distinct properties of the texts in this microblog. First, the terminology in StockTwits employs everyday English words but in ways that carry specific financial and investment meanings (Oliveira, Cortez, & Areal, 2016). Second, StockTwits is also characterized by the use of non-text characters to convey feelings and beliefs, such as emojis and emoticons (Novak, Smailović, Sluban, & Mozetič, 2015). Moreover, the posts made by users (investors) comprise a more prominent use of negation, sarcasm, and domain-specific analogies that are very hard to extract by hand-crafted features (Shirani-Mehr, 2014).

The contributions of this study are three-fold. First, the inclusion of non-text features, emojis, is shown to improve investor sentiment classification. Second, this study evaluates GloVe and Word2Vec to analyze their ability in capturing domain-specific word similarities compared with domain-general word embeddings. This is carried out through an entirely novel domain-specific evaluation method called the FinSim Index, which represents the similarity between two words in the finance context. Finally, different types of deep neural networks are constructed, which are able to detect abstract-level feature types such as sarcasm and achieve the highest accuracy in investor sentiment classification.

This paper continues with a literature review in section 2 and a discussion of the methodology in section 3. The results are reported in section 4, including a discussion of emojis, word embeddings, and deep neural networks, while section 5 gives the conclusion, describes the limitations, and foreshadows future work.

2. Related Studies

News websites, social networks, and weblogs provide modern investors with the opportunity to exchange information and opinions about financial markets with high frequency (Sun, Belatreche, Coleman, McGinnity, & Li, 2014). Since the advent of the Internet, various techniques have been utilized by researchers in order to use this information to extract measures of investor sentiment. These methods can be classified into two main groups: lexicon-based techniques and machine learning techniques. The use of these lexicon-based approaches in the financial domain was initiated by Tetlock, Saar-Tsechansky, & Macskassy (2008), who constructed a daily measure of the sentiment using daily content from a popular Wall Street Journal (WSJ) column. This measure is called the *pessimism factor* since it is highly related to words with negative polarity.

A key limitation with the pessimism factor is that it is constructed by categorizing words according to the General Inquirer's Harvard IV-4 dictionary. Such dictionaries may be limited in their ability to assign sentiment to words used in the financial context, due to the use of domain-specific language. In order to overcome this drawback, Dougal, Engelberg, García, & Parsons (2012) and García (2013) have constructed an alternative investor sentiment measure by using a domain-specific dictionary that is developed from a large sample of 10k financial reports (Loughran & McDonald, 2011). Written language has recently undergone a rapid transformation through the Internet, which has resulted in a range of new ways to express a specific idea. Emojis and emoticons, Internet slang, acronyms, and sarcasm are examples of the entities and linguistic structures that have become pervasively common among people, especially the users of social networks. Lexicon-based sentiment classification approaches are not able to capture these new features.

The vast amount of investment sentiment-related data that is publicly available on the Internet has resulted in a tremendous growth in the use of machine learning-based approaches for investor sentiment classification. Antweiler & Frank (2004) applied Naïve Bayes and Support Vector Machines for classifying texts from stock message board postings on Yahoo Finance, using uni-grams that have the highest average mutual information with class labels to categorize sentiment. The same methodology was adopted by Sprenger, Tumasjan, Sandner, & Welpe (2014) to classify investor sentiment using data from Twitter. Subsequent analyses of Twitter posts have classified investor sentiment using a training dataset that was manually labeled by ten finance experts (Ranco et al., 2015) and through the application of the Naive Bayes classifying on a balanced dataset that took negation and emoticons into account (T. Li, van Dalen, & van Rees, 2017).

Researchers have recently directed their attention to the extraction of investor sentiment from StockTwits (Oliveira et al., 2013). For the first time on texts from StockTwits, (Oh & Sheng, 2011) utilized the bag of words (BoW) feature transformation to train the decision tree classifier. Various machine learning algorithms have also been adapted to measure investor sentiment via StockTwits such as Naïve Bayes, Decision Trees, and Bayesian Networks (Al Nasseri, Tucker, & de Cesare, 2014, 2015; See-To & Yang, 2017). To boost these approaches, they have been combined with different feature selection and extraction techniques including information gain criteria and a TF-IDF weighting scheme with simple uni-grams. (T. Wang et al., 2017) identify that superior performance in StockTwits sentiment classification is achieved by using uni-grams as the features and Support Vector Machine as the classifier.

Deep learning is one of the popular machine learning techniques that has commanded attention in various complex artificial intelligence problems including computer vision (Krizhevsky, Sutskever, & Hinton, 2017), speech recognition (Hinton et al., 2012), and machine translation (Luong, Kayser, & Manning, 2015). With different architecture, it has been successfully adapted for natural language processing problems, specifically sentiment analysis. For the first time, Kalchbrenner et al. (2014) implemented Convolutional Neural Network (CNN) on various sentence modeling problems and classify sentiment across a Twitter dataset, resulting in a 25% reduction in error compared with the state-of-the-art traditional classification systems. Furthermore, different versions of Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho, van Merrienboer, Bahdanau, & Bengio, 2014), have also been tested on various NLP tasks including sentiment classification (Yin, Kann, Yu, & Schütze, 2017). These models can capture long-term semantic and syntactic dependency of the texts, whereas CNNs focus on the local features through the convolutional and pooling layers embedded inside them. Wang et al. (2016) have introduced joint CNN and RNN architecture to combine their advantageous characteristics of simultaneously extracting local and long-term features respectively.

Before feeding the texts to a DNN for any NLP task, the words are transformed into high-dimensional embedding vectors that capture semantic and syntactic similarities between words. The embedding vectors are usually imported from pre-trained word embeddings that are optimized over a large unlabeled corpus rather than randomly initialized ones, since the former can represent the semantic and syntactic relationships between the words better. Two widely used algorithms, Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) and GloVe (Pennington et al., 2014), take a large corpus as an input and produce a high-dimensional vector space working as unsupervised learning algorithms. GloVe examines the co-occurrence matrix of the words in constructing the word embeddings, whereas Word2Vec trains a simple neural network with one hidden layer. The word embeddings trained by these algorithms on massive corpora have been commonly used for various NLP tasks as they tend to lead DNNs in solving the semantic and syntactic sparsity (X. Wang et al., 2016).

In an intuitive study, Kim (2014) takes advantage of pre-trained word embeddings in training a simple CNN model and shows that this model outperforms the model with random embeddings initialization. It is also shown that CNN with trainable word embeddings is capable of fine-tuning these embeddings based on the problem at hand. Moreover, the results show that combined CNN-RNN models outperform both RNN and CNN types of models, especially when the pre-trained word embeddings are fed into models (Wang et al., 2016). However, these domain-general word embeddings are not able to perfectly capture domain-specific similarities, especially in StockTwits where investors have constructed their own language. In order to overcome this domain-specificity, Li & Shah (2017) have trained domain-specific word embedding over a large dataset from StockTwits with the aim of building a finance sentiment lexicon. It is shown that the domain-specific word embeddings result in significantly better sentiment lexicons than the domain-general word2vec embeddings.

3. Methodology

In this section, we will discuss traditional classification algorithms and deep neural networks in detail.

3.1. The Traditional Machine Learning Paradigm

Traditional classification methods generally consist of two main parts: feature engineering/extraction and classification algorithm development. These sub-processes, proposed in this research, will be reviewed briefly in the following sections.

3.1.1. Feature Extraction

Various feature types have been examined in traditional sentiment classification problems (Pang & Lee, 2008) and particularly in the problem of investor sentiment, some of which are as follows:

- ***N-grams:*** In this research, we have tested the presence of uni-grams, bi-grams, and tri-grams to capture some of the complex linguistic structures such as negation (e.g. “not good”) and phrases (e.g. “very happ”). During the construction of *n*-grams features, various values have been examined for minimum document frequency and maximum document frequency in order to eliminate the effect of unusual terms (e.g. misspelt words) and common terms (e.g. stop-words).
- ***Negation:*** In order to empower the feature extraction mechanism, we have followed (Pang, Lee, & Vaithyanathan, 2002) and tagged the words that come after the negation words, words such as “not”, “no”, “never”, “nothing”, “nowhere”, and “none” until clause-level punctuation.

- ***Emojis:*** Emojis play a decisive role in determining the sentiment polarity of informal and short texts in social media, weblogs, or comments. The language used by people in StockTwits differs significantly from other social networks. For example, the word "red" carries a pessimistic meaning in StockTwits while it would be interpreted simply as a color in other social networks. Emojis also have different usage patterns in StockTwits. 🚀 (rocket), 💰 (money bag), 🐻 (bear face), 💩 (pile of poo), 🐂 (ox), 📈 (chart increasing), and 📉 (chart decreasing) are some of the emojis commonly used by investors on the StockTwits platform in order to express their feelings and ideas. For the first time in the field of investor sentiment classification, we explore the effect of emojis on financial text labeling.

3.1.2. Classification Algorithms

Three relatively popular and high-performing classification algorithms, including Naïve Bayes (NB), Maximum Entropy (MaxEnt), and Support Vector Machine (SVM), will be implemented to identify the most successful one. The best traditional algorithm chosen here will play a baseline role for the rest of analysis undertaken in this study.

Naïve Bayes (NB) (Manning, Raghavan, & Schütze, 2008) is one of the most frequently used algorithms in text categorization and information retrieval problems. NB is a simple probabilistic classifier which is based on Bayesian Theorem and assumes independence among features of the observations. Starting from Bayesian theorem for a document from class $C_k, \forall\, k = 1, 2, \dots, K$ and with feature vector $(x_1, x_2, \dots, x_n)$, we have:

$$\hat{y} = \underset{k \in 1,2,\dots,K}{argmax}\, P(C_k) * \prod_{i=1}^{n} P(x_i|C_k) \tag{1}$$

where $\hat{y} = C_k$ for $k = 1, 2, \dots, K$.

Maximum Entropy (MaxEnt) (Berger, Pietra, & Pietra, 1996) estimates the $P(C_k|x_1, x_2, \dots, x_n)$ based on the following equation:

$$P_{ME}(C_k|x_1, x_2, \dots, x_n) = \frac{1}{Z(x_1, x_2, \dots, x_n)} exp\left(\sum_{i=1}^{n} \lambda_{x_i, C_k} F_{x_i, C_k}(x_1, x_2, \dots, x_n, C_k)\right) \tag{2}$$

where $Z(x_1, x_2, \dots, x_n)$ is normalization function, λ_{x_i, C_k} is feature-weight parameters, and F_{x_i, C_k} is feature/class function for feature x_i and class C_k and defined as:

$$F_{x_i, C_k}(x_1, x_2, \dots, x_n, C'_k) = \begin{cases} 1, & n_{x_i}(x_1, x_2, \dots, x_n) > 0 \text{ } and \text{ } C'_k = C_k \\ 0, & otherwise \end{cases} \tag{3}$$

Support Vector Machine (SVM) (Scholkopf & Smola, 2001) is a larger marginal classifier rather than a probabilistic classifier. It separates the observations in different classes, optimally keeping the margin as great as possible and reaching the optimal **w** by the following linear optimization problem:

$$min\, ||\boldsymbol{w}||^2_{\mathcal{H}} + C \sum_{i=1}^{n} \xi_i \tag{4}$$

subject to:

$$y_i(\mathbf{w}^T x_i + b) \geq 1 - \xi_i\, , for\ i = 1, 2, \dots, n \tag{5}$$

$$\xi_i \geq 0 \tag{6}$$

where x_i is the observations, $y_i \in \pm 1$ is the class labels for $i = 1,2,\dots,n$, ξ_i is the slack variable for linearly not separable observations, $\mathcal{H}$ shows feature space, and C is the regularization parameter, in that the bigger it is the more the errors are penalized.

3.2. Deep Learning Paradigm

"Deep learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level" (LeCun, Bengio, & Hinton, 2015, p. 1). Deep neural networks automatically capture the representation of words, conferred contextual information, from a raw input corpus for a particular task, independent of any hand-crafted features. We will discuss word embeddings as well as each of these architectures in the following sub-sections.

3.2.1. Word Embedding

The word vectors, also called word embeddings, capture semantic and syntactic characteristics of words over the corpus. Thus, semantically and syntactically similar words will be mapped to nearby points. There are two principal approaches to develop pre-trained word embeddings, GloVe and Word2Vec. Applying these two algorithms, we have also constructed new word embeddings using unlabeled messages collected from StockTwits in order to evaluate their performance in capturing domain-specific similarities in finance.

3.2.1.1. Skip-gram with Negative Sampling (SGNS)

Briefly, SGNS (Mikolov, Sutskever, et al., 2013) is a predictive approach that tries to find context words surrounding a given target word. Using a fully connected neural network with a single hidden layer, it aims to maximize the average of the sum of log probabilities through the following objective function:

$$J_\theta = \frac{1}{T}\sum_{t=1}^{T}\sum_{-n\le j\le n, j\ne 0}\log p(w_{t+j}|w_t) \tag{7}$$

where T is corpus size, n is context size, and $p(w_{t+j}|w_t)$ is calculated by the following softmax function:

$$p(w_O|w_I) = \frac{\exp {v'_{w_O}}^{\mathrm{T}} v_{w_I}}{\sum_{w\in W} {v'_w}^{\mathrm{T}} v_{w_I}} \tag{8}$$

where W is vocabulary size and v_w and v'_w are "input" and "output" embedding vector of word w. However, this will lead to a very high computation cost of $\nabla p(w_O|w_I)$ due to the size of W, which is usually large, and, therefore, two options have been introduced in order to make it computationally efficient (Mikolov, Sutskever, et al., 2013).

First, the **sub-sampling** scheme has been proposed to deal with frequent words such as "in", "the", and "a", as they usually provide less information than rare words. Second, the **negative sampling** has been presented based on the skip-gram model but with a different objective function to approximate the loss of softmax with the aim of reducing computation time.

3.2.1.2. Global Vectors (GloVe)

On the other hand, GloVe (Pennington et al., 2014) forms the co-occurrence matrix X each of whose elements, X_{ij}, represents the number of times word j appears in the context of word i. The word context is defined by a variable window size. During construction of the co-occurrence matrix, the decreasing weighting function of $1/d$ applies for the pairs that appear d words away from the center word as they may carry less relevant information.

The soft constraint for each word pair is defined as follow:

$$w_i^T w_j + b_i + b_j = \log(X_{ij}) \tag{9}$$

where w_i is vector for center word and w_j is vector for the context word and b_i and b_j are their scalar biases, respectively. In the end, the cost function below, a weighted least squares regression model, will be minmized:

$$J = \sum_{i=1}^{V}\sum_{j=1}^{V} f(X_{ij})(w_i^T w_j + b_i + b_j - \log(X_{ij}))^2 \tag{10}$$

where V is the size of the vocabulary and f is weighting function designed to reduce the effect of extremely common word pairs. The authors have chosen the following function:

$$f(X_{ij}) = \begin{cases}(x/x_{max})^{\alpha}, & if\ x < x_{max}\\ 1, & otherwise\end{cases} \tag{11}$$

where $x_{max} = 100$ and $\alpha = 0.75$, suggested in the corresponding paper (Pennington et al., 2014).

3.2.2. *The Convolutional Neural Network (CNN)*

The CNN implemented for investor sentiment classification in this study, following Kim's study (Kim, 2014), consists of four primary layers. The corresponding CNN is constructed by an input layer, convolution layer, max-pooling layer, and fully connected layer (see Figure ***1***). Below, we have discussed each layer with the relative mathematical formulation.

Input Layer treats the input sentence (tweet on StockTwits in our case) as a sequence of n words, each of which is represented by a d-dimensional vector of embedding: $[x_1, x_2, x_3, \dots, x_n]$ where $x_i \in \mathbb{R}^d \ \forall\ i = 1,2,3, \dots, n$.

Convolution Layer aims to capture local features that concurrently appear in the previous layer by a set of learnable filters called convolution kernels. Mathematically, the weight matrix for the convolution filter is $\mathbf{w} \in \mathbb{R}^{h \times d}$, which will be applied to the window of h words with an embedding dimension of d.

After convolving every possible window of words, the feature map c becomes:

$$c = [c_1, c_2, c_3, \dots, c_{n-h+1}] \tag{9}$$

where $c \in \mathbb{R}^{n-h+1}$ and the convolution filter c_i for position i in the sentence is calculated by:

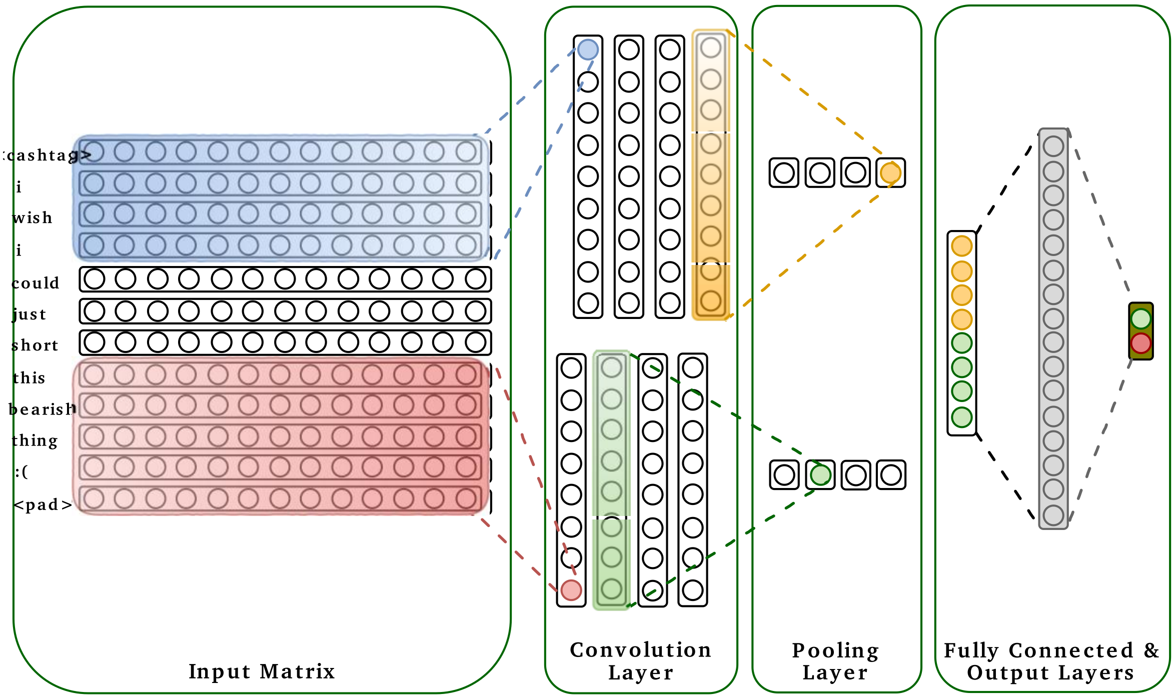
$$c_i = f(\mathbf{w} \cdot x_{i:i+h-1} + b) \tag{10}$$

where $b \in \mathbb{R}$ is bias and f is a non-linear activation function[1].

Max-pooling Layer addresses the most important features by pooling over every feature map bearing a close resemblance to the process of feature selection in natural language processing. Thus, the pooled feature map, p, will be calculated by:

$$\mathrm{p} = [\max(c_1, c_2, c_3, \dots, c_{n-h+1})] \tag{11}$$

Figure 1: Downscaled Convolutional Neural Network (CNN) for investor sentiment classification.



[1] In this study, the Rectified Linear Unit (ReLU) (Hinton et al., 2012) has been chosen as the activation function as it speeds up the training process and leads better performance in many cases.

Finally, the concatenated and flattened pooled feature maps are passed through a high dimensional dense layer - known as \textbf{\textit{the fully connected layer}} and fed into the output layer whose output is the class probabilities. The output layer computes these probabilities by soft-max activation as follows:

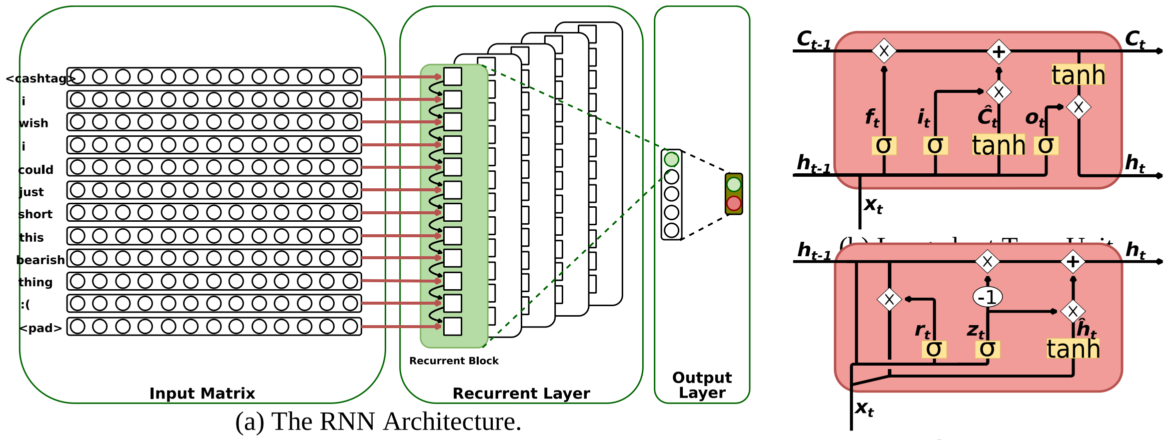
$$P(y = j|\mathbf{x}, \mathbf{w}, b) = \text{softmax}_j(\mathbf{x}^T\mathbf{w} + b) = \frac{e^{\mathbf{x}^T\mathbf{w}_j + b_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^T\mathbf{w}_k + b_k}} \quad (12)$$

where $\mathbf{w}_k$ and b_k are the weight vector and bias of the k-th class.

3.2.3. The Recurrent Neural Network (RNN)

The recurrent neural network (RNN), as an extension of feed-forward neural networks, can handle variable-length sequences, having a recurrent hidden state whose activation on the current time-step is dependent on what it has seen on the earlier time step (see
Figure ***2*** (a)). Despite excellent performance on various problems such as speech recognition, language modeling, and image captioning, the original RNN is not practically able to learn long-term dependencies in the sequences (Bengio, Simard, & Frasconi, 1994). Two recent versions of RNNs have been proposed: Long Short-Term Memory (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Unit (Cho et al., 2014). The input and output layers are the same as in CNN, so we skip re-explaining them here. The following subsections will discuss the LSTM and GRU units (shown in
Figure ***2*** (b) and
Figure ***2*** (c) respectively).

Figure 2: Downscaled Recurrent Neural Networks implemented on investor sentiment classification. a) RNN Structure, b) a LSTM unit, and c) a GRU unit.



(a) The RNN Architecture.

(c) Gated Recurrent Unit.

3.2.3.1. Long-short Term Memory (LSTM)

Incorporating the cell state C_t at time step t, the LSTM unit controls the flow of information from the previous time step. This enables it to store relevant information from early time steps and carry it over long time steps to employ in later time steps. This process takes place through three gates; the forget gate, the input gate, and the output gate (see Figuer 4(b)). The parameters are updated through the following equations:

$$f_t = \sigma(W_f \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}] + b_f) \quad (13)$$

$$i_t = \sigma(W_i \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}] + b_i) \quad (14)$$

$$\tilde{C}_t = \tanh(W_C \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}]) \quad (15)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (16)$$
$$o_t = \sigma(W_o \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}] + b_o) \quad (17)$$
$$h_t = o_t * \tanh(C_t) \quad (18)$$

3.2.3.2. Gated Recurrent Unit (GRU)

Like the LSTM unit, the GRU is designed to capture long-term dependencies of the input sequences but without carrying the cell state from one time step to the next. Moreover, it merges the input gate and forget gate into a single update gate that controls the degree to which past information should matter in the current time step. This is determined by:

$$z_t = \sigma(W_z \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}]) \quad (19)$$
$$r_t = \sigma(W_r \cdot [\mathbf{x}_t, \mathbf{h}_{t-1}]) \quad (20)$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * \mathbf{h}_{t-1}, \mathbf{x}_t]) \quad (21)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (22)$$

4. Results and Discussions

In this section we discuss the performance of both traditional classifiers that use various feature types and deep neural networks for classifying investor sentiment.

4.1. Data from StockTwits

StockTwits is a social media platform designed for investors wherein they share ideas, beliefs, and/or feelings about financial markets behavior. It is a place for users to observe traders and investors, produce posts and contribute to conversations related to the market and individual stocks. Here, amateur investors can meet and interact with professionals freely. The streams in StockTwits contain ideas, links, charts, and financial data expressed within 140 characters. By the end of 2016, more than 63 million messages had been posted by 250,000 users.

Table 1 provides a brief statistics of collected messages from StockTwits between June 2008 and December 2016. As shown in Table 1, users have posted 63,647,533 messages in total between June 2008 and December 2016, which include 82.20% unlabeled messages,

	Volume (Percentation)
Positive	9,161,337 (14.39%)
Negative	2,173,180 (3.41%)
Unlabelled	52,313,016 (82.20%)
Total	**63,647,533 (100.0%)**

14.39% positive messages, and 3.41% negative messages. unlabeled messages have been fed.

Table 1: Statistics of collected messages from StockTwits.com.

For the purpose of this research, all into word embedding generation algorithms, GloVe and Word2Vec, to map the words into the high-dimensional space of embeddings. Moreover, a "balanced dataset" of messages has been randomly chosen that contains 217,712 bullish and 216,924 bearish messages. All classifiers and algorithms have been trained and tested on this dataset in order to have a consistent comparison. All messages have been put through some general pre-processing tasks, including the replacement of URLs with <url>, cashtags with <cashtag>, hashtags with <hashtag>, user mentions with <usertag>, and real numbers with <number>, collapsing letter repetitions (e.g. "haaaaappppppy" and "Coooool" will become "haaapppy" and "Coool", respectively), expanding contractions (e.g. "I've" will be replaced with "I have"), and discarding tokens with occurrences less than 5.

4.2. Traditional Investor Sentiment Classification

Which Feature Type, Which Algorithm
In our method of constructing the baseline to evaluate the effect of emojis and deep neural networks in the financial context, four different feature types with three classification techniques have been incorporated.
 briefly illustrates the experimental setups for traditional classification approaches including feature extraction and classifier development. The raw messages are transformed to binary feature vectors whose element i is set after pre-processing to one if feature f_i exists in the corresponding message and zero otherwise.

4.2.1. Which Feature Type, Which Algorithm

In our method of constructing the baseline to evaluate the effect of emojis and deep neural networks in the financial context, four different feature types with three classification techniques have been incorporated.

By removing infrequent and useless features such as misspellings, we have discarded the features that appear in less than five messages to reduce the sparsity of the input. Instead of using built-in stop-words, we have eliminated the features that appear in over 75% of the messages to remove less informative but highly frequent features. During the classifier development process, we have implemented SVM with linear kernel, MaxEnt with liblinear solver, and Multinomial NB classifier.

We have carried out stratified 10-fold cross-validation while default values are set for all other parameters of both vectorizers and classifiers. Moreover, we have implemented a Wilcoxon Sum-Rank Test (WSuRT) (Vidakovic, 2013) to statistically compare the independent and random sample[2].

Figure 3: Performance of various classifiers on different feature types.

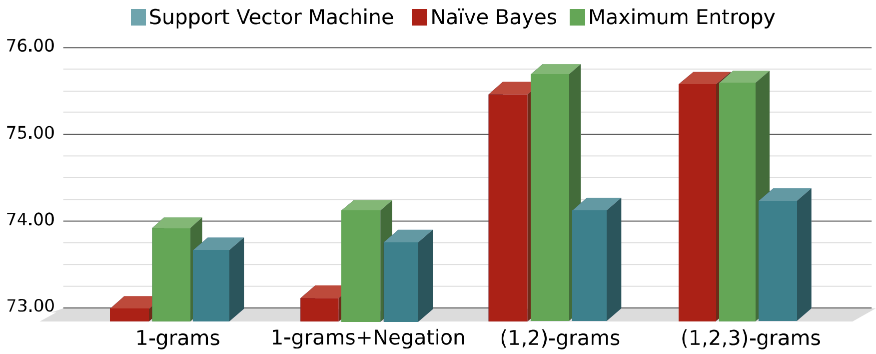


Figure *1***Figure 3: Performance of various classifiers on different feature types.** shows the performance of the classifiers, Naïve Bayes (NB), Maximum Entropy (MaxEnt), and Support Vector Machine (SVM), incorporating four different feature sets, 1-

Table 2: Chosen parameter values in feature extraction process and classification process.

[2] We have used Scikit-Learn API steps of developing the traditional classifier.

grams, 1-grams with negation tag, (1,2)-grams, and (1,2,3)-grams. MaxEnt out-performs SVM on every feature type (WSuRT with p-value≤2.1e-32). However, it does not show significant out-performance compared to NB over (1,2)-grams and (1,2,3)-grams with Wilcoxon p-value of 0.7959 and 0.0524, respectively. It can be seen that the negation tag does not lead to significant improvement in the performance of classifiers. This result implies that more sophisticated feature engineering mechanisms are required to capture such a complex linguistic structure, whereas, combining bi-grams or tri-grams with uni-grams boosts the performance regardless of the classification algorithm (WSuRT with p-value≤10.8e-6). Moreover, SVM demonstrates the least responsiveness or improvement inaccuracy with respect to changes in feature types, perhaps because of the complex pattern of the dataset, which is hard for the model to capture using simple a linear kernel.

Figure 4: Statistics of messages with emojis and performance of MaxEnt with emojis and without emojis.

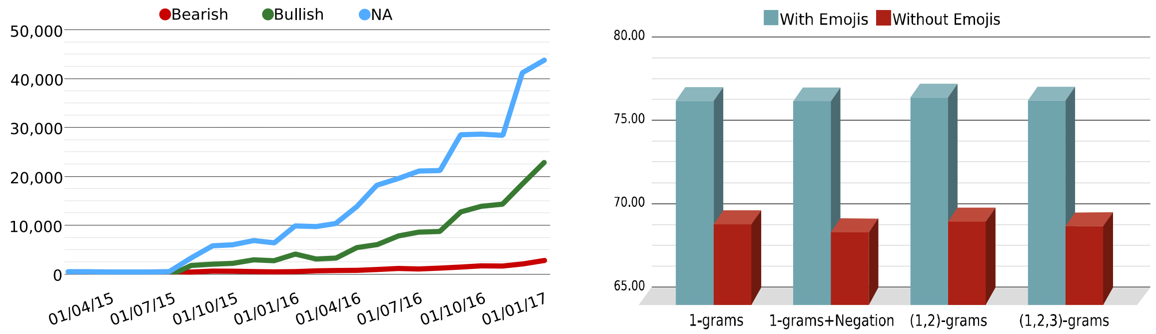


Figure 5: Emoji cloud for popular emojis in a) bullish messages and b) bearish messages.

Feature Extraction	Value	Classifier	Value
Minimum Document Frequency	5	Kernel (Support Vector Machine)	Linear
Maximum Document Frequency	75%	Solver (Maximum Entropy)	LibLinear
Feature Transformation	Binary	Model (Na¨ıve Bayes)	Multinomial NB

(a)

(b)

4.2.2. Emojis and Investors

Emojis are dynamic and dominant entities of financial social networks, in our case StockTwits, where investors actively express their feelings and opinions. This website has provided emojis for the use of investors since mid-2015. Therefore, the limited number of messages, approximately 0.8% of messages, contain at least one emoji. Emojis are now becoming exponentially more popular as a means of expressing feelings and emotions in the financial context (see **Error! Reference source not found.** (a)). Briefly, there are 1,658 unique emojis that have been used 1,032,352 times overall by investors in 508,097 messages. Of the sentiment-labeled messages, there are only 19,376 bearish and 144,166 bullish messages in which at least one emoji has appeared[3]. Therefore, to better demonstrate and analyze emojis' discriminative power in investor sentiment classification, we have constructed another balanced dataset that contains 38,752 messages, half bullish and half bearish. **Error! Reference source not found.** (b) shows the performance of MaxEnt classifier, the best classifier from the previous section, on this dataset. The expected results reveal that the emojis lead to 7.5% higher classification accuracy[4] through manipulating the discriminative power that is somehow hidden in their usage pattern. However, the presence of other feature types, bi-grams, tri-grams, and negation, does not have significant impact on the accuracy in the corresponding dataset with p-value≤0.3843.

4.3. Deep Learning

Before moving on to the results gained from use of the deep learning paradigm, we will briefly discuss the experimental settings for word-embedding constructions algorithms, GloVe and Word2Vec, and deep neural networks, CNN, GRU, and LSTM.

4.3.1. Word Embeddings, Domain-specific vs. Domain-general

Error! Reference source not found. presents values for the parameters of GloVe and SGNS in constructing domain-specific word embeddings, GloVeST and Word2VecST, and brief information about the corpus on which they have been trained. Without padding, the corpus contains 52,313,016 unlabeled messages from StockTwits which are composed of more than 838 million tokens after pre-processing. The domain-specific word embeddings are trained for 50 iterations to construct a 300-dimensional embedding vector of the words that appeared more than five times, taking a

[3] **Error! Reference source not found.** shows the emoji cloud for common emojis in bullish and bearish messages.

[4] This is confirmed by WSuRT with p-value≤1.82e-4.

Table 4: Examples of word pairs with FinSim index and cosine similarities taken from word embeddings.

Word Pairs	FinSim Index	GloVeST	GloVe	Word2VecST	Word2Vec
Bearish & Negative	0.8	0.6202	0.4193	0.4000	0.4869
Bullish & Positive	0.8	0.6527	0.4551	0.3676	0.5112
Bought & Long	0.7	0.8347	0.5936	0.7090	0.5211
Scalp & Swing	0.75	0.7765	0.0114	0.6385	0.0911
Mutual & Reciprocal	0.2	-0.0855	0.4390	-0.0181	0.5800
Hedge & Mutual	0.75	0.5088	0.3813	0.5276	0.0991
Correlation with FinSim Index		**0.7274**	**0.2524**	**0.7731**	**0.4268**

Table 3: Brief info about general word emebeddings and parameter setup to train domain-specific embeddings, GloVeST and Word2VecST

window size of eight. By way of comparison, GloVe and Word2Vec are trained on the general datasets, Wikipedia 2014 plus Gigaword 5 with 6 billion tokens and Google News with 100 billion tokens respectively.

Two main schemes have been introduced to evaluate the word embeddings: extrinsic and intrinsic. Extrinsic evaluation methods use word embeddings as an input for another task such as named-entity recognition, part-of-speech tagging, or sentiment classification (Pennington et

	GloVe	*Word2Vec*	*GloVeST and Word2VecST*
Token Size	400,000	1,000,000	263,306
Vector Dimension	300	300	300
Number of Tokens	6,000,000,000	100,000,000,000	838,009,514

GloVe (Pennington et al., 2014)		*SGNS (Mikolov et al., 2013b)*	
Minimum Frequency	5	Minimum Frequency	5
Window Size (Symmetric)	8	Window Size	8
Xmax	100	*k*	25
A	0.75	*t*	.0001
Number of Iterations	50	Number of Iterations	50

al., 2014) with their particular performance measure. Intrinsic methods assess the quality of word embeddings by evaluating syntactic and semantic relationships between words by use of a set of pre-selected query terms (Schnabel et al., 2015). Similarity (or relatedness) is an example of intrinsic approaches where the aim is to measure the correlation between the similarity scores of query terms and cosine similarity as computed by the word embeddings. However, existing query datasets are not suitable to evaluate finance word embeddings as they

do not cover any domain-specific query terms. Therefore, we have constructed a dataset of 158 query terms for finance, called *FinSim*, to assess the word embeddings intrinsically.
Table ***4*** shows a few examples of queries using the FinSim Index and cosine similarities calculated from GloVeST and Word2VecST (domain-specific word embeddings) and GloVe[5] and Word2Vec[6] (domain-general word embeddings). The FinSim index represents the similarity of words in the finance context, scored by five finance experts, and scales between [-1, 1].

Then, the Pearson correlation between this FinSim index and cosine similarities reveals the

Common Parameters			
Maximum length		30	
Unknown embedding vector		U[-1,1]	
Kernel and bias initializer		Normal (He, Zhang, Ren, & Sun, 2015)	
Activation		ReLU (Hinton et al., 2012)	
Loss function		Binary cross-entropy	
Optimizer		Adadelta (Zeiler, 2012)	
Back size and epochs		500 & 100	
CNN		***LSTM and GRU***	
Filter size	200	Hidden units	100
Kernel size	[2, 3]	Dropout	0.2
Pooling	Global max-pooling		
Dropout	0.5		

quality of word embeddings and their ability to represent finance-related syntactic and semantic relationships. The last row in
Table ***4*** displays the correlation score between the FinSim index and the cosine similarity calculated by each pre-trained word embedding. It can be easily seen that domain-specific word embeddings out-perform the domain-general word embeddings in capturing the finance context similarities. It is also shown that the SGNS has produced more high-quality word embeddings to interpret the finance-specific language used by investors than GloVe. Among domain-general word embeddings, Word2Vec performs better than GloVe because it has been trained over an extremely large dataset that enables it to capture some level of finance syntactic and semantic relationships. This demonstrates the notably high demand for domain-specific word embeddings, especially among finance and investment communities.

4.3.2. Deep Learning Algorithms

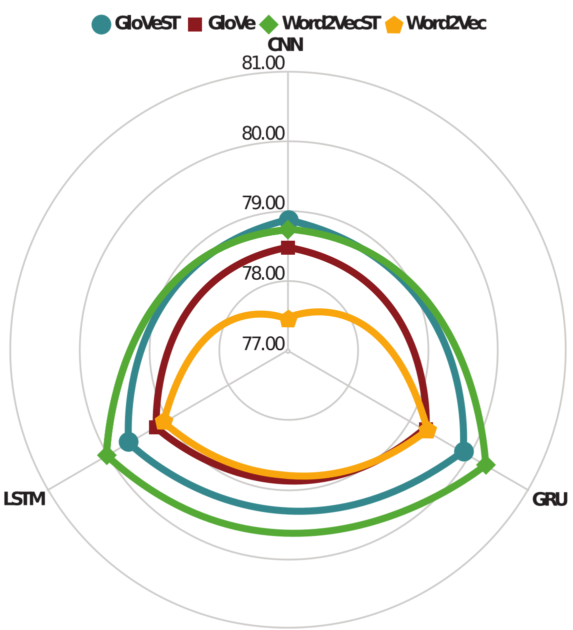
This section discusses the performance of deep neural networks and the three-fold effect of word embeddings: convolutional versus recurrent neural networks, domain-specific versus domain-general word embeddings, and static versus non-static word embeddings. Before

Table 5: Parameter setup for CNN, GRU, and LSTM.

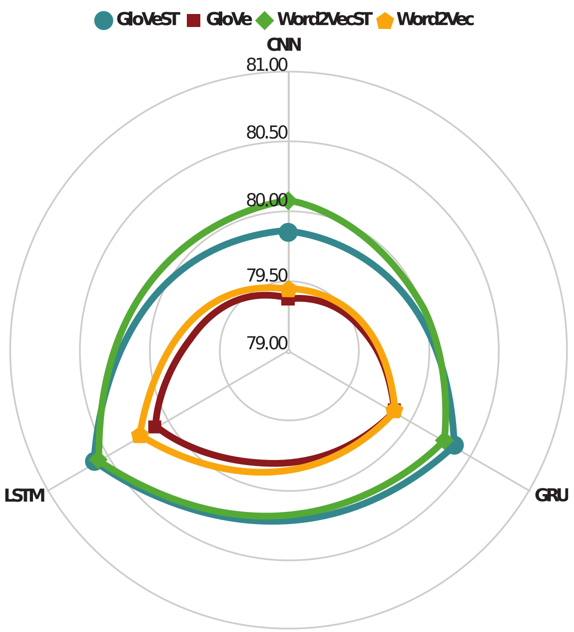
discussing the results, note that deep neural networks have millions of weights other than word embeddings to fit the problem at hand. The models used in this study reflect the quality of the word embeddings while performing sentiment classification. However, the difference might not be huge, especially when a reasonable amount of data is provided for them to train.

[5] Available on http://nlp.stanford.edu/data/glove.6B.zip
[6] Available on https://code.google.com/archive/p/word2vec/

Figure 4: Performance of Deep Learning Algorithms with Static and Non-static Word Embeddings.



(a) Static



(b) Non-static

4.3.2.1. Convolution vs. Recurrent

Error! Reference source not found. provides an overview of the hyper-parameter setup of the deep neural networks, CNN, GRU, and LSTM[7]. Again, stratified 10-fold cross-validation and Wilcoxon Sum-Rank Test have been undertaken to evaluate the performance of the DNNs.

Figure 4 illustrates the performance of deep neural networks trained over various word embeddings with static or non-static states. CNNs and RNNs both outperform the best traditional approach[8], extracting a greater number of hidden sentimental and semantic overtones of messages. The CNNs handle local features at different positions using convolutional filters and handle long-range relationships using pooling operations. In contrast, RNNs try to capture long-term dependencies through memory and forget gates. As the tweets are fairly short, we expected CNN to perform as well as the GRU and LSTM. However, it has underperformed the RNNs more specifically when static word embeddings are fed into the models (WSuRT with p-value≤0.0015). Moreover, GRU performs the best with static word embeddings whereas the LSTM outperforms in non-static word embeddings where models have the chance to back-propagate word embeddings.

4.3.2.2. Domain-specific vs. Domain-general

As Figure 4 shows, domain-specific word embeddings lead to higher accuracy than the domain-general ones even in the non-static state where the network updates them over the problem at hand. As shown in subsection 4.3.1, this is because words have a different pattern of usage in the finance context where investors have developed their own language. Therefore, compared with general-domain word embeddings, domain-specific word embeddings are more efficient universal feature extractors that help deep neural networks better understand financial language (WSuRT with p-value≤0.0115). This result is particularly noteworthy given deep neural

[7] The models have been constructed in Keras API (https://keras.io/) with Tensorflow backend and trained on single NVidia K80 GPU.

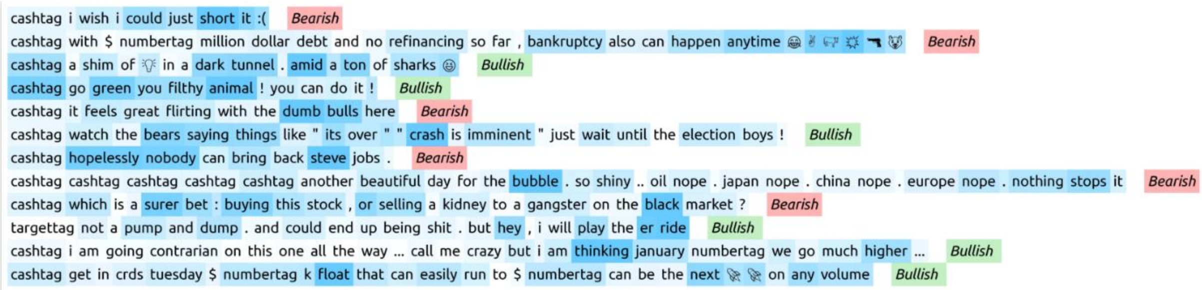
[8] MaxEnt with (1,2)-grams plus emojis gives 77.68% accuracy on the same dataset (WSuRT with p-value≤10.8e-6).

networks have millions of weights other than word embeddings to update during the training process.

4.3.2.3. Static vs. Non-static

Initial word embeddings carry any information about syntactic and semantic properties of every token in the corpus where the words with similar syntactical and semantic characteristics appear close to each other. However, they do not entail any information about the characteristics of the words for the problem at hand. Therefore, the deep neural network with non-static word embedding provides a valuable opportunity to adjust word vectors and make them more specific to the problem at issue here, investor sentiment classification.

Error! Reference source not found. shows the location[9] of the top 10 polar words in the stock market context before (static state) and after (non-static state) training the LSTM with GloVeST (the best combination). Predictably, the neural network fine-tunes the word embeddings in such a way that they become distinguishable based on their sentiment too. Thus, this provides an ideal chance for the neural networks to calibrate the word embeddings to reach their highest performance. We can observe significant out-performance of DNNs with non-



static word embeddings compared to static ones, as confirmed by WSuRT with a p-value≤0.000487. As illustrated in Figure 4, CNN with non-static word embeddings shows a 2% boost in its classification performance compared with static ones, and yet GRU and LSTM perform better regardless of word embedding types.

4.3.3. Qualitative Analysis

In order to better understand the performance of deep neural networks, LSTM with GloVeST, we have extracted the saliency of some of the input texts where the goal is to visualize the units

that contribute most to the final classification. By computing the gradient of output category with respect to the input, the saliency score demonstrates how output value changes with respect to a small change in the input (Simonyan, Vedaldi, & Zisserman, 2013).

In other words, the saliency score is the absolute value of the derivative of the loss function with respect to each dimension of all input words in the corresponding sentence (J. Li, Chen, Hovy, & Jurafsky, 2015). **Error! Reference source not found.** illustrates the gradient concentration of all input words in 12 sentences with

Figure 7: Location of sentiment related words before (black) and after (positives & negatives).

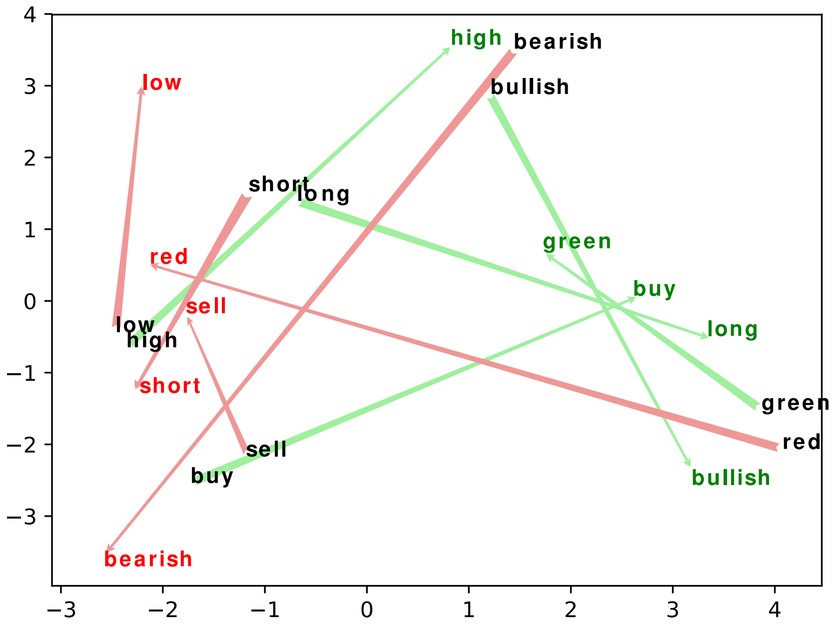
[9] Principal Component Analysis (PCA) (Wold, Esbensen, & Geladi, 1987) is used to reduce the dimension of word embedding vectors in order to ease visualization.

Figure 8: Saliency concentration over every token in the sentences extracted from LSTM with GloVeST (the best combination).

Darker shadows show intense saliency concentration variable length and various types of structures. With these few examples, the aim is to show how LSTM reflects different properties such as negation, sarcasm, irony, joke, and/or emojis. For the short messages, the LSTM relies mostly on discriminative features, such as emojis or sentimental words. Not surprisingly, it is able to understand some level of jokes and sarcasm by assigning higher saliency to the relevant tokens. On the other hand, it disregardsinsubstantial parts with lower saliency score and accumulates key information over lengthy sentences capturing long-term discriminative dependencies.

5. Conclusion, Limitations, and Future Work

The development of an accurate classifier of investor sentiment is required to facilitate empirical investigations of the role that sentiment plays in financial markets. Using data from StockTwits, it is shown that MaxEnt and NB outperform SVM despite their simple classification foundation with a strong independence assumption of the features. Moreover, bi-grams and tri-grams robustly boost the classification performance of investor sentiment, capturing long-range dependencies to some extent in the tweets. Although negation is one of the key grammatical rules that inverts the meaning and polarity of a sentence in multiple ways, the implemented negation tagging mechanism does not lead to significant improvement in the performance of classifiers (see Figure 3). As discussed in Section 14.2.2, this study reveals that emojis carry very strong discriminative power in the finance context in spite of their domain-specific pattern of usage. Thus, the existence of emojis in the financial texts has contributed substantially to classification performance. However, in the presence of emojis, the bi-grams and tri-grams retain their significant out-performance. However, the feature preparation process plays a crucial role in the performance of traditional classification methods such as SVM, as debated in the literature.



In general, deep neural networks outperform traditional methods with 1-4% improvement in classification accuracy, depending on the topology and word embeddings (see Figure 4). As we have discussed previously, LSTM and GRU unexpectedly perform better than CNN, although the StockTwits messages are quite short and therefore suitable for CNN to learn local features. LSTM demonstrates robust ability to capture long-term discriminative dependencies without any feature engineering. It is able to focus on the linguistic entities such as emojis, negation, and sarcasm to some degree. Domain-specific word embeddings produce better DNN models of investor sentiment classification and achieve higher accuracy. The domain-specific word embeddings have presented a consistent performance in capturing finance-context similarities compared with general-domain word embeddings, as shown by the intrinsic evaluation method. They illustrate this performance through a higher Pearson correlation with the FinSim score of word pairs, which is indexed by finance experts.

There are a number of methodological limitations that propose new lines for further investigation. First, we tested n-grams, emojis and emoticons, and negation in developing a valid baseline of traditional investor sentiment classification. The rule-based feature engineering mechanisms will help to capture domain-specific properties of texts from StockTwits to some degree and lead to the development of robust baseline for DNNs in future studies. Second, we extracted hyper-parameters of the deep neural networks from related studies on general-purpose sentiment classification problems. Therefore, the hyper-parameter tuning, which will undoubtedly lead to higher performance, is recommended for DNNs in the investor sentiment classification problem. Moreover, the DNNs with more complex topologies might be considered to see if they can improve accuracy, since they will lead to a higher computation cost. Combining CNN with an RNN is another option to boost the classification accuracy that enables the model to capture both local features and long-term dependencies of complicated texts from StockTwits. Third, the dataset for training domain-specific word embeddings is limited to unlabeled messages posted on StockTwits with a small corpus size.
Although these word embeddings show outstanding performance in capturing semantic and syntactic similarities of the finance context, other resources are available to extract highly reliable word embeddings that can well represent finance context similarities. Finally, we have created a query dataset of similar finance words including the limited number of highly frequent words in StockTwits. In order to have a robust intrinsic evaluation method, we recommend extending the query dataset and including less-frequently occurring word pairs.

References

Aggarwal, C. C., & Zhai, C. X. (2012). Mining Text Data. New York, NY, USA: Springer.

Al Nasseri, A., Tucker, A., & de Cesare, S. (2014). Big Data Analysis of StockTwits to Predict Sentiments in the Stock Market. In S. Džeroski, P. Panov, D. Kocev, & L. Todorovski (Eds.), Discovery Science: 17th International Conference, DS 2014, Bled, Slovenia, October 8-10, 2014. Proceedings (pp. 13–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-11812-3_2

Al Nasseri, A., Tucker, A., & de Cesare, S. (2015). Quantifying StockTwits semantic terms' trading behavior in financial markets: An effective application of decision tree algorithms. Expert Systems with Applications, 42(23), 9192–9210. https://doi.org/10.1016/j.eswa.2015.08.008

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance, 59(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). {S}enti{W}ord{N}et 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, … D. Tapias (Eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valletta, Malta: European Language Resources Association (ELRA).

Baker, M., & Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. The Journal of Finance, 55(5), 2219–2257. https://doi.org/10.1111/0022-1082.00285

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166.

Berger, A. L., Pietra, V. J. Della, & Pietra, S. A. Della. (1996). A maximum entropy approach to natural language processing. Computational Linguistics, 22(1), 39–71.

Black, F. (1986). Noise. The Journal of Finance, 41(3), 528–543.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8.

Brace, I. (2008). Questionnaire design: How to plan, structure and write survey material for effective market research (2nd ed.). London, UK: Kogan Page Publishers.

Chan, S. W. K., & Chong, M. W. C. (2017). Sentiment analysis in financial texts. Decision Support Systems, 94, 53–64. https://doi.org/https://doi.org/10.1016/j.dss.2016.10.006

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder--Decoder Approaches. In Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (pp. 103–111). Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-4012

Cochrane, J. H. (2000). New facts in finance. Economic Perspectives. Federal Reserve Bank of Chicago, 23(3), 36–58.

Deng, S., Sinha, A. P., & Zhao, H. (2017). Adapting sentiment lexicons to domain-specific social media texts. Decision Support Systems, 94, 65–76. https://doi.org/https://doi.org/10.1016/j.dss.2016.11.001

Dougal, C., Engelberg, J., García, D., & Parsons, C. A. (2012). Journalists and the Stock Market. The Review of Financial Studies, 25(3), 639–679. https://doi.org/10.1093/rfs/hhr133

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383–417.

García, D. (2013). Sentiment during Recessions. The Journal of Finance, 68(3), 1267–1300. https://doi.org/10.1111/jofi.12027

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-

Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (pp. 1026–1034). Santiago, Chile: IEEE Computer Society. https://doi.org/10.1109/ICCV.2015.123

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. r., Jaitly, N., … Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine, 29(6), 82–97. https://doi.org/10.1109/msp.2012.2205597

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (pp. 655–665). Baltimore, MD, USA: Association for Computational Linguistics. https://doi.org/10.3115/v1/P14-1062

Kim, Y. (2014). Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (pp. 1746–1751). Doha, Qatar: Association for Computational Linguistics.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Image{N}et Classification with Deep Convolutional Neural Networks. Communications of ACM, 60(6), 84–90. https://doi.org/10.1145/3065386

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2015). Visualizing and understanding neural models in {NLP}. arXiv Preprint arXiv:1506.01066.

Li, Q., & Shah, S. (2017). Learning Stock Market Sentiment Lexicon and Sentiment-Oriented Word Vector from StockTwits. In Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017) (pp. 301–310). Vancouver, BC, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/K17-1031

Li, T., van Dalen, J., & van Rees, P. J. (2017). More than just noise? {E}xamining the information content of stock microblogs on financial markets. Journal of Information Technology, 1–20. https://doi.org/10.1057/s41265-016-0034-2

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? {T}extual Analysis, Dictionaries, and 10-{K}s. The Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Luong, T., Kayser, M., & Manning, C. D. (2015). Deep Neural Language Models for Machine Translation. In Proceedings of the 19th Conference on Computational Natural Language Learning (pp. 305–309). Beijing, China: Association for Computational Linguistics.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. New York, NY, USA: Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Proceedings of International Conference on Learning Representation Workshop Track (pp. 1–12). Scottsdale, AZ, USA.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems (pp. 3111–3119). Lake Tahoe, NV, USA: Curran Associates Inc.

Nofsinger, J. R. (2005). Social Mood and Financial Economics. Journal of Behavioral Finance, 6(3), 144–160. https://doi.org/10.1207/s15427579jpfm0603_4

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. PloS ONE, 10(12), 1–22. https://doi.org/10.1371/journal.pone.0144296

Oh, C., & Sheng, O. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In Proceediings of Thirty Second International Conference on Information Systems (pp. 1–19). Shanghai, China: Association for Information Systems.

Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. In L. Correia, L. P. Reis, & J. Cascalho (Eds.), Progress in Artificial Intelligence: 16th Portuguese Conference on Artificial Intelligence, EPIA 2013, Angra do Hero'{\i}smo, Azores, Portugal, September 9-12, 2013. Proceedings (pp. 355–365). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-40669-0_31

Oliveira, N., Cortez, P., & Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. Decision Support Systems, 85, 62–73. https://doi.org/10.1016/j.dss.2016.02.013

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (pp. 79–86). Philadelphia, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/1118693.1118704

Pennington, J., Socher, R., & Manning, C. D. (2014). Glo{V}e: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (pp. 1532–1543). Doha, Qatar: Association for Computational Linguistics.

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The Effects of Twitter Sentiment on Stock Price Returns. PLoS ONE, 10(9), 1–21. https://doi.org/10.1371/journal.pone.0138441

Schnabel, T., Labutov, I., Mimno, D., & Joachims, T. (2015). Evaluation methods for unsupervised word embeddings. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 298–307). Lisbon, Portugal: Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1036

Scholkopf, B., & Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Cambridge, MA, USA: MIT Press.

See-To, E. W. K., & Yang, Y. (2017). Market Sentiment Dispersion and Its Effects on Stock Return and Volatility. Electronic Markets, 27(3), 283–296. https://doi.org/10.1007/s12525-017-0254-5

Shirani-Mehr, H. (2014). Applications of deep learning to sentiment analysis of movie reviews.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv Preprint arXiv:1312.6034.

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In M. G. Robert, A. D. Don, L. E. John, & R. J. A. Little (Eds.), Survey Nonresponse (pp. 163–177). New York, NY, USA: John Wiley & Sons, Ltd.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and Trades: the Information Content of Stock Microblogs. European Financial Management, 20(5), 926–957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Sun, F., Belatreche, A., Coleman, S., McGinnity, T. M., & Li, Y. (2014). Pre-processing Online Financial Text for Sentiment Classification: A Natural Language Processing Approach. In 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (pp. 122–129). London, United Kingdom: IEEE Computer Society. https://doi.org/10.1109/CIFEr.2014.6924063

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying

Language to Measure Firms' Fundamentals. The Journal of Finance, 63(3), 1437–1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Vidakovic, B. (2013). Engineering Biostatistics: An Introduction using MATLAB and WinBUGS. John Wiley & Sons, Ltd.

Wang, T., Wang, G., Wang, B., Sambasivan, D., Zhang, Z., Li, X., … Zhao, B. Y. (2017). Value and Misinformation in Collaborative Investing Platforms. ACM Transaction on the Web, 11(2), 8:1-8:32. https://doi.org/10.1145/3027487

Wang, X., Jiang, W., & Luo, Z. (2016). Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts. In Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers (pp. 2428–2437). Osaka, Japan: The COLING 2016 Organizing Committee.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2(1), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative Study of {CNN} and {RNN} for Natural Language Processing. arXiv Preprint arXiv:1702.01923.

Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. arXiv Preprint arXiv:1212.5701.

Cluster Analysis and Firm Patterns: A New Approach

Son Tran, Massey University

Jianguo Chen, Massey University

Carolyn G. Wirth, Massey University

ABSTRACT

This paper introduces a novel clustering procedure called regression oriented-weighted K-means clustering (ROWK) to address the heterogeneous-group specific coefficients. Moreover, ROWK employs the regression mean absolute residuals to inform the cluster analysis identification of optimal weights of cluster features. Simulation results show that ROWK works well by (1) placing more (less) weights on relevant (irrelevant) clustering features, (2) identifying weights of features by their contributions to not only cluster recognition but also regression estimation, and (3) reducing the influence of the multi-collinearity problem.

Keywords: Cluster analysis, K-means, feature weightings, group-specific coefficients, firm patterns, earnings persistence

1. Introduction

When conducting regression analysis to forecast and/or test hypotheses, researchers usually make some assumptions, as in the case of the Gauss-Markov assumption. Nevertheless, low predictive power remains a phenomenon that challenges researchers. A reasonable explanation stems from the violation of an implicit assumption of regression analysis, i.e. constant coefficients. Indeed, it is not difficult to find evidence against parameter homogeneity (Lin & Ng, 2012).

While knowledge of underlying sources breaching the constant-coefficients assumption is well addressed in existing research, solutions developed in order to improve forecasting results are still restricted to either including these sources into predictive regressions, or ad-hoc partitioning techniques or in the extreme case, running individual time-series analyses. For example, industry classifications are routinely employed by researchers as a standard criterion for partitioning when running regression analysis (e.g. Cohen & Zarowin, 2008; Hribar & Collins, 2002).

However, this practice generally leads to imprecise estimates due to the different relations between the dependent variable and its determinants within each industry (Fairfield, Whisenant, & Yohn, 2003). Furthermore, Bernard & Skinner (1996) find that discretionary accruals estimates are even less precise in time-series estimates. For those concerns, a partitioning technique that is able to incorporate several potential partitioning factors, identity the appropriate number of groups to assign and achieve the smallest within-cluster variance is critical. However, existing studies of earnings forecasts rarely develop such technique (Richardson, Tuna, & Wysocki, 2010).

This paper proposes to address these issues by employing a data clustering technique, particularly K-means. It is an unsupervised learning procedure that organizes observations into different clusters such that observations in the same cluster are homogeneous to each other but are different from those in other groups (Fred & Jain, 2005). Therefore, K-means could offer a potential solution to deal with the problem of heterogeneous group-specific coefficients.

Despite its popularity and successful use in many applications, K-means still has some substantial inherent shortcomings (Qian, 2006). The most challenging issue of cluster analysis relates to the feature's (or dimension) weightings[10]. Features have differential abilities to define cluster patterns. As a result, strengthening the highly relevant features while lessening the effect of irrelevant or less important features can be essential to discovering the true cluster membership (Brusco & Cradit, 2001). These issues have not been sufficiently recognized and addressed in much of the past research (e.g.Epure, Kerstens, & Prior, 2011; Lee, Lee, & Wicks, 2004).

This paper addresses this issue of clustering technique by connecting clustering and regression analysis. We call this regression oriented-weighted K-means clustering (hereafter, ROWK). It mitigates the violation of homogeneous coefficients by placing firms into different groups such that firms in the same group are homogeneous to each other but different from those in other groups. Furthermore, regression analysis with its mean (absolute) square residuals may

[10] In this paper, "feature" or "clustering feature" or "dimension" denotes a characteristic that is used to distinguish clusters. For example, sepal length, sepal width, petal length and petal width are features that is used to classify the well-known Iris dataset of 150 flower specimens (Amorim & Mirkin, 2012).

potentially provide an external criterion to address the deficiencies of K-means, particularly with respect to the problem of finding optimal weights.

Using simulated data, this study documents that ROWK successfully mitigates the effect of highly correlated variables. Specifically, it aims to lower the weights of those dimensions that are less important and have high within-cluster correlation with relevant features. With the external objective to minimize the mean (absolute) square residuals of the regression model, ROWK demonstrates superior performance to identify the correct weights of cluster features. Irrelevant and random factors are more likely to receive low or zero weights after executing ROWK. As a result, the precision cluster membership increases significantly compared to the standard un-weighted K-means.

This study contributes to both cluster analysis and financial literature in several important ways. First, to the best of our knowledge, this study is the first to simultaneously address the issues of inherent drawbacks of K-means and the constant-coefficients assumption violation. Some studies attempt to deal with cluster feature weighting (e.g. Chiang & Mirkin, 2010) or heterogeneous group-specific coefficients (e.g. Ando & Bai, 2015), but they do so independently. This study proposes combining cluster analysis and regression analysis to enhance the performance of both optimal weights in clustering and heterogeneous parameters in regression.

Second, this study develops and introduces a new, standard procedure to conduct ROWK in order to connect clustering and regression. It has the advantage of being easy to understand and execute using typical data programs such as SAS. Hence ROWK equips researchers with a powerful tool to enhance regression results whenever there are indications of heterogeneous coefficients, which are typically problematic in financial disciplines.

Third, despite recent efforts to address the issue of heterogeneous parameters, most studies focus merely on regression side, ignoring underlying reasons for the problem, i.e. cluster patterns (e.g Ando & Bai, 2015; Lin & Ng, 2012). To find the optimal weights of clustering features, the ROWK procedure proposed in this study helps to distinguish those factors that are essential to identity the cluster pattern. It will provide empirical results that shed more light on why certain features are more important. For example, firm size, is documented in both theoretical and empirical studies to be an important factor that moderates the relationship between independent and dependent variables (e.g.Cooper, Gulen, & Schill, 2008; F.Fama & R.French, 1992; Samuels, 1965).

The remainder of the paper is organized as follows. Section 2 reviews prior literature and develops testable hypotheses. Section 3 describes the proposed ROWK procedure. Section 4 presents details of a data simulation. These simulated data are then used to demonstrate the performance of ROWK in section 5. Conclusions and future work are then discussed in section 6.

2. Literature review

This section gives a brief summary on evidence of different behaviors across grouping firms and suggests reasons why cluster analysis emerges as a potential weapon to combat the heterogeneous coefficients problem.

2.1 Violation of the underlying constant β assumption of regression analysis

Regression models using cross-sectional or panel data often take coefficients' homogeneity as an assumption. However, it is not difficult to find evidence against it.[11] These could make overall regression results invalid. Hence, instability of coefficients over time and/or across firm groups, low R-squared in-sample estimation and poor out-of-sample predictive performance are likely to be observed when this violation occurs. Appropriate partitioning of data is suggested as an essential solution for addressing this problem by Ou and Penman (1989) and Nissim and Penman (2001).

Approaches to partitioning data have tended to be simplistic, such as by dividing a whole sample into different quintiles of certain firm features at which we expect to observe different relationships. Common partitioning variables are: firm size and book to market ratio in the asset pricing model (e.g. Fama & French, 1993); industry competitiveness in corporate governance (Giroud & Mueller, 2011); earnings volatility (Dichev & Tang, 2009), firm life cycle (Dickinson, 2011), and business strategy (Little, Little, & Coffee, 2009) in equity valuation. By dividing data into sub-samples, researchers can gain a deeper understanding of dynamic relationships between concepts, achieve better estimations, increase predictive power, and observe better consistency of estimates across time.

However, this approach suffers from two shortcomings. First, it is ad-hoc and does not take into account data patterns. Dividing data (firms) into quintiles based on some proposed factors (e.g. size, market to book ratio, earnings volatility, etc.) without paying attention to the nature of the firm data is not an optimal solution. For instance, consider the case where we divide firms into 5x5 portfolios based on firm size and book to market ratio. If the nature of the data is such that firms are best clustered into 2x3 groups, then employing 5x5 portfolios will not provide an optimal partition. A limit in the number of factors used in the partitioning process is the second flaw. For example, it is infeasible to split firms by ten potential features using their, say, quintiles as thresholds to partition the sample. Even if we could, it is a challenge to present the results in tabular form.

There is a strand of research mainly in the field of econometrics that delves deeper into the heterogeneous group-specific coefficients issue. Goldfeld and Quandt (1973) present the first attempt to use threshold variables to form clusters. The threshold variables are determined as a linear function of several transition variables. Lin and Ng (2012) develop a similar threshold method called two-step Pseudo Threshold approach. Unlike Goldfeld and Quandt (1973), the threshold variables and their corresponding threshold values are identified totally within the regression model without knowledge of the true transition variables. Recently, researchers have started to recognize the usefulness of cluster-alike algorithms to identify clusters. However, their algorithms, which aim to minimize a regression's sum of square residuals (e.g. Lin & Ng,

[11] For more evidence of dissimilar magnitudes in the way key financial ratios predict earnings, see Amor-Tapia and Tascón Fernández (2014), Nunes, Serrasqueiro, and Leitao (2010), Bauman (2014), and Dichev and Tang (2009).

For reviews of evidence of heterogeneous group-specific coefficients in panel data, see Lin and Ng (2012) and Hsiao and Tahmiscioglu (1997). For evidence of heterogeneous industry-specific coefficients, see Burnside (1996), Cohen and Zarowin (2008), Hribar and Collins (2002), and Amor-Tapia and Tascón Fernández (2014).

For evidence of intra-industry spillover effects of capital investment announcements, see Chen, Ho, and Shih (2007). For evidence of heterogeneous industry-specific coefficients, see Burnside (1996), Cohen and Zarowin (2008), Hribar and Collins (2002), and Amor-Tapia and Tascón Fernández (2014). For evidence of intra-industry spillover effects of capital investment announcements, see Chen, Ho, and Shih (2007).

2012) or add a penalty term (e.g. Ando & Bai, 2015), are still merely based on regression analysis while ignoring the richness of information from cluster patterns [12].

For the above reasons, a technique that is able to utilize several partitioning features and splits data (such as firm observations) into meaningful/useful groups could help researchers to gain new insights into the important features that cluster the data, and consequently improve the performance of statistical tests[13]. Clustering, in the form of un-supervised classification assigning objects into unlabeled classes, is such a technique. Aims and shortcomings of cluster analysis are discussed in the next section.

2.2 Cluster analysis and feature weightings

The aim of cluster analysis is to place observations into different clusters such that observations in the same cluster are homogeneous to each other but are different from ones in other groups (Fred & Jain, 2005). K-means clustering is the most popular method in the family of centroid approaches. It is credited with simplicity, low computational resources and high popularity among several clustering methods, and accordingly is employed in this paper as the core technique to explore firm patterns. Among of segmentation studies explored by Dolnicar (2002), it accounts for 37% (68 out of 184) of all the clustering methods used[14].

The ultimate goal of cluster analysis is to discover the true cluster structure. In this regard, choosing relevant features and deciding upon their weights are critical parts to ensure success (Brusco & Cradit, 2001). A number of variable-weighting methods have been proposed and developed. These methods try to find the most appropriate weighting variables in order to eliminate the irrelevant variables and consequently strengthen the cluster results (Brusco & Cradit, 2001). Among the notable variable-weighting methods applied for K-means analysis is 'synthesized clustering' introduced by Desarbo, Carroll, Clark, and Green (1984). Through an iterative fitting process, variable weights are generated using a weighted K-means procedure. However, in a study comparing the performance of various variable-weighting methods, Gnanadesikan, Kettenring, and Tsao (1995) find that the Synthesized Clustering procedure is less effective than simpler methods such as equal-weight scaling, standardization, and range-scaling.

---

[12] An example may clarify this statement. The target of these papers is to assign observations into clusters such that after running a regression within each cluster, the total square of residuals is minimized. To achieve this, Lin and Ng (2012) propose an algorithm which repeatedly assigns observation i to group $g*$ if $SSR_i^{g^*} = \arg min_g \, (y_i - \widehat{y_{ig}})^2$, g=1,…,G, and $\widehat{y_{ig}}$ is the estimation of y_i using coefficients estimated within group g. Now, suppose there are two groups: Group 1 and group 2. These groups are represented by two regression models: $y_i = 0.1 + 0.3X_i + \varepsilon_i$ and $y_i = 0.1 + 0.5X_i + \varepsilon_i$ accordingly. Let A and B represent two observations which belong to the same group, such as group1. X_A=1, X_B=1.5, ε_A=1, ε_B=-1. So, even if the estimated $\widehat{\beta_g}\,, g = 1{,}2$ are correctly estimated (i.e. 0.3 and 0.5 for group 1 and 2 respectively), according to the algorithm, while A is correctly identified to group1, B is not ($SSR_A^{g=1} = 1$; $SSR_A^{g=2} = 1.44$; $SSR_B^{g=1} = 1; SSR_B^{g=1} = 0.49$). This example illustrates that the performance of grouping methods which merely depend on regression analysis is highly sensitive to interactions between the sign and magnitude of error terms and discrepancies of coefficients across clusters. Now cluster analysis shows their power. If observation A and B are close points in the space (as is a usual case) whose dimensions are partitioning features, this meaningful information will be captured by cluster analysis

[13] The purpose of clustering is for either understanding (meaningful clusters) or utility (useful clusters) (Tan, Steinbach, & Kumar, 2005). For the utility purpose, each group (cluster) could be represented by a cluster prototype. Then these prototypes could facilitate the subsequent data analysis or data processing technique such as summarization or compression. In contrast, for understanding, cluster analysis uncovers meaningful groups whose members share common characteristics. These clusters would help us analyze and describe what are the true structures underlying the data.

[14] For interests on K-means algorithm, see Amorim and Mirkin (2012)

A number of variable-weighting methods have been proposed and developed. Among the notable variable-weighting methods applied for K-means analysis is 'synthesized clustering' introduced by Desarbo, Carroll, Clark, and Green (1984). Through an iterative fitting process, variable weights are generated using a weighted K-means procedure. However, in a study comparing the performance of various variable-weighting methods, Gnanadesikan, Kettenring, and Tsao (1995) find that the Synthesized Clustering procedure is less effective than simpler methods such as equal-weight scaling, standardization, and range-scaling.

A recent effort on weighted K-means is from Amorim and Mirkin (2012) who introduce the so called Intelligent Minkowski metric Weighted K-Means (iMWK-Means, for short). This is a closed form algorithm which is analogous to that of Huang et al. (2008) with an adjustment to the distance formulae. Particularly, instead of using the Euclidean metric in the criterion, they utilize the Minkowski metric and sketch out the searching procedure for Minkowski centers as a process of minimization of a convex function. By simulation, iMWK is shown to outperform both K-means and weighted K-means.

However, there are still important deficiencies of iMWK. First, the criterion used to derive optimal weights is totally drawn from the clustering itself. Put another way, its objective is to minimize within cluster distances (measured by Minkowski metric) given that the weights are supposed to be non-negative and sum to unity[15]. The objective is intuitive, but given that it is internally-derived, it fails to define the exponent parameter (β) within the model. The exponent parameter is instead user-defined before running the clustering procedure. Consequently, an optimal β is only identified through the supervised or semi-supervised process (Amorim & Mirkin, 2012).

Second, iMWK only address the problem of noise or irrelevant features, and in their simulated data, each of the clusters is spherical. Problems associated with elongated clusters or correlated dimensions are not considered. Finally, optimal weights as estimated by iMWK do not necessarily coincide with the weights that best improve the regression analysis. As a result, it should be regression analysis which provides the ultimate criteria to guide the cluster analysis and adjust the weights of features, not the internal target of the clustering itself.

In summary, variable weighting is still a challenging issue and processes need to be refined further in order to strengthen clustering performance. The next section will discuss our novel clustering method which is proposed to address problems of regression (i.e. heterogeneous group-specific coefficients) and clustering (i.e. feature's weightings) simultaneously.

3. New approach for K-means: Regression oriented Weighted K-means

This section presents the econometric framework. Subsequently, a procedure to implement ROWK is introduced. Then, it proposes hypotheses which will be tested in the next section.

3.1 Models

Let *i*=1,…, *N* representing an index of observations. For simplicity, the paper only considers the case of a cross-section data. For panel data, nothing changes except that "*i*" is replaced by

[15] Particularly, it minimizes $J = \sum_{k=1}^{K}\sum_{v=1}^{V}\sum_{x_i \in \xi_k} \| w_v x_{iv} - w_v c_{kv} \|^{\beta}$ where K is number of clusters; set of V features v, and x_{iv}, c_{kv} are the value of feature v at entity i and centroids $k \in \xi_k$ accordingly. w_v denote feature weights. The exponent β is a pre-defined parameter presenting the rate of effect of the weights on its contribution to the distance.

"*it*" where *t* is an index of time[16]. There is a multivariate input data set *X* that is represented as an *N* x *V* matrix, where *V* are the number of dimensions of clustering data or the number of cluster features. $x_i = (x_{i1}, x_{i2} \dots x_{iV})^T$, $i=1,\dots,N$. Let K^0 represents the true number of clusters (which is unknown and fixed). Denote $\xi_1^0, \xi_2^0, \dots, \xi_K^0$ as the corresponding true K^0 clusters with centers $c_1, \dots, c_{K^0}$ respectively where $c_k = (c_{k1}, c_{k2} \dots c_{kV})^T$, and $k=1,\dots, K^0$. Let N_k^0 be the true number of cross-sectional units within group *k* (*k*=1,…, *K*) so that $N = \sum_{k=1}^{K} N_k^0$. The response variable of the *i*th unit, y_i is expressed as

$$y_i = \alpha_{(i)} + z_i' \beta_{(i)} + \varepsilon_i, i = 1, \dots, N \qquad (1)$$

where z_i is a *P*x1 vector of explanatory variables and ε_i is the unit-specific error. $\alpha_{(i)} = (\alpha_{(i)1}, \dots, \alpha_{(i)p})'$ and $\beta_{(i)} = (\beta_{(i)1}, \dots, \beta_{(i)p})'$ are *P* x 1vectors of intercepts and slope coefficients for unit *i* respectively. A group effect is modelled by allowing $\alpha_{\xi_k^0} = (\alpha_{\xi_k^0,1}, \dots, \alpha_{\xi_k^0,p})'$ and $\beta_{\xi_k^0} = (\beta_{\xi_k^0,1}, \dots, \beta_{\xi_k^0,p})'$ be *P* x 1vectors of group-specific intercept and slope coefficients such that $\alpha_{(i)}$ and $\beta_{(i)}$ equal or closely approximate $\alpha_{\xi_k^0}$ and $\beta_{\xi_k^0}$ respectively for all *i*'s in ξ_k^0 [17].Further assume that observations exhibit similar characteristics to others within the same group. Each characteristic is represented by one dimension (feature), i.e. x_{iv}.

The following assumptions will be made: (i) $\varepsilon_i \sim (0, \sigma^2)$ has finite fourth moments and has cross-sectional and serial independence, i.e. $\varepsilon_{ij} = \sigma^2 I$ where *I* is the identify matrix; (ii) 0< σ^2<∞ and (iii) ε_i is independent of z_i' for all *k*=1,…, K^0. Our objective is to estimate $\beta_{\xi_k^0}$ (and $\alpha_{\xi_k^0}$) without knowing ξ_k^0. This can be achieved by the proposed ROWK procedure.

3.2 The Regression oriented Weighted K-means (ROWK)

The ROWK procedure includes three steps. First, the regression model is identified. Second, features are selected. Finally, optimal weights are identified.

3.2.1 Specifying the regression model

The regression model is represented as in equation 1. Assumptions (i), (ii), and (iii) above imply that the model is correctly specified and can be consistently estimated within each true cluster.

3.2.2 Feature selection

This paper argues that for robust clustering, the selection process should strongly connect with underlying theory. Features selected as inputs for clustering have to be characteristics that contribute to the distinct behavior of clusters (Ketchen & Shook, 1996). Furthermore, since the ROWK procedure aims to place low or zero weights on irrelevant variables, the problem of the inductive approaches could be mitigated[18]. For the next part, assume that throughout the feature

---

[16] Note that the threshold approach is only executed on panel data because it has to run individual time-series regression. Our framework can apply for both cross-sectional and panel data, so it can be applied in case of unavailability of individual time-series. Additionally, it also allows for the case that a firm *i* can move to different clusters overtime.

[17] We allow for both the intercept and slope coefficients to be group-specifics. For panel data with unobserved heterogeneity (α_i), we can transform the original data into demeaned data (i.e. $\overline{y_{it}} = y_{it} - \frac{1}{T}\sum_{t=1}^{N} y_{it}$ and $\overline{z_{it}} = z_{it} - \frac{1}{T}\sum_{t=1}^{N} z_{it}$). Then we have a model with no intercept and only group-specific slope coefficients.

[18] Feature selection can be based on two different approaches. Enhancement of the selection process can be achieved by the deductive approach which emphasizes a strong link between the selection process and theory, leading to a priori expectations regarding the employed variables and the nature of the clusters (Ketchen & Shook, 1996). The second approach is an inductive process which does not require any such a priori expectations, resulting in employment of as many variables as possible (Epure et al., 2011). The latter approach can cause problems of irrelevant variables and high dimensionalities.

selection process, there are *V* cluster features[19]. The input variables for the clustering process will be represented by an *N* x *V* matrix, $x_i = (x_{i1}, x_{i2} \dots x_{iV})^T$, $i=1,\dots,N$.

3.2.3 Optimal weights

Let's w_v $(v = 1, \dots, V)$ denote the corresponding weight of feature *v*. Then, the weighted K-means clustering attempts to assign *N* data points into *K* disjoint clusters such that the sum-of-squares criterion, J, is minimized:

$$J = \sum_{k=1}^{K} \sum_{v=1}^{V} \sum_{X_i \in \xi_k} w_v{}^{\gamma} \| x_{iv} - c_{kv} \|^{\theta} \quad \text{subject to } w_v \geq 0 \text{ and } \sum_{v=1}^{V} w_v = 1.$$

where γ is the parameter presenting the rate of effect of the weights on its contribution to the distance. ‖.‖ is a norm. In this paper, the norm is chosen as the Euclidean metric with θ=2 and γ=1. The main innovation of ROWK is that it does not seek the set of optimal weights $\{w_1^*, \dots, w_V^*\}$ which minimizes the function J as in previous research on weighted K-means. Instead, it seeks to identify a set of $w_v^*, v = 1, \dots, V$ such that when applying these weights to K-means clustering, finding clusters, and regressing using equation 1, the absolute sum of the residuals is minimized[20]:

$$SAE = \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

where $\hat{y}_i$ is the estimation of y_i. Applying the weights into K-means implies that features are rescaled based on the squared root of corresponding weights. For the algorithm to find the optimal weights, see Appendix A.

To identify the true number of clusters, i.e. K^0, this paper uses two approaches. The first approach is an informal approach which graphs the values of the sum of absolute residuals for a given *k* against *k*. Then, the chosen *k* is the 'knee point' at which the graph starts to flatten. The second approach is to use the modified BIC criterion as in the work of Lin and Ng (2012):

$$BIC(k) = \log\left(\sum_{i=1}^{N} (y_i - \hat{y}_i)^2\right) + k(P+1)\frac{\log(N)}{N} + (k-1)\frac{\log(N^2)}{N^2}$$

where $\hat{y}_i$is estimated y_i; *k* is number of clusters, *P* is number of regressors, and *N* is the number of observations. The number of clusters with the least modified BIC is chosen.

3.3 Computation time issues

Basically, the ROWK procedure repeatedly runs the K-means algorithm with different sets of feature weights. Hence, the computation time of this procedure depends on (1) the time to run each K-means algorithm, and (2) the number of K-means algorithms to run. The time to run each K-mean algorithm is proportional to the product of the number of observations (*N*) and the dimension of the variable space (*V*). Adding each features, one by one, and running steps 2.1 to 2.4 drastically could slow down the running time, especially in case of high dimensions. An alternative option is to pick all features at the same time and run steps 2.1 to 2.4. However, while simulation results show that it easily reaches the local optimal, the result is much worse compared to the approach of adding each feature one by one.

3.4 Hypotheses

[19] For simplicity, this paper assumes that after the feature selection process, exactly *V* features are identified. In real cases, the number of features that are used in clustering tends to greater than *V*. Our assumption is acceptable and is less harmful than the case of omitting relevant features because ROWK is built to place low or zero weights on irrelevant features.

[20] This paper uses the sum of absolute residuals instead of the sum of squared residuals to mitigate the effect of outliers. In simulation results, it moderately improves the performance of ROWK.

ROWK procedure employs regression mean absolute (square) residuals as a guide to adjust the features' weights when there exist differences between degrees of contribution of clustering features to identify clusters. As a result, the weight of a feature reflects its importance not only to cluster identification, but also to improve the regression analysis. For example, suppose there is a regression model with five group-specific coefficients. Among these five clusters, only regression coefficients in cluster 1 significantly differ from those of other clusters. Further assume that there are five relevant features for clustering, say $X_1, \dots, X_5$. Among them, X_1 and X_2 are more relevant to distinguish the clusters and when running through the weighted K-means algorithm, they receive higher weights, say $w_1 = w_2$, and $w_1, w_2 > w_3, w_4, w_5$. Additionally, while X_1 provides relevant information to distinguish between cluster 1 and the rest of other four clusters, X_2 helps to distinguish between all clusters, except cluster 1. It turns out that ROWK will put more weight on X_1 than X_2 because its objective is to minimize the sum of absolute (squared) residuals. This is rational because the ROWK procedure's ultimate goal is to improve regression results through clustering.

Hypothesis: *When features have different degree of contribution to cluster identification and regression estimation, ROWK outperforms generic K-means (both standardized and un-standardized) with regard to precision of cluster recognition and regression estimation. The mechanisms underlying the outperformance of ROWK are through these channels. Specifically, ROWK:*

i, Places more (less) weights on more (less) relevant features.

ii, Reduces the influence of the multi-collinearity problem by reducing the weights of irrelevant features which are highly correlated with relevant features.

iii, In the ROWK context, relevance is captured not only by contribution to cluster recognition but also by regression estimation.

4. Simulation

To test proposed hypotheses, different sets of simulated parameters are used, and are described immediately before the results in the next section. The set of parameters includes[21]:

- *N*: number of observations; N_k^0: number of members of cluster k ; K^0: number of cluster; *P*: number of independent variables; *V*; number of features;
- λ : extent of differences between cluster 1 and other clusters;
- $\sigma_{k,v}^2$: level of density of cluster *k* with regard to feature *v*. It depends on two parameters, i.e. w_{den_v} and θ;
- $\sum_{v1v2} = cov(\epsilon_{i,v1}^k, \epsilon_{i,v2}^k), v1, v2 = 1, \dots, V$: within covariance matrix of features. This is assumed to be the same across clusters;
- $w_v, v = 1, \dots, V$: weights of clustering features; $\sum_{v=1}^V w_v = 1$;
- $Z_{i,p} \sim N(0, \sigma_{i,p}^2), p = 1, \dots, P$: independent variables;
- $\alpha_{(k)}$ *and* $\beta_{(k)}$: group-specific intercept and slope coefficients of regression models respectively;
- ε_i: unit-specific error.

5. Performance of the proposed ROWK in simulated data

Three simulated cases is used to test part i), ii), and iii) of the Hypothesis. Each case is created to shed light on channels leading to the outperformance of ROWK as compared to K-means. Case 1 includes a simple set of simulated data with uncorrelated features. Case 2 employs the same data as Case 1, but with correlated features. Case 3 analyzes a situation where features'

[21] See Appendix B for descriptions and formulae of simulated parameters

weights come from two sources, i.e. contributions to cluster recognition and to regression estimations.

For clustering validation, this paper uses the purity index, a measure of precision of assigning entities to clusters. From this section onward, whenever the term 'class' is used, it indicates the true cluster. Let K' and K^0 equal the number of clusters identified through clustering and the number of true clusters (classes), respectively. Also, $p_{k'k^o}$ is denoted as the probability that a member of cluster k' belongs to class k^0, $p_{k'k^o} = m_{k'k^o}/m_{k'}$, and $\overline{p_{k'k^o}}$ is the probability that a member of class k^0 belongs to cluster k as $\overline{p_{k'k^o}} = m_{k'k^o}/\overline{m_{k^o}}$, where $m_{k'}$ ($\overline{m_{k^o}}$) is the number of entities in cluster *i* (class *j*) and $m_{k'k^0}$ is the number of entities of cluster k' in class k^0. Then, the purity index of class k^0 and all classes are calculated as $p_{k^0} = \max_{k'} p_{k'k^0}$ and $purity = \sum_{k^0}^{K^0} \frac{\overline{m_{k^0}}}{N} p_{k^0}$, respectively.

For each set of parameters, 100 simulated data samples are generated. Then K-means with unstandardized features is run using these 100 samples. The average of class purities and mean squared residuals (MSE) are calculated[22]. T-tests are used to test for significant differences of means.

5.1 Case 1

Panel A of Table 1 displays simulated parameters for Case 1. There are 5000 observations which belong to five classes. Each class has 1000 members. There are four features, X_i, i=1,…,4 and a random variable (~*N*(0,1)) which is used as an irrelevant clustering feature. For simplicity, only X_3 has more weights relative to others. For the regression model, there are two independent variables Z_1 and Z_2 which are also features of clustering, i.e. $Z_1 = X_1$ and $Z_2 = X_2$. Z_1 and Z_2 satisfy assumptions in section 2.3.1.

Panel B of Table 1 presents descriptive statistics of clustering features for Case 1. Standard deviations of the four features are similar ranging from 1.49 for X_2 to 1.74 for X_3. Panel C of Table 1 exhibits the whole-sample (lower triangle) and within class 1 (upper triangle) correlation matrix of clustering features. While, all pairs of features, as expected, display no significant correlation within class 1, there are significant correlations (five out of six pairs of correlations) between features for the whole sample.

[Insert Table 1 here]

The first step of the ROWK procedure relates to ranking feature based on regression MSEs calculated by running K-means only for each feature. The results of feature ranking at K^{max}=10 are presented in Figure 1 [23]. As expected, X_3 has lowest MSEs (0.837), and accordingly ranks first. Noticeably, X_2, which is generated as a clustering feature, has a lower ranking (5th) than X_5, a random feature (ranked 4th). Panel A of Table 2 presents the results of ROWK at K^{max}=10 for each number of features. Panel B of Table 2 exhibits the optimal weights at different numbers of clusters. Using modified BICs (not reported) and the graph in Figure 2, the optimal number of clusters is found to be five. Accordingly, the resulted optimal weights are (X_1, X_2, X_3, X_4, X_5) = (1.05, 1.1, 2.145, 1.2, 0). Hence, using ROWK procedure, the most important feature (i.e. X_3) receives the highest weight. In contrast, the irrelevant (random) feature X_5

[22] For the sake of conciseness, only the mean squared residuals are presented when testing the hypotheses. The findings remain unchanged when the mean absolute residuals is used.

[23] Assume that the possible maximum number of clusters is 10.

receives no weight after running ROWK. This evidence supports parts i) of the hypothesis which posit that the mechanisms underlying the outperformance of ROWK are through placing more (less) weights on more (less) relevant features.

[Insert Figure 1, Table 2 and Figure 2 here]

Panel A of Table 3 compares the performance of ROWK relative other methods. When using optimal weights, the MAE is significantly lower (0.8224) compared to methods using whole-sample regression (0.9513), standardized K-means (0.8455) and unstandardized K-means (0.8399). With respect to class purity, 71.3% of members are correctly assigned. In contrast when using typical standardized K-means, only 50.8% members are correctly assigned. For robustness, 100 simulated data are generated with the same set of parameters. Then, K-means is run using the optimal weights found in Case 1, and the averages are calculated. Panel B presents the performance of ROWK for the 100 out-of-sample data. Consistent with the in-sample results, the performance of ROWK is superior as compared to other methods.

[Insert Table 3 here]

5.2 Case 2

Next, simulated data are generated with the same set of simulated parameters as in Case 1 with an exception that features are within-class correlated. Recall that in Case 1, X_4 has the second-highest weight (i.e. 1.2). In Case 2, X_4 is generated to be strongly positively correlated with X_3, the most relevant feature. Panel A of Table 4 documents the simulated parameters which are basically identical to those of Case 1. The difference between Case 2 and Case 1 is apparent in Panel B. The within-class 1 correlation between X_3 and X_4 is highly positive (ρ=0.537), while other within-class correlations are still insignificantly different from zero.

[Insert Table 4 here]

Given the high correlation between X_3, the most important feature, and X_4, it is expected that when running K-means using X_4 alone, the MSE will be lower than that of Case 1. Figure 3 supports this statement (i.e. 0.929 vs. 0.925 for Case 1 and Case 2, respectively). Table 5 exhibits the optimal weights for each different number of clusters. Using both modified BICs and the graph (not reported) the optimal number of clusters is again found to be five. Accordingly, the resulting optimal weights are (X_1, X_2, X_3, X_4, X_5) = (1.5, 0.6, 2.64, 0.277, 0). Recall in the Case 1 set of optimal weights, ROWK assigns the second highest weight (1.2) to X_4. In Case 2, however, X_4 is strongly correlated with X_3, so ROWK addresses this problem to mitigate the effect of multicollinearity by lessening the weight of X4 (i.e. 0.277). This evidence is consistent with part *ii*) of the Hypothesis which states that ROWK reduces the influence of the multi-collinearity problem by reducing the weights of irrelevant features which are highly correlated with relevant features.

[Insert Figure 3 and Table 5 here]

The performance of ROWK (not reported here) relative to other methods is similar to Case 1. Using optimal weights, MAE is significantly lower (0.828) than those of whole-sample regression (0.9513), standardized K-means (0.862) and unstandardized K-means (0.86). Note that relative to the case of uncorrelated features, the performance of K-means (both standardized and unstandardized), is significantly worse when features are correlated. These findings are consistent with the argument that K-means does not address the problem of multicollinearity. In contrast, the performance of ROWK is unimpaired with or without multicollinearity. For example, the difference in MAE between Case 1 and Case 2 is only 0.006, and statistically is no different from zero. Results from out-of-sample data remain unchanged.

5.3 Case 3

For Case 3, some adjustments of simulated parameters are made as follows. X_3 is generated to be a highly relevant feature to recognize class 1's membership, and X_4 is simulated to provide rich information to distinguish membership of other classes. Clustering features are generated to be uncorrelated. The regression coefficients for Z_2 are also adjusted to correctly reflect relative positions between classes[24]. Unlike Case 1 and 2, in Case 3, Z_1 and Z_2 are not set to be clustering features.

Figure 4 presents the results of the feature ranking at K^{max}=10. Given X_3 is created as the relevant feature to identify class 1 (i.e.as the most outstanding class), when running K-means for each individual feature, X_3 has the lowest MSE (0.894), and accordingly ranks first. Although X_4 is generated to contribute to cluster recognition in the same degree as X_3, it ranks second. This is due to the fact that its contribution only recognizes classes other than class 1; hence it does not help to reduce MAE as much as X_3 does.

[Insert Figure 4 here]

Table 6 presents the optimal weights for each different number of clusters. Using modified BICs (not reported) and the graph in Figure 5, the optimal number of clusters is identified as four. Accordingly, the resulting optimal weights are (X_1, X_2, X_3, X_4, X_5) = (0.285, 0, 0.556, 0.47, 0). Consistent with part *iii)* of the Hypothesis, ROWK assigns a higher weight to X_3 than to X_4. The optimal weights that minimize a regression's MAE reflect its importance not only to cluster identification, but also to improving the regression analysis. As a robustness check (untabulated), 100 simulated samples are created with the same set of parameters and the MAEs are calculated for two sets of weights: one with the above set of optimal weights (0.285, 0, 0.556, 0.47, 0) and the other with equal weights of X_3 and X_4 (0.285, 0, 0.556, 0.556, 0). The results show that the average MSE for the optimal set of weights is significant lower than the set with equal X_3 and X_4 weights. The performances of ROWK (untabulated) relative to other methods are similar to Case 1 and Case 2.

[Insert Figure 5 and Table 6 here]

6. Conclusion and future work

This paper proposes a novel clustering procedure that connects clustering and regression analysis in order to address problems of feature weightings associated with K-means and issues of regression estimations with group-specific coefficients. The new regression oriented-weighted K-means (ROWK) procedure employs regression mean absolute (square) residuals to guide the adjustment of the features' weights when differences exist between degrees of contribution of clustering features to identify clusters.

Simulation results demonstrate that proposed method successfully mitigates the effect of highly correlated variables. In addition, features' weights identified by ROWK reflect features' contributions to not only cluster recognition but also regression estimation. Consequently,

[24] See Appendix C for detail of distances in classes' centers.

ROWK correctly place more weight on more relevant features and lower weight on less relevant features. More importantly, the results from clustering can successfully be used to resolve the problem of heterogeneous group-specific coefficients in regression estimations, leading to lower MSE and consistent coefficient estimates.

The problem of heterogeneous group-specific coefficients in regression estimations is well-documented in finance research (Lin & Ng, 2012). Given the superior performance of ROWK as demonstrated above, future finance research may benefit from the application of the ROWK procedure whenever there are suspicions that regression coefficients are group-specific. Some potential applications of ROWK in finance are proposed as follows.

Research efforts to develop a robust approach to measure firm life cycle stages are sparse and constrained to simple identities (Dickinson, 2011). For example, firms falling into the same phase of life cycle are likely to have the same age and size, hence these are common proxies for life cycle. Dickinson (2011) further employs cash flow patterns to proxy for firm life cycle. This paper conjectures that firms exhibit some similarities as they evolve across their life cycles. Moreover, regression estimations relating to earnings forecasts based on firm life cycles exhibit group-specific coefficients. Therefore, cluster analysis, and specifically ROWK may be a potential solution to identify the stages within firm life cycles.

Discretionary accruals estimates could also be improved by the use of ROWK. Researchers typically estimate discretionary accruals by running a regression of non-discretionary accruals within each industry. However, this practice generally leads to imprecise estimates due to the different relations between the dependent variable and its determinants within each industry (Fairfield et al., 2003). Therefore, ROWK can be used to combine the model of non-discretionary accruals and weighted K-means to identify clusters where the coefficients in the regression model are homogeneous.

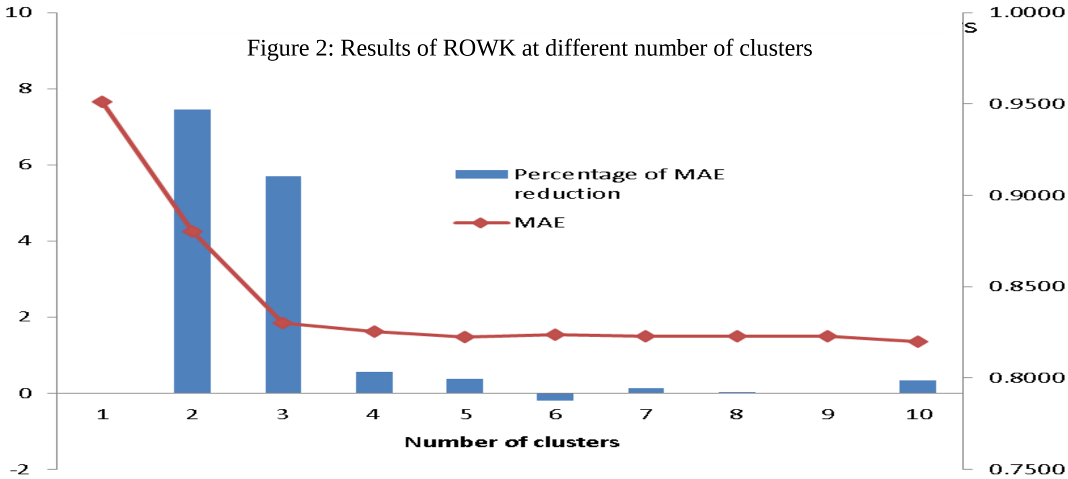


Figure 2: Results of ROWK at different number of clusters

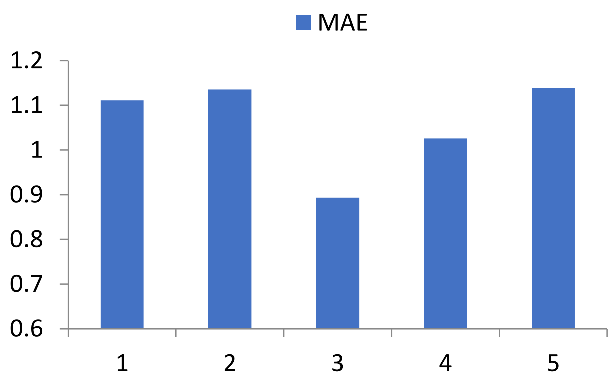
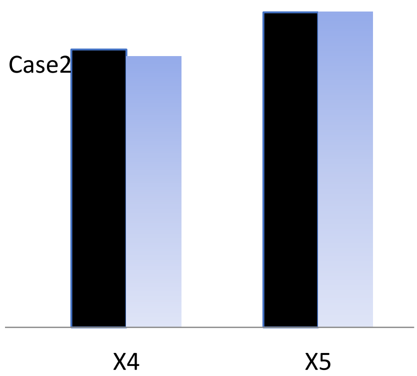


Figure 4: MSEs for each Feature at K'=10 (Case 3)



(case 2.vs case 1)

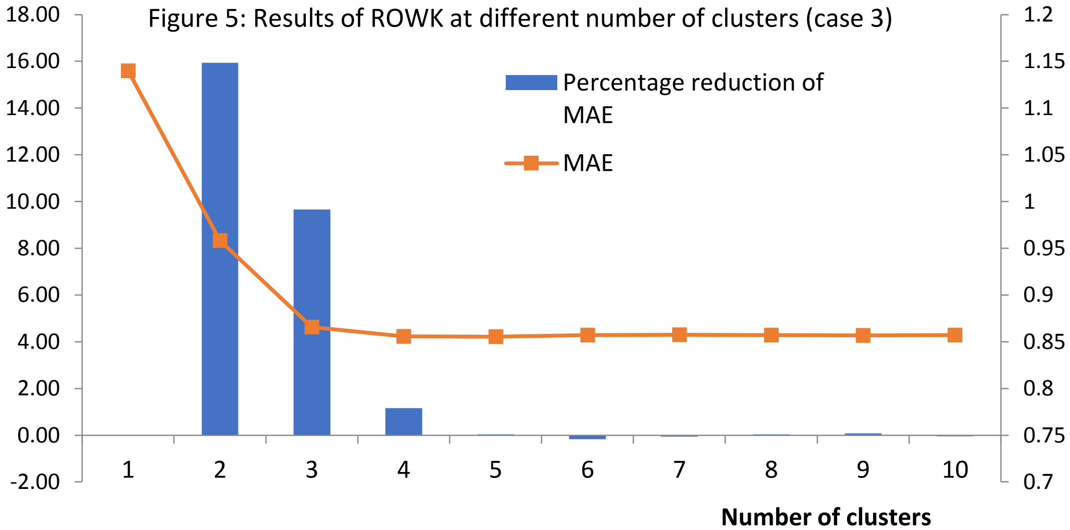


Figure 5: Results of ROWK at different number of clusters (case 3)

Table 1: Descriptive statistics of clustering features (case 1)

– **Regression model:**

$\boldsymbol{y_i = \alpha_{\xi_k^0} + z_i' \beta_{\xi_k^0} + \varepsilon_i,\ i = 1, \dots, N; k = 1, \dots, 5}$ (equation 5)

where $\boldsymbol{\alpha_{\xi_k^0}}$ and $\boldsymbol{\beta_{\xi_k^0}}$ are 2 x 1 vectors of heterogeneity group-specific intercepts and slope coefficients for unit i respectively. $\boldsymbol{\varepsilon_i \sim N(0,1)}$, $\boldsymbol{cov(\varepsilon_i, \varepsilon_j) = I}$ (the identity matrix) and $\boldsymbol{cov\ (\varepsilon, z') = 0}$. $\boldsymbol{\alpha_{\xi_k^0} = 1}$, and $\boldsymbol{\beta_{\xi_{k,1}^0} = 0.5}$. For all k=1,..,5. Each class has 1000 observations ($\boldsymbol{N_k^0 = 1000}$).

– **Clusters' membership:**

There are 4 features $\boldsymbol{X_v, v = 1, \dots, 4}$ and a random variable $\boldsymbol{X_5 \sim N(0,1)}$. $\boldsymbol{X_{i,v}^k = (c_{k,v} + \epsilon_{i,v}^k), v = 1, \dots, 4}$ where $\boldsymbol{c_{k,v}}$ denotes the center of class *k* measured by feature v, $\boldsymbol{\epsilon_{i,v}^k \sim N(0, \sigma_{k,v}^2)}$ where $\boldsymbol{\sigma_{k,v}^2 = w_{k,den\_v}^2 \frac{\theta_k^2}{\sum_{v=1}^{V} w_{k,den\_v}^2}}$; $\boldsymbol{cov(\epsilon_{i,v1}^k, \epsilon_{i,v2}^k) = 0\ if\ v1 \neq v2}$. $\boldsymbol{d_{C_{1j}} > \lambda d_{C_{ij}}, i \neq 1}$ where $\boldsymbol{d_{C_{ij}}}$ is distance between class i and class j; $\boldsymbol{d_{C_{ij}}^3 > \lambda_3 d_{C_{ij}}^v, v \neq 3}$ where $\boldsymbol{d_{C_{ij}}^v}$ denotes the distance between class i and class j as measured by feature *v*. $Z_1 = X_1$ and $Z_2 = X_2$.

Panel A: Simulated parameters

N=5000, $\boldsymbol{N_k^0 = 1000, k = 1, \dots, 5}$, P=2, V=5

$\boldsymbol{w_{k,den\_3} = 0.5}$; $\boldsymbol{w_{k,den\_v} = 1, v \neq 3}$; $\boldsymbol{\lambda = \lambda_3 = 0.9}$, $\boldsymbol{\theta = 2.5}$

$\boldsymbol{\alpha_{\xi_k^0} = 1, \beta_{\xi_{k,Z2}^0} = 0.5, \beta_{\xi_{1,Z1}^0} = -1, \beta_{\xi_{2,Z1}^0} = 0.5, \beta_{\xi_{3,Z1}^0} = 0.2, \beta_{\xi_{4,Z1}^0} = 0.1, \beta_{\xi_{5,Z1}^0} = 0}$

Panel B: Summary statistics of clustering features

	Mean	Std	Skewness	Percentiles		
				1%	50%	99%
X1	0.2317	1.5167	-0.0416	3.7233	0.2366	-3.3355
X2	0.4127	1.4939	-0.0715	3.8130	0.4445	-3.1268
X3	-0.1668	1.7429	0.4557	3.6322	-0.5002	-3.1200
X4	0.2100	1.5855	0.0622	4.0103	0.1836	-3.3534
X5	0.0044	1.0160	-0.0255	2.3451	0.0202	-2.2985

Panel C: Correlation matrix

The upper triangle displays correlations within class 1(N=1000), and the lower triangle displays correlations for whole data sample (N=5000). Correlations with 1% significant level is bolded.

	X1	X2	X3	X4	X5
X1	1.0000	-0.0632	-0.0327	0.0381	0.0522
X2	-0.0070	1.0000	0.01993	-0.0095	0.0123
X3	**-0.0766**	**0.1703**	1.0000	-0.0316	0.0083
X4	**0.0380**	**0.0893**	**0.3372**	1.0000	0.0183
X5	0.0097	0.0042	0.0008	-0.0003	1.0000

Table 2: Optimal weights by ROWK procedures-Case1

Panel A: ROWK results at K^{max}=10 for each number of features (j=1,…,5)						
No. of features (*j*)	MAE	X1	X2	X3	X4	X5
1	0.8375	0	0	1	0	0
2	0.8323	0	0	1.8	1	0
3	0.8238	1.2	0	1.8	1	0
4	0.8208	1.2	0	1.8	1	0.0741
5	**0.8201**	**1.2**	**0.5**	**1.6364**	**1**	**0**
Panel B: ROWK results at each number of cluster k', k=1,…,10						
Number of clusters (k')	MAE	X1	X2	X3	X4	X5
1	0.9513	x	x	x	x	x
2	0.8803	1.4	0	1.1	1	0
3	0.8302	1.3	0.5	2.22	1.1	0
4	0.8254	1.2	1.3	1.9	1.1	0
5	**0.8223**	**1.05**	**1.1**	**2.145**	**1.2**	**0**
6	0.8239	0.91	0.533	1.76	0.867	0
7	0.8228	0.91	0.167	1.76	0.867	0
8	0.8227	1.365	0.12	1.76	0.867	0
9	0.8228	1.365	0.12	1.467	0.867	0
10	0.8201	1.2	0.5	1.64	1	0

Table 3: Performance of ROWK (case 1)

"Dif" denotes differences of MAE between a certain approach and the optimal weights found in ROWK. T-tests are used to test for mean differences. Purity denotes the version 1 overall purity index. Var Y denotes total variance of dependent variables. R_sq denotes the index of model fitness (=1-MSE/Var Y). vs.ideal = (MAE_i – MAE_{ideal})/(MAE_{whole}-MAE_{ideal}). *, **, *** denote significance at 10%, 5% and 1%, respectively.

Panel A: In-sample (N=5000)

	MAE	Dif	X1	X2	X3	X4	X5	Purity	MSE	Var Y	R_sq	vs. ideal
Ideal	0.7837	-	-	-	-	-	-	1.000	0.962		55.53	0.00
Whole	0.9513	0.1289***	1	1	1	1	1	0.200	1.494		30.91	100.00
Optimal	**0.8224**	0.0000	**1.05**	**1.1**	**2.145**	**1.2**	**0**	**0.713**	**1.074**	2.1626	**50.35**	**23.04**
Stand_K-means	0.8455	0.0232***	1	1	1	1	1	0.508	1.160		46.37	36.87
Unstand_K-means	0.8399	0.0176***	1	1	1	1	1	0.586	1.135		47.51	33.53

Panel B: Out-of-sample (N=5000, 100 out-of-sample simulations)

	MAE	Dif	X1	X2	X3	X4	X5	Purity	MSE	Var Y	R_sq	vs. ideal
Ideal	0.7842	-	-	-	-	-	-	1.000	0.963		54.42	0.00
Whole	0.9416	0.1233***	1	1	1	1	1	0.200	1.454		31.19	100.00
Optimal	**0.8183**	0.0000	**1.05**	**1.1**	**2.145**	**1.2**	**0**	**0.720**	**1.068**	2.1124	**49.44**	**21.63**
Stand_K-means	0.8490	0.0307***	1	1	1	1	1	0.518	1.175		44.35	41.16
Unstand_K-means	0.8389	0.0206***	1	1	1	1	1	0.621	1.139		46.08	34.72

Table 4: Descriptive statistics of clustering features (case 2)

Regression model and clusters' members are generated to be identical to those of case 1, with an exception that the within-class correlation matrix is not the identity matrix. See panel C Table 8 for details of the correlation matrix.

Panel A: Simulated parameters

N=5000, $\boldsymbol{N_k^0 = 1000, k = 1, \dots, 5}$, P=2, V=5
$\boldsymbol{w_{k,den\_3} = 0.5, w_{k,den\_v} = 1, v \neq 3}, \lambda = \lambda_3 = \boldsymbol{0.9}, \theta = \boldsymbol{2.5}$
$\boldsymbol{\alpha_{\xi_k^0} = 1, \beta_{\xi_{k,Z2}^0} = 0.5, k = 1, \dots, 5, \beta_{\xi_{1,Z1}^0} = -1, \beta_{\xi_{2,Z1}^0} = 0.5, \beta_{\xi_{3,Z1}^0} = 0.2, \beta_{\xi_{4,Z1}^0} = 0.1, \beta_{\xi_{5,Z1}^0} = 0}$

Panel B: Correlation matrix

The upper triangle displays correlations within class 1(N=1000), and the lower triangle displays correlations for whole data sample (N=5000). Correlations with 1% significant level is bolded.

	X1	X2	X3	X4	X5
X1		0.0397	-0.0327	0.0137	0.0522
X2	**0.0794**		0.0166	0.0047	0.0177
X3	**-0.0766**	**0.1700**		**0.5366**	0.0083
X4	0.0285	**0.0743**	**0.5464**		0.0201
X5	0.0097	0.0060	0.0008	-0.0082	

Table 5: Optimal weights by ROWK procedures-Case2

ROWK results at each number of cluster k', k=1,…,10

Number of clusters (k')	MAE	X1	X2	X3	X4	X5
1	0.9513	1	1	1	1	1
2	0.8912	1.4	0.85	1.1	0.8265	0
3	0.8369	1.5	0.7692	2.5385	0.6666	0
4	0.8313	1.3	0.8	2.64	0.2439	0
5	**0.8283**	**1.5**	**0.6**	**2.64**	**0.277**	**0**
6	0.8259	0.95	0.9	1.958	0.1058	0
7	0.8283	1.575	0	2.299	0.1736	0
8	0.8281	0.4	0	1	0.5218	0
9	0.8258	0.75	0	2	1	0
10	0.8244	0.9091	0.1786	2.47	1.2727	0

Table 6: Optimal weights by ROWK procedures-Case 3

The regression model and clusters' members are generated as described in table 5. Simulated parameters are presented as follows:.

N=5000, $N_k^0 = 1000, k = 1, \dots, 5$, P=2, V=5

$w_{k,den\_1} = 0.9$, $w_{k,den\_2} = 1$, $w_{k,den\_3} = 0.5$, $w_{k,den\_4} = 0.6$, $\lambda = \lambda_3 = 0.7$, $\theta = 1.7$

$\alpha_{\xi_k^0}$=1, $\beta_{\xi_{k,Z2}^0} = 0.5$, $\beta_{\xi_{1,Z1}^0} = -1$, $\beta_{\xi_{2,Z1}^0} = 0.6$, $\beta_{\xi_{3,Z1}^0} = 0.2$, $\beta_{\xi_{4,Z1}^0} = 0.4$, $\beta_{\xi_{5,Z1}^0} = 0$

ROWK results at each number of cluster k', k=1,…,10

Number of clusters (k')	MAE	X1	X2	X3	X4	X5
1	1.1400	1	1	1	1	1
2	0.9583	3.619	0	3.3	1.468	0
3	0.8658	0.7	0	1.3	1.1	0
4	**0.8557**	**0.285**	**0**	**0.556**	**0.470**	**0**
5	0.8555	0.542	0	1	0.733	0
6	0.8570	0.95	0	1.5	1	0
7	0.8575	0.75	0	0.917	0.588	0
8	0.8572	0.714	0	1.061	1	0
9	0.8566	0.75	0	2.073	1	0
10	0.8570	0.7	0	2.1	1	0

References

Amor-Tapia, B., & Tascón Fernández, M. T. (2014). Estimation of Future Levels and Changes in Profitability: The Effect of the Relative Position of the Firm in Its Industry and the Operating-Financing Disaggregation. Revista de Contabilidad, 17(1), 30-46.

Amorim, Renato Cordeiro ee, & Mirkin, Boris. (2012). Minkowski Metric, Feature Weighting and Anomalous Cluster Initializing in K-Means Clustering. Pattern Recognition, 45(3), 1061-1075.

Ando, Tomohiro, & Bai, Jushan. (2015). Panel Data Models with Grouped Factor Structure under Unknown Group Membership. Journal of Applied Econometrics, n/a-n/a.

Bauman, Mark P. (2014). Forecasting Operating Profitability with Dupont Analysis: Further Evidence. Review of Accounting and Finance, 13(22), 191-205.

Brusco, M. J., & Cradit, J. D. (2001). A Variable-Selection Heuristic for K-Means Clustering. Psychometrika, 66(2), 249-270

Burnside, C. (1996). Production Function Regressions, Returns to Scale, and Externalities. Journal of Monetary Economics, 37(2), 177-201

Cohen, Daniel A., & Zarowin, Paul. (2008). Accrual-Based and Real Earnings Management Activities around Seasoned Equity Offerings

Cooper, Michael J., Gulen, Huseyin, & Schill, Michael J. (2008). Asset Growth and the Cross-Section of Stock Returns. Journal of Finance, 63(4), 1609-1651.

Chen, Sheng-Syan, Ho, Lan-Chih, & Shih, Yi-Cheng. (2007). Intra-Industry Effects of Corporate Capital Investment Announcements. Financial Management, 36(2), 125-145.

Chiang, Mark Ming-Tso, & Mirkin, Boris. (2010). Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. Journal of Classification, 27(1), 3-40.

Desarbo, W. S., Carroll, J. D., Clark, L. A., & Green, P. E. (1984). Synthesized Clustering- a Method for Amalgamating Alternative Clustering Bases with Different Weighting of Variables Psychometrika, 49(1), 57-78.

Dickinson, V. (2011). Cash Flow Patterns as a Proxy for Firm Life Cycle. Accounting Review, 86(6), 1969-1994.

Dichev, I. D., & Tang, V. W. (2009). Earnings Volatility and Earnings Predictability. Journal of Accounting and Economics, 47(1-2), 160-181.

Dolnicar, Sara. (2002). A Review of Unquestioned Standards in Using Cluster Analysis for Data-Driven Market Segmentation. CD Conference Proceedings of the Australian and New Zealand Marketing Academy Conference 2002.

Epure, Mircea, Kerstens, Kristiaan, & Prior, Diego. (2011). Bank Productivity and Performance Groups: A Decomposition Approach Based Upon the Luenberger Productivity Indicator. European Journal of Operational Research, 211(3), 630-641.

F.Fama, Eugene, & R.French, Kenneth. (1992). The Cross Section of Expected Stock Returns. Journal of Finance, XLVII(2), 427-465.

Fairfield, P. M., Whisenant, J. S., & Yohn, T. L. (2003). Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing. Accounting Review, 78(1), 353-371.

Fama, E. F., & French, K. R. (1993). Common Risk-Factors in the Returns on Stocks and Bonds Journal of Financial Economics, 33(1), 3-56.

Gnanadesikan, R., Kettenring, J. R., & Tsao, S. L. (1995). Weighting and Selection of Variables for Cluster Analysis Journal of Classification, 12(1), 113-136.

Goldfeld, Stephen, & Quandt, Richard. (1973). The Estimation of Structural Shifts by Switching Regressions Annals of Economic and Social Measurement, Volume 2, Number 4 (pp. 475-485): National Bureau of Economic Research, Inc.

Giroud, Xavier, & Mueller, Holger M. (2011). Corporate Governance, Product Market Competition, and Equity Prices. The Journal of Finance, LXVI(2), 563-600.

Hribar, P., & Collins, D. W. (2002). Errors in Estimating Accruals: Implications for Empirical Research. Journal of Accounting Research, 40(1), 105-134.

Hsiao, C., & Tahmiscioglu, A. K. (1997). A Panel Analysis of Liquidity Constraints and Firm Investment. Journal of the American Statistical Association, 92(438), 455-465.

Huang, J.Z., Xu, J., Ng, M., & Ye, Y. (2008). Weighting Method for Feature Selection in K-Means. in: H. Liu, H. Motoda (Eds.), Computational Methods of Feature Selection, Chapman & Hall/CRC, 193-209.

Ketchen, D. J., & Shook, C. L. (1996). The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal, 17(6), 441-458.

Lee, C. K., Lee, Y. K., & Wicks, B. E. (2004). Segmentation of Festival Motivation by Nationality and Satisfaction. Tourism Management, 25(1), 61-70.

Lin, Chang-Ching, & Ng, Serena. (2012). Estimation of Panel Data Models with Parameter Heterogeneity When Group Membership Is Unknown. Journal of Econometric Methods, 1(1).

Little, P. L., Little, B. L., & Coffee, D. (2009). The Du Pont Model: Evaluating Alternative Strategies in the Retail Industry. Academy of Strategic Management Journal, 8, 71-80.

Nissim, Doron, & Penman, Stephen H. (2001). Ratio Analysis and Equity Valuation: From Research to Practice. Review of Accounting Studies, 6, 109-154.

Nunes, Paulo Macas, Serrasqueiro, Zelia Silva, & Leitao, Joao. (2010). Are There Nonlinear Relationships between the Profitability of Portuguese Service Sme and Its Specific Determinants? Service Industries Journal, 30(8), 1313-1341.

Ou, Jane A., & Penman, Stephen H. (1989). Financial Statement Analysis and the Prediction of Stock Return. Journal of Accounting and Economics, 11, 295-329.

Qian, Y. (2006). K-Means Algorithm and Its Application for Clustering Companies Listed in Zhejiang Province. In A. Zanasi, C. A. Brebbia & N. F. F. Ebecken (Eds.), Data Mining Vii: Data, Text and Web Mining and Their Business Applications (Vol. 37, pp. 35-44).

Richardson, Scott, Tuna, Irem, & Wysocki, Peter. (2010). Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances. Journal of Accounting & Economics, 50(2-3), 410-454.

Samuels, J. M. (1965). Size and the Growth of Firms. Review of Economic Studies, 32, 105-112.

Tan, Pang-Ning, Steinbach, Michael, & Kumar, Vipin. (2005). Introduction to Data Mining, (First Edition): Addison-Wesley Longman Publishing Co., Inc.

Appendix

Appendix A: The algorithm for finding optimal weights

Let K' represents the number of clusters used during the clustering process, K^{max} is the suspected maximum number of clusters, and $\xi'_k\,; k = 1, \dots, K'$ is the cluster k identified during the clustering process. Optimal weights of clustering features are estimated through the following algorithm:

For each number of clusters K'=1,…,K^{max}, run:

Step 1: Ranking features.

- Step 1.1: For each feature $X_{i,v}, v = 1, \dots, V; i = 1, \dots, N$,
  - Run K-means with only one feature- $X_{i,v}$, get $\{\xi'_1, \xi'_2, \dots, \xi'_{K'}\}$
  - Run a regression of equation 1 for each $\xi'_k, k = 1, \dots, K'$, and calculate the sum of the absolute residuals: $SAE = \sum_{i=1}^{N} |y_i - \hat{y}_i|$.
- Step 1.2: Rank features based on SAE. The first ranking is the feature having the lowest SAE and vice versa. Without loss of generality, assume $X_{i,1} - rank\ 1^{st}$, $X_{i,2} - rank\ 2^{nd}$, …, $X_{i,V} - rank\ V^{th}$.

Step 2: Finding optimal weights. For the ordered list of clustering features $\{X_{i,1}, \dots, X_{i,V}\}$. Let $w^*_{1,2} = 1$. Repeat for j=2,…, V.

- Step 2.1: Pick the first j features from the list, i.e. X_1,…, X_j; set w_1= $w^*_{1,j}$;w_2=$w^*_{2,j}$; …;w_{j-1}=$w^*_{j-1,j}$; $w_{j,j} = w^*_{j,j} = 1$. Run K-means with this set of $w_v, v = 1, \dots, V$. Save the clusters $\{\xi'_1, \xi'_2, \dots, \xi'_{K'}\}$ then run the equation 1 regression for each cluster, and calculate the sum of the absolute residual, denoted by SAE^*_j.
- Step 2.2: Run K-means with w_1= $w^*_{1,j}$;w_2=$w^*_{2,j}$; …;w_{j-1}=$w^*_{j-1,j}$; $w_{j,j} = w^*_{j,j}$. Save the clusters $\{\xi'_1, \xi'_2, \dots, \xi'_{K'}\}$ and then run the equation 1 regression for each cluster, and calculate the sum of the absolute residual, denoted by SAE^0_j.
- Step 2.3: Denote Δ as the minimum percentage change of weights. Set Δ= 10%. Repeat for η=1,…,j.
  - Fix all weights except w_η, (w_v , $v = 1, \dots, j, and\ v \# \eta$). Set $w_\eta = w^*_{\eta,j}(1+\Delta)$ $and\ w_\eta = w^*_{\eta,j}/(1+\Delta)$. Run K-means with these sets of weights; save sets of clusters$\{\xi'_1, \xi'_2, \dots, \xi'_{K'}\}$. Run the regression in equation 1 for each set of clusters, and calculate corresponding *SAE*s, denoted by $SAEs_{w_\eta}$.
- Step 2.4:
  - If $min\left\{SAE^0_j;\ SAEs_{w_\eta} : \eta = 1, \dots, j\right\} \neq SAE^0_j$ .Lets the weights corresponding to the lowest SAE are $\ddot{w}_{1,j}, \dots, \ddot{w}_{j,j}$, update $w^*_{1,j} = \ddot{w}_{1,j}, \dots, w^*_{j,j} = \ddot{w}_{j,j}$. Then return to step 2.2.

- If $min\left\{SAE_j^0;\ SAEs_{w_\eta} : \eta = 1, \dots, j\right\} = SAE_j^0$ then repeat step 2.3, but replace Δ by $\lambda * \Delta$; $\lambda = 2, \dots, 100$.
  - If at $\lambda = \lambda^* \in [2; 100]$ that $min\left\{SAE_j^0;\ SAEs_{w_\eta} : \eta = 1, \dots, j\right\} \neq SAE^0$, stop at $\lambda = \lambda^*$. Let the weights corresponding to the lowest SAE be $\dddot{w}_{1,J}, \dots, \dddot{w}_{J,J}$, and update $w_{1,j}^* = \dddot{w}_{1,J}, \dots, w_{j,j}^* = \dddot{w}_{J,J}$. Then return to step 2.2.
  - If $min\left\{SAE_j^0;\ SAEs_{w_\eta} : \eta = 1, \dots, j\right\} = SAE_j^0$ for all $\lambda = 2, \dots, 100$, then:
    - If $SAE_j^0 < SAE_j^*$, update $w_{1,\boldsymbol{j+1}}^* = w_{1,j}^*, \dots,\ w_{j,\boldsymbol{j+1}}^* = w_{j,j}^*$.
    - If $SAE_j^0 \geq SAE_j^*$, update $w_{1,\boldsymbol{j+1}}^* = w_{1,j}^*, \dots,\ w_{j-1,\boldsymbol{j+1}}^* = w_{j-1,\boldsymbol{j}}^*;\ w_{j,\boldsymbol{j+1}}^* = 0$.

At the end of this algorithm, for each k ($k = 1, \dots, K^{max}$), a set of optimal weights is identified $\{w_{1,}^*, \dots,\ w_{V,}^*\}$ along with the corresponding set of clusters $\{\xi_1^*, \dots, \xi_K^*\}$.

Appendix B: Descriptions and formulae of simulated parameters

Symbol	Descriptions	Formulae
c_{iv}	cluster centre *i* by feature *v*	
d_{Cij}	the distance between the center of cluster *i* and the center of cluster *j*	$d_{Cij} = \left[\sum_{v=1}^{V} \left(c_{iv} - c_{jv}\right)^2\right]^2$
λ	extent of differences between cluster 1 and other clusters	$d_{C_{1j}} > \lambda\, d_{C_{ij}}, for\ all\ i\#1$
$X_{i,v}^k$	cluster k's members	$X_{i,v}^k = \left(c_{k,v} + \epsilon_{i,v}^k\right), \epsilon_{i,v}^k \sim N(0, \sigma_{k,v}^2)$
$\sigma_{k,v}^2$	the degree of density of cluster *k* measured by feature *v*	$\sigma_{k,v}^2 = w_{k,den\_v}^2 * \frac{\theta_k^2}{\sum_{v=1}^{V} w_{k,den\_v}^2}$
w_{k,den_v}	the degree density of cluster *k* measured by feature *v*	
θ_k^2	the mean squared distances between members of a cluster *k* to its center	

Appendix C: Distances between classes' centers (case 3)

Appendix C

Distances between classes' centers (case 3)

Class	1	2	3	4	5
1	x				
2	3.856	x			
3	2.864	2.906	x		
4	3.655	0.694	2.919	x	
5	2.305	2.035	1.680	2.192	x

How markets will drive the transition to a low carbon economy

Jianlei Han, Faculty of Business and Economics, Macquarie University, North Ryde, New South Wales, Australia

Martina K. Linnenluecke, Faculty of Business and Economics, Macquarie University, North Ryde, New South Wales, Australia

Zheyao (Terry) Pan, Faculty of Business and Economics, Macquarie University, North Ryde, New South Wales, Australia

Tom Smith, Faculty of Business and Economics, Macquarie University, North Ryde, New South Wales, Australia

ABSTRACT

This paper makes the case that markets will now drive the transition to a low carbon economy, especially those with favorable institutional settings. Here, we extend the notion of 'path creation' to the country-level of analysis to map out different pathways for cleantech development within a real options framework and offer a corresponding valuation of cleantech patents. Results from our analysis suggest a significant potential for the development of cleantech patents, particularly if their development and commercialization pathways can be fostered through a supporting institutional environment that promotes innovation and low-carbon development through carbon pricing policies as well as country-level public R&D expenditure and human capital. Our estimates of total wealth creation through the development of cleantech patents range from US$10.16 to US$15.49 trillion dollars with investment growth from US$2.93 to US$3.71 trillion. The paper concludes by outlining implications for firms and policy, and offers suggestions for future research.

Keywords: Low carbon economy; Transition; Clean technology; Patents; Commercialization Markets.

1. Introduction

Scientists have long been concerned with the level and rate of global environmental change caused by human activity. Recent scientific findings show that there is an urgency to undertake immediate action to either slow down or reverse adverse trends on global scales. One key framework, referred to as "planetary boundaries", stipulates that there are nine key boundaries in Earth system processes which should not be transgressed if we want to sustain human life on the planet (Rockström et al., 2009; Steffen et al., 2015). The nine boundaries concern upper limits for climate change, ocean acidification, stratospheric ozone depletion, biochemical flows (e.g., nitrogen and phosphorus), atmospheric aerosol loading, freshwater use, land-system change, biosphere integrity and novel entities (e.g., micro-plastics or other types of pollution). These boundaries are seen as interlinked, such that if one boundary is transgressed, safe levels for other processes could also be under serious risk. Four boundaries (climate change, biosphere integrity, biogeochemical flows, and land-system change) have already been exceeded. The rate of global environmental change and resulting environmental and social impacts have led to calls for a rapid transition to a low-carbon economy and the decarbonization of carbon-intensive sectors (such as energy, transportation) through immediate action and investments in cleaner products and processes on massive scales.[25]

Investments in the development and uptake of new technology are typically fraught with uncertainty. This uncertainty stems from uncertainty about the technology's commercial success, but also from political uncertainty. However, in the case of clean technology development, commentators are now observing a heightened level of international policy action on climate change, and a confluence of policy, business and grassroots support for a low carbon economy (Linnenluecke, Meath, Rekker, Sidhu, & Smith, 2015). The "Paris Agreement", negotiated in December 2015 by 197 Parties to the United National Framework Convention on Climate Change (UNFCCC), sent a clear signal to international markets that a technological transformation with possible breakthroughs in clean technology and related areas such as bio- and geo-sequestration is likely. This global policy action coincides with policy and technological developments that are taking place on national levels. Several highly-developed countries (e.g., Germany) have adopted expansionist strategies, making significant investments in R&D and cleantech to foster technological innovation (Klaassen, Miketa, Larsen, & Sundqvist, 2005). Studies now show a growing number of firms producing cleantech inventions, leading to a rapid growth in cleantech patents (Rudyk, Owens, Volpe, & Ondhowe, 2015).

This paper makes the case that markets will now drive the transition to a low carbon economy, especially those with favorable institutional settings. Previous work on technological breakthroughs – which has largely followed the work of Schumpeter (1934; 1942) – has portrayed innovation as processes of 'creative destruction' and 'creative accumulation' (Pavitt, 1986). Creative destruction has been conceptualized as a process of competence-destroying innovation through which 'disruptive' innovations by entrepreneurial firms make existing products, processes or business models obsolete (Christensen, 1997), while creative

[25] For the climate change boundary, Rockström and colleagues suggest boundary values of 350 parts per million CO_2 (a measure of the concentration of CO_2 in the atmosphere) and 1 W m^{-2} (a measure of the radiative forcing or imbalance in the Earth's energy budget in watts per square meter) above pre-industrial levels, respectively. Since current levels of CO_2 in the atmosphere are already at over 400 ppm, this approach indicates that the continued use of fossil fuel reserves places society at great risks.

accumulation allows incumbent firms to develop competence-enhancing innovations based on existing knowledge (Bergek, Berggren, Magnusson, & Hobday, 2013; Tushman & Anderson, 1986). Waves of discontinuous technological change have occurred before in sectors such as transportation, telecommunication and biotechnology[26], each leading to the demise of 'old' industries and the rise of new ones (Senge & Carstedt, 2001). Much academic work has focused on understanding the actors and mechanisms behind these shifts from old to new regimes and has mapped the trajectories of entrepreneurial and incumbent firms, institutions as well as stakeholders (e.g., Kostoff, Boylan, & Simons, 2004; Rothaermel, 2001; Walsh, 2004). 'Path dependence' has emerged a popular concept in this context – referring to the issue that certain technologies become dominant, locking society into the adoption of certain development pathways over others (e.g., Liebowitz & Margolis, 1995). The issue can be illustrated using the fossil fuel-driven energy system as an example – the system has remained dominant despite known environmental externalities and cost-competitive alternatives (Unruh, 2000). Clean technologies such as solar or wind face the challenge of 'path creation' (i.e., the development of critical mass to become cost-competitive) while prevailing technologies are still dominant (Dijk & Yarime, 2010).

By looking at the role of market activity, the paper answers the call for further research on the role of cumulative investments to achieve a transition to a low-carbon economy (e.g., Busch, Bauer, & Orlitzky, 2016). Estimates suggest that significant cumulative investments in energy supply and energy efficiency are required over the next decades to bring about a transition to a low-carbon future, and range from US$48 trillion to US$53 trillion (International Energy Agency, 2014). Here, we extend the notion of 'path creation' to the country-level of analysis to map out different pathways for cleantech development within a real options framework and offer a corresponding valuation of cleantech patents. Our aim is to provide both a conceptual framework and empirical assessment of how corporate patenting activity is leading to value creation. Real Options theory is a powerful tool for quantifying investment outcomes and investment risk (Copeland & Weiner, 1990), and is an extension of standard financial appraisal methods. Options theory was originally developed in the 1970s for valuing financial options (Black & Scholes, 1973; Merton, 1973). Real Options theory is an extension of Options theory and is primarily used to value flexibility and strategic options (Brennan & Schwartz, 1985; McDonald & Siegel, 1986; McDonald & Siegel, 1985; Titman, 1985). It is particularly useful in the context of this research, as it supports policy analysis and provides insights into how policy decisions relate to investment behavior and strategic decisions (Blyth et al., 2007; Trigeorgis & Reuer, 2017).

The paper is structured as follows. First, we offer a literature review on prior research on the transition to a low carbon economy. Next, we outline our method which covers (1) a Real Options framework to map out pathways for cleantech development and to offer a corresponding valuation of cleantech patents; and (2) a panel regression which examines the impact of country-level economic variables (real GDP, market return, and turnover) and country-level institutional variables on patent intensity. For purposes of the Real Options

[26] Other historical technological breakthroughs are (1) railroads, (2) electricity, (3) automobiles, (4) radio, (5) microelectronics, (6) personal computers, (7) biotechnology, and (8) the Internet. See Harrison Hong, Jose Scheinkman, Wei Xiong "Advisors and asset prices: A model of the origins of bubbles", Journal of Financial Economics 89 (2008) 268– 287.

framework, we use a decision-tree approach with one stage per year. We begin our analysis from the present and extend the time horizon until 2050; populating the Real Options framework with data from a patent dataset of 286,924 granted cleantech patents spanning 85 countries. Cleantech is now a fast-growing patent class, pointing to a significant potential to create the next technological breakthrough. Our analysis estimates that the total wealth created by the development of cleantech patents ranges from US$10.16 to US$15.49 trillion dollars and will involve an additional investment stimulus to the economy from US$2.93 to US$3.71 trillion; however, this wealth creation also comes with a significant amount of investment requirements and investment risk. For purposes of the panel regression, we draw on insights from institutional theory to discuss how factors such as carbon pricing policies (in the form of a carbon tax or Emission Trading Scheme) and well as country-level public R&D expenditure and human capital create key national parameters that support how the cleantech revolution will unfold.[27] We discuss how the Real Options framework applies differently to settings with good institutional support, and offer a valuation under a different (more favorable) institutional setting compared to our initial analysis. The paper concludes by outlining implications for firms and policy, and offer suggestions for future research.

2. Transitioning to a Low Carbon Economy: Literature Review

Supporting research on the transition to a low-carbon economy has been undertaken from a number of disciplinary perspectives. In the organization and management field, researchers have analyzed the costs and strategic benefits of firms implementing mitigation strategies (either on a voluntary basis or in response to regulatory requirements such as a carbon pricing policy) and concluded that competitive benefits result for those firms that can successfully innovate and translate mitigation efforts into a competitive advantage. The organization and management literature has also offered significant contributions for understanding innovation processes (e.g., Cohen & Levinthal, 1990), life-cycles (e.g., Audretsch & Feldman, 1996) and discontinuous technological change (e.g., Tushman & Anderson, 1986) on firm and industry levels. The foundations for many innovation studies can be traced back to Schumpeter (1934; 1942)'s work on how technological innovation drives economic growth, and how waves of discontinuous technological change disrupts existing industries and lead to their 'creative destruction'. These processes can be illustrated using the fossil fuel industry. Since the industrial revolution, this industry underwent a phase of rapid growth driven by the exploitation of finite fossil fuel resources. This phase has been followed by a long conservation stage, during which adaptive change has focused on developing existing technology and increasing efficiency. However, there are now growing signals that fossil fuels will not remain the standard for energy generation due to environmental concerns, resource limits and rapid developments in alternative technologies.

Related work on transitions to low-carbon energy systems has been undertaken in a number of disciplinary areas, for example, in the areas of energy transitions, governance and social change. In the area of energy transitions, an often-articulated perspective is that the degrading state of the environment creates an imperative to alter how society creates and utilizes energy (Araújo, 2014). Given the urgency with which the transition would need to take place to avoid further environmental degradation as well as the scale of the undertaking, the literature has focused on analyzing how different national and regional systems of innovation (i.e., networks

[27] Due to data availability, our panel regression includes 32 out of the 85 countries included in the patent dataset.

of government, industry and other actors) can foster the development, diffusion and uptake of clean technologies (e.g., Tan, 2010) and allow countries to embark on transitions to a low-carbon economy. Cases that have received much interest include the rapid introduction of natural gas in the Netherlands and the UK as a replacement for coal and oil (Verbong & Geels, 2007; Winskel, 2002). More recently, research has also focused extensively on the German 'Energiewende' (German for energy transition) (Hake, Fischer, Venghaus, & Weckenbrock, 2015), the strong position of wind power in the Danish electricity market (Munksgaard & Morthorst, 2008), as well as the pioneering use of bioenergy in Finland and Sweden, hydroelectricity in Norway and geothermal energy in Iceland (Sovacool, 2017). Extensive research has also focused on analyzing the feasibility and market potential of various technologies (Ashina, Fujino, Masui, Ehara, & Hibino, 2012; Elliston, Diesendorf, & MacGill, 2012; Geels, Berkhout, & van Vuuren, 2016)

In the area of governance and social change, planning for transitions (also referred to as 'transitions management') refers to a governance approach towards sustainable development. Transitions management (e.g., Kemp, Loorbach, & Rotmans, 2007; Loorbach, 2010; Loorbach & Rotmans, 2006; Rotmans & Loorbach, 2009) assumes that societies are characterized by complex governance structures, that is, by diversity, heterogeneity, uncertainties about future developments and consequently a limited possibility for the government alone to induce long-term change. Transitions management focuses on the analysis of different actors and on the complexity of relationships between them to understand how long-term sustainability and a transition to a low-carbon society can be achieved. Transitions management identifies four types of governance activities relevant to societal transitions: strategic, tactical, operational, and reflexive activities (Loorbach, 2010). *Strategic (or long-term) activities* include processes of vision development driving the development trajectories of a community. *Tactical (or mid-term) activities* include steering activities such as supporting institutions, programs, funding opportunities, regulations and infrastructure commitments, but also actors who can help to shape transition agendas and coalitions. *Operational (short-term) activities* include so-called experiments – these are innovative projects carried out in business and industry, in politics and in civil society to reach short term goals within five years. Lastly, *reflexive activities* include monitoring, assessments and evaluation of the ongoing policies, and ongoing societal change. In this reflexive type of governance the role of science is important as to provide deep analysis of the processes and dynamics, and to translate it into the respective agendas (Loorbach, 2010).

Policy action can either support the existing technology base (i.e., reinforce the dominance of existing fossil fuel technologies), or facilitate its renewal (i.e., an expansion in cleantech). Research on the transition to a low-carbon economy has extensively discussed the role of institutional factors (as opposed to firm-level capabilities and competencies) in fostering cleantech and low carbon developments (e.g., Aguilera-Caracuel & Ortiz-de-Mandojana, 2013; Bürer & Wüstenhagen, 2009). While there is ample evidence that firms are important primary actors for driving the cleantech revolution, the role of a supporting institutional environment is regarded as equally important to promote innovation and low-carbon development. Institutional theory holds that systems of private and public institutions as well as surrounding organizations shape innovation outcomes through both formal and informal structures (Hoskisson, Eden, Lau, & Wright, 2000). Formal structures include formal regulatory and administrative mechanisms, including policies (e.g., carbon taxes, emissions trading schemes), regulations, laws and codes of conduct (Marcus & Aragon-Correa, 2011) which can either foster or hinder sustainable development. Informal structures include societal and cultural mechanisms such as norms, values or established ways of operating that organizations and individuals adhere to (Bruton, Ahlstrom, & Li, 2010; Scott, 1987). Taken together, these strands of thought indicate that transitions to a low-carbon future will be driven by market interest, but also by an enabling institutional environment. We consequently factor in variables such as carbon pricing and carbon taxes and well as country-level public R&D expenditure into our analysis in the following sections.

3. Method

Part 1 – Real Options Framework

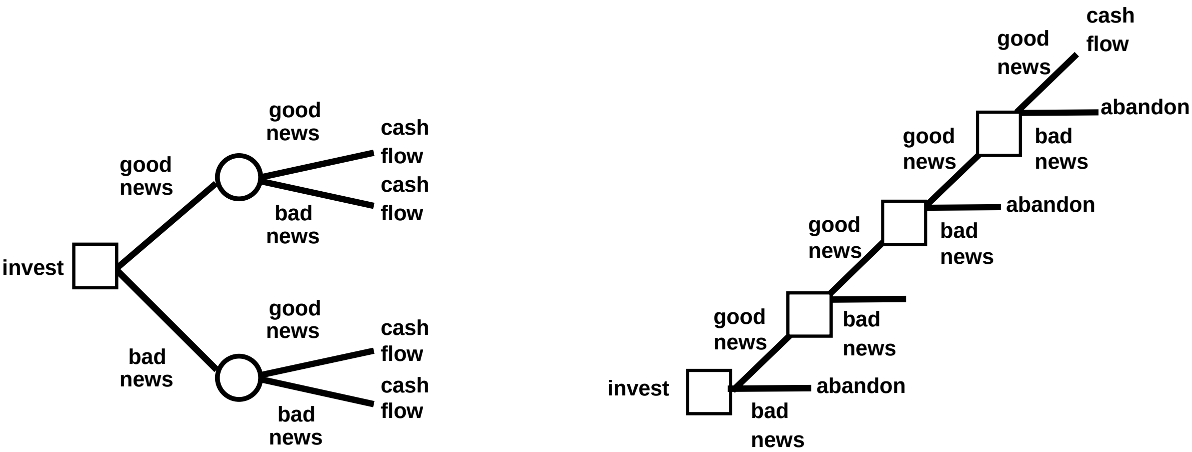
The first part of the analysis builds on a Real Options framework to map out pathways for cleantech development and to offer a corresponding valuation of cleantech patents. Historically, valuation is rooted in the idea of the time value of money. The idea is centered on Fisher's (1930) book "The Theory of Interest". The underlying idea is simple – a dollar today is worth more than a dollar tomorrow. To value any asset, the information needed comprises the asset's cash flows, and the present value of the asset's cash flows. The value of the asset is the sum of the present values. This approach can be used to find the value of any asset including mortgages, stocks, and bonds – in fact all financing alternatives facing the firm - and investment proposals. The Net Present Value (NPV) rule indicates whether a particular investment opportunity is value adding. NPV is calculated as the difference in the present value of cash flows and investment costs. The NPV rule is the basis of other rules currently appearing in the popular press, such as Economic Value Added (EVA), Value Based Management, Economic Profit and Cost Benefit analysis. Diversification is a second key idea, based on the works of Markowitz (1952) and Sharpe (1964), and helps determine the return that an investor should require for any asset, which can be expressed as Required Return = Risk Free Rate + Beta [Market Risk Premium]. It is this required return that is used in calculating present values under the time value of money idea.

One problem in applying net present value analysis to real options is that the possible cash flows and associated probabilities are difficult and complex to specify. NPV does not takes into account the timing of the investment, the uncertainty surrounding the future cash flows from the investment, and the possibility that an investment might be abandoned (i.e., cleantech patents may not be successful). In many traditional type companies (sometimes called Brick and Mortar companies) the use of discounted cash flows does a reasonable job in approximating value. This is because for these companies there is little flexibility and hence the real option value is low. However, under different circumstances, the NPV approach is not sufficient to recognize the value of the options (i.e., business opportunities) created through an investment. In any technology sector, the majority of value is thought to come from real option value (Schwartz & Moon, 2000). Clean technology patents are therefore a case for the application of real options. There are many embedded options involved, including the option to abandon and the option to expand. If a patent shows up as a failure at an early stage, further investments can be abandoned. This effectively truncates losses to the amount of funds spent until this point. If the patent is successful, there is an option to inject further funds and apply for approval for the commercial roll-out of the technology. Finally, when the firm goes to market there is an option to curtail production at an early stage if the new technology proves to be a commercial failure. The optimal use of each of these options effectively truncates the down side exposure and maximizes the up side gains.

One valuation approach is to apply existing option pricing models, such as the Black-Scholes option pricing model (Black & Scholes, 1973). However, there are several problems in applying existing option pricing models to real options. Contrary to model requirements, the present value of the project, the required outlay and the time left before the investment decision must be made may not be certain for the real options. The required outlay and the time left before the investment decision must be made will vary according to the actions of the firm's competitors. Real options may not be proprietary; that is, the firm may not be the only one able to exercise the option, such as the development of microcomputers. One approach to valuing real options is the use of simulation (Schwartz & Moon, 2000). The value of the asset underlying the option is varied and its impact on the value of the project or firm is determined, where the operations of the project are varied as if the firm optimally utilized managerial flexibility. The resultant range represents the value added by the embedded real options. Another approach to valuing real options (used in this paper) is decision-tree analysis which helps to overcome the problems of both simple net present value analysis and the difficulties in applying the option pricing models to real options (Copeland & Weiner, 1990; Kellogg & Charnes, 2000). The application of decision-tree analysis to the valuation of biotechnology is illustrated in Kellogg and Charnes (2000).

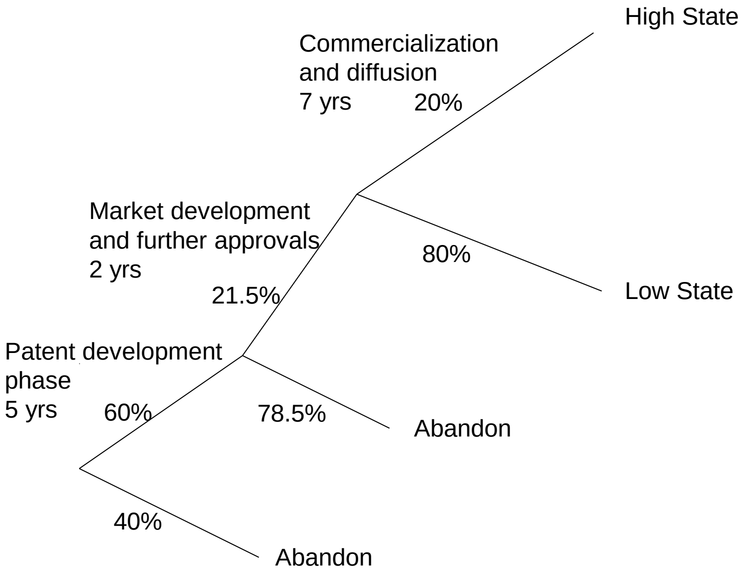
The diagram below shows the embedded options that exist in companies with real options (right side). The existence of abandonment options mean that the distribution of cash flows is truncated on the down side. The presence of expansion options means that the distribution of cash flows is magnified on the up side. In more traditional companies (left side), there is a full distribution of cash flows, both on the up side and on the down side. The difference in cash flows we see with traditional firms and clean technology firms is very similar to the difference in cash flows we see with stocks and options. The discounted cash flow method is fundamentally flawed when applied to companies in which there are embedded real options, and might even result in a value of zero (Kellogg & Charnes, 2000).

Figure 1: Valuing options



Here, we use the decision-tree approach detailed above to value cleantech patent development. The aim of the analysis is to determine an aggregate measure of the growth in wealth from the clean technology breakthrough and in addition to derive an estimation of the required investment to achieve that result. Using estimates by the US Congress (1993) and Lund and Jensen (2016), a typical patent development valuation involves a decision-tree as follows:

Figure 2: Decision-tree approach to map pathways for cleantech development



Part 2 – Panel Regression

The second part of the analysis offers a panel regression which examines the impact of country-level economic variables (*real GDP*, *market return* and *turnover*) and country-level institutional variables on patent intensity. While there have been much research on the determinants of innovation, there is considerably less known about the drivers of cleantech innovations. In this section, our objective is to determine the economic and institutional factors driving the cross-country cleantech innovation activities. The following 32 countries[28] are included in this analysis: Australia, Austria, Belgium, Brazil, Canada, China, Denmark, Finland, France, Germany, Greece, Hungary, India, Indonesia, Ireland, Israel, Japan, Korea, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Russia, South Africa, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States. The data set is an unbalanced panel from 1996 to 2010.

We measure the dependent variable, *patent intensity*, by dividing the number of cleantech patents by population. We include the logarithm of this variable in our regression analysis to ensure a more normal distribution. The country-level economic variables included in the analysis are *real GDP*, *market return* and *turnover*. The country-level institutional variables include R&D intensity (*R&D*), a carbon pricing policy variable (*CarbonETS*) to reflect whether a country has imposed a carbon tax or emission trading scheme (ETS), an antidirector rights index (*ad_rights*) which measures shareholder protection (Djankov, La Porta, Lopez-de-Silanes, & Shleifer, 2008), and a measure for human capital. The literature has long recognized human capital as an engine of innovation (e.g., Benhabib & Spiegel, 2005; Subramaniam & Youndt, 2005). Following Barro and Lee (2013) among others, we measure human capital in the country level as the average years of schooling (*yrs_sch*) of males and females above 15 years old. To control for worldwide factors, we first control for the oil price (*oil*) and its squared

[28] Due to data availability, our panel regression includes 32 out of the 85 countries included in the patent dataset

term (*oil.sq*), because oil price development is regarded as a key drivers of cleantech innovation activity (see e.g., Cumming, Henriques, & Sadorsky, 2016; Popp, 2002). Second, we use a time dummy to count for overall world-wide macroeconomic dynamics (the oil price and its squared term are not included in this estimation as both measures would be correlated with the time dummy). Third, given that cleantech patent activity has significant geographic features (see also the descriptive statistics below), we also include a continent-fixed effect for cleantech patent activities. Table 1 introduces the definitions and data sources of these variables. We present summary statistics for the variables further below in the results section.

Table 1: Variable definitions and data sources

Variable	Definition	Data source
patent intensity	Number of granted cleantech patents divided by the population	PATSTAT database
real GDP	Natural logarithm of GDP, PPP (constant 2010 international $) per capita	World Bank Development Indicators, seehttp://data.worldbank.org.
market return	S&P Global Equity Indices (annual % change)	World Bank Development Indicators
turnover	Natural logarithm of stock market turnover of listed companies (% of GDP)	World Bank Development Indicators
R&D	Percentage of RD expenditure scaled by GDP	World Bank Development Indicators
CarbonETS	Carbon pricing policy variable; it equals to 0 if there is no carbon pricing scheme (i.e., a carbon tax or an emission trading scheme, ETS) in a certain country in a certain year, and equals to 1 otherwise	Kossoy et al. (2015)
ad_rights	The number of shareholder protection mechanisms based on a country's laws	Djankov et al. (2008)
yr_sch	Average years of schooling of males and females above 15 years of age	Barro and Lee (2013); see also http://barrolee.com
oil	Natural logarithm of real 2010 oil prices. Real oil prices are calculated as Brent spot prices ($US) deflated by US CPI	BP Statistical Review of World Energy (2016)
oil.sq	Squared natural logarithm of real oil prices	BP Statistical Review of World Energy (2016)

4. Data Sources

We use data from the Worldwide Patent Statistical Database (PATSTAT) maintained by the European Patent Office (EPO) to construct our patent variables. PATSTAT offers a comprehensive coverage of patent applications worldwide from 1850 onwards; filed by public and private firms through 94 regional, national and international patent offices. Previous studies (e.g., Gao & Zhang, 2016; Moshirian, Tian, Zhang, & Zhang, 2015) attest that the data quality of PATSTAT is comparable to other databases widely used in the innovation literature, such as the US-based National Bureau of Economic Research (NBER) Patent and Citation database. However, compared to the NBER patent database, PATSTAT has a much wider coverage since the former only records patent filings to the US Patent Office (USPTO).

Therefore, the NBER database may underestimate the innovations of non-US firms, especially for firms in emerging countries which typically do not file patent applications to the USPTO (Chang, McLean, Zhang, & Zhang, 2013). In addition, the NBER database ended on 2006 while the PATSTAT database keeps recording granted patents and patent applications.

Our sample period ranges from 1980 to 2010, covering 85 countries. While we do have data available up to 2017, we decided to truncate the sample period up to 2010 as there are significant reporting lags in patent databases (Gao & Zhang, 2016). We use the International Patent Classification (IPC) provided by World Intellectual Property Organization (WIPO) to identify the classification of patents. The IPC for cleantech is obtained from the WIPO IPC Green Inventory,[29] and we include bio-fuels, fuel cells, hydro energy, wind energy, solar energy, geothermal energy, integrated gasification combined cycle, energy from manmade waste, and 'other use of heat' for cleantech. For purposes of our descriptive analysis (see next section), we obtain patents from other industries. We use the IPC codes for biotech and pharmaceuticals from Harhoff and Reitzig (2004), and the IPC codes for robotics from Keisner, Raffo, and Wunsch-Vincent (2015).[30]

5. Results: Descriptive Statistics

To obtain overall statistics about the dispersion of cleantech patents and the worldwide value of their commercialization, we compile a list of all patent grants worldwide (based on our sample of 85 countries) and then compare their growth over time, also in comparison to other industries (biotech, pharmaceuticals, and robotics) (see Figure 2).

The left column in Figure 2 shows the total number of patent applications and granted patents for cleantech, biotech, pharmaceuticals and robotics from 1980 to 2010. Pharmaceutical patents have the highest level of applications and grants (a total of 2,013,645 applications and 954,470 grants), followed by biotech (a total of 880,234 applications and 429,651 grants) and cleantech (a total of 514,307 applications and 286,924 grants). Robotics has the lowest level of applications and grants (a total of 55,937 applications and 31,369 grants). To compare the growth rates of these four groups of patents, we rescale the patent numbers. For each group, we set up an index of 100 in 1980, and then keep the same growth rate to the actual patent number. The right column shows the patent index for both applications and grants.
To show a more recent trend of patent growth, we plot the number and index of patent applications and grants from 1990 to 2010 in Figure 3.

[29]See http://www.wipo.int/classifications/ipc/en/est/

[30] The biotech patents codes are IPC C07G, C12M, C12N, C12P, C12Q, C12R, and C12S. The pharmaceuticals patent codes are IPC A61K, excluding the subclass A61K7. The list of robotics patent codes are: B25J 9/16, B25J 9/20, B25J 9/0003, B25J 11/0005, B25J 11/0015, B60W 30, B60W2030, Y10S 901, G05D 1/0088,G05D 1/02, G05D 1/03, G05D2201/0207, and G05D 2201/0212.

Figure 2: Worldwide patents 1980 to 2010

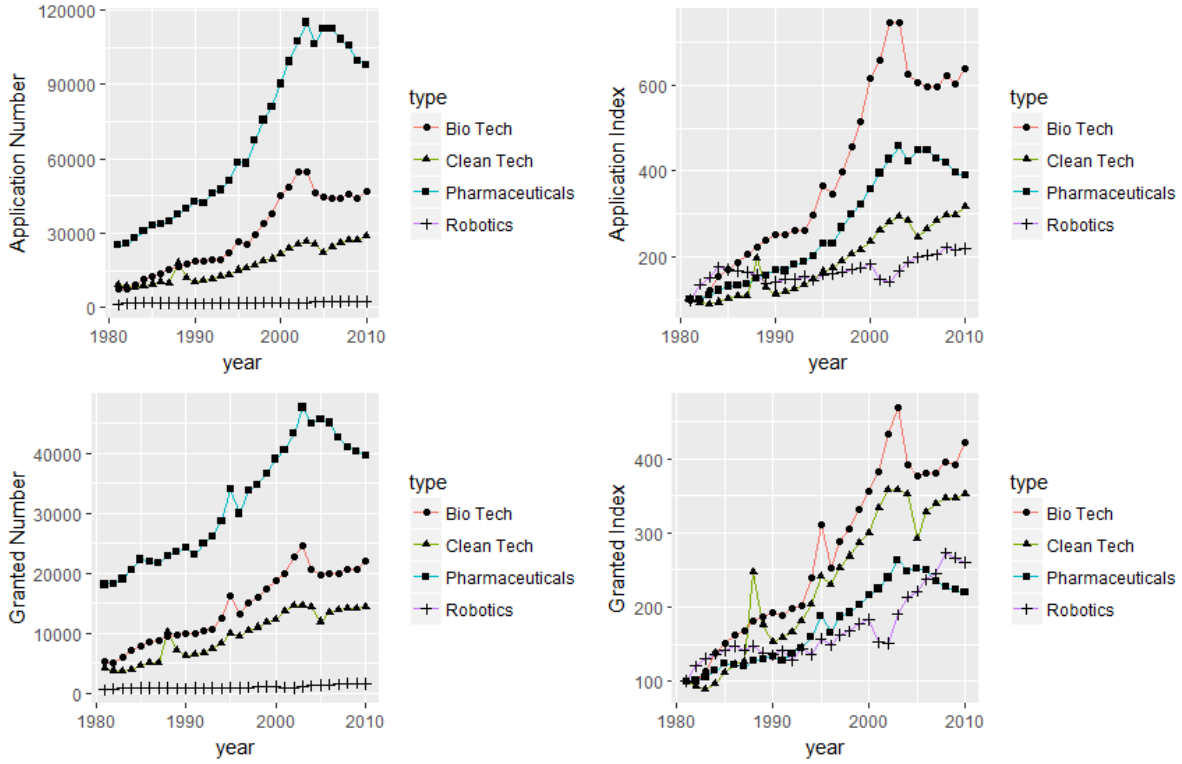
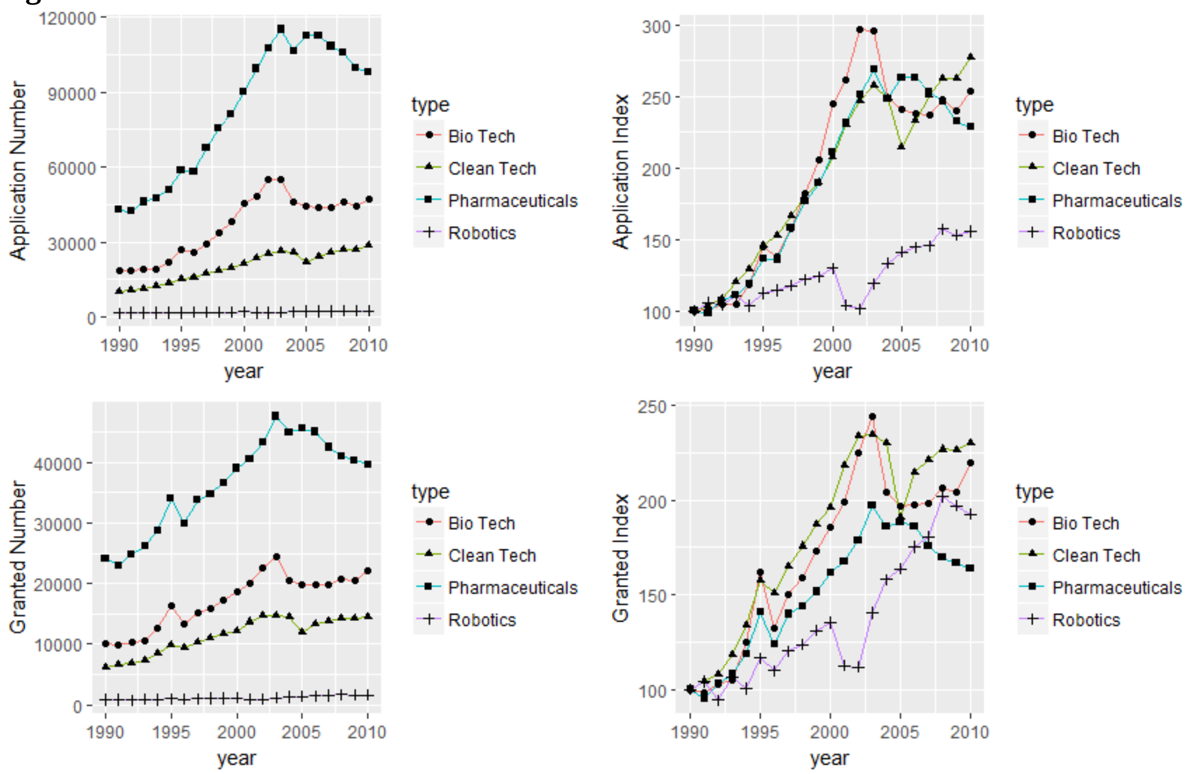


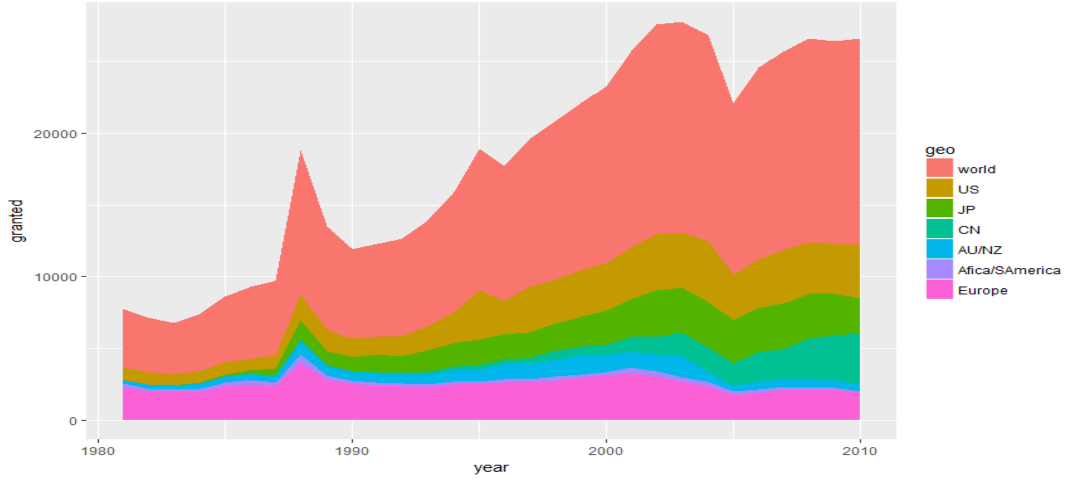
Figure 3: Worldwide Patents 1990 to 2010



From the right column of Figure 3, it is evident that cleantech has the highest growth rate, both in terms of applications and grants, especially in the most recent period from 2005 to 2010. It is not surprising that biotech also maintains an overall high growth rate since biotechnology has been one of the previous technological breakthroughs (Hong, Scheinkman, & Xiong, 2008). However, it is decreasing in the most recent period from 2005 to 2010. This result suggests that cleantech is now a fast-growing patent class, pointing to significant potential to create the next technological breakthrough.

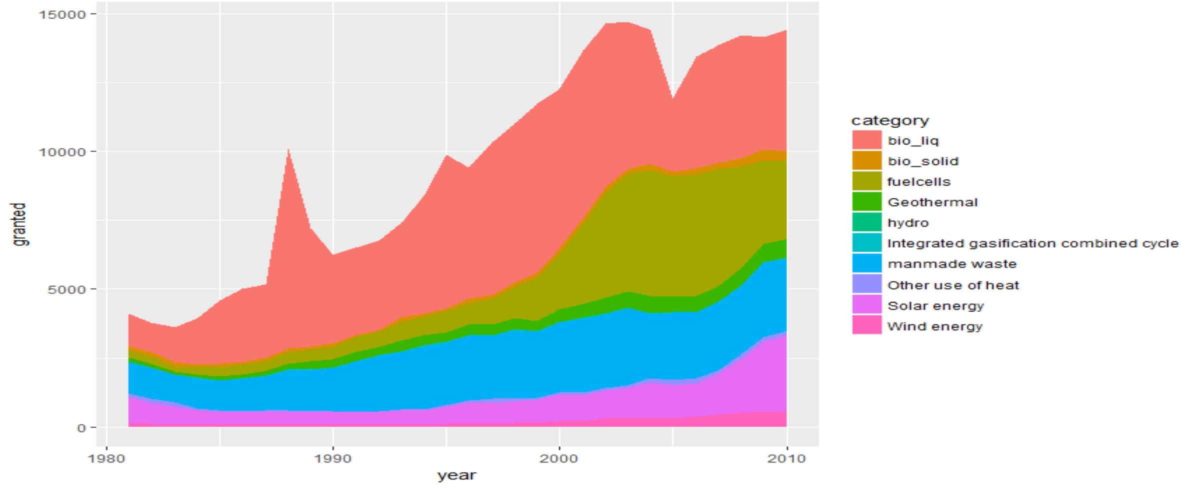
Figure 4 shows the cleantech patent distribution in several major countries/regions, including the United States (US), China (CN), Japan (JP), Australia and New Zealand (AU/NZ), Africa and South America (Africa/SAmerica), and Europe. The figure shows that there are both high levels and high growth rates of cleantech patents in the US, China and Japan.

Figure 4: Cleantech patent distributions by geography



In addition, Figure 5 plots the cleantech patent distribution by categories. Cleantech patents can be divided into following categories: biofuel (liquid), biofuel (solid), fuel cells, geothermal, hydro, integrated gasification combined cycle, manmade waste, solar energy, wind energy and other use of heat. From the figure, we can see that biofuel, fuel cells, solar energy, and wind energy experience a fast growing in the sample period.

Figure 5: Cleantech Patent distribution by categories



6. Results of the Decision-Tree Analysis

Using the decision-tree above, we provide a valuation of cleantech patent development and the corresponding required investment (see also Table 2). Data stems from our dataset of 286,924 granted cleantech patents spanning 85 countries described above. Detailed calculations and the R code for the analysis are shown in the appendix. We assume 4 cycles of patent development from now (2017) to 2050, and provide results for both a 100% penetration of the energy market and 66% penetration as follows. The result for 100% market penetration is a total wealth increase of US$15.490 trillion. This estimate is similar to estimates by external sources (i.e., the International Energy Agency, the International Renewable Energy Agency and Bloomberg) which use different modelling approaches and conclude that the wealth increase is about US$19 trillion.[31] In terms of investment, our results show an investment of US$0.259 trillion up front and over time US$2.023 trillion for a total of US$2.282 trillion. For the other 99% of patents that do not see further development, we assume an associated expenditure of US$0.005 billion ($5 million) each which gives an investment amount of US$1.422 trillion for a total investment of US$3.704 trillion. In comparison the International Energy Agency, the International Renewable Energy Agency and Bloomberg which puts the new investment figure at US$29 trillion. For the case of renewables being 66% of the total energy market we find that the wealth creation is US$10.157 trillion and the total new investment is US$2.923 trillion (Table 2).

Table 2: Results of the decision-tree analysis

Penetration	**100%**	**66%**
	US$Billion	US$Billion
Wealth Created	15,490	10,159
Investment initial	259.2	170
Investment contd.	2023	1327
Total Investment	2282	1497
99% of patents not developed	1422	1426
Total	3,704	2,923

[31] International Energy Agency and International Renewable Energy Agency, 2017, Perspectives for the Energy Transition: Investment needs for a low carbon energy system. OECD/IEA and IRENA

7. Results of the Panel Regression for Institutional Factors

Table 3 provides summary statistics for the institutional factors included in our analysis. The patent intensity is, on average, 7.5 cleantech patents per one million people. Countries invest on average 1.722% of GDP on R&D and 37.6% of observations in the dataset have a carbon tax or emission trading scheme.

Table 3: Summary statistics

Panel A: Mean and Standard Deviation

Variable	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
patent intensity	7.551	12.727	0.973	2.632	8.002
real GDP (US$)	9.927	0.822	9.449	10.179	10.504
market return (%)	13.173	35.521	-10.287	14.970	31.005
turnover	4.193	0.739	3.722	4.226	4.666
R&D (%)	1.722	0.835	0.987	1.715	2.377
CarbonETS	0.376	0.485	0	0	1
ad_rights	3.422	1.040	2.500	3.500	4.000
yr_sch	10.036	1.574	8.910	10.260	11.190
oil	3.764	0.505	3.404	3.586	4.255
oil.sq	14.421	3.817	11.590	12.859	18.105

Panel B: Correlation Table

Variable	*patent intensity*	*real GDP*	*market return*	*turnover*	*R&D (%)*	*CarbonETS*	*ad_rights*	*yr_sch*	*oil*	*oil.sq*
patent intensity										
real GDP (US$)	0.42									
market return (%)	-0.08	-0.15								
turnover	0.24	0.27	-0.24							
R&D (%)	0.57	0.59	-0.06	0.43						
CarbonETS	0.01	0.04	-0.04	0.05	0.04					
ad_rights	0.11	0.18	-0.01	-0.14	0.02	-0.15				
yr_sch	0.61	0.59	-0.10	0.32	0.46	0.23	0.06			
oil	-0.09	0.15	-0.15	0.21	-0.02	0.37	-0.08	0.19		
oil.sq	-0.10	0.15	-0.16	0.22	-0.02	0.38	-0.08	0.19		

We present the estimation results of our panel regression in Table 4. Column (1) controls the continent dummy, while column (2) controls both the continent dummy and year dummy. In Column (1), the estimated coefficient for the oil price (*oil*) is positive and significant while the estimated coefficient for squared oil price (*oil.sq*) is negative and significant, indicating a U-shaped relation between cleantech innovation activity and the oil price. This finding is consistent with Cumming et al.'s (2016) study on cleantech venture capital. The intuition of this U-shape pattern is that higher energy prices initially encourage more renewable energy innovations to substitute for fossil fuels (Popp, 2002). However, when the energy price continues to increase, it creates a strong incentive for incumbent firms to exploit higher cost oil deposits which were economically infeasible with low energy prices. The estimates for R&D are both statistically significant in Column (1) and Column (2). In terms of economic significance, a one-standard deviation increase of R&D would give rise to 2.6% increase in cleantech patents. Unlike other technologies, cleantech is a 'public' good as there are no clearly defined property rights for environmental sources and thereby needs the "visible hand" from government (Demsetz, 1970). Consistent with economic theory and the fact that government funding is the primary source for the early development of cleantech innovation (Cumming et al., 2016), our results highlight the prominent role of government R&D policy in driving the cleantech innovation.

Table 4: Determinants of cleantech patent activities

Notes: This table reports coefficients of regressions of country-level factors on cleantech patents. The dependent variable *patent intensity* is the natural logarithm of (number of granted cleantech patents divided by the population). The independent variables are defined as follows: *R&D* is the percentage of R&D expenditure scaled by GDP; *CarbonETS* is a dummy variable which equals 0 if there is no carbon tax or emission trading scheme, and 1 otherwise; *oil* is the natural logarithm of the oil price; *oil.sq* is the squared natural logarithm of the real oil price; *real GDP* is the natural logarithm of GDP per capita; *turnover* is the natural logarithm of the stock market turnover of listed companies (% of GDP); *ad_rights* is the number of shareholder protection mechanisms based on a country's laws; *yr_sch* is average years of schooling of males and females above 15 years of age. We report the t-statistics in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively.

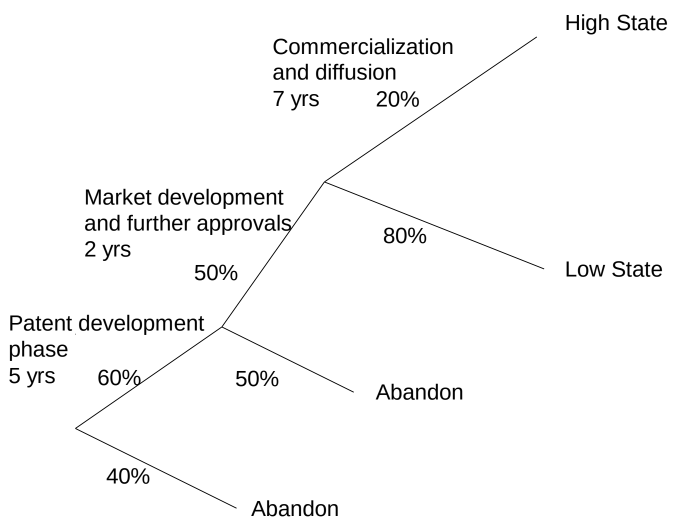
	Dependent variable:	
	patent intensity	
	(1)	(2)
R&D	0.371***	0.347***
	(3.932)	(3.580)
CarbonETS	0.518***	0.545***
	(3.848)	(3.915)
oil	4.037**	
	(2.521)	
oil sq	-0.616***	
	(-2.860)	
real GDP	0.196*	0.232*
	(1.689)	(1.925)
turnover	-0.013	-0.048
	(-0.159)	(-0.557)
market return	-0.002	0.00001
	(-1.544)	(0.004)
ad_rights	0.082	0.073
	(1.456)	(1.285)
yr_sch	0.107**	0.110**
	(2.265)	(2.278)
Trend	-0.007	
	(-0.258)	
Constant	-21.984***	-15.672***
	(-6.894)	(-16.762)
Continent dummy	Yes	Yes
Year dummy		Yes
Observations	429	429
R^2	0.767	0.771
Adjusted R^2	0.759	0.754

The CarbonETS variable is positive and statistically significant. On average, countries with a carbon pricing policy (carbon tax or emissions trading scheme) have a cleantech patent ratio that is 4.2% higher than that for countries without a carbon pricing policy. Examining the level of schooling as a proxy for human capital, we find strong evidence that human capital plays a positive role in driving the cleantech innovations. A standard deviation increase in years of schooling increases cleantech patents by 1.3%. We also find that a country's real GDP is positively related to the cleantech patents. This is consistent with previous findings in the innovation literature which indicate that wealthier countries innovate more (Moshirian et al., 2015). The shareholder protection measure (*ad_ rights*) is positive but not significant, which could be due to the control variable for continental fixed effects. The effects of market return and turnover are both insignificant. Overall, our results suggest that institutional factors such as carbon pricing policies as well as country-level public R&D expenditure and human capital are key parameters that support cleantech development and create a supporting institutional environment to promote innovation and low-carbon development.

8. Modifications to the Decision-Tree Analysis for Different Institutional Environments

To show the effect of institutional factors on the above valuation approach we show another version of the decision-tree (see Figure 6) in an environment with more favorable institutional conditions going forward. The parameters in the analysis are the same in the following decision-tree except that the probability of approval is now 50% and investment in working capital at the commencement of commercialization is now 14% in this more favorable environment.[32]

Figure 6: Modified decision-tree



The wealth created is now US$15.976 trillion in the 100% renewable scenario and US$10.556 trillion in the 66% renewable scenario. This is a marked increase from the less institutionally

[32] Since the probability of approval is higher in this case, it will take fewer patents to make 100% of the market at 2% favorable penetration. We change the total number of patents to 280 (100%) and 185 (66%) in this scenario.

favorable scenario. Investment is now much less at US$2.611 trillion in the 66% renewable scenario and US$3.215 trillion in the 100% renewable case.

Table 5: Results of the modified decision-tree analysis

Penetration	**100%**	**66%**
	US$Billion	US$Billion
Wealth Created	15,976	10,556
Investment initial	112	74
Investment contd.	1674	1106
Total Investment	1786	1180
99% of patents not developed	1429	1431
Total	3,215	2,611

9. Discussion and Future Research Directions

The rate of global environmental change and resulting environmental and social impacts make a rapid transition to a low-carbon economy an immediate priority. This paper has argued that markets will now drive the transition to a low carbon economy, especially those with favourable institutional settings. By looking at the role of market activity, the paper answers the call for further research on the role of cumulative investments to achieve a transition to a low-carbon economy. In particular, we have built upon the notion of 'path creation' (e.g., Dijk & Yarime, 2010) on the country-level of analysis to map out different pathways for cleantech development within a real options framework (specifically a decision-tree), and have offered a corresponding valuation of the development of cleantech patents. We have supplemented this analysis by also examining the impact of country-level economic variables (real GDP, market return and turnover) and country-level institutional variables on patent intensity. Results from this analysis suggest an enormous potential for cleantech patents, particularly if their development and commercialization pathways can be fostered through a supporting institutional environment that promotes innovation and low-carbon development.

The aim of our analysis was to provide both a conceptual framework and empirical assessment of how corporate patenting activity is leading to value creation. To do so we drew upon Real Options theory – which has a long history in the economics and finance disciplines as the basis for mathematical models to value options (e.g., Black & Scholes, 1973). Real Options have also been discussed in strategic management research; however, here the concept has mostly been used for purposes of theoretical reasoning to think about various opportunities open for firms and industries (Trigeorgis & Reuer, 2017). In this paper, we have expanded the Real Options concept to map out future commercialization pathways for cleantech patents on a country level, also including an explicit attempt at valuing the commercialization of clean tech patents. While transformations to a low carbon society are certainly complex and involve economic, social and technological change on many levels (Turnheim et al., 2015), prior research has demonstrated the importance of patenting activity for achieving breakthrough innovations (Phene, Fladmoe- Lindquist, & Marsh, 2006) and for creating wealth, growth and prosperity. The Real Options approach in this paper supports policy analysis and provides insights into how policy decisions relate to investment behavior and strategic decisions (Blyth et al., 2007; Trigeorgis & Reuer, 2017).

Our results show greater patent activity in contexts characterized by more stringent carbon policies (in form of carbon taxes or emissions trading schemes), as well as greater investments into country-level public R&D and higher human capital. These results are consistent with Aguilera-Caracuel and Ortiz-de-Mandojana (2013) who use an institutional approach and we find that green innovative firms are located in areas with stronger environmental regulations. While environmental policies can introduce additional costs for firms to do business (e.g., Jaffe, Peterson, Portney, & Stavins, 1995), a more stringent regulatory environment is seen as conducive to the development of environmentally friendly technologies, as it effectively forces firms to innovate (Porter & Van der Linde, 1995). Greater investments into country-level public R&D and higher human capital are certainly correlated with other development indicators, but create nonetheless favourable conditions to support societal transitions (Loorbach, 2010; Loorbach & Rotmans, 2006). However, several trillion dollars of new investment are going to be needed to drive the transition to a low carbon economy and to further develop and commercialize cleantech patents. Our analysis of the potential for wealth creation through the development of cleantech patents should therefore not just be interpreted as opportunity-only case, as the opportunities created have also extreme investment risks. In the baseline scenario, the chance of technical success 60%, approval 21.5% and high market penetration 20% for an overall chance of high market penetration of 0.0258 and this is only for the quarter of one percent of patents that make it to further development at each phase. Taking this additional probability into account means that the success rate for a sample patent that is very close to zero. In addition to the extremely low risk of success, the investment cost is very high.

Our findings have shown that, for 66% clean energy penetration by 2050, there is a US$170 billion up-front investment at the beginning of the discovery stage and a further US$1.327 trillion investment at the beginning of the 7 year commercialization stage. The form and the source of this investment funding may well be an interesting topic for further study. Will future investment be primarily equity-based because of the high risks of R&D development and the sometimes slow progress to positive cash flows? Or will debt be the primary financing mechanism so that the patent developers maintain full ownership of their proprietary intellectual capital? In addition, future research can focus on comparing and contrasting the transition to cleantech in countries that have been at the forefront of policy and practical implementation of clean technology (e.g., Denmark, Germany) with a country such as Australia which has largely delayed implementation of cleantech while at the same time extracting and profiting from fossil fuels.

Future studies can also offer updated information regarding the stages in the decision-tree as further information becomes available about future development pathways, and can thereby refine the stages and probabilities of our analysis. Another avenue for future research is to verify findings from previous technological breakthroughs. The proliferation of the Internet has shown how breakthrough innovations can lead to a boom and bust cycle (referred to as the so-called dot.com crisis). It may therefore be interesting to study whether cleantech will go through a similar cycle. Hong et al. (2008) provide a model of the boom and bust cycles surrounding technological breakthroughs that can be used for this analysis. Finally, further studies can examine even more fully the effect of institutional factors on cleantech development and the transition to a cleantech powered economy. Cumming et al. (2016) provide an analysis of venture capital financing of cleantech investments and this work could be extended to include equity, debt and hybrid security fundraising.

Cleantech is now a fast-growing patent class, showing signs of becoming the next technological breakthrough. Our analysis above has estimated that the total wealth created by the development of cleantech patents ranges from US$10.16 to US$15.49 trillion dollars and will involve an additional investment stimulus to the economy from US$2.93 to US$3.71 trillion; however, it is clear that this wealth creation also comes with a significant amount of investment requirements and investment risk. The question is what policy-makers and decision-makers can do in the face of such extreme investment risk. One possibility is a phased investment approach where the US$170 billion up-front investment is spread over the 5 years of the discovery process with the opportunity to abandon if the project is not showing sufficient progress. Similarly, the additional investment at commercialization could be phased so that the full roll out does not occur in the first year, giving management a chance to test the market before the full investment is made. Our projections show that there are vast sums of money to be made by investment in cleantech. Estimates of wealth creation of between US$10.16 to US$15.49 trillion mean that there are enormous opportunities for business to create wealth and drive GDP for decades to come. Governments, policy makers and grass root support has got us to this stage in history, it is now business that will drive the transition to a cleantech future.

Appendix A

Energy Data

We first estimate the world market in energy. Energy as a percentage of GDP varies by country with for example the US having energy at 8% of GDP. We estimate the global average at 4% of GDP and factoring in world GDP at US$74.152 trillion[33], which gives us a world market of US$2.966 trillion.

We estimate the number of cleantech patents worldwide from 1980-2010 to be 286,924. Out of these patents only a small number will go to the development stage.[34] We assume conservatively that only 1% get developed and that this happens in 4 waves between now (2017 and 2050). We assume a high state at commercialisation of 2% market penetration with a probability of 20% and a low state penetration of 1% with an 80% probability. Production, marketing and general costs are as shown below. Upfront investment costs are US$0.05 billion and an additional outlay of 17% of revenue is required at the commercialisation. Thus there are two phases of investment: First, at the discovery phase, and second, at the commercialisation phase. Discount rates are assumed to be the government bond rate of 1% at discovery and approval and a risk adjusted rate of 9% at commercialisation.

[33] http://data.worldbank.org/indicator/NY.GDP.MKTP.CD
[34] See a related report at: https://www.forbes.com/sites/danielfisher/2014/06/18/13633/#6413d1c36f1c.

Appendix Table 1

Costs	% of	
Production	25.5%	pa
Marketing	20%	pa
General additional	10%	pa
Working capital	17%	incremental
Tax rate	0.3	
	US$Billion	
Initial outlay	0.05	

Revenue	US$Billion
World Energy Market	2966

Market share	
High	0.02
Low	0.01
High market share probability	0.2

Discount rate	
Development	0.01
Approval	0.01
Commercialisation	0.09

The development of 1% of patents in 4 waves at the probabilities given by the decision-tree would yield an expected 100% renewable energy market. The development of .66% of patents in 4 waves would yield a 66% market coverage. We provide results for both:

Appendix Table 2

Penetration	**100%**	**66%**
	US$Billion	US$Billion
Wealth Created	15,490	10,159
Investment initial	259.2	170
investment contd.	2023	1327
Total Invest	2282	1497
99% not developed	1422	1426
Total	3,704	2,923

Using the real option decision-tree model with the above parameters yields a value of each patent of US$5.98 billion. We assume 4 cycles of patent development from now till 2050 (ie .25% 648 now and 648 repeated every 7 years) for a total wealth increase of US$15.490 trillion. In terms of investment our results show an investment of US$0.259 trillion up front and over time US$2.023 trillion for a total of US$2.282 trillion. For the other 99% of patents that don't get developed we assume an associated expenditure of US$0.005 Billion ($5 mil) each which gives an investment amount of US$1.422 trillion for a total investment of US$3.704 trillion.

For 66% penetration of the energy market the figures are the second column above. For the case of renewables being 66% of the total energy market we find that the wealth improvement would be US$10.159 trillion and the total investment will be US$2.923 trillion.

To show the effect of institutional factors on these valuations we show another version of the decision-tree in a country/environment with more favourable institutional conditions. Everything is the same in the following decision-tree except that the probability of approval is now 50% and investment in working capital at the commencement of commercialisation is now 14% in this more favourable environment. The wealth added and required investment now changes to:

Appendix Table 3

Penetration	**100%**	**66%**
	US$Billion	US$Billion
Wealth Created	15,976	10,556
Investment initial	112	74
investment contd.	1674	1106
Total Invest	1786	1180
99% not developed	1429	1431
Total	3,215	2,611

Appendix B
R-Code

```
##R code for real options valuation using the decision-tree method##
##Assumptions##
c_pr=0.255                    ##cost of production (% of revenue)##
c_mkt=0.2                     ##cost of marketing  (% of revenue)##
c_gen=0.1                     ##general cost (% of revenue)##
y_rd=5                        ##life of R&D stage##
y_ap=2                        ##life of approval##
y_com=7                       ##life of commercialisation##
p_rd=0.6                      ##probability of R&D success##
p_ap=0.215                    ##probability of Approval##
p_hs=0.2                      ##probability of High market share##
d_rd=0.01                     ##discount rate of development##
d_ap=0.01                     ##discount rate of approval##
d_com=0.09                    ##discount rate of commercialisation#
t=0.3                         ##tax rate##
wc=0.17                       ##working capital (% of revenue)##
inv=0.1                       ##initial investment outlay (in
billions)##
rev=2966                      ##current world energy market value
(4% of
                              world gdp, in billions)##
s_h=0.02                      ## market share in high stage##
s_l=0.01                      ## market share in low stage##
```

```
n_pat1=648                                            ##number of patents (100% penetration case)
                                                       ##developed##
n_pat2=425                                            ##number of patents (66% penetration case)
                                                       ##developed##
nt_pat=286924                                         ##number of total patents##
inv_nd=0.005                                          ##investment for non-developed patents##
n_wave=4                                              ##number of patent development waves##

##cash flow in year 8 of  (year 1 of commerialisation)##
cash8_h=s_h*(rev*(1-c_pr-c_mkt-c_gen)*(1-t)-rev*wc)    ##cash flow in high market share stage##
cash8_l=s_l*(rev*(1-c_pr-c_mkt-c_gen)*(1-t)-rev*wc)    ##cash flow in low market share stage##

##cash flow per year from year 9 to year 14  (year 2 to year 7 of of commerialisation)##
cash9_h=s_h*(rev*(1-c_pr-c_mkt-c_gen)*(1-t))           ##cash flow in high market share stage##
cash9_l=s_l*(rev*(1-c_pr-c_mkt-c_gen)*(1-t))           ##cash flow in high market share stage##

##PV of cash flows in year 7 (year 0 of commerialisation ) ##
PV7_h=cash8_h/(1+d_com)+((cash9_h/d_com)*(1-(1/(1+d_com)^(y_com-1))))/(1+d_com)
                                                       ##PV in year 7 (high market stage)##
PV7_l=cash8_l/(1+d_com)+((cash9_l/d_com)*(1-(1/(1+d_com)^(y_com-1))))/(1+d_com)
                                                       ##PV in year 7 (low market stage)##
PV7_ap=PV7_h*p_hs+PV7_l*(1-p_hs);                      ##PV in year 7 of approval##
PV7_dl=0;                                              ##PV in year 7 of decline##
##PV in year 5 (year 0 of application)##
PV5_rd=(p_ap*PV7_ap+(1-p_ap)*PV7_dl)/((1+d_ap)^y_ap)
                                                       ##PV in year 5 of R$D success##
PV5_f=0                                                ##PV in year 5 of R$D success##

##PV in year 0 (year 0 of R&D)##
PV0=(p_rd*PV5_rd+(1-p_rd)*PV5_f)/((1+d_rd)^y_rd)
##NPV in year 0##
NPV=PV0-inv*(1-t);



##summary##
```

```
##case 1: 648 patents##
value1=n_pat1*NPV*n_wave                                    ##total value added ##
inv_ini1=n_pat1*inv*n_wave                                  ##initial total investment##
inv_cont1=wc*rev*p_rd*p_ap*n_pat1*n_wave* (p_hs*s_h+(1- p_hs)*s_l)
                                                            ##total continuing investment##
inv_dv1=inv_ini1+inv_cont1                                  ##total investment for developed patents##
inv_rest1=(nt_pat-nwave*n_pat1)*inv_nd                      ##investment for the non-developed patents##
inv_total1=inv_dv1+inv_rest1                                ##total investment

##case 2: 425 patents##
value1=2=n_pat2*NPV*n_wave                                  ##total value added ##
inv_ini2=n_pat2*inv*n_wave                                  ##initial total investment##
inv_cont2=wc*rev*p_rd*p_ap*n_pat2*n_wave* (p_hs*s_h+(1- p_hs)*s_l)
                                                            ##total continuing investment##
inv_dv2=inv_ini2+inv_cont2                                  ##total investment for developed patents##
inv_rest2=(nt_pat-n_pat2*n_wave)*inv_nd                     ##investment for the non-developed patents##
inv_total2=inv_dv2+inv_rest2                                ##total investment
```

References

Aguilera-Caracuel, J., & Ortiz-de-Mandojana, N. (2013). Green innovation and financial performance: An institutional approach. Organization & Environment, 26(4), 365-385.

Araújo, K. (2014). The emerging field of energy transitions: Progress, challenges, and opportunities. Energy Research & Social Science, 1, 112-121.

Ashina, S., Fujino, J., Masui, T., Ehara, T., & Hibino, G. (2012). A roadmap towards a low-carbon society in Japan using backcasting methodology: feasible pathways for achieving an 80% reduction in CO 2 emissions by 2050. Energy policy, 41, 584-598.

Audretsch, D. B., & Feldman, M. P. (1996). Innovative clusters and the industry life cycle. Review of industrial organization, 11(2), 253-273.

Barro, R., & Lee, J.-W. (2013). A new data set of educational attainment in the world, 1950-2010. Journal of Development Economics, 104, 184-198.

Benhabib, J., & Spiegel, M. M. (2005). Human capital and technology diffusion. Handbook of economic growth, 1, 935-966.

Bergek, A., Berggren, C., Magnusson, T., & Hobday, M. (2013). Technological discontinuities and the challenge for incumbent firms: Destruction, disruption or creative accumulation? Research Policy, 42(6), 1210-1224.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of political economy, 81(3), 637-654.

Blyth, W., Bradley, R., Bunn, D., Clarke, C., Wilson, T., & Yang, M. (2007). Investment risks under uncertain climate change policy. Energy policy, 35(11), 5766-5773.

BP Statistical Review of World Energy. (2016). Available at: https://www.bp.com/content/dam/bp/pdf/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-full-report.pdf.

Brennan, M. J., & Schwartz, E. S. (1985). Evaluating natural resource investments. Journal of business, 58(2), 135-157.

Bruton, G. D., Ahlstrom, D., & Li, H. L. (2010). Institutional theory and entrepreneurship: Where are we now and where do we need to move in the future? Entrepreneurship Theory and Practice, 34(3), 421-440.

Bürer, M. J., & Wüstenhagen, R. (2009). Which renewable energy policy is a venture capitalist's best friend? Empirical evidence from a survey of international cleantech investors. Energy policy, 37(12), 4997-5006.

Busch, T., Bauer, R., & Orlitzky, M. (2016). Sustainable development and financial markets: Old paths and new avenues. Business & Society, 55(3), 303-329.

Chang, X., McLean, R., Zhang, B., & Zhang, W. (2013). Patents and productivity growth: Evidence from global patent awards. Unpublished working paper.

Christensen, C. M. (1997). The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Boston, MA: Harvard Business School Press.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative science quarterly, 128-152.

Copeland, T., & Weiner, J. (1990). Proactive management of uncertainty. The McKinsey Quarterly, 4, 133-152.

Cumming, D., Henriques, I., & Sadorsky, P. (2016). 'Cleantech' venture capital around the world. International Review of Financial Analysis, 44, 86-97.

Demsetz, H. (1970). The private production of public goods. The Journal of Law and Economics, 13(2), 293-306.

Dijk, M., & Yarime, M. (2010). The emergence of hybrid-electric cars: Innovation path creation through co-evolution of supply and demand. Technological Forecasting and Social Change, 77(8), 1371-1390.

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The law and economics of self-dealing. Journal of financial economics, 88(3), 430-465.

Elliston, B., Diesendorf, M., & MacGill, I. (2012). Simulations of scenarios with 100% renewable electricity in the Australian National Electricity Market. Energy policy, 45, 606-613.

Gao, H., & Zhang, W. (2016). Employment nondiscrimination acts and corporate innovation. Management Science, online first, 1-18.

Geels, F. W., Berkhout, F., & van Vuuren, D. P. (2016). Bridging analytical approaches for low-carbon transitions. Nature Climate Change, 6(6), 576-583. doi: 10.1038/nclimate2980

Hake, J.-F., Fischer, W., Venghaus, S., & Weckenbrock, C. (2015). The German Energiewende–history and status quo. Energy, 92, 532-546.

Harhoff, D., & Reitzig, M. (2004). Determinants of opposition against EPO patent grants—the case of biotechnology and pharmaceuticals. International journal of industrial organization, 22(4), 443-480.

Hong, H., Scheinkman, J., & Xiong, W. (2008). Advisors and asset prices: A model of the origins of bubbles. Journal of financial economics, 89(2), 268-287.

Hoskisson, R. E., Eden, L., Lau, C. M., & Wright, M. (2000). Strategy in emerging economies. Academy of management journal, 43(3), 249-267.

International Energy Agency. (2014). World energy investment outlook: Special report. Available at: https://www.iea.org/publications/freepublications/publication/WEIO2014.pdf

Jaffe, A. B., Peterson, S. R., Portney, P. R., & Stavins, R. N. (1995). Environmental regulation and the competitiveness of US manufacturing: what does the evidence tell us? Journal of Economic literature, 33(1), 132-163.

Keisner, A., Raffo, J., & Wunsch-Vincent, S. (2015). Breakthrough technologies-Robotics, innovation and intellectual property: World Intellectual Property Organization-Economics and Statistics Division.

Kellogg, D., & Charnes, J. M. (2000). Real-options valuation for a biotechnology company. Financial Analysts Journal, 56(3), 76-84.

Kemp, R., Loorbach, D., & Rotmans, J. (2007). Transition management as a model for managing processes of co-evolution towards sustainable development. The International Journal of Sustainable Development & World Ecology, 14(1), 78-91.

Klaassen, G., Miketa, A., Larsen, K., & Sundqvist, T. (2005). The impact of R&D on innovation for wind energy in Denmark, Germany and the United Kingdom. Ecological economics, 54(2), 227-240.

Kossoy, A., Peszko, G., Oppermann, K., Prytz, N., Gilbert, A., Klein, N., . . . Wong, L. (2015). Carbon Pricing Watch 2015: The World Bank.

Kostoff, R. N., Boylan, R., & Simons, G. R. (2004). Disruptive technology roadmaps. Technological Forecasting and Social Change, 71(1), 141-159.

Liebowitz, S. J., & Margolis, S. E. (1995). Path dependence, lock-in, and history. Journal of Law, Economics and Organization, 11(2), 205-226.

Linnenluecke, M. K., Meath, C., Rekker, S., Sidhu, B. K., & Smith, T. (2015). Divestment from fossil fuel companies: Confluence between policy and strategic viewpoints. Australian Journal of Management, 40(3), 478-487.

Loorbach, D. (2010). Transition management for sustainable development: A prescriptive, complexity- based governance framework. Governance, 23(1), 161-183.

Loorbach, D., & Rotmans, J. (2006). Managing transitions for sustainable development. Understanding Industrial Transformation, 187-206.

Lund, M., & Jensen, J. D. (2016). A real options approach to biotechnology investment policy. Preventive Veterinary Medicine, 128, 58-69.

Marcus, A., & Aragon-Correa, J. A. (2011). Firms, regulatory uncertainty, and the natural environment. California management review, 54(1), 5-16. doi: 10.1525/cmr.2011.54.1.5

Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77-91.

McDonald, R. L., & Siegel, D. (1986). The value of waiting to invest. The Quarterly Journal of Economics, 101(4), 707-727.

McDonald, R. L., & Siegel, D. R. (1985). Investment and the valuation of firms when there is an option to shut down. International economic review, 26(2), 331-349.

Merton, R. C. (1973). Theory of rational option pricing. The Bell Journal of economics and management science, 4(1), 141-183.

Moshirian, F., Tian, X., Zhang, B., & Zhang, W. (2015). Financial liberalization and innovation. Unpublished working paper.

Munksgaard, J., & Morthorst, P. E. (2008). Wind power in the Danish liberalised power market—Policy measures, price impact and investor incentives. Energy policy, 36(10), 3940-3947.

Pavitt, K. (1986). ‘Chips’ and ‘trajectories’: How does the semiconductor influence the sources and directions of technical change? In R. MacLeod (Ed.), Technology and the Human Prospect (pp. 31-54). London Frances Pinter.

Phene, A., Fladmoe- Lindquist, K., & Marsh, L. (2006). Breakthrough innovations in the US biotechnology industry: the effects of technological space and geographic origin. Strategic Management Journal, 27(4), 369-388.

Popp, D. (2002). Induced innovation and energy prices. The American Economic Review, 92(1), 160-180.

Porter, M. E., & Van der Linde, C. (1995). Green and competitive: Ending the stalemate. Harvard business review, 73(5), 120-134.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S., Lambin, E. F., . . . Foley, J. A. (2009). A safe operating space for humanity. nature, 461(7263), 472-475.

Rothaermel, F. T. (2001). Incumbent's advantage through exploiting complementary assets via interfirm cooperation. Strategic Management Journal, 22(6- 7), 687-699.

Rotmans, J., & Loorbach, D. (2009). Complexity and transition management. Journal of Industrial Ecology, 13(2), 184-196.

Rudyk, I., Owens, G., Volpe, A., & Ondhowe, R. (2015). Climate change mitigation technologies in Europe - Evidence from patent and economic data: The United Nations Environment Programme (UNEP) and the European Patent Office (EPO).

Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Cambridge, MA: Harvard University Press.

Schumpeter, J. A. (1942). Capitalism, Socialism and Democracy. New York: Harper & Brothers.
Schwartz, E. S., & Moon, M. (2000). Rational pricing of internet companies. Financial Analysts Journal, 56(3), 62-75.
Scott, W. R. (1987). The adolescence of institutional theory. Administrative science quarterly, 493-511.
Senge, P. M., & Carstedt, G. (2001). Innovating our way to the next industrial revolution. MIT Sloan management review, 42(2), 24-38.
Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425-442.
Sovacool, B. K. (2017). Contestation, contingency, and justice in the Nordic low-carbon energy transition. Energy policy, 102, 569-582.
Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., . . . Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. Science, 347(6223). doi: 10.1126/science.1259855
Subramaniam, M., & Youndt, M. A. (2005). The influence of intellectual capital on the types of innovative capabilities. Academy of management journal, 48(3), 450-463.
Tan, X. (2010). Clean technology R&D and innovation in emerging countries - Experience from China. Energy policy, 38(6), 2916-2926.
Titman, S. (1985). Urban land prices under uncertainty. The American Economic Review, 75(3), 505-514.
Trigeorgis, L., & Reuer, J. J. (2017). Real options theory in strategic management. Strategic Management Journal, 38(1), 42-63.
Turnheim, B., Berkhout, F., Geels, F., Hof, A., McMeekin, A., Nykvist, B., & van Vuuren, D. (2015). Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. Global Environmental Change, 35, 239-253.
Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. Administrative science quarterly, 31(3), 439-465.
Unruh, G. C. (2000). Understanding carbon lock-in. Energy policy, 28(12), 817-830.
Verbong, G., & Geels, F. (2007). The ongoing energy transition: Lessons from a socio-technical, multi-level analysis of the Dutch electricity system (1960–2004). Energy policy, 35(2), 1025-1037.
Walsh, S. T. (2004). Roadmapping a disruptive technology: A case study: The emerging microsystems and top-down nanosystems industry. Technological Forecasting and Social Change, 71(1), 161-185.
Winskel, M. (2002). When systems are overthrown: The 'Dash for Gas' in the British electricity supply industry. Social Studies of Science, 32(4), 563-598.

Emerging Stock Market Co-movements and Third-Country Effects

Hideaki Hirata, Hosei University and Japan Center for Economic Research

Sunghyun Henry Kim, Sungkyunkwan University

ABSTRACT

This paper investigates the effects of financial globalization – in particular cross-border capital flows in financial markets – on pairwise excess stock return co-movements in Emerging Asia during 2001-2012. The analysis shows that increased co-movements are explained not by bilateral capital flows between Asian countries but by capital flows from the G7 countries. That is, the high correlation of stock returns in Emerging Asia is the result of synchronized capital flows from the G7 countries into Asian financial markets. Despite a recent surge in regional capital flows within Emerging Asia, no stock return "de-coupling" from the G7 countries has taken place.

JEL classification: F3, G1
Keywords: *Stock Market; Co-movement; Emerging Asia; Synchronization; Financial linkages*

1. Introduction

Since the 2000s, as compared to the other emerging economies such as Latin America and MENA, Emerging Asia (EA) has witnessed a substantial increase in the volume of international capital flows that have overshadowed the increase in international trade flows. And during the same period of time, in EA, a steady rise in the co-movement of stock markets has been observed and the correlation in excess stock returns among EA in the early 2000s was close to zero but reached 0.4 in 2012 (Figure 1).[35,36] While both EA's intra-regional and inter-regional capital flows have rapidly grown, one can observe two interesting facts. First, the growth of intra-regional financial linkages has become much stronger than that of inter-regional financial linkages since the 2000s.[37] Second, there has also been a significant rise and share in inflows from advanced countries to EA countries.[38] Given these facts, it is worth exploring what are the underlying reasons for the greater synchronization in stock market returns in EA.

Figure 1. Average Stock Return Correlation in the G7 and the EA Countries

Notes: The figure shows the equally-weighted average annual pairwise correlation coefficients of excess stock returns among the G7 countries and among 10 EA countries. See Table 1 for a list of the countries included.

Is the greater synchronization the result of growing regional economic and financial integration among EA countries or increased capital flows between EA and advanced economies? Does the greater synchronization provide evidence of the Asian equity markets being coupled with the equity markets of advanced economies, i.e., do the EA's stock markets with sizable markets on their own withstand pressure from the developed economies' stock markets or are the EA's stock markets subject to the influence from the developed economies' stock markets.

The aim of this study is to document the evidence of stock market synchronization in EA and examine the sources of stock return co-movements. Stock market co-movements shown in EA may indicate that there are increasing bilateral capital flows among EA economies. However,

[35] The average correlation among G7 countries was still higher at around 0.6 in 2012.

[36] EA includes China, Hong Kong, India, Indonesia, Korea, Malaysia, Philippines, Singapore, Taiwan, and Thailand. This definition is consistent with IMF's Regional Economic Outlook of Asia.

[37] See Figure 2 explained in the next section.

[38] See Lane and Milesi-Ferretti (2007) and Milesi-Ferretti, Strobbe, and Tamirisa (2010).

even without bilateral capital transactions, stock returns in EA countries can move together if capital simultaneously flows from the advanced economies to EA economies. We call those originating from the advanced countries on synchronization "third-country effects." Identifying the underlying reasons for stock market synchronization in EA countries is important for understanding the nature of synchronization in EA financial markets and to evaluate the impact of regional economic cooperation such as the Chiang Mai initiative and the Asian Bond Markets Initiative (ABMI) (Bekaert and Harvey, 2014).

Unlike previous studies that have mainly relied on price data to extract common factors, this study uses direct measures of cross-border financial flows taken from the IMF's *Coordinated Portfolio Investment Survey* (CPIS). Using these measures, we estimate the effects of bilateral capital flows among EA economies versus third-country effects of capital flows from the G7 countries on stock return co-movements in EA countries.[39] The impact of shocks to capital flows from advanced economies to the EA countries should differ across countries due to different degrees of integration with global financial markets.[40] We capture these time-varying and country-specific effects of capital flows on stock returns by running static and dynamic panel regression models.[41]

Since we focus on the post-financial liberalization period (2001-2012), during which EA financial markets were likely to be highly integrated with the rest of the world, we can capture the impact of non-institutional changes in economic globalization on stock market synchronization. Most previous research has focused on the impact of institutional liberalization of financial markets such as removal of legal restrictions in international capital flows during a period when financial markets were not completely open.[42] In contrast, most of our sample countries (except for China) had already fully liberalized their international financial markets by the start of our estimation period (Table 1) and therefore, we can capture the impact of cross-border capital flows arising from non-institutional economic reasons.

The baseline empirical analysis focuses on annual observation for the pairs of 10 EA countries (yielding 45 pairwise correlations a year) over the period 2001–2012. In the regressions, we control for cross-sectional dependence, heteroscedasticity, and the possibility of serial correlation. We also control for possible endogeneity arising from the dynamic nature of stock market co-movements across countries (King et al., 1994; Bekaert et al., 2009).

[39] Most previous studies use static or dynamic factor models to identify the contribution of national or global common factors to variations in prices. For example, Forbes and Chinn (2004) run regressions of the computed country-specific factor loadings on several indicators of bilateral linkages between each pair of small and large countries. Bekaert et al. (2009) use an asset pricing model and run various estimations with (excess) stock returns of each country on the left hand side and returns on global or developed countries' portfolios on the right hand side.

[40] Bekaert and Harvey (1997) argue that the correlations across national stock markets are directly linked to the degree with which countries are integrated with global capital markets.

[41] There are some preceding studies that have used quantitative data on capital flows such as Flavin et al. (2002), Froot and Ramadorai (2008), and Dellas and Hess (2005). However, these studies use cross-section or pooled regressions that neglect the time-dimension of economic integration in the 2000s and beyond. Beine and Candelon (2011) and Bekaert and Wang (2009) use both time and cross-sectional dimensions simultaneously, but their focus is limited to the effects of the degree of economic liberalization and openness.

[42] See, for example, Bekaert and Harvey (1997, 2000), Bekaert et al. (2002), Dellas and Hess (2005), and Beine and Candelon (2011).

Table 1. Sample Countries and the Stock Market Indices

EA	China (02, Shanghai Stock Exchange: Index: A Shares), Hong Kong (*, Hong Kong Hang Seng Index), India (92, Bombay Stock Exchange: Index: SENSEX), Indonesia (89, Jakarta Stock Exchange Composite Index), Korea (92, Korea Stock Exchange KOSPI 200 Index), Malaysia (88, FTSE Bursa Malaysia EMAS Index), Philippines (91, Philippines Stock Exchange All Share Index), Singapore (*, Straits Times Index STI), Taiwan (91, Taiwan TPEx Exchange Index), Thailand (87, Stock Exchange of Thailand SET Index)
G7	United States of America (*, S&P 500 Index), Canada (*, S&P/TSX Composite Index), Germany (*, Deutsche Boerse AG German Stock Index DAX), France (*, CAC 40 Index), Italy (*, FTSE MIB Index), United Kingdom (*, FTSE 100 Index), Japan (83, Nikkei 225)

Notes: Numbers in parentheses show the year in which the domestic stock market was opened to foreign investors (Bekaert and Harvey, 2000, 2002; Bekaert, et al., 2005). * indicates that the country's domestic stock market was already fully liberalized before the start of our estimation period. In addition, the names of national stock indices used in this study are shown in parentheses.

When third-country effects are not included, the regression results seem to suggest that stock return co-movements in EA countries can be explained by bilateral portfolio investment flows. However, once third-country effects are taken into account, the effects of bilateral flows become insignificant. Third-country effects are highly significant and positive in most cases. Capital flows from the G7 countries significantly affect the stock return movements in EA countries even after controlling for potentially important factors such as trade agreements, industry differences, inflation, economic development, and financial depth. The main conclusions of the regression results remain unchanged even when we extend the sample to include the non-Asian BRICS countries (Brazil, Russia, and South Africa) and FDI data. Therefore, it could be argued that in terms of stock returns in EA countries, no "de-coupling" has taken place.

This study is related to the literature using asset pricing models to measure third-country effects. Globally integrated financial markets result in domestic stock returns being partly determined by global returns. That is, global common shocks explain part of the variation of domestic stock returns. Global shocks can be empirically identified by using factor models (Forbes and Chinn, 2004; Brooks and Del Negro, 2006) or arbitrage pricing theory or asset pricing models such as Fama-French, or Heston-Rouwenhorst models (Bekaert et al., 2009; Dutt and Mihov, 2013; Brooks and Del Negro, 2004, 2005).[43] The advantage of this approach is that one can identify global factors, country-specific factors, and other potential factors such as sector-specific and regional factors that determine market returns in each country without using quantitative measures of cross-border transactions.

The remainder of this study is organized as follows. Section 2 provides an overview of the related literature and of recent developments in the financial globalization of EA countries. Section 3 then outlines the models and variables used for the estimation in this study, while Section 4 presents the results of the empirical analysis. Finally, Section 5 concludes.

2. Financial Globalization in Emerging Asia

[43] Another strand of studies uses GARCH models and their variants and measures the share of stock return variation explained by global common factors as a measure of the degree of integration with global markets. See, for example, Gérard et al. (2003).

A large body of theoretical and empirical studies has focused on the role of real and financial linkages in explaining economic co-movements in emerging markets. With regard to excess stock return co-movements, previous studies in the 1990s and the early 2000s have found that the degree of co-movements in emerging markets with the rest of the world is generally low, implying that developed countries with large stock markets have only a limited impact on developing countries with small financial markets (Bekaert and Harvey, 1997; De Santis and Imrohoroglu, 1997; Forbes and Chinn, 2004; Chi et a., 2006). Reasons for the limited impact include the presence of transaction costs, restrictions on cross-country capital flows (Bekaert and Harvey, 2000), and home bias in international investment (Karolyi and Stulz, 1996). More recent studies, however, suggest that regional convergence in Asia has (at least gradually) emerged but the degree of impact from the global markets has got greater as well (Boubakri and Guillaumin, 2016; Chien et al., 2015; Dewandaru et al., 2015; Lee and Jeong, 2016; Mandigma, 2014; Tiwari et al. 2013; Jiang et al., 2017). In fact, in EA, a steady rise in the co-movement of stock markets has been observed and the correlation in excess stock returns among EA in the early 2000s was close to zero but reached 0.4 in 2012 (Figure 1).

Under the recent development of EA's stock markets in global markets, one should consider not only bilateral capital flows among emerging markets but also third-country effects from the developed economies in order to properly analyze the reasons behind the stock market co-movements. However, few studies have focused on the role of bilateral flows in emerging markets, mainly because of the lack of data on bilateral financial flows and their limited size. In this study, we focus on both bilateral capital flows and third-country effects based on quantitative data of financial flows.

From a theoretical perspective, the effect of financial integration on co-movements a priori is indeterminate. On the one hand, financial integration may lead to greater synchronization through demand-side effects. On the other, it may lead to greater specialization in production through the reallocation of capital across sectors, which could reduce co-movements. The literature on international business cycles suggests that financial globalization can result in greater exposure to non-global shocks such as country-specific or sector-specific shocks, which can lower co-movements (see, e.g., Kalemli-Ozcan et al., 2013).
A good example why a priori the effect is indeterminate is provided by Forbes and Chinn (2004)." Consider the case that a negative shock in large country *g* leads to investor pessimism which drives down stock returns in country *g*. One possible scenario is that this pessimism leads investors in country *g* to reduce their investment in small country *x* to ensure their liquidity, which lowers stock returns in country *x* (resulting in higher co-movement). The other scenario is that investors in country *g* increase their exposure to relatively better positioned country *y*, potentially driving up stock returns in country *y* (resulting in lower co-movement).

Table 2 shows the total stock market capitalization of the 10 EA countries that our study focuses on and of the G7 countries. As can be seen, the share of the 10 EA countries in global stock market capitalization increased by more than 10 percentage points from 2001 to 2012, while the share of the G7 countries decreased by 25 percentage points. The share of the 10 EA countries and the G7 countries together in global stock market capitalization was 88% in 2001 and 75% in 2012. On the other hand, the share of the G7 countries alone shrank from 81% in 2001 to 57% in 2012. In sum, while the weight of the 10 EA countries in global stock market capitalization has increased and that of the G7 countries has decreased, capital inflows in the EA countries from the G7 countries are nevertheless still large.

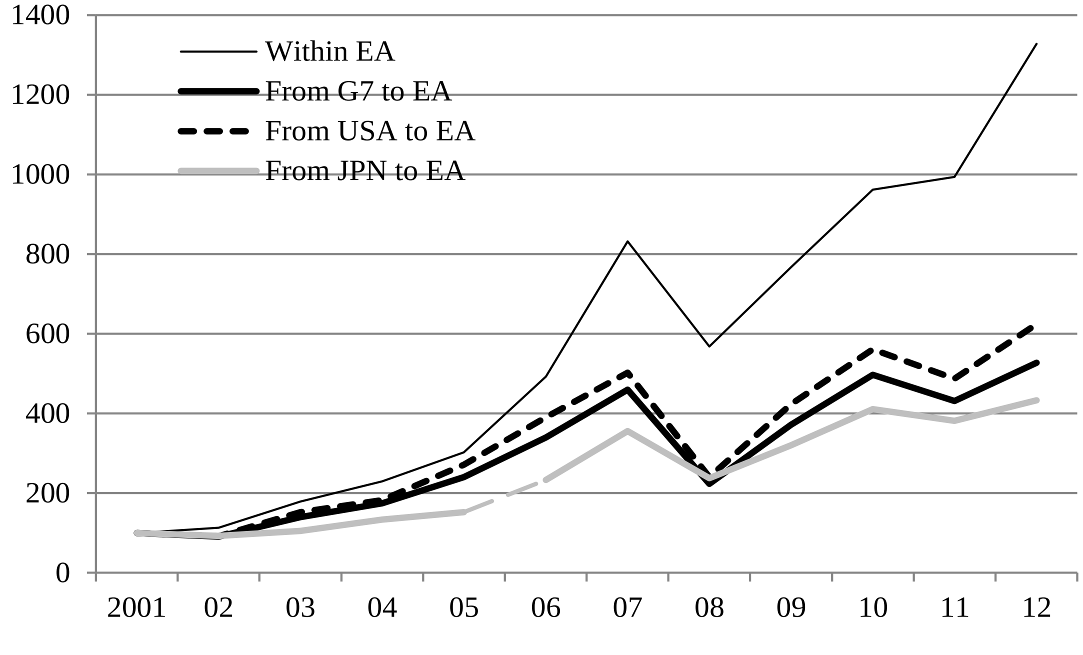
Table 2. Share in Global Stock Market Capitalization

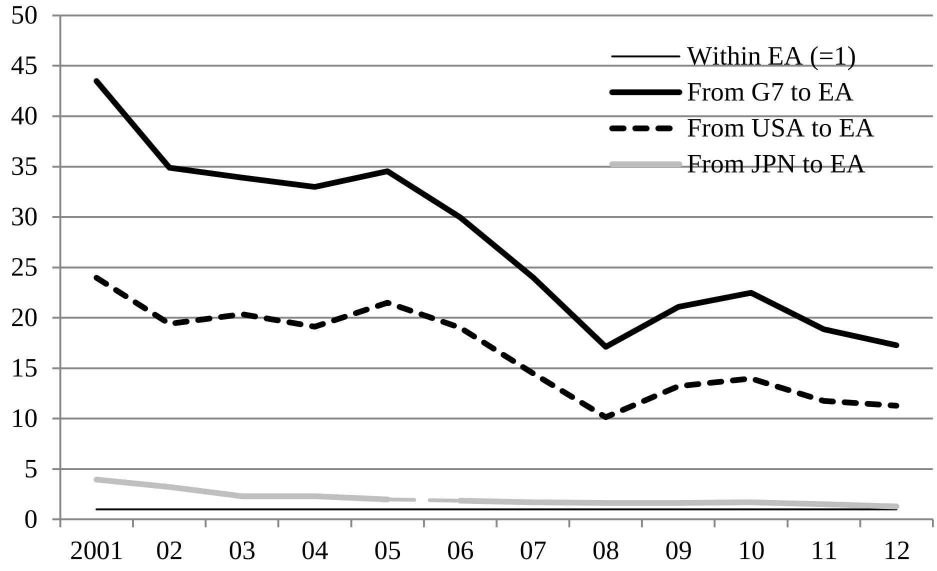
	EA	BRICS	G7		
				USA	Japan
2001	7.4%	3.9%	81.2%	51.7%	8.4%
2005	9.7%	7.1%	73.3%	41.4%	11.6%
2008	19.2%	15.2%	64.5%	36.2%	9.9%
2010	19.9%	17.4%	55.0%	31.2%	7.5%
2012	18.2%	14.1%	57.1%	34.2%	6.7%

Notes: Data sources are International Financial Statistics, IMF and Taiwan Stock Exchange. The BRICS include not only the non-EA BRICS (Brazil, Russia, and South Africa) but also the EA BRICS (China and India).

Figure 2 shows relative trends in bilateral financial flows within EA countries and financial inflows from the G7 countries into EA countries from 2001 to 2012. The upper figure shows the increase of bilateral portfolio flows among the 10 EA countries and financial inflows from the G7 countries to 10 EA countries, with all flows are set to 100 in the base year 2001. The lower chart shows the relative volumes of inflows from G7 countries to EA to the volume of aggregated bilateral financial flows among EA countries in each year. As can be seen, nominal bilateral financial flows among EA countries rose by a factor of more than 13, while financial inflows from the G7 rose by a factor of more than 5 during this period. That being said, the amount of financial inflows from the G7 still remains much larger than bilateral financial flows among the EA countries, suggesting that inflows from the G7 (particularly the US) still have an important impact on the EA countries.

Figure 2. Capital Flows within the EA Countries and Inflows from the G7





Notes: The figure shows bilateral portfolio flows among the 10 EA countries and financial inflows from the G7 countries. With all flows are set to 100 in the base year 2001, the upper chart shows the increase of the flows, while the lower chart shows the relative volumes of inflows from G7 countries to EA to the volume of aggregated bilateral financial flows among EA countries in each year.

3. Empirical Estimation

3.1. Estimation Models

We first estimate the following static regression model:

$$\boldsymbol{\rho}_{jk,t} = \alpha + \beta \boldsymbol{X}_{jk,t} + \gamma \boldsymbol{Z}_{jk,t} + u_{jkt} \quad , \tag{1}$$

where $\boldsymbol{\rho}_{jkt}$ is the pairwise excess stock return correlation, $\boldsymbol{X}_{jk,t}$ is a vector of bilateral capital flows between countries j and k and the capital flows from large country g to small countries j and k (third-country effect) and $\boldsymbol{Z}_{jk,t}$ is a set of control variables. The error terms are $u_{jkt} = \eta_{jk} + v_t + \varepsilon_{jkt}$, where η_{jk} represents country-pairwise fixed effects that capture country-pair specific factors explaining co-movements, v_t is a set of year dummies, and ε_{jkt} represents pure error terms.

This static model, however, does not capture the potential dynamics of stock return co-movements. Therefore, we also use the following dynamic model with lagged values of the dependent variable on the right hand side:

$$\boldsymbol{\rho}_{jk,t} = \alpha + \beta \boldsymbol{X}_{jk,t} + \gamma \boldsymbol{Z}_{jk,t} + \theta \boldsymbol{\rho}_{jk,t-1} + u_{jkt}. \tag{2}$$

As discussed by Blundell and Bond (1998), Rioja and Valev (2004), and Wintoki et al. (2012), this type of model potentially suffers from biased and inconsistent estimators as well as possible simultaneity of explanatory variables. To resolve these problems, we use the system generalized method of moments (GMM) to estimate the following model:[44]

$$\begin{bmatrix}\boldsymbol{\rho}_{jk,t}\\ \boldsymbol{\Delta\rho}_{jk,t}\end{bmatrix} = \alpha + \beta\begin{bmatrix}\boldsymbol{X}_{jk,t}\\ \boldsymbol{\Delta X}_{jk,t}\end{bmatrix} + \gamma\begin{bmatrix}\boldsymbol{Z}_{jk,t}\\ \boldsymbol{\Delta Z}_{jk,t}\end{bmatrix} + \theta\begin{bmatrix}\boldsymbol{\rho}_{jk,t-1}\\ \boldsymbol{\Delta\rho}_{jk,t-1}\end{bmatrix} + \begin{bmatrix}\eta_{jk}\\ 0\end{bmatrix} + \begin{bmatrix}v_t\\ \boldsymbol{\Delta}v_t\end{bmatrix} + \begin{bmatrix}\varepsilon_{jkt}\\ \boldsymbol{\Delta}\varepsilon_{jkt}\end{bmatrix}, \tag{3}$$

assuming following orthogonality conditions:

$$E(\boldsymbol{\rho}_{\boldsymbol{jk},t-s}\varepsilon_{jkt}) = E(\boldsymbol{X}_{jk,t-s}\varepsilon_{jkt}) = E(\boldsymbol{Z}_{jk,t-s}\varepsilon_{jkt}) = 0,$$

[44] See Blundell and Bond (1998) and Arellano and Bover (1995) for details of the system GMM.

$E(\boldsymbol{\rho}_{\boldsymbol{jk},t-s}(\eta_{jk} + \varepsilon_{jkt})) = E(\boldsymbol{X}_{jk,t-s}(\eta_{jk} + \varepsilon_{jkt})) = E(\boldsymbol{Z}_{jk,t-s}(\eta_{jk} + \varepsilon_{jkt})) = 0,$ for s>1. (4)

The system GMM estimator controls for unobservable heterogeneity bias, inconsistency, and simultaneity, which enables us to produce efficient estimates. We use lagged variables as instruments for estimating the system. We use lagged levels and lagged first differences of predetermined and endogenous variables as instruments. The model is estimated using two-step GMM, which procures asymptotically more efficient estimates than one-step GMM.

3.2. Measures of Excess Stock Return Correlation

Variables used in the estimation and data sources are shown in Table 3. Excess stock returns are measured as U.S. dollar denominated stock returns minus the risk free rate, for which the three-month U.S. Treasury bill rate is used. Following Bekaert et al. (2009), we use weekly stock returns computed from national stock indices in order to avoid potential econometric problems resulting from the non-synchronous trading of securities when using very high frequency data. The indices are chosen from Bloomberg's list of *Indexes by Location* and their names are shown in Table 1. If multiple indices are listed for one country, one of them is chosen based on data availability and frequency of use in the academic literature. Using the computed weekly excess returns, we then calculate pairwise correlations coefficients for each year. The pairwise correlation coefficients (ρ_{jkt}) are all Fisher's z transformed to avoid the limited dependent variable problem.[45]

Table 3. Variables, Summary Statistics, and Data Sources

Variable	N	Mean	SD	Data Source
Stock Return Correlation	540	0.374	0.464	Bloomberg, CEIC, FRED
Bilateral Capital Flows	540	0.006	0.013	CPIS, World Development Indicators
Third-country Capital Flows	540	0.162	0.132	CPIS, World Development Indicators
RTA	540	0.391	0.488	CEPII
Economic Development	540	9.620	0.986	Penn World Table, World Development Indicators
Inflation Difference	540	3.030	2.795	World Development Indicators
Financial Depth	540	1.960	0.644	Global Financial Development Database
Krugman Index	540	0.075	0.041	UNIDO

3.3. Measures of Cross-Border Financial Flows

Measuring the degree of bilateral financial integration has been a long-standing challenge to economists. Some studies have used the degree of restrictions on cross-border financial transactions (e.g., Kose et al., 2009) or non-bilateral measures of financial openness (e.g., Dellas and Hess, 2005). However, these measures capture de jure restrictions on financial flows (e.g., Imbs, 2006) and make it difficult to identify the origin of financial flows. Moreover, as mentioned earlier, there are relatively few de jure restrictions on capital flows during our observation period. Therefore, the approach taken in this study is to use data on

[45] Simple correlation coefficients can be non-constant over time as they may be subject to amplification during periods of high market volatility (Forbes and Rigobon, 2002). One way to tackle this problem is to use conditional correlations. Bekaert et al. (2009) argue that factor models capture the expected correlation and the residual error terms (if >0) can be considered as the effect of contagion, which hikes stock return volatility (and simple correlation coefficients). Another way is to control for the impact of time-variant interdependence among equity markets, which is the most important time-variant transmission channel of stock returns that can cause volatility (Longin and Solnik, 1995). Conceptually, the approach taken in this study is similar to the latter approach.

direct bilateral asset holdings from the IMF's CPIS as a quantitative measure of financial integration. The data are available from 2001, which restricts our observation period to the period from 2001 to 2012.[46] The IMF compiles data not only on portfolio investment but, since 2009, also data on foreign direct investment (FDI).[47] In the analysis below, we mainly rely on the data on portfolio capital flows but also use capital flow data including FDI to check the sensitivity of our baseline results.

Bilateral capital flows between countries *j* and *k* are measured by $\frac{F_{jkt}+F_{kjt}}{Y_{jt}+Y_{kt}}$, where F_{jkt} denotes country *j* residents' portfolio investment assets held in country *k* and, conversely, F_{kjt} denotes country k residents' portfolio investment assets held in country *j*. *Y* denotes the GDP of each country. We also use a measure that includes both portfolio investment and FDI, which is defined as $\frac{F_{jkt}+F_{kjt}}{Y_{jt}+Y_{kt}} + \frac{D_{jkt'}+D_{kjt'}}{Y_{jt'}+Y_{kt'}}$, where $D_{jkt'}$ denotes country *j*'s direct investment assets held in country k, and vice versa. Since direct investment are available only from 2009 onward, we use the average of $\frac{D_{jkt'}+D_{kjt'}}{Y_{jt'}+Y_{kt'}}$ during period *t'* (from 2009 to 2012) for the FDI measure for all periods.

Capital flows from the G7 countries (labeled *g*) to a pair of EA countries *j* and *k* are measured by $\frac{F_{gkt}+F_{gjt}}{Y_{jt}+Y_{kt}}$, where F_{gjt} (F_{gkt}) denotes country g's portfolio investment assets held in country *j* (*k*). Note that we do not include the EA countries' portfolio investment assets held in the G7 countries because many data points are missing and even if they exist, the absolute amount is small.[48] Capital flows data including FDI from the G7 countries to EA are constructed by the same method as above.

3.4. Control Variables

Stock market co-movements are the result of not only capital flows but also a number of other factors. In order to avoid the omitted variable bias that would result from ignoring such factors, we include a vector of control variables in the regression.

First, the literature often stresses the importance of economic fundamentals, particularly the role of industry structure in explaining international stock return synchronization.[49] Roll (1992) argues that similarities in the industrial structure can lead to a high correlation in stock returns. However, examining data for 12 European countries, Heston and Rouwenhorst (1994) find that industrial structure does not appear to play a significant role in stock return co-movements.

---

[46] While the CPIS reports bilateral equity holdings and debt securities holdings separately, we use aggregate portfolio investment data due to numerous gaps in the separate data series. As Imbs (2006) documented, the components of the portfolio investment data in the CPIS (equity and debt investments) are strongly correlated with each other and the amount of equity transaction is much larger than debt transaction in general, which rationalizes the use of aggregate portfolio investment data instead of equity flows data for this study.

[47] The survey is called *Coordinated Direct Investment Survey*.

[48] In the empirical estimation, we also use data that include capital flows in both directions (from the G7 to EA, vice versa). The results are similar to the case when we use portfolio investment assets only.

[49] In addition to economic fundamentals, recent literature evidences that investor demand in addition to the effect of economic fundamentals can be a driver of the synchronization (e.g., Greenwood, 2005; Greenwood, 2008; Boyer, 2011; Hau and Lai, 2016; Bartram, Griffin, Lim, and Ng, 2015). Bartram, Griffin, Lim, and Ng (2015)

More recently, Dutt and Mihov (2013) use time-varying country-pair-specific industrial composition measures to test if Roll (1992) or Heston and Rouwenhorst (1994) stands and confirm the findings of Roll (1992). In this study, following Imbs (2006), we use the Krugman index (Krugman, 1991) to measure similarities in industrial specialization (*Krugman Index*). We define the index as follows: $S_{jkt}^{1} = \sum_{n=1}^{7}\left|s_{njt} - s_{nkt}\right|$, where s_{njt} and s_{nkt} denote the output share of ISIC 1 digit-level industry *n* in country *j*'s and country *k*'s total output respectively. The data are taken from the United Nations' *Statistical Yearbook*. The expected sign of the estimated coefficient of this variable is negative. If countries *j* and *k* have similar industrial structures (so that the *Krugman index* is small), sector-specific shocks will move stock returns in both countries in the same direction and therefore create a high correlation of stock returns.

Second, the role of multilateral trade liberalization is considered. From a theoretical perspective, participating in regional trade agreements, by lowering the cost of imported goods, is likely to increase the expected future stock returns of member countries, thereby increasing synchronization of stock returns (Basu and Morey, 2005). Previous research suggests that this theoretical prediction is empirically supported (Henry, 2000; Berben and Jansen, 2005). We use a dummy variable that takes 1 when a pair of countries has a bilateral trade agreement or belongs to the same regional trade agreement otherwise 0 (*RTA*). The expected sign of the coefficient is positive.

Third, we use three variables to control for different macroeconomic fundamentals of countries in each pair: (1) the pairwise sum of the logged real per capita GDP in U.S. dollars as a proxy for the economic development of each pair of countries (*Economic Development*); (2) the absolute difference in annual changes in the CPI as the proxy for differences in inflation rates of each pair of countries (*Inflation Difference*); and (3) the sum of the ratios of domestic credit to the private sector to output as a proxy for the availability of domestic financial intermediation (*Financial Depth*). The sign for the *Economic Development* variable is expected to be positive, that for the *Inflation Difference* variable negative, and that for the *Financial Depth* variable positive.

4. Estimation Results

4.1. Test for Strict Exogeneity

Before estimating the model, we test the strict exogeneity of the capital flows data by examining whether the effects of bilateral capital flows among EA economies and third-country effects of capital flows from the G7 countries are related to past stock return co-movements. Theoretically, stock return co-movements could lead to increased or decreased third-country effects. From the perspective of portfolio diversification, if two countries exhibit similar stock return movements, there is less incentive for investors to invest in both countries at the same time, implying that stock return co-movements have a negative effect on capital inflows from the G7 countries. However, theories on crisis contagion focusing on information cascades suggest that investors in advanced economies may classify two small countries that show similar stock return movements in the same investment category and therefore change their investment in these countries simultaneously, which means that stock return co-movements would be positively correlated with capital inflows from the G7 countries.

Using the method described in Wooldridge (2002), we run the following panel regression to test strict exogeneity:

$$\boldsymbol{Y_t} = \alpha + \beta \boldsymbol{X}_{t+1} + \gamma \boldsymbol{Z}_t + \eta_{jk} + \varepsilon_{jkt}, \quad (6)$$

where $\boldsymbol{Y_t}$ is the pairwise correlation of stock returns at time *t*, $\boldsymbol{X}_{t+1}$ is a subset of the bilateral and third-country capital flows and control variables at time *t*+1, and $\boldsymbol{Z_t}$ is the bilateral and third-country capital flows and control variables at time *t*. The null hypothesis of strict exogeneity is that β is near zero and insignificant, since stock return co-movements should not be correlated with the future realization of a subset of the bilateral and third-country capital flows and control variables.

Table 4 shows that the coefficient estimates for the future values of the bilateral and third-country capital flows are all statistically insignificant, indicating that they are strictly exogenous. Coefficients on most control variables are also insignificant, also implying strict exogeneity. Note that the future values of *RTA* and *Inflation Difference* variables are significantly different from zero, but their signs are opposite to the theoretically predicted value. Given these results, all explanatory variables are assumed to be strictly exogenous and lagged $\boldsymbol{Y}$ is endogenous, so that GMM-type instruments are used only for lagged dependent variable $\boldsymbol{Y}$ for dynamic models.[50]

4.2. Baseline Estimation

Table 5 reports the regression results of the baseline model. We first examine the model with bilateral capital flows only (first four columns) and then the model with both bilateral and third-country capital flows (last four columns). We use both static and dynamic panel regression models for the two sets of control variables (with and without the *Financial Depth* and *Krugman Index* variables, while the *RTA*, *Economic Development*, and *Inflation Difference* variables are always included). For the static models, the standard Hausman test supports the use of a random effects model. For the dynamic models, a one-year lag of the dependent variable is included in the regression, while the set of two- and three-year lags of the dependent variable (GMM-type) and one-year lags of all explanatory variables (IV-type) are used as instruments.

The regression results show the following observations. The coefficients on bilateral portfolio investment flows are marginally significant when third-country capital flows are excluded. The coefficients are positive, implying that more bilateral financial flows increase stock return co-movements in EA countries. However, the positive effect of bilateral financial flows disappears when third-country effects are included. In the regressions with both bilateral and third-country capital flows, bilateral financial flows all become insignificant and in some cases have negative signs, while third-country effects are all positive and significant at the 1% level. These results are consistent across both the static and dynamic models and both sets of control variables.

This result strongly suggests that positive excess stock return co-movements in EA countries are mainly due to capital flows from the G7 countries and not due to bilateral financial flows among EA countries. This result is similar to findings of previous studies using different approaches (Forbes and Chinn, 2004; Dellas and Hess, 2005; Froot and Ramadorai, 2008).

[50] In the sensitivity analysis, we examine the case assuming that bilateral capital flows among EA economies are predetermined.

When all five control variables are used, the positive effects of G7 capital flows are stronger than in the case with only three control variables.

The most plausible explanation of the insignificant coefficient on bilateral capital flows is as follows: regional and bilateral integration of financial markets in EA countries is still incomplete and the size of financial flows among EA countries is quite small compared to capital flows from the G7 countries (Figure 2). That is, EA financial markets are more integrated with the United States and other G7 markets than with each other.

Table 4. Testing Strict Exogeneity

	1	2	3	4
Third-country Capital Flows (t+1)	0.318		0.154	0.389
	(0.711)		(0.351)	(0.902)
Bilateral Capital Flows (t+1)		6.020	5.744	5.367
		(0.911)	(0.845)	(0.807)
RTA (t+1)				-0.207 ***
				(-3.228)
Economic Development (t+1)				-1.250
				(-1.467)
Inflation Difference (t+1)				0.036 ***
				(3.397)
Financial Depth (t+1)				0.001
				(0.308)
Krugman Index (t+1)				-1.291
				(-0.287)
Third-country Capital Flows (t)	0.751 *	1.067 ***	0.923 **	0.658
	(1.874)	(5.203)	(2.153)	(1.531)
Bilateral Capital Flows (t)	0.625	-5.329	-4.999	-4.132
	(0.448)	(-0.830)	(-0.758)	(-0.642)
RTA (t)	0.079 *	0.077 *	0.072 *	0.279 ***
	(1.872)	(1.840)	(1.727)	(4.450)
Economic Development (t)	0.071 ***	0.068 **	0.068 **	1.305
	(2.366)	(2.304)	(2.281)	(1.540)
Inflation Difference (t)	0.002	0.002	0.002	-0.018 ***
	(0.330)	(0.376)	(0.463)	(-2.400)
Financial Depth (t)	0.000	0.000	0.000	0.000
	(0.891)	(0.937)	(0.901)	(-0.067)
Krugman Index (t)	-1.378 ***	-1.382 ***	-1.409 ***	-0.367
	(-2.797)	(-2.861)	(-2.874)	(-0.081)

Notes: Numbers in the brackets are t-statistics.

Therefore, bilateral capital flows among EA countries do not explain stock return correlations, while capital flows from third countries (i.e., the G7 countries) do play a significant role in stock return correlations.

The coefficients on the control variables are plausible in most cases. The coefficients on *Economic Development* are positive and significant, implying that the stock return correlation tends to be higher for richer country pairs in the region. The coefficients on the *Krugman index* are, as expected, negative, although most of them are insignificant.[51] The coefficients on the *RTA* variable are positive and significant, which is consistent with the theoretical prediction that participation in the same regional and bilateral trade agreements should lead to a higher stock return correlation (Dutt and Mihov, 2013).[52] Finally, most of the coefficients on the *Financial Development* variable are insignificant.

4.3. Sensitivity Analysis

Having established that capital flows from the G7 countries play a significant role in stock return co-movements in the EA region, we are interested in which country or countries have the most important effect. Consequently, Tables 6 shows the regression results when we replace capital flows from the G7 countries overall with capital flows from the United States only, Japan only, and the sum of four European countries (France, Germany, Italy, and the United Kingdom) only. In all cases, the coefficients on third-country effects are significant and positive, implying that all three country blocks have significant effects. Bilateral capital flows are all insignificant and the signs of the coefficients are positive in the static models but negative in the dynamic models.

The coefficients on the control variables in all cases are similar to those of the baseline result. Next, we extend the analysis to other emerging markets in other regions. The first four columns of Table 7 show the case when we extend our sample which includes non-EA BRICS countries, i.e., Brazil, Russia, and South Africa, in addition to the 10 EA countries.

Now, with 13 countries, we have 78 country pairs for a 12-year observation period. The results in Table 7 show that the main conclusion still holds even for this extended sample: the coefficients on third-country capital flows are still positive and significant. Moreover, the coefficients on bilateral flows are insignificant in most cases. One interesting result is that the coefficient of the *Krugman index* is negative and significant. Because the newly included countries have very different industrial structures from the EA countries, sectoral differences among the countries in the sample are much more pronounced, explaining the significantly negative coefficient on the *Krugman index*.

The last four columns in Table 7 report the case when we expand the capital flow data to include FDI. The inclusion of FDI in capital flow data is important both from a theoretical and an empirical perspective as shown in Imbs (2006) and Otto et al. (2001). Ideally, it would be preferable to consider portfolio investment and FDI separately. The empirical results show that the main result still holds with FDI data included: third-country effects are significant and positive. The actual size of the coefficients decreases, but this is due to the fact that the absolute size of capital flows is now larger, since FDI is included.

[51] A possible reason is the rough sectoral classification that we used. Introducing more detailed classification as in Dutt and Mihov (2013) may produce different results.

[52] Several previous studies such as Forbes and Chinn (2004) and Walti (2011) have used trade flows as explanatory variables. However, the coefficients reported in those studies are not significant and in many cases negative (not shown). From a theoretical perspective, trade flows could have a positive or negative effect on stock return co-movements, depending on the type of trade. In this study, we do not explicitly include trade flows because of potential endogeneity problems arising from simultaneity with RTA and industry structure. The potential endogeneity problems arising when including trade flow are well documented in Beine and Candelon (2011) and Lane and Milesi-Ferretti (2008).

Table 5. Stock Market Correlation Regressions

	w/o Third-country Capital Flows				w/ Third-country Capital Flows			
	Static Models		*Dynamic Models*		*Static Models*		*Dynamic Models*	
Portfolio Investment								
Third-country Capital Flows					1.066 ***	1.000 ***	1.119 ***	1.087 ***
					(6.428)	(5.251)	(5.974)	(5.255)
Bilateral Capital Flows	3.653 **	3.330 *	3.146	2.921 *	0.277	-0.101	-0.987	-1.154
	(2.099)	(1.849)	(1.610)	(1.750)	(0.228)	(-0.087)	(-0.760)	(-0.985)
Controls								
RTA	0.101 *	0.109 **	0.093	0.114 **	0.120 ***	0.139 ***	0.157 ***	0.167 ***
	(1.851)	(2.289)	(1.479)	(2.269)	(2.983)	(3.375)	(3.518)	(3.783)
Economic Development	0.127 ***	0.128 ***	0.156 ***	0.137 ***	0.068 ***	0.069 ***	0.085 ***	0.078 ***
	(4.224)	(5.569)	(4.617)	(4.886)	(2.366)	(2.829)	(2.645)	(2.713)
Inflation Difference	-0.004	-0.007	-0.008	-0.018 ***	0.001	-0.005	-0.014 *	-0.015 **
	(-0.676)	(-1.243)	(-0.394)	(-2.476)	(0.112)	(-0.836)	(-1.942)	(-2.255)
Financial Depth	0.032		-0.028		0.033		0.002	
	(0.713)		(-0.452)		(0.904)		(0.057)	
Krugman Index	-0.705		-0.450		-1.113 ***		-0.609	
	(-1.045)		(-0.518)		(-2.567)		(-1.119)	
Lagged Dependent Variable								
Dep Var (t-1)			-0.139 ***	-0.118 ***			-0.113 ***	-0.113 ***
			(-2.480)	(-2.794)			(-2.560)	(-2.607)
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	540	540	495	495	540	540	495	495
R-squared	0.293	0.291	0.273	0.275	0.337	0.331	0.271	0.314
AR(1) test (p-value)			0.000	0.000			0.000	0.000
AR(2) test (p-value)			0.318	0.288			0.232	0.291
Hansen test of over-identification (p-value)			0.292	0.668			0.707	0.667
Diff-in-Hansen test of exogeneity (p-value)			0.109	0.591			0.604	0.591

Notes: Numbers in the brackets are t-statistics.

Finally, Table 8 displays two additional sensitivity analyses. In the first case, we exclude time fixed effects but include financial crisis dummies (2008, 2009=1, otherwise 0); in the second case, we assume bilateral capital flows among EA economies are predetermined (= correlated with past errors, and not correlated with current and future errors). Both cases show that the main conclusion still stands.

Table 6. Third-Country Effects by Country/Region

	Third-country Capital Flows from					
	USA		Japan		Europe	
	Static	*Dynamic*	*Static*	*Dynamic*	*Static*	*Dynamic*
Portfolio Investment						
Third-country Capital Flows (USA)	1.799 *** (6.514)	1.755 *** (5.629)				
Third-country Capital Flows (Japan)			9.120 *** (3.750)	13.573 *** (4.618)		
Third-country Capital Flows (Europe)					3.021 *** (5.966)	4.353 *** (5.261)
Bilateral Capital Flows	0.385 (0.327)	-0.684 (-0.566)	0.059 (0.045)	-2.559 (-1.415)	1.242 (0.958)	-1.072 (-0.748)
Controls						
RTA	0.125 *** (3.108)	0.163 *** (3.547)	0.099 ** (2.276)	0.121 *** (2.747)	0.106 *** (2.573)	0.149 *** (3.648)
Economic Development	0.064 ** (2.171)	0.085 *** (2.579)	0.090 *** (2.736)	0.088 ** (2.297)	0.084 *** (3.222)	0.087 *** (2.753)
Inflation Difference	0.001 (0.232)	-0.013 * (-1.926)	0.001 (0.084)	-0.011 (-1.609)	-0.002 (-0.270)	-0.015 ** (-2.126)
Financial Depth	0.033 (0.909)	0.004 (0.095)	0.049 (1.172)	0.021 (0.426)	0.026 (0.724)	-0.007 (-0.159)
Krugman Index	-0.975 *** (-2.332)	-0.455 (-0.837)	-1.222 *** (-2.523)	-0.968 * (-1.648)	-1.187 *** (-2.596)	-0.842 (-1.539)
Lagged Dependent Variable						
Dep Var (t-1)		-0.112 *** (-2.546)		-0.113 *** (-2.465)		-0.114 *** (-2.622)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	540	495	540	495	540	495
R-squared	0.335	0.319	0.316	0.302	0.327	0.321
AR(1) test (p-value)		0.000		0.000		0.000
AR(2) test (p-value)		0.288		0.345		0.265
Hansen test of over-identification (p-value)		0.629		0.670		0.771
Diff-in-Hansen test of exogeneity (p-value)		0.565		0.632		0.665

Notes: Numbers in the brackets are t-statistics.

Table 7. Sensitivity Analysis (Sample Countries, Definition of Investment)

	Emerging Asia + BRICS				Portfolio Investments + FDI			
	Bilateral		Third+Bilateral		Bilateral		Third+Bilateral	
	Static	Dynamic	Static	Dynamic	Static	Dynamic	Static	Dynamic
Portfolio Investment								
Third-Country Effects			0.835 ***	0.845 ***				
			(4.278)	(3.308)				
Bilateral Capital Flows	3.197 *	1.792	0.379	0.480				
	(1.926)	(0.680)	(0.293)	(0.268)				
Third-Country Effects (w/ FDI)							0.466 ***	0.491 ***
							(3.873)	(3.589)
Bilateral Capital Flows (w/ FDI)					0.163	0.014	-0.980	-1.385 **
					(0.126)	(0.010)	(-1.563)	(-2.089)
Controls								
RTA	0.003	0.011	0.025	0.047	0.128 **	0.117 *	0.119 ***	0.138 ***
	(0.061)	(0.161)	(0.585)	(0.988)	(2.041)	(1.694)	(2.381)	(2.591)
Economic Development	0.158 ***	0.184 ***	0.117 ***	0.132 ***	0.145 ***	0.167 ***	0.080 ***	0.095 ***
	(6.031)	(5.442)	(4.147)	(3.902)	(4.859)	(5.099)	(2.702)	(2.865)
Inflation Difference	-0.005	0.000	-0.003	-0.002	-0.006	-0.020 ***	-0.002	-0.016 **
	(-0.919)	(0.037)	(-0.672)	(-0.255)	(-0.958)	(-2.674)	(-0.275)	(-2.145)
Financial Depth	0.008	0.024	-0.012	-0.013	0.037	-0.028	0.055	0.014
	(0.243)	(0.558)	(-0.357)	(-0.329)	(0.800)	(-0.541)	(1.379)	(0.307)
Krugman Index	-1.578 ***	-2.047 ***	-1.762 ***	-1.698 ***	-0.588	-0.081	-0.651	-0.109
	(-3.135)	(-2.536)	(-4.010)	(-3.194)	(-0.868)	(-0.099)	(-1.323)	(-0.181)
Lagged Dependent Variable								
Dep Var (t-1)		-0.132 ***		-0.150 ***		-0.111 ***		-0.112 ***
		(-4.643)		(-4.925)		(-2.485)		(-2.572)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	936	858	936	858	540	495	540	495
R-squared	0.249	0.238	0.263	0.253	0.283	0.262	0.316	0.302
AR(1) test (p-value)		0.000		0.000		0.000		0.000
AR(2) test (p-value)		0.009		0.270		0.222		0.251
Hansen test of over-identification (p-value)		0.897		0.412		0.778		0.740
Diff-in-Hansen test of exogeneity (p-value)		0.012		0.720		0.719		0.668

Notes: Numbers in the brackets are t-statistics

Table 8. Sensitivity Analysis (Time Dummies Only for the Crisis Period)

	Crisis				Predetermined Bilateral Effects	
	Bilateral		Third+Bilateral		Bilateral	Third+Bilateral
	Static	Dynamic	Static	Dynamic	Dynamic	Dynamic
Portfolio Investment						
Third-Country Effects			1.128 ***	0.863 ***		1.189 ***
			(6.282)	(4.412)		(3.820)
Bilateral Capital Flows	3.371 *	1.851	-0.361	-1.437	-1.549	-3.559
	(1.796)	(1.233)	(-0.276)	(-1.157)	(-0.050)	(-0.900)
Controls						
RTA	0.137 ***	0.137 **	0.144 ***	0.173 ***	0.140 **	0.182 ***
	(2.392)	(2.253)	(3.222)	(3.531)	(2.280)	(3.290)
Economic Development	0.145 ***	0.167 ***	0.074 ***	0.113 ***	0.167 ***	0.088 ***
	(4.609)	(4.805)	(2.460)	(3.238)	(4.481)	(2.510)
Inflation Difference	-0.015 **	-0.038 ***	-0.008	-0.034 ***	-0.018 ***	-0.012 **
	(-2.012)	(-4.660)	(-1.200)	(-4.422)	(-2.685)	(-1.980)
Financial Depth	0.018	-0.036	0.028	-0.016	-0.024	0.006
	(0.373)	(-0.665)	(0.701)	(-0.353)	(-0.460)	(0.140)
Krugman Index	-0.630	0.000	-1.031 **	-0.197	0.021	-0.498
	(-0.813)	(0.000)	(-2.159)	(-0.300)	(0.030)	(-0.900)
Dummy 2008	-0.032	-0.109	0.038	-0.030		
	-0.387	-1.175	0.468	-0.327		
Dummy 2009	0.003	0.002	-0.013	0.009		
	0.048	0.033	-0.198	0.146		
Lagged Dependent Variable						
Dep Var (t-1)		-0.119 ***		-0.116 ***	-0.113 ***	-0.108 ***
		(-2.763)		(-2.699)	(-2.380)	(-2.330)
Time Fixed Effects	No	No	No	No	Yes	Yes
N	540	495	540	495	495	495
R-squared	0.140	0.112	0.188	0.157	0.112	0.157
AR(1) test (p-value)		0.000		0.000	0.000	0.000
AR(2) test (p-value)		0.489		0.478	0.489	0.478
Hansen test of over-identification		0.870		0.739	0.870	0.739
Diff-in-Hansen test of exogeneity (p-value)		0.915		0.789	0.915	0.789

Notes: Numbers in the brackets are t-statistics.

5. Conclusion

The objective of this paper is to analyze the sources of stock return synchronization in EA countries – that is, whether such synchronization is due to increased bilateral capital flows among EA countries or due to synchronized capital flows from the G7 economies into EA countries.

The regression results show that the main force behind stock return co-movements in EA is the third-country effect, not bilateral capital flows. Although there has been considerable progress in Asian financial market integration in recent years as a result of initiatives for regional economic and financial cooperation, capital flows among EA countries are still comparatively small and, as shown in the empirical analysis, do not play a significant role in the co-movement of stock market returns. Instead, stock market co-movements are still largely explained by capital flows from the G7 countries.

The results of the various models in this study highlight the need for a more in-depth examination of the sources of stock return co-movements in the countries of Emerging Asia. First, as highlighted by Kalemli-Ozcan et al. (2013), in addition to portfolio investment and FDI, cross-border bank lending may play an important role in stock market co-movements, so that ideally these should be included in the analysis in order to gain a comprehensive understanding of the impact of capital flows from third countries. Next, uncertainty shocks can play an important role in explaining asset price co-movements (Hirata et al., 2013) and therefore should also be included in the analysis. However, creating uncertainty measures for emerging economies presents a challenge.

References

Arellano, M. and O. Bover. 1995. "Another Look at the Instrumental Variable Estimation of Error-Components Models." Journal of Econometrics, 68: 29-51.

Bartram, S.M., J. M. Griffin, T-H. Lim, D. T. Ng. 2015. "How Important Are Foreign Ownership Linkages for International Stock Returns?" Review of Financial Studies, 28(11), 3036–3072.

Basu, P., and M. R. Morey. 2005. “Trade Opening and the Behavior of Emerging Stock Market Prices.” Journal of Economic Integration, 20(1): 68–92.

Beine, M., and B. Candelon. 2011. “Liberalisation and Stock Market Co-Movement between Emerging Economies.” Quantitative Finance, 11(2): 299–312.

Bekaert, G. 1995. “Market Integration and Investment Barriers in Emerging Equity Markets.” World Bank Economic Review, 9(1): 75–107.

Bekaert, G., and C. R. Harvey. 1997. “Emerging Equity Market Volatility.” Journal of Financial Economics, 43(1): 29–77.

Bekaert, G., and C. R. Harvey. 2000. “Foreign Speculators and Emerging Equity Markets.” The Journal of Finance, 55(2): 565–613.

Bekaert, G., and C. R. Harvey. 2002. “Research in Emerging Markets Finance: Looking to the Future.” Emerging Markets Review, 3(4): 429–448.

Bekaert, G., and C. R. Harvey. 2014. “Emerging Equity Markets in a Globalizing World.” Mimeo.

Bekaert, G., and X. S. Wang. 2009. “Globalization and Asset Prices.” Mimeo.

Bekaert, G., C. R. Harvey, and C. Lundblad. 2005. “Does Financial Liberalization Spur Growth?” Journal of Financial Economics, 77(1): 3–55.

Bekaert, G., C. R. Harvey, and R. L. Lumsdaine. 2002. “The Dynamics of Emerging Market Equity Flows.” Journal of International Money and Finance, 21(3): 295–350.

Bekaert, G., R. J. Hodrick, and X. Zhang. 2009. “International Stock Return Co-movements.” Journal of Finance, 64(6): 2591–2626.

Berben, R. P., and W. J. Jansen. 2005. “Co-movement in International Equity Markets: A Sectoral View.” Journal of International Money and Finance, 24(5): 832–857.

Blundell, R. and S. Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” Journal of Econometrics, 87(1): 115–143.

Boubakri, S. and C. Guillaumin. 2015. “Regional Integration of the East Asian Stock Markets: An Empirical Assessment.” Journal of International Money and Finance, 57: 136-160.

Boyer B. 2011. “Style-related Co-movement: Fundamentals or Labels?” Journal of Finance, 66: 307-32.

Brooks, R., and M. Del Negro. 2004. “The Rise in Co-movement Across National Stock Markets: Market Integration or IT Bubble?” Journal of Empirical Finance, 11(5): 659–680.

Brooks, R., and M. Del Negro. 2005. “Country versus Region Effects in International Stock Returns.” The Journal of Portfolio Management, 31(4): 67–72.

Brooks, R., and M. Del Negro. 2006. “Firm-Level Evidence on International Stock Market Co-movement.” Review of Finance, 10(1): 69–98.

Chien, M.-S., C.-C. Lee, T.-C. Hu, H.-T., Hu. 2015. “Dynamic Asian Stock Market Convergence: Evidence from Dynamic Cointegration Analysis among China and ASEAN-5.” Economic Modelling, 51: 84-98.

Davis, J. S. 2014. “Financial Integration and International Business Cycle Co-movement.” Journal of Monetary Economics, 64: 99–111.

De Santis, G., and S. Imrohoroglu. 1997. "Stock Returns and Volatility in Emerging Financial Markets." Journal of International Money and Finance, 16(4): 561–579.

Dellas, H., and M. Hess. 2005. "Financial Development and Stock Returns: A Cross-Country Analysis." Journal of International Money and Finance, 24(6): 891–912.

Dewandaru,G., R. Masih,and A.M.M. Masih. 2015. "Why is No Financial Crisis a Dress Rehearsal for the Next? Exploring Contagious Heterogeneities across Major Asian Stock Markets." Physica A, 419: 241-259.

Dutt, P., and I. Mihov, 2013. "Stock Market Co-movements and Industrial Structure." Journal of Money, Credit and Banking, 45(5): 891–911.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton. 1988. "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty." Quarterly Journal of Economics, 103(1): 51–78.

Flavin, T. J., M. J. Hurley, and F. Rousseau. 2002. "Explaining Stock Market Correlation: A Gravity Model Approach." The Manchester School, 70(S1): 87–106.

Forbes, K. J., and M. D. Chinn. 2004. "A Decomposition of Global Linkages in Financial Markets Over Time." The Review of Economics and Statistics, 86(3): 705–722.

Forbes, K. J., and R. Rigobon. 2002. "No Contagion, Only Interdependence: Measuring Stock Market Co-movements." The Journal of Finance, 57(5): 2223–2261.

Froot, K. A., and T. Ramadorai. 2008. "Institutional Portfolio Flows and International Investments." The Review of Financial Studies, 21(2): 937–971.

Gérard, B., K. Thanyalakpark, and J. A. Batten. 2003. "Are the East Asian Markets Integrated? Evidence from the ICAPM." Journal of Economics and Business, 55(5–6): 585–607.

Greenwood R.. 2008. "Excess Co-movement of Stock Returns: Evidence from Cross-sectional Variation in Nikkei 225 Weights." Review of Financial Studies, 21: 1153-1186.

Greenwood, R.. 2005. "Short and Long-term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage." Journal of Financial Economics, 75: 607-649.

Hau, Harald, and S. Lai. 2017. "The Role of Equity Funds in the Financial Crisis Propagation." Review of Finance, 21(1): 77–108,

Henry, P. B. 2000. "Stock Market Liberalization, Economic Reform, and Emerging Market Equity Prices." Journal of Finance, 55(2): 529–564.

Heston, S. L., and K. G. Rouwenhorst. 1994. "Does Industrial Structure Explain the Benefits of International Diversification?" Journal of Financial Economics36(1): 3-27.

Hirata, H., M. A. Kose, C. Otrok, and M. Terrones. 2013. "Global House Price Fluctuations: Synchronization and Determinants." NBER International Seminar on Macroeconomics 2012, University of Chicago Press, 119–166.

Imbs, J. 2006. "The Real Effects of Financial Integration." Journal of International Economics, 68(2): 296–324.

Jiang,Y., H. Nie, and J.Y. Monginsidi. 2017. "Co-movement of ASEAN Stock Markets: New Evidence from Wavelet and VMD-based Copula Tests." Economic Modelling, 6(4): 384-398.

Kalemli-Ozcan, S., E. Papaioannou, and J.-L. Peydro. 2013. "Financial Regulation, Financial Globalization, and the Synchronization of Economic Activity." Journal of Finance, 68(3): 1179–1228.

Karolyi, G. A., and R. M. Stulz. 1996. "Why Do Markets Move Together? An Investigation of US-Japan Stock Return Co-movements." Journal of Finance, 51(3): 951–986.

King, M., E. Sentana, and S. Wadhwani. 1994. "Volatility and Links between National Stock Markets." Econometrica, 62(4): 901–933.

Kose, M. A., E. S. Prasad, and M. E. Terrones. 2009. "Does Openness to International Financial Flows Raise Productivity Growth?" Journal of International Money and Finance, 28(4): 554–580.

Krugman, P. R. 1991. "Increasing Returns and Economic Geography." Journal of Political Economy, 99(3): 483–499.

Lane, P. R, and G. M. Milesi-Ferretti. 2007. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities: 1970–2004."Journal of International Economics, 73(2): 223–50.

Lane, P. R., and G. M. Milesi-Ferretti. 2008. "The Drivers of Financial Globalization." American Economic Review, 98(2): 327–332.

Lee, G. and J. Jeong. 2016. "An Investigation of Global and Regional Integration of ASEAN Economic Community Stock Market: Dynamic Risk Decomposition Approach." Emerging Markets Finance and Trade, 52: 2069-2086.

Longin, F. and Bruno Solnik. 1995. "Is the Correlation in International Equity Returns Constant: 1960–1990?" Journal of International Money and Finance, 14(1): 3-26.

Mandigma, M. B. S. 2014. "Stock Market Linkages among the ASEAN 5+3 Countries and US: Further Evidence." Management and Administrative Sciences Review, 3(1): 53-68.

Milesi-Ferretti, G.M., F. Strobbe, and N. T. Tamirisa. 2010. "Bilateral Financial Linkages and Global Imbalances: A View on the Eve of the Financial Crisis." IMF Working Paper, No.10/257.

Otto, G., G. Voss, and L. Willard. 2001. "Understanding OECD Output Correlations." RBA Research Discussion Papers, 2001-05.

Rioja, F., and N. Valev. 2004. "Does One Size Fit All? A Reexamination of the Finance and Growth Relationship." Journal of Development Economics, 74(2): 429–447.

Roll, R. 1992. "Industrial Structure and the Comparative Behavior of International Stock Market Indices." Journal of Finance, 47(1): 3–41.

Tiwari, A. K., A., A. B. Dar, N. Bhanja, and A. Shah. 2013. "Stock Market Integration in Asian Countries: Evidence from Wavelet Multiple Correlations." Journal of Economic Integration 28(3): 441-56.

Walti, S. 2011. "Stock Market Synchronization and Monetary Integration." Journal of International Money and Finance, 30(1): 96–110.

Wintoki, M. B., J. S. Linck, and J. M. Netter. 2012. "Endogeneity and the Dynamics of Internal Corporate Governance." Journal of Financial Economics, 105(3): 581–606.

Wooldridge, J. M. 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge

Assessing the Degree of Financial Integration in ASEAN– A Perspective of Banking Competitiveness

Tiantian Zhang, Nottingham University Business School China, University of Nottingham Ningbo China

Kent Matthews, Cardiff Business School, Cardiff University

ABSTRACT

This paper assesses the degree of the ASEAN regional financial integration by investigating the co-movements of the interest rates and the convergence properties of banking market competitiveness over the period of 1994-2016. The banking competitiveness is modelled by the non-structural Panzar-Rosse (PR) model, and the estimated *H-statistics* are used to test for β- and σ- convergence. The financial integration process does not necessarily foster competition, but a convergence trend toward monopolistic competitive market has been seen. The ASEAN banking markets remain sensitive to the global environment and the convergence tendency is easily distorted by the financial crises. further regional coordination is still required to establish a strong financial market that better withstands shocks from outside the region.

Keywords: financial integration; banking competitiveness; convergence; H-statistics.

1. Introduction

After nearly 5 decades of development, the Association of Southeast Asian Nations (ASEAN) bloc has increased in economic significance, and achieved notable progress in regional co-operation and integration. The economic integration is where the bloc has made the most progress with the creation of the ASEAN Economic Community (AEC) on 31 December 2015. Further economic integration and acceleration of intra-regional single market and trading activities require a stable financing framework for the region. The trend towards financial integration in the region began following the 1997/98 Asian financial crisis, which revealed the inherent weakness of the financial system in ASEAN countries, especially in the banking systems. The process was initiated by the Chiang Mai Initiative (CMI) (2000), aims to establish a co-operative framework of liquidity support among member countries during periods of economic/financial distress, and help to avoid a future recurrence of the 1997 Asian crisis. Moreover, when the Euro was launched as a single currency since 1999, the notion of a similar currency union in the ASEAN bloc was raised and became a long-term strategy for policy makers[53], in the hope of promoting trade and investment across the member countries by reducing cross-border transaction costs and the risk of exchange rate volatilities.

In most ASEAN countries, capital markets remain underdeveloped and the banking system is still the principal vehicle of financial intermediation and the channel of monetary policy pass-through. The ASEAN banking sector has had significant structural changes since the 1997/98 Asian crisis and the global financial crisis of 2008-2009. The ASEAN banking market is characterized by privatization, deregulation, M&A, foreign firm entrance, as well as market integration (Khan, Ahmad and Gee, 2016). On the one hand, penetration of foreign banks may foster competition. The ASEAN Banking Integration Framework (ABIF) of 2014 also requires member countries to further liberalize their banking markets and achieve a semi-integrated banking market by 2020. On the other hand, national governments have been trying to consolidate their internal banking market through mergers and acquisitions, which may increase concentration and reduce competition. The issue is of concern to policy makers as well as financial participants as the state of competition influences financial stability, capital accessibility and monetary policy effectiveness. One of the main research objectives of this paper is to investigate to what extent the financial integration process has affected the banking markets in ASEAN countries, by examining the evolution in banking market competitiveness of the 5 founder countries (also known as ASEAN-5)[54].

If the regional coordination toward financial integration has been effective, it should have at least propelled the competitive structure of member countries' banking market toward a level that is compatible with each other. Individual banks then can be better prepared for a more liberalized reginal market under full integration, in which all individual banks are expected to charge the same price for similar products to avoid unbalanced capital flows. Economic efficiency will also be improved through price reductions if the integration process improves competition. Therefore, the convergence properties of banking market competitiveness could serve as a valuable indictor to assess the effectiveness of banking market integration. Similar

[53] The regionalism and economic integration also exist to large extent in larger geographical areas, e.g. the East Asia area, including all ASEAN countries plus People's Republic of China, Japan, South Korea and Taiwan. However, among the alternative geographical areas, ASEAN has the most possibility for a truly integrated financial market and regional currency arrangement, primarily because of the political wills (Bayoumi and Mauro, 2001).

[54] Namely, Indonesia, Malaysia, the Philippines, Singapore and Thailand. The 5 countries together account for 87% of the total GDP, 73.51% of the total population of ASEAN, and is a representative sample of the ASEAN.

notion has been used to study financial integration in the EU countries, e.g. Weill (2013), Andrieş and Căpraru (2014), but to the best of our knowledge, no such research has been done for ASEAN countries. The present study aims to fill this gap by providing a convergence analysis on ASEAN-5's banking market competitiveness over a long period of 1994-2016. The degree of competition in one country's market is measured by the *H-statistic*, obtained by using a non-structural approach, known as the Panzar-Rosse (PR) reduced-form revenue model. The *H-statistics* are then used to test for convergence properties which is used as an indicator for the financial market integration.

The rest of this paper is organised as follows. Section 2 reviews the literature on economic and financial integration indictors, and how banking market competitiveness has been empirically examined in the literature. Section 3 outlines the methodology used, i.e. the Panzar-Rosse (PR) model along with some recent improvements and the test procedures for β- and σ- convergence. Section 4 provides information on the data and Section 5 provides some preliminary analysis using price-based indicators for banking integration. Section 6 reports the estimated results for *H-statistics* and Section 7 analyses the convergence properties of ASEAN banking competitiveness. The final section summarises our findings.

2. Related literature

Conventionally, the degree of financial integration is assessed by either the "quantity-based" indicators, such as increasing cross-border lending activities and increasing share of investors' holding of non-domestic assets in the investment portfolio, or more often is assessed by the "price-based" indicators, which are either computed or model-based measures of (evolutions of) dispersions in assets returns, and a declining trend in dispersions, i.e. existence of convergence, is a signal of financial integration as the assets returns should be more influenced by common factors rather than country-specific factors. For example, banking market integration is indicated by a narrowing dispersion of interest rates on consumer credit, mortgages and deposits with agreed maturity.

The computed price-based measures are broadly used in policy researches, such as the annual reports on "Financial Integration in Europe issued by the European Central Bank (ECB) since 2005 and the "ASEAN Integration Report 2015" by the ASEAN Secretariat. The convergence tendencies are also tested by model-based methods, such as cointegration analysis (Centeno and Mello, 1999; Schuler and Heinemann, 2002) or other time-series techniques (Fratzscher, 2002). The concepts of "*β-convergence"* and *"σ-convergence"* (Sala-i-Martin, 1996) and the Phillips and Sul (2007) panel convergence tests have also been applied to study financial integration, especially banking market integration in European countries (Affinito and Farabullini, 2006; Rughoo and Saranties, 2012, 2014). Similar indicators have also been applied to Asian financial markets. The extant literature provides mixed results, some studies find evidence for short-term or partial financial integration in the ASEAN markets, although there is little evidence for sustainable long-term convergence (Rizavi et al. 2011, Tang, 2011, De Truchis and Keddad, 2013, Guesmi, Teulon and Muzaffar, 2014). Other studies argue that the intra-regional financial integration is outweighed by external integration with other major economies (e.g. Boresztain and Loungani, 2011), and the apparent intra-regional integration is actually driven by the external forces. Nevertheless, based on the traditional measures, the degree of integration in ASEAN financial markets, seems improving since the 1997 Asian crisis in despite of the driving forces, although not as strong as the real sector integration.

The convergence property of the 'price-based" measures are prevalent indicators for financial integration, as it satisfies the belief that market integration should lead to an "one-pricing" behaviour, normally through fostered competition brought by removal of cross-border barriers. However, for countries under the transitional process toward full market integration, what is also of great importance is to ensure financial stability and hedge against potential systematic risks during the transitional period. As argued by Karim and Zaini (2001), the increased competitive pressures will "affect banks (differently) depends in part on their ability to adapt and operate efficiently in the new environment". Banks, from countries in which the financial intermediaries are operating in a more competitive environment, and those in which the financial intermediation are conducted in a more efficient manner would be more productive and profitable given the available resources. The less competitive and less efficient banks would lose their market shares and face the risk of being driven out of the market or being taken-over eventually. Banks from countries with significant differences in economic fundamentals[55] operate under different technological and cost structures, therefore charging similar prices does not necessarily imply convergence in their profitability and productive efficiency. The price-based measures may only serve as good indicators for integration when the involved economies are already highly integrated in real sectors, e.g. an integrated labour market which allows for free labour flow and therefore compatible cost of labour; an integrated goods and services market which ensures compatible living standards and living costs across countries; an integrated regulatory framework which ensures banks face the same regulatory pressure regarding their pricing behaviour and so on. On the other hand, institutional and operational convergence in terms of competitiveness provides valuable information on how well the banks are prepared for challenges associated with the ongoing integrating process.

Studies of the convergence properties at the institutional and operational level are relatively limited in general, especially on for banking competitiveness. Weill (2013) estimates Lerner indices and PR *H-statistics* for the EU banking market during the early 2000s and find that there is no apparent increase in the market competitiveness. Andrieş and Căpraru (2014) investigate the level of banking competitiveness in EU27 from 2004 to 2010, also using non-structural PR *H-statistics*, and find evidence for Competition-Efficiency Hypothesis. Both studies find some evidence for existence of *β-convergence* and *σ-convergence* in the state of banking competitiveness of the EU economies. For Asian countries, Matthews and Zhang (2010) test for convergence in bank efficiency and productivity in China; Zhang and Matthews (2012) apply these tests to investigate the convergence properties in bank efficiency in Indonesia. Cross-country studies on convergence properties in bank competitiveness has not been seen for the ASEAN.

Majority of the studies on bank competitiveness in ASEAN countries investigate relationship between bank competition and many other factors but financial integration. For example, banking competitiveness and monetary policies transmission (Khan, Ahmad and Gee, 2016); the impacts of competitiveness on the bank risk taking actions (Liu, Molyneux and Nguyen, 2012); and the relationship between bank competitiveness and financial stability (Fu, Lin and Molyneux, 2014). Nevertheless, most of the studies found that banking markets in ASEAN countries are under monopolistic competition. The literature also largely support the positive impacts of foreign bank entry on the competition in ASEAN countries, e.g. Olivero *et al.* (2011), Mulyaningsih Daly and Miranti (2015). The present study contributes to this stream of the literature by providing empirical evidence on the convergence properties of ASEAN

[55] Empirical studies on macroeconomic convergence in ASEAN countries provide mixed evidence and generally conclude with a relatively low degree of integration comparing with the EU countries (Bayoumi and Mauro, 2001; Tan, 2016).

banking competitiveness, which are then used as an institutional level indicator for the degree of financial integration of the region.

3. Methodology

3.1 Measures for banking market competitiveness

The degree of competition and market structures are empirically tested by using various methods, which could be divided into two major streams: the traditional structural approach and the newly emerged non-structural approaches based on the New Empirical Industrial Organisation (NEIO) approach. Most of the early studies in banking use structural approach, such as structure-conduct-performance (SCP) paradigm. The strong assumption of exogenously shaped market structures and the one-way causality from market structure to banks' performance in SCP paradigm was then challenged by the Efficient-Structure Hypothesis (ESH), raised by Demsetz (1974) and Peltzman (1977). ESH assumes endogenous market structure, which is formed as a result of exogenous firm-specific efficiencies, and the concentration is a result of efficiency instead of collusion behaviours. However, the microeconomic foundation for this positive relationship between profitability and concentration in both theories is weak[56], and could be undermined easily by other theories, e.g. contestable market and price competition between non-collusive oligopolists would also lead to efficient outcomes.

The NEIO theories challenge the theoretical and empirical problems of traditional structural approaches, and examine the market structure directly through the firms' price-marginal cost margin without including any explicit market structure measurement[57]. Two most important techniques in NEIO studies include Bresnahan-Lau (BL) mark-up model , developed by Bresnahan (1982) and Lau (1982), and Panzar-Rosse (PR) reduced-form revenue model developed by Rosse and Panzar (1977) and Panzar and Rosse (1982, 1987)[58]. Unlike the BL model, which is usually applied to aggregate industry data, the PR model utilises firm-specific data, allowing for differences in production function of individual firms. The PR model investigates the relationship between input factor prices and firm's revenue, without requiring information on equilibrium output price and quantities of the industry. Also as pointed by Shaffer (2004), the BL model estimates are more likely to show an anticompetitive bias in small samples, but the PR approach provides robust results in small sample case.

The PR model examines the extent to which changes in input factor prices are reflected in (equilibrium) revenues of a specific firm. The key argument is that, an increase in input factor prices will increase the marginal cost for all kind of firms, but the reactions to this change is different in different type of markets. In perfect competitive market or contestable market, the marginal revenues will increase by the same amount as marginal costs so that the zero economic profit is still maintained. Therefore the total revenue should increase proportionally with the increase in factor prices in a competitive market. On the other hand, monopoly or perfect collusions with full market power, who operate on the elastic part of the demand curve, would bear all the reduction in equilibrium output demand due to increases in price, and the total revenue is reduced in this case.

---

56 Bikker and Haaf (2000) presented some theoretical derivation of the positive relationship between market concentration and market performance, but valid under strong assumptions.

57 For more discussion on the distinguishment between NEIO from SCP and ESH, see Vasala (1995).

58 The Iwata model, Iwata (1974), is also a popular approach of NEIO, but has not been used in banking industry extensively. Only very few studies of banking industry use this approach, for example, Shaffer and DiSalvo (1994).

The PR model is empirically tested by estimating a log linear reduced form revenue function in terms of input factor prices and other exogenous variables as follows:

$$lnR_{it} = \alpha_0 + \sum_{j=1}^{J} \alpha_j lnW_{jit} + \sum_{k=1}^{K} \beta_k lnZ_{kit} + \sum_{n=1}^{N} \gamma_n lnX_{nt} + \mu_{it} \quad (1)$$

This is a typical panel estimation model that contains observations of i banks over t periods, where $i = 1,2,...,I$ and $t = 1,2,...,T$. W_{jit} represents the price for j^{th} input for i^{th} bank in period t. The variables in Z are bank-specific exogenous variables that influence bank's revenue and cost functions. Some macroeconomic variables may also exogenously affect bank's revenue through demand for credit, and therefore are included in the vector X, and μ_{it} is a random error term. Since all terms are expressed in the log form, the coefficients α_j can be interpreted as the price elasticities of revenue, and the PR H-statistic is defined as the sum of the estimates for α's:

$$H = \sum_{j=1}^{J} \alpha_j \quad (2)$$

This *H-statistic* measures the extent to which total revenue responds to a change in input factor prices. According to the theorem and two propositions of Panzar and Rosse (1987), the *H-statistic* for a monopolist or perfect cartel must be nonpositive ($H \le 0$), indicating that an increase in input factor prices will reduce bank's total revenue. In symmetric Chamberlinian monopolistic competitive equilibrium, the *H-statistic* is less than or equal to unity, indicating that the reduction in revenue is less than proportion with the increases in input prices ($0 < H < 1$). The *H-statistic* equals to unity ($H = 1$) when the banking market is in long-run competitive equilibrium, implying that bank's total revenue will increase by exactly the same proportion as costs. The PR H-statistic is not only used to reject certain market types. Panzar and Rosse (1987) and Vesala (1995) proved that, under certain conditions, the magnitude of H-statistic can be interpreted as a measure of the degree of competition[59].

The dependent variable used in the present study is total bank revenue (*TR*)[60], including both interest revenue (*IR*) from the traditional bank business, and other operating income (*OYY*) accounting for the increasingly important non-traditional banking activities, such as fee-based products and other off-balance sheet activities. Following the intermediation approach of Sealey and Lindley (1977), the input variables are the number of employees, fixed assets and total deposits, and their prices (*PL, PFA* and *PD)* are included as the W variables in equation (1). The bank-specific control variables include the size of the bank (*SIZE*) measured by the natural logarithm of their total assets, the ratio of loans to deposit ($R_{L/D}$) measuring the leverage and the ratio of loans to assets ($R_{L/A}$) measuring the loan intensity. These two ratios could be considered as a bank's risk preference measurement. The growth rate of GDP (ΔGDP_{it}) is included to capture the effect of macroeconomic variations. The equation estimated for each country is as following:

---

[59] Panzar and Rosse (1987) also attempted a model for oligopoly, and showed that the *H-statistic* is negative, however, there is no evidence of generality and in general the relationship is indeterminate.

[60] Most studies, e.g. Such as Shaffer (1982), Claessens and Leaven (2004) among others, have chosen to scale the dependent variable by bank's total assets, or include a scale variable on the right-hand side of the equation to capture the economies of scale. However, as argued by Bikker *et al.* (2011) that that scaling the dependent variable or including scaling variables as explanatory variable would essentially transform the revenue equation into a 'price' equation, and such practice may distort the nature of the revenue equation and lead to a systematic bias of *H-statistic* towards unity under monopoly or monopolistic competition models. For this reason, in the present study, the absolute (total) revenue measure are used and there is no size/scale variable included on the right-hand side of the equation.

$$lnTR_{it} = \alpha_0 + \alpha_1 lnPFA_{it} + \alpha_2 lnPL_{it} + \alpha_3 lnPD_{it} + \beta_1 SIZE_{it} + \beta_2 lnRL/D_{it} + \beta_3 lnRL/A_{it} + \gamma ln\Delta GDP_{it} + \mu_{it} \quad (3)$$

and the *H*-statistic is defined as $H = \alpha_1 + \alpha_2 + \alpha_3$ (4)

The validity of PR model and its *H-statistic* depends heavily on the market equilibrium assumption. The predicting power of the *H-statistic* is only valid, especially for monopolistic and perfect competition type of market, when the market is in its long-run equilibrium. This assumption can be tested empirically by estimating a reduce-form profit equation. The idea of this test is that the bank's profit should be equalised under the competitive pressure and no bank can make supernormal return in equilibrium, therefore the profit should not be affected by changes in input prices if the market is in long-run equilibrium. This test is usually based on an equation that replaces the dependent variable in equation (1) with pre-tax profit measurements, e.g. return on assets (ROA), as following:

$$ln\pi_{it} = \alpha'_0 + \sum_{j=1}^{J} \alpha'_j lnW_{jit} + \sum_{k=1}^{K} \beta'_k lnZ_{kit} + \sum_{n=1}^{N} \gamma'_n lnX_{nt} + \mu'_{it} \quad (5)$$

Where π_{it} is the pre-tax profit measure[61], and all the other variables are the same as defined in equation (1). The test statistic, *E*, is defined as the sum of the price elasticities of profit with respect to each input factor price:

$$E = \sum_{j=1}^{J} \alpha'_j \quad (6)$$

When *E*=0, the market is in long-run equilibrium; while *E*<0 implies the market disequilibrium in which increases in factor price would lead to decrease in profit. All explanatory variables on the right-hand side are the same as those used in equation (3), and the equation is also estimated by GLS with cross-sectional fixed effects panel estimation method:

$$lnROA*_{it} = \alpha'_0 + \alpha'_1 lnPFA_{it} + \alpha'_2 lnPL_{it} + \alpha'_3 lnPD_{it} + \beta'_1 SIZE_{it} + \beta'_2 lnRL/D_{it} + \beta'_3 lnRL/A_{it} + \gamma' ln\Delta GDP_{it} + \mu'_{it} \quad (7)$$

and the *E-statistic* is the sum of the coefficients on logarithms of input prices:

$$E = \alpha'_1 + \alpha'_2 + \alpha'_3 \quad (8)$$

For the market that is not always in equilibrium when the data are observed, which is very likely to be the case for transitional economies, the PR model may not be able to provide a good indicator for the market competitiveness if the adjustment towards market equilibrium is not instantaneous. This issue has been discussed thoroughly by Goddard and Wilson (2009), and suggested that for unbiased *H-statistic* estimates when market is off equilibrium, one should apply an appropriate dynamic panel estimator, such as Arellano and Bond's (1991) generalized method of moments (GMM) procedure, by including a lagged dependent variable[62]. Also the speed of adjustment towards equilibrium can be directly assessed through the coefficient estimates on the lagged depend variable. The equation with the dynamic adjustment is as following, which is estimated by Arellano and Bond's (1991) GMM procedure:

$$\Delta lnTR_{it} = \alpha''_0 + \alpha''_1 \Delta lnPFA_{it} + \alpha''_2 \Delta lnPL_{it} + \alpha''_3 \Delta lnPD_{it} + \beta''_1 \Delta size_{it} + \beta''_2 \Delta lnRL/D_{it} + \beta''_3 \Delta lnRL/A_{it} + \gamma'' \Delta ln\Delta GDP_{it} + \eta \Delta lnTR_{it-1} + \Delta\mu_{it} \quad (9)$$

[61] Following Claessens and Laeven (2004) among others, the dependent variable (ROA*) is adjusted, such as ROA* =1+ROA, to deal with negative profits.

[62] In Goddard and Wilson (2009), they also suggest that the dynamic adjustment should be made to profit equation too when testing market equilibrium. However, this was not done in the present study since the validity of *E-statistic* does not depend on the market long-run equilibrium assumption.

Where parameter η is the so-called "persistence coefficient", which measures the adjustment speed towards market equilibrium and plays a crucial role in estimating unbiased *H-statistic*. The cross-bank fixed effects are eliminated by using first-order differences of all variables. The GMM estimator for unbiased *H-statistic* under market disequilibrium is defined as:

$$H' = (\alpha''_1 + \alpha''_2 + \alpha''_3)/(1-\eta) \quad (10)$$

This recommendation is used with caution in the present study, and only used when strong evidence of market disequilibrium is found.

3.2 Tests for convergence

In the growth literature, *β-convergence* exists when the economy of low-income countries grows faster than that of high-income countries, in other words, the low-income countries are catching up with the high-income countries. An alternative concept is *σ-convergence*, which relates to the dispersion of interested measures across groups of economies and is achieved when the dispersion narrows over time. The two concepts of convergence are related but they are conceptually different: *σ-convergence* studies how the distribution of income evolves over time whereas *β-convergence* studies the mobility of income within the same distribution. "*β-convergence* is a necessary, but not sufficient condition for *σ-convergence*" (Sala-i-Martin, 1996). Applying to the banking competitiveness analysis, a *β-convergence* means that the less competitive markets are catching up with more competitive markets. Following the specification in Weill (2013) among others, the unconditional *β-convergence* is tested by the following equation:

$$lnH_{i,t} - lnH_{i,t-1} = \alpha + \beta lnH_{i,t-1} + \varepsilon_{i,t} \quad (11)$$

where $H_{i,t}$ [63]is the *H-statistic* for country *i* during time period *t*. $H_{i,t-1}$ denotes the *H-statistic* for country *i* in the previous period *t-1*. α and β are the parameters to be estimated. Then there is *β-convergence* if the relationship between competition growth rate and initial competition level is negative and significant. The greater the absolute value of β, the greater the tendency for competition convergence. However, the results of this general test does not indicate whether the 'catching-up' effect is driven by the improvement from less competitive banking markets or from the deterioration from the more competitive banking markets[64]. Therefore, equation 11 is modified to test for the disaggregated *β-convergence* of banking market competitiveness, as follows[65]:

$$lnH_{i,t} - lnH_{i,t-1} = \alpha + \beta_1 lnHhigh_{i,t-1} + \beta_2 lnHlow_{i,t-1} + \varepsilon_{i,t} \quad (12)$$

$$lnHhigh_{i,t-1} = \begin{cases} lnH_{i,t-1}, & if\ lnH_{i,t-1} > \overline{H}_t \\ 0, & if\ lnH_{i,t-1} < \overline{H}_t \end{cases}$$

and

$$lnHlow_{i,t-1} = \begin{cases} lnH_{i,t-1}, & if\ lnH_{i,t-1} < \overline{H}_t \\ 0, & if\ lnH_{i,t-1} > \overline{H}_t \end{cases}$$

where $\overline{H}_t$ represents the average competition level of ASEAN-5 banks at time period t; $\beta_1$and β_2 are the convergence parameters for countries with higher competition level above the regional average and countries with a lower competition level correspondingly, and a negative and significant estimate implies a force of convergence from the sub-group.

The *σ-convergence* exists when there is reduction of the dispersion of bank competition level among ASEAN-5 economies. The *σ-convergence* is tested through the following equation:

---

[63] The *H-statistics* is adjusted, such as $H^* = 2+H$, to deal with negative values.

[64] This is also the reason why β- and σ-convergence test is preferred to the Phillips and Sul (2007) panel convergence test, which tests convergence relative to the panel cross-section average over time without providing information on the direction of the convergence.

[65] Similar modification is used by Andrieş and Căpraru (2014).

$$\Delta W_{i,t} = \theta + \sigma W_{i,t-1} + \nu_{i,t} \quad (13)$$

where $\Delta W_{i,t} = W_{i,t} - W_{i,t-1}$; $W_{i,t} = lnH_{i,t} - ln\overline{H}_t$, and θ and σ are coefficients to be estimated. $\nu_{i,t}$ is the error term. If the coefficient σ is negative and statistically significant, there is *σ-convergence*, implying a narrowing dispersion over time in ASEAN-5's banking competitiveness.

Both convergence tests require a time series of *H-statistics* that shows the dynamic changes in banking competitiveness over time, however, the PR model using panel data only provides information on the overall competitiveness for the sample period. Also the PR model depends heavily on the assumption of long-run market disequilibrium condition, therefore is not directly applicable to measure market competitiveness in a single year[66]. Therefore, in order to investigate the evolution of banking market competitiveness over time, we apply the GMM dynamic estimator to estimate PR *H-statistics* for 5-year rolling-windows of the sample, using equations (9) and (10). We obtain 19 *H-statistics* over 23 years for each country.

4. Data

In this study, we choose commercial banks of the ASEAN-5 during the period of 1994-2016, which covers the 1997/98 Asian financial crisis and the global financial crisis of 2008-2009. These ASEAN-5 are the 5 largest economies in this sub-region with dominating economic importance, and their attitudes toward regional co-operations and integrations, would be consequently, crucial for the regional agreements. The economic importance of them has made the 5 countries together a reasonable representative for the whole region in many aspects.

The data set is primarily drawn from the balance sheet and income statement of individual bank from the ORBIS Bank Focus[67], which reports published financial statements from financial institutions worldwide. For missing data, other resources including annual reports of individual bank, central bank reports and internet web resources are used. Only commercial banks are considered in this study as they are carrying out relatively similar banking business, and comprise the largest segment of depository institutions. The unconsolidated financial reports are used, where available, to avoid double-counting. Consolidated reports are used wherever unconsolidated reports are not available. After adjusting the data for missing values, reported errors and outliers, it ends up with an unbalanced panel data set of 2809 bank-year observations, in which each bank exists for at least 5 consecutive years[68]. Macroeconomic data are collected from the International Monetary Fund (IMF) website.

The price of labour is calculated as the ratio of personnel expenses to the fixed assets[69]. The price of deposits is calculated as the ratio of interest expenses to total deposits. The price of fixed assets is measured by the ratio of other operating expenditures to the fixed assets, here

[66] Although this practice has been seen in a few studies, e.g. Weill (2013) and Andrieş and Căpraru (2014), it is theoretically incorrect.

[67] Used to be known as BankScope database of Bureau van Dijk

[68] Only a limited number of bank reports with limited time periods are available for Singapore. According to Monetary Authority of Singapore, although there are 120 commercial banks operating in Singapore by March of 2011, only 6 of them are local banks and only one-fifth (26 out of 114) of foreign banks offer full banking services. Financial reports of foreign subsidiaries or representative offices are not publicly available. However, given that the Singaporean banking industry is fairly competitive, surviving banks must represent the general operational level, otherwise it can be argued that they would have been eroded quickly under the competitive pressure. The small sample of Singapore is treated as a representative of the Singaporean banking industry.

[69] As the number of employee is not available in most cases.

the operating expense can be interpreted as capital maintenance. The output is the total revenue, including both interest revenue and non-interest revenue. For the long-run market equilibrium test, the dependent variable is the rate of return on assets (ROA). All the monetary variables are adjusted by the Purchasing Power Parity exchange rate and converted into inflation-adjusted US dollars. Table 1 presents a brief comparison on the mean values of key variables across ASEAN-5 over the whole sample periods and sub-periods. Observations during the two financial crisis periods, i.e. 1997-1998 and 2008-2009, are excluded from the descriptive statistics to avoid distortions. After adjusting inflation and exchange rate, many of the key variables show significant differences in value among these 5 countries, but with similar evolving patterns. Total revenues, labour cost and maintenance cost of fixed assets increase overtime for all countries, and the unit cost of deposit has shown a slight decreasing trend in most of the countries.

Table 1. Descriptive statistics of key variables

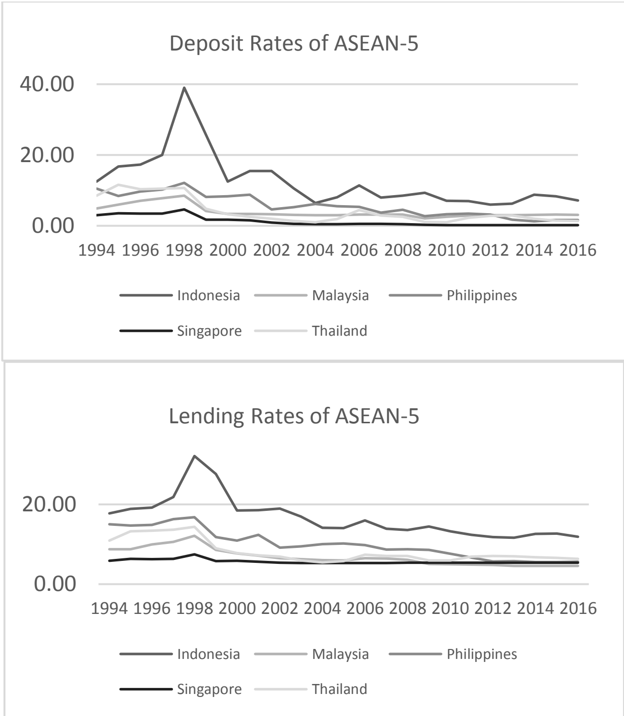
	country	No. Obs.	TR	ROA	PL	PFA	PD	L/A	L/D	ΔGDP
1994 - 2016	**I**	1,117	1235.48	0.57	2.05	2.57	0.10	0.57	0.84	6.59
	M	639	1600.36	1.10	2.40	2.57	0.15	0.52	0.70	7.20
	P	510	628.48	1.14	1.04	1.68	0.04	0.46	0.60	6.80
	S	139	2456.35	1.15	3.86	5.28	0.02	0.49	0.66	7.45
	T	404	2649.67	-0.16	1.10	1.42	0.04	0.70	0.88	5.61
	St. Dev.		843.63	0.57	1.15	1.53	0.05	0.09	0.15	0.71
1994 - 1996	**I**	107	362.10	1.63	1.82	2.67	0.21	0.71	0.83	10.04
	M	80	439.93	1.26	1.11	0.94	0.04	0.60	0.76	11.95
	P	50	313.33	2.01	0.55	1.04	0.05	0.54	0.70	7.14
	S	7	1443.20	1.46	0.39	0.55	0.03	0.56	0.67	10.00
	T	43	3040.92	1.53	0.36	0.48	0.08	0.84	1.05	9.38
	St. Dev.		1170.69	0.28	0.62	0.89	0.07	0.13	0.49	1.73
1999 - 2007	**I**	426	1061.18	1.26	1.92	2.66	0.08	0.47	0.70	7.11
	M	247	1779.32	1.24	1.90	2.13	0.23	0.51	0.66	8.12
	P	223	519.66	0.74	0.86	1.50	0.05	0.43	0.57	7.19
	S	64	1757.83	1.70	4.12	5.68	0.03	0.46	0.68	8.93
	T	160	1854.31	-0.24	0.70	1.44	0.03	0.64	0.83	7.70
	St. Dev.		584.80	0.74	1.37	1.75	0.08	0.08	0.10	0.75
2010 - 2016	**I**	389	1787.22	1.30	2.33	2.24	0.03	0.65	0.91	7.22
	M	187	2040.73	1.04	3.94	4.22	0.13	0.47	0.73	7.23
	P	143	964.32	1.46	1.51	2.20	0.02	0.44	0.56	7.97
	S	39	4139.40	0.93	5.14	6.55	0.01	0.49	0.63	7.77
	T	130	3413.83	0.81	1.97	1.71	0.02	0.70	0.89	5.30
	St. Dev.		1284.06	0.27	1.52	2.02	0.05	0.12	0.15	1.06

Notes: I, M, P, S and T stands for Indonesia, Malaysia, the Philippines, Singapore and Thailand respectively. No.Obs = number of observations; TR=Total Revenue; ROA= Return on Assets (unadjusted); PL= Price of Labour; PFA = Price of Fixed Assets; PD = Price of Deposit; L/A= loan to asset ratio; L/D = loan to deposit ratio. ΔGDP = GDP growth rate. All Monetary variables are measured in million US dollar.

5. Preliminary analysis using 'price-based' indicators

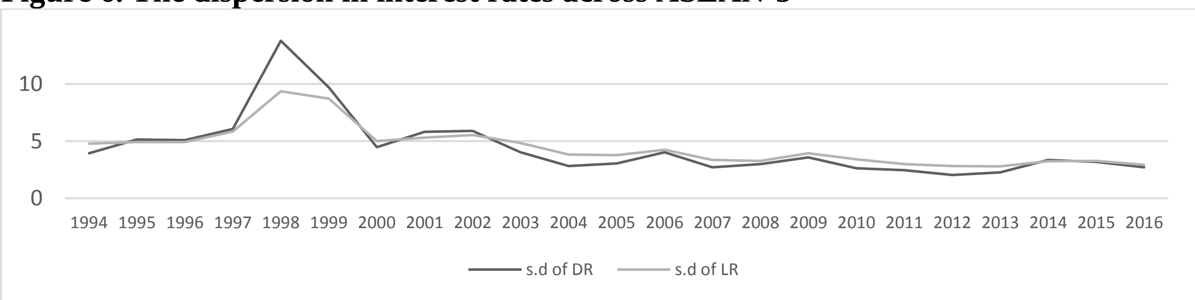
We also collected the information of interest rates of ASEAN-5. A visual examination on the interest rates across the ASEAN-5 provides primary evidence of some degree of convergence in pricing behaviour of banks in the region. Figure 1 provides a plot on the deposits rate and lending rate, which are the interests on the two main products of commercial banks. A gradual declining trend has been seen over the past two decades, especially after the1997/98 Asian financial crisis. The declining trend in lending rate in particular implies cheaper credit which encourage greater investment and economic growth. The trend towards the one-price behaviour after the1997 Asian financial crisis is further evidenced by the narrowing dispersion shown in Figure 2. The standard deviation of deposits rate and lending rate has decrease from their peak (13.75 for deposit rate, 9.36 for lending rate) in 1998 to less than 3 in recent years.

Figure 5. Interest rates across ASEAN-5 for 1994-2016



Data source: IMF International Financial Statistics (IFS)

Figure 6. The dispersion in interest rates across ASEAN-5



Data source: IMF International Financial Statistics (IFS). Notes: DR = deposit interest rate, LR = Lending interest rate

The results of Johansen multivariate Co-integration test[70] finds at least three co-integrating relationships among the interest rates on deposit and two co-integrating relationships among the lending interest rates of ASEAN-5, implying that the interest rates in ASEAN-5 share similar stochastic trends. The test results are reported in Table 2. This preliminary finding on co-movement of interests rates provides some evidence of the trend toward “one-price” behaviour of ASEAN banks, thus a signal of certain degree of integration of the financial markets in ASEAN.

Nevertheless, as argued earlier, the price-based convergence indicator does not necessarily imply convergence in banks’profitability and productive efficiency, which are more important for countries under the transitional process toward full market integration. Institutional and operational convergence in terms of competitiveness add valuable information on how well the banks are prepared for the ongoing integrating process, and assist policy consideration regarding financial stability during.

6. Empirical results of PR *H-statistics*

Equations (7) and (8) are estimated to test for market equilibrium condition for the whole sample period, along with a Wald test statistic for the null hypothesis that H0: E=0. The null hypothesis of market disequilibrium is only rejected in Indonesia. In fact, most ASEAN countries are developing economies where banking market might involve reforms and structural changes constantly, and there are two financial crises contained in the sample period.

Table 2. **Johansen Test for Co-integration of interest rates among ASEAN-5 (1994-2016)**[71]

Cointegration test based on Johansen's maximum likelihood method: Deposit rates (DR)					
λmax rank tests			λmax rank value		95% critical values
H0: r=0	Ha: r>0		51.74**		30.04
H0: r≤1	Ha: r>1		27.28**		23.80
H0: r≤2	Ha: r>2		23.21**		17.89
H0: r≤3	Ha: r>3		9.09		11.44
H0: r≤4	Ha: r>4		2.23		3.84
λtrace rank tests			λtrace rank value		
H0: r=0	Ha: r=1		113.55**		59.46
H0: r=1	Ha: r=2		61.81**		39.89
H0: r=2	Ha: r=3		34.53**		24.31
H0: r=3	Ha: r=4		11.32		12.53
H0: r=4	Ha: r=5		2.23		3.84
Normalised ecm:	DR1	DR2	DR3	DR4	DR5
	1	-8.88E-16	-1.78E-15	54.51	-13.82
	1.04E-17	1	-2.22E-16	16.83	-4.45
	-9.71E-17	-1.11E-15	1	13.05	-3.86
Cointegration test based on Johansen's maximum likelihood method: Loan rates (LR)					
λmax rank tests			λmax rank value		95% critical values
H0: r=0	Ha: r>0		45.96*		30.04
H0: r≤1	Ha: r>1		29.05*		23.80
H0: r≤2	Ha: r>2		14.59		17.89
H0: r≤3	Ha: r>3		7.17		11.44
H0: r≤4	Ha: r>4		0.15		3.84
λtrace rank tests			λtrace rank value		
H0: r=0	Ha: r=1		96.92*		59.46

[70] Augmented Dickey-Fuller tests for unit root show that the interest rates of ASEAN-5 from 1994- 2016 are all I(1) processes.

[71] The optimal lag length in VAR estimation is selected based on the minimization of the AIC and the SBC criteria. The selection of the appropriate model regarding the deterministic components in the multivariate system is based on the ‘Pantula Principle’, suggested by Johansen (1992).

H0: r=1	Ha: r=2		50.96*	39.89	
H0: r=2	Ha: r=3		21.91	24.31	
H0: r=3	Ha: r=4		7.32	12.53	
H0: r=4	Ha: r=5		0.15	3.84	
Normalised ecm:	LR1	LR2	LR3	LR4	LR5
	1	2.22E-16	-0.93	-2.16	0.82
	0	1	-0.43	-0.21	-0.14

** denotes significance at 5%.

A stable long-run equilibrium is unlikely to exist for most ASEAN countries in the sample period. A Chow structural break test at 1998 and 2009 is carried out for each country to test for the stability of the parameters, and the results confirm that all countries have experienced at least one, in most case two, significant structural changes following the financial crises. Therefore, the parameters in the estimating equation are not stable, neither the estimates for the E-statistics, and the assumption of long-run market equilibrium for the whole period does not hold.

Table 3. Long-run Market Equilibrium Test Results (1994-2016)[72]

ROA*	**Indonesia**	**Malaysia**	**Philippines**	**Singapore**	**Thailand**
Intercept	0.1766	4.5016	4.3536	4.5897	2.7523
	(1.88)	(0.07)***	(0.19)***	(0.08)***	(0.50)***
P_{FA}	-0.1048	-0.0002	-0.0044	-0.0038	-0.0002
	(0.09)	(0.00)	(0.00)	(0.00)	(0.01)
P_L	0.0745	0.0013	0.0049	0.0025	0.0116
	(0.07)	(0.00)*	(0.00)	(0.00)	(0.01)
P_D	-0.0787	-0.0006	0.0012	0.0017	-0.0056
	(0.04)*	(0.00)	(0.00)	(0.00)	(0.00)
$R_{L/D}$	0.0909	0.0003	0.0312	0.0519	0.0222
	(0.04)**	(0.00)	(0.01)***	(0.01)***	(0.02)
$R_{L/A}$	-0.1336	-0.0003	-0.0286	-0.0590	-0.0250
	(0.07)*	(0.00)	(0.01)***	(0.01)***	(0.02)
SIZE	-0.0144	-0.0000	0.0013	0.0000	0.0091
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)***
ΔGDP_t	0.9447	0.0483	0.0785	0.0273	0.3982
	(0.39)**	(0.01)***	(0.04)*	(0.02)	(0.11)***
R^2	0.124	0.0495	0.1257	0.3603	0.2628
***E*-statistic**	**-0.109**	**0.0005**	**0.0017**	**0.0004**	**-0.0308**
H_0:E=0	F(1,80)=3.14*	F(1,43)=0.30	F(1,34)=0.12	F(1,11)=0.05	F(1,22)=1.01
Chow-test (1998)	F(7,80)=1.07	F(7,43)=5.02***	F(7,34)=5.42***	F(7,11)=145.26***	F(7,22)=10.26***
Chow-test (2009)	F(7,80)=4.11***	F(7,43)=4.94***	F(7,34)=2.97**	F(7,11)=136.20***	F(7,22)=2.55**

***, **, * denote significance at 1%, 5% and 10% level accordingly. Significance test-statistics are reported in ().

Therefore, in light of Goddard and Wilson (2009), the Arellano and Bond's (1991) GMM dynamic estimation is applied in estimating the PR model for the whole sample period. Given the significant structural changes following the two financial crises, dynamic *H-statistics* for sub-periods, 1994-1998, 1999-2009 and 2010-2016 are also estimated by using Equation (9) and (10). Results are presented in Table 4.

The banks specific variables generally show consistent marginal effects on a bank's revenue, where banks operates with lower leverage (($R_{L/D}$) and higher loan intensity ($R_{L/A}$) can earn

[72] The *F*-test (H_0:η_i=0) for bank-specific effect confirms that the cross-bank heterogeneities are significant in Malaysia and the Philippines, but not significant for the other three countries. To keep consistency in estimators, a pooled model with Cluster-robust standard errors, which captures the possible unobserved characteristics of each individual bank, is used. Fixed effect estimation for Malaysia and the Philippines produces consistent results.

higher revenue in most cases. The lower the value for $R_{L/D}$ the more prudential the bank is, and higher revenue is rewarded to more prudential banks. Positive sign on $R_{L/A}$ implies that the higher proportion of loan in total assets, the higher revenue a bank could generate, which indicates that interest income from loans are still the main income source for most banks in ASEAN. Bigger banks have some advantages in revenue generating which may imply some degree of market power. The effect of macroeconomic condition has shown negative effect in the sub-period 1999-2009, which may suggest an unstable economic environment which is a side effect of high growth.

After adjusting the dynamics of transition, the estimated *H-statistics* are generally lower than what has been typically found in the literature using static fixed effect panel estimations[73], but still in the range of monopolistic competition for most of the cases. Using the values of *H-statistics* as a measurement of competition level, the countries can be ranked from the least competitive to the most competitive banking market as: Thailand, the Philippines, Indonesia, Singapore and Malaysia. The discrepancies in banking market competition levels of the ASEAN-5 are still quite significant. The sub-period results show that an improvement in market competitiveness is only seen in Malaysia and Thailand.

The GMM dynamic *H-statistics* provide information on the overall competitiveness for the sample periods under market disequilibrium situation, but it does not show the changing patterns over time and the information needed for the convergence analysis, and a 5-year rolling-window of PR *H-statistics* for each country is also estimated by using Equation (9) and (10) to show the changing patterns over time. The estimates are then used for the convergence analysis. Results are summarized in the Table 5.
The 5-year rolling window estimates of *H-statistics* confirms the substantial differences among the ASEAN-5 banking market competitiveness and reveals the evolution process of each country's bank competition. Malaysia has the most competitive and relatively stable banking market, while market competitiveness fluctuates the most in Singapore and Thailand. The result does not show clear improvement trend over the whole period, rather periodical declines appear at various occasions. Nevertheless, the member countries so exhibit similar responses to global shocks to different extents.

The competition level dropped sharply for all countries straight after the two financial crises, i.e. roughly around 1998-2003 and 2008-2012, and a relatively steady competition level has been seen during the period between the two financial crises from 2000 to 2006. The structural reforms of individual banking markets, the regional financial market liberalization and integration process generally impose stabilising impacts on the banking market competitiveness in ASEAN. However, the ASEAN banking markets are heavily influenced by the global environment, and are easily distorted by the global economic downturn, especially for countries with higher degree of openness, like Singapore and Thailand[74].

[73] For example in Claessens and Laeven (2004), the estimates of *H*-statistic for Indonesia, Malaysia and the Philippines are 0.62, 0.68, and 0.66 accordingly for the years 1994-2001.

[74] The dramatic decline in Singaporean bank competition level in recent years may also attribute to the enhancement of the regulatory role of Monetary Authority of Singapore (MAS) in the banking market after the 2006. Studies for periods before the 2007-08 global financial crisis typically find that Singapore has the most competitive banking market, e.g. Jeon *et al.* (2011); Olivero *et al.* (2011), which is consistent with our results for the same period.

Table 4. GMM Dynamic Estimation of PR H-statistics (1994-2016)

	1994-2016					**1994-1998**					**1999-2009**					**2010-2016**				
TR	I	M	P	S	T	I	M	P	S	T[75]	I	M	P	S	T	I	M	P	S	T
Constant	9.15	0.73	-0.13	-2.66	-0.86	7.14	0.21	-4.87	-	-	9.31	2.57	6.71	-0.34	3.56	4.00	1.07	-2.57	-4.00	1.08
	(0.56)***	(0.86)	(1.30)	(1.25)**	(1.04)	(2.49)***	(2.00)	(3.46)	-	-	(0.83)***	(0.90)***	(1.56)***	(1.28)	(1.64)**	(2.57)	(1.60)	(2.11)	(2.52)	(1.33)
P_{FA}	0.05	0.04	0.08	0.03	-0.06	0.05	0.07	-0.08	-	-	0.06	0.03	0.10	0.14	0.05	0.02	0.11	0.11	0.04	-0.06
	(0.01)***	(0.01)***	(0.03)**	(0.02)	(0.02)**	(0.05)	(0.13)	(0.09)	-	-	(0.01)***	(0.01)***	(0.04)**	(0.03)***	(0.04)	(0.03)	(0.03)***	(0.05)**	(0.06)	(0.03)*
P_L	-0.02	-0.01	-0.03	-0.01	0.04	0.06	-0.01	0.09	-	-	0.00	0.03	0.02	-0.07	0.01	0.00	-0.07	-0.04	-0.11	0.03
	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.05)	(0.13)	(0.09)	-	-	(0.02)	(0.03)	(0.04)	(0.03)**	(0.05)	(0.03)	(0.04)*	(0.04)	(0.07)*	(0.04)
P_D	0.38	0.46	0.25	0.32	0.32	0.43	0.17	0.42	-	-	0.41	0.40	0.35	0.40	0.24	0.29	0.63	0.17	0.04	0.33
	(0.01)***	(0.02)***	(0.02)***	(0.02)	(0.02)***	(0.09)***	(0.15)	(0.11)***	-	-	(0.02)***	(0.03)***	(0.03)***	(0.02)***	(0.03)***	(0.02)***	(0.03)***	(0.03)***	(0.06)	(0.04)***
$R_{L/D}$	-0.33	-0.40	-0.09	-0.19	0.05	-0.29	-0.32	-0.56	-	-	-0.30	0.23	-0.17	-0.19	0.38	-0.31	-0.72	-0.22	-0.25	-0.55
	(0.02)***	(0.03)***	(0.05)*	(0.10)	(0.07)	(0.12)**	(0.26)	(0.42)	-	-	(0.05)***	(0.11)**	(0.07)**	(0.09)**	(0.10)***	(0.02)***	(0.05)***	(0.14)	(0.27)	(0.11)***
$R_{L/A}$	0.47	0.54	0.06	0.38	0.54	0.37	0.51	0.53	-	-	0.42	-0.16	0.10	0.42	0.08	0.25	0.90	0.14	0.26	1.51
	(0.03)***	(0.03)***	(0.06)	(0.09)	(0.07)***	(0.13)***	(0.41)	(0.46)	-	-	(0.05)***	(0.12)	(0.07)	(0.09)***	(0.11)	(0.04)***	(0.05)***	(0.16)	(0.33)	(0.12)***
SIZE	0.81	0.65	0.67	0.66	0.85	0.83	0.47	0.76	-	-	0.84	0.77	0.81	0.69	0.96	0.70	0.56	0.33	0.66	0.72
	(0.02)***	(0.03)***	(0.03)***	(0.05)	(0.03)***	(0.11)***	(0.22)**	(0.16)***	-	-	(0.02)***	(0.04)***	(0.04)***	(0.05)***	(0.05)***	(0.02)***	(0.05)***	(0.05)***	(0.08)***	(0.05)***
ΔGDP_t	-1.87	0.01	-0.05	0.55	0.00	-1.52	-0.03	1.04	-	-	-1.94	-0.41	-1.50	0.24	-1.17	-0.85	0.61	0.77	0.63	-0.09
	(0.12)***	(0.18)	(0.28)	(0.25)	(0.22)	(0.49)***	(0.49)	(0.82)	-	-	(0.19)***	(0.19)**	(0.33)***	(0.25)	(0.35)***	(0.55)	(0.36)*	(0.47)*	(0.44)	(0.28)
TRt-1	0.02	0.23	0.24	0.26	0.12	0.05	0.46	0.20	-	-	0.00	0.01	0.07	0.10	0.06	0.18	0.01	0.45	0.20	0.14
	(0.02)	(0.02)***	(0.03)***	(0.04)	(0.02)***	(0.11)	(0.19)**	(0.11)*	-	-	(0.03)	(0.04)	(0.05)	(0.04)**	(0.03)	(0.03)***	(0.04)	(0.06)***	(0.11)*	(0.04)***
H statistic	**0.42**	**0.64**	**0.38**	**0.47**	**0.34**	**0.57**	**0.42**	**0.54**	-	-	**0.47**	**0.47**	**0.50**	**0.52**	**0.31**	**0.37**	**0.67**	**0.43**	**-0.04**	**0.36**
Wald test H=0	chi2(1)= 553.64***	chi2(1) = 336.79***	chi2(1) = 94.70***	chi2(1) = 112.88***	chi2(1) = 81.21***	chi2(1) = 30.71***	chi2(1) = 2.26*	chi2(1) = 8.78***	-	-	chi2(1) = 206.11***	chi2(1) = 134.37***	chi2(1) = 114.88***	chi2(1) = 154.07***	chi2(1) = 25.02***	chi2(1) = 232.24***	chi2(1) = 217.29***	chi2(1) = 13.41***	chi2(1) = 0.16	chi2(1) = 39.31***
Wald test H=1	chi2(1) = 633.78***	chi2(1) = 91.23***	chi2(1) = 143.20***	chi2(1) = 67.51***	chi2(1) = 312.46***	chi2(1) = 9.07***	chi2(1) = 3.52*	chi2(1) = 8.63***	-	-	chi2(1) = 223.34***	chi2(1) = 122.96***	chi2(1) = 59.97***	chi2(1) = 75.26***	chi2(1) = 130.89***	chi2(1) = 243.29***	chi2(1) = 49.88***	chi2(1) = 11.30***	chi2(1) = 31.55***	chi2(1) = 106.32***
Market Condition	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	-	-	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	**MC**	**M**	**MC**

Notes: ***, **, * denote significance at 1%, 5% and 10% level accordingly. Estimated standard errors are reported in ( ). M – Monopoly or perfect collusive; MC -- Monopolistic Competition; PC – Perfect Competition.
I, M, P, S and T stands for Indonesia, Malaysia, the Philippines, Singapore and Thailand respectively.

75 Singapore and Thailand do not have enough observations to carry out the dynamic estimation in this period.

***Table 5. Summary of 5-year rolling window estimates of* H-statistics.**

	Indonesia	Malaysia	Philippines	Singapore	Thailand	Regional average
1994-1998	0.57	0.42	0.54			0.51
1995-1999	0.55	0.55	0.66	0.72	1.10	0.72
1996-2000	0.56	0.55	0.56	0.76	0.92	0.67
1997-2001	0.52	0.65	0.56	0.85	0.62	0.64
1998-2002	0.50	0.47	0.43	0.62	0.27	0.46
1999-2003	0.48	0.61	0.47	0.56	-0.12	0.40
2000-2004	0.49	0.48	0.49	0.55	-0.21	0.36
2001-2005	0.54	0.41	0.45	0.47	0.07	0.39
2002-2006	0.74	0.37	0.39	0.46	0.48	0.49
2003-2007	0.56	0.52	0.37	0.37	0.50	0.46
2004-2008	0.53	0.48	0.37	0.35	0.51	0.45
2005-2009	0.52	0.37	0.31	0.43	0.49	0.42
2006-2010	0.42	0.63	0.20	0.42	0.46	0.43
2007-2011	0.36	0.60	0.27	0.22	0.27	0.35
2008-2012	0.31	0.59	0.17	0.20	0.25	0.30
2009-2013	0.35	0.63	0.09	0.09	0.22	0.27
2010-2014	0.39	0.63	0.33	-0.03	0.27	0.32
2011-2015	0.39	0.52	0.28	0.23	0.28	0.34
2012-2016	0.36	0.51	0.28	0.06	0.66	0.37
Country Average	0.48	0.53	0.38	0.41	0.39	0.44

7. The convergence properties of PR *H-statistics*

Equation (11), (12) and (13) are estimated[76] using 5-year rolling window *H-statistics* reported in Table 5, and Table 6 exhibits the results for the convergence tests. The tests results show strong evidence for unconditional *β-convergence* and *σ-convergence* with negative and significant coefficients, indicating the existence of both type of convergences in ASEAN banking market competitiveness over the past two decades in general. However, this convergence phenomenon of banking market competitiveness need to be interpreted with caution, as there are still significant discrepancies existing among ASEAN-5's banking markets. The last five-year *H-statistics* in the rolling window estimation are 0.36, 0.53, 0.38, 0.06 and 0.66 for Indonesia, Malaysia, Philippines, Singapore and Thailand respectively.

Additionally, it has also been noticed that the convergence speed is negatively affected by the two financial crises. Both *β-convergence* and *σ-convergence* are tested for the whole sample period and three sub-periods, divided by the common turning points in the dynamic paths of *H-statistics*, namely period 1-4 which coincides with year 1994-2001, period 5-11 which coincides with year 1998-2008 and period 12-19 which coincides with year 2005-2016, and this is also roughly the period around the 1997/98 Asian crisis, the period between the two financial crises and the period of the aftermath of the 2008/09 financial crisis. Although all sub periods are still characterized by converging trend in banking market, the speed of convergence has slowed down, evidenced by a significantly reduced absolute value of the coefficient on the lagged dependent variable in period 5-11, i.e. after the 1997/98 Asian crisis, and a less significant drop in the period after the 2008/09 global crisis.

[76] Estimated by a pooled model with Cluster-robust standard errors, which captures the possible unobserved characteristics of each individual country.

In terms of the disaggregated *β-convergence*, the convergence coefficients for both groups of countries are negative and significant in most cases. This indicates that the convergences is happening from both directions. The financial integration process has not improved the overall competitiveness of the ASEAN banking markets, instead, the highly competitive market has become less competitive while the competiveness of the originally less competitive markets has been enhanced. The *σ-convergence* further confirms that the banking competition levels of these 5 countries are moving towards a sample average rather than the best-practice. The sub-period estimation also shows that the convergence speed from both groups all slowed down overtime and the less competitive market show insignificant converging tendency in recent years (2005-2016).

Table 6. Test of β-convergence of H-Statistics

Unconditional β-convergence					
	1994-2016		Period 1-4	Period 5-11	Period 12-19
Constant	0.08		0.23	0.10	0.06
	$(0.02)^{**}$		$(0.05)^{***}$	$(0.02)^{***}$	$(0.02)^{*}$
$lnH_{i.t-1}$	-0.25		-0.48	-0.33	-0.23
	$(0.06)^{**}$		$(0.11)^{**}$	$(0.02)^{***}$	$(0.08)^{*}$
		Wald test		$F(1,4) = 36.68^{***}$	$F(1, 4) = 1.45$
R-squared	0.13		0.51	0.20	0.09
Disaggregated β-convergence					
	1994-2016		Period 1-4	Period 5-11	Period 12-19
Constant	0.08		0.33	0.10	0.08
	$(0.02)^{**}$		$(0.07)^{***}$	$(0.02)^{***}$	$(0.03)^{*}$
$lnHhigh_{i.t-1}$	-0.24		-0.62	-0.30	-0.26
	$(0.06)^{**}$		$(0.10)^{***}$	$(0.07)^{**}$	$(0.10)^{*}$
		Wald test		$F(1, 4)=21.36^{***}$	$F(1,4)= 0.15$
$lnHlow_{i.t-1}$	-0.30		-0.74	-0.42	-0.35
	$(0.09)^{**}$		$(0.15)^{***}$	$(0.03)^{***}$	(0.22)
		Wald test		$F(1,4)=91.29^{***}$	$F(1, 4)= 0.09$
R-squared	0.14		0.60	0.23	0.10
σ-convergence					
	1994-2016		Period 1-4	Period 5-11	Period 12-19
Constant	-0.005		-0.02	-0.002	-0.002
	(0.01)		(0.02)	(0.01)	$(0.01)^{*}$
$W_{i.t-1}$	-0.25		-0.44	-0.26	-0.21
	$(0.03)^{***}$		$(0.09)^{***}$	$(0.03)^{***}$	$(0.08)^{*}$
		Wald test		$F(1,4)=45.47^{***}$	$F(1,4)= 0.35$
R-squared	0.12		0.32	0.13	0.07
No. of obs.	88		13	35	40

Notes: ***, **, * denote significance at 1%, 5% and 10% level accordingly. Estimated standard errors are reported in ( ). The Wald test reports the test results of whether the coefficient, i.e. the convergence speed, is significantly different from the previous sample period.

8. Conclusions

The aim of this paper is to assess the degree of integration in the ASEAN's financial markets, with a focus on the convergence properties of the most important element of its financial system, namely the retail banking markets. Theoretically, the market integration should foster competition, therefore promote efficiency and welfare through reduced prices of financial services and credit, and ideally the prices will converge to a common level under highly integrated markets. One of the interesting findings in this paper is that, the financial integration has obvious promoted convergence of banking market based on both price-based indicators and institutional level indicators of market competitiveness, but its impact on improving banking market competitiveness level is less clear. The ASEAN banking markets are converging toward monopolistic competition rather than competitive market structure as expected. Interestingly, similar phenomenon has also been seen in the EU banking markets, which is considered to be highly integrated (Weill, 2013; Andrieş and Căpraru, 2014).

The optimal competitive structure is always a debatable question, especially for banking market. In a competitive market, banks are price-takers and maximise profit by supplying the greatest quantity of credit and minimising costs. On the other hand, banks with market power can charge a price above its marginal cost level and make supernormal profit; as a result they are normally better capitalized to withstand shocks and relatively more stable. Some degree of market power, hence profitability may help maintain stability in the financial sector (Northcott, 2004). However, the market ends up with less available credit and at a higher price. The trade-off between economic growth and financial stability highlights that both of the two extremes of market structure could have positive implications, and neither of them is ideal for the banking market. However, the notable difference in banking market competitiveness between member countries will surely impose difficulties on the process of banking market integration. Opening up the cross-border market barriers and allowing free capital movements will definitely introduce shocks to the domestic markets. The consequences are difficult to predict. Banks in competitive markets are forced to operate on the lowest point of cost curves, and to implement efficient production technology and efficient resources allocation. These banks may have the advantages of crack internal cooperate governance structure and better experiences on quick adaptation to new environment. Therefore, banks from countries with competitive banking market may dominate banks from countries with less competitive market. However it could also be the other way around, banks with high degree of market power in one particular country may be better capitalised through years of accumulation of supernormal profits, and may have established solid relation networks with the cooperate sectors that the foreign competitors hardly break into. Either way it goes, successful integration of the banking markets should encourage the banking market competitiveness converge toward a common standard, if it does not foster competition as expected, for banks to survive in a more open environment.

Banking markets in the ASEAN-5 countries are under monopolistic competition in most cases, and the banking markets show some level of symmetric response to common shocks. Some evidence of both β- and σ-*convergence* has been found, but the convergence process is easily distorted by the external economic environment. The cross-country comparison shows that the actual degree of competitiveness still vary significantly from country to country and the ongoing process of ASEAN financial integration has not improved banking market competitiveness in general. The ASEAN banking markets are still quite sensitive to and heavily influenced by the global environment, further regional integration is still required to establish a strong banking market that better withstands shocks from outside the region.

References

Affinito, M. & Farabullini, F., 2006. ‘Does the law of one price hold in retail banking? An analysis of national interest rate differentials in the euro area’. Bank of Italy, Economic Research Department (August).

Andrieş, A, M & Căpraru, B 2014, ‘The nexus between competition and efficiency: The European banking industries experience’, International Business Review, vol. 23(3), pp. 566-579.

Arellano, M. & Bond, S. 1991, ‘Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations’, Review of Economic Studies, vol. 58, pp. 277-297.

ASEAN Integration Report 2015, the ASEAN Secretariat, viewed 25 July 2016, <http://asean.org/storage/2015/12/ASEAN-Integration-Report-2015.pdf>.

Barro, R, J & Sala-I-Martin, X 1991, ‘Convergence Across States and Regions’, Brookings Papers on Economic Activity, vol. 1991, no. 1, pp. 107-182.

Bayoumi, T. and P. Mauro. 2001. ‘The Suitability of ASEAN for a Regional Currency Arrangement.’ The World Economy, vol. 24, pp. 933–54.

Bikker, J. A. & Haaf, K. 2000. ‘Competition, Concentration and Their Relationship: An Empirical Analysis of the Banking Industry’, Journal of Banking and Finance, vol. 26, pp. 2191-2214.

Bikker, J, A, Shaffer, S & Spierdijk, L 2011, ‘Assessing competition with the Panzar–Rosse model: the role of scale, costs, and equilibrium’, Review of Economics Statistics, vol. 94(4), pp. 1025-1044.

Borensztein, E. and P. Loungani. 2011. "Asian Financial Integration: Trends and Interruptions", IMF Working Paper WP/11/4.

Bresnahan, T. F. 1982, ‘The Oligopoly Solution Concept Is Identified’, Economics Letters, vol. 10, pp. 87-92.

Centeno, M. and A. S. Mello. 1999, "How Integrated are the Money Market and the Bank Loans Market within the European Union?" Journal of International Money and Finance, vol.18, 75-106.

Claessens, S. & Laeven, L 2004, ‘What drives bank competition? Some international evidence’, Journal of Money, Credit, and Banking, vol. 36(3), pp.563-583.

Demsetz, H. 1974, ‘Where is the new industrial state?’, Economic Inquiry, vol. 12(1), pp. 1-12.

De Truchis, G & Keddad, B 2013, ‘Southeast Asian monetary integration: New evidences from fractional cointegration of real exchange rates’, Journal of International Financial Markets, Institutions and Money, vol. 26, pp. 394-412.

F1 u, X, Lin, Y & Molyneux, P 204, ‘Bank competition and financial stability in Asia Pacific’, ournal of Banking & Finance, vol. 38, pp. 64-77.

Fratzscher, M. 2002, "On Currency Crises and Contagion", European Central Bank Working Paper No. 139.

Goddard, J & Wilson, J, O 2009, ‘Competition in banking: a disequilibrium approach’, Journal of Banking & Finance, vol.33(12), pp. 2282-2292.

Guesmi, K, Teulon, F & Muzaffar, A, T 2014, ‘The evolution of risk premium as a measure for intra-regional equity market integration’, International Review of Financial Analysis, vol. 35, pp. 13-19.

International Monetary Fund (IMF) 2016, IMF Data Mapper, viewed 13 August 2016, < http://www.imf.org/external/datamapper/index.php>.

Iwata, G. 1974. "Measurement of Conjectural Variations in Oligopoly." Econometrica, vol.42(5), pp. 947-66.

Jeon, B. N. Olivero, M. P. & Wu, J. 2011, ‘Do foreign banks increase competition? Evidence from emerging Asian and Latin American banking markets’, Journal of Banking & Finance, vol. 35(4), pp 856-875.

Karim, A. and M. Zaini. 2001. "Comparative Bank Efficiency across Select ASEAN Countries." ASEAN Economic Bulletin, vol. 18(3), 289.

Khan, H. H., Ahmad, R. B. & Gee, C, S 2016, ‘Bank competition and monetary policy transmission through the bank lending channel: Evidence from ASEAN’, International Review of Economics & Finance, vol. 44, pp. 19-39.

Lau, L 1982, ‘On Identifying the Degree of Competitiveness from Industry Price and Output Data’, Economics Letters, vol. 10, pp. 93-99.

Liu, H, Molyneux, P & Nguyen, L, H 2012, ‘Competition and risk in South East Asian commercial banking’, Applied Economics, vol. 44, no. 28, pp. 3627-3644.

Matthews, K. and Zhang, N. 2010, ‘Bank productivity in China 1997-2007: Measurement and convergence’. China Economic Review, vol.21(4), pp.617-628.

Mulyaningsih, T., Daly, A. & Miranti, R. 2015, ‘Foreign participation and banking competition: Evidence from the Indonesian banking industry’, Journal of Financial Stability, vol. 19, pp. 70-82.

Northcott, C. A. 2004. ‘Competition in Banking: A Review of the Literature,’ Bank of Canada Working Paper 2004-24. Bank of Canada.

Olivero, M. P., Li, Y. & Jeon, B. N., 2011, ‘Competition in banking and the lending channel: Evidence from bank-level data in Asia and Latin America’, Journal of Banking & Finance, vol. 35(3), pp. 560-571.

Panzar, J. C. & Rosse, J. N. 1982. "Structure, Conduct, and Comparative Statics," Bell Laboratories economic discussion paper No. 248.

Panzar, J. C. & Rosse, J. N. 1987, ‘Testing for Monopoly Equilibrium’, The Journal of Industrial Economics, vol. 35, pp. 443-456.

Peltzman, S. 1977. ‘The Gains and Losses from Industrial Concentration.’ Journal of Law & Economics, vol. 20(2), pp. 229-63.

Phillips, P.C.B.& Sul, D., 2007. ‘Transition Modelling and Econometric Convergence tests’, Econometrica, vol. 75(6), pp.1771-1855.

Rizavi, S. S.; B. Naqvi and S. K. A. Rizvi. 2011. "Global and Regional Financial Integration of Asian Stock Markets." International Journal of Business and Social Science, vol. 2(9), 82-93.

Rosse, J. N. & Panzar, J. C. 1977, ‘Chamberlin vs. Robinson: An Empirical Test for Monopoly Rents’, Stanford University Studies in Industry Economics, no. 77.

Rughoo, A. & Saranties, N., 2012, ‘Integration in European Retail Banking: Evidence from savings and lending rates to non-financial corporations’, Journal of International Financial Markets, Institutions and Money, vol. 22, pp.1307-1327.

Rughoo, A. & Saranties, N., 2014, ‘The Global Financial Crisis and Integration in European Retail Banking’, Journal of Banking and Finance, vol. 40, pp.28-41.

Sala-i-Martin, X. X. 1996. ‘The Classical Approach to Convergence Analysis’, The Economic Journal, vol. 106(437), 1019-36.

Sealey, G. W. & Lindley, J. T. 1977, ‘Inputs, outputs and a theory of production and cost at depository financial institutions’, Journal of Finance, vol. 32, pp. 1251-1266.

Schuler, M. & Heinemann, F., 2002. ‘How Integrated are European Retail Financial Markets? A Cointegration Analysis’. ZEW Papers, Germany.

Shaffer, S. 1982, ‘Competition, Conduct and Demand Elasticity’, Economics Letters, vol. 10, pp. 167-171.

Shaffer, S 2004, ‘Patterns of competition in banking’, Journal of Economics and Business, vol. 56(4), pp. 287-313.

Shaffer, S. and DiSalvo, J. 1994. ‘Conduct in a Banking Duopoly.’ Journal of Banking & Finance, vol.18(6), 1063-82.

Tan, M, S 2016, ‘Policy coordination among theASEAN-5: A global VAR analysis’, Journal of Asian Economics, vol. 44, pp. 20-40.

Tang, K 2011, ‘The precise form of uncovered interest parity: A heterogeneous panel application in ASEAN-5 countries’, Economic Modelling, vol. 28(1-2), pp. 568-573.

The Association of Southeast Asian Nations 2016, About ASEAN, viewed 25 July 2016, <http://asean.org/asean/about-asean/>.

Vesala, J. 1995. "Testing for Competition in Banking: Behavioral Evidence from Finland," Bank of Finland Studies, Working Paper, No. E:1.

Weill, L 2013, ‘Bank competition in the EU: How has it evolved?’, Journal of International Financial Markets, Institutions and Money, vol. 25, pp. 100-112.

Zhang, T. and Matthews, K. 2012, ‘Efficiency convergence properties of Indonesian banks 1992-2007’, Applied Financial Economics, vol.22(17), pp.1456-1478.

The Impact of Realized Jumps and Continuous Variance on Variance Risk Premium

R.L. Shankar, Great Lakes Institute of Management, Chennai, India

Jijo Lukose P.J., Indian Institute of Management Kozhikode, India

Ganesh Sankar, Institute for Financial Management and Research, Chennai, India

ABSTRACT

Realized variance has two components: jumps and continuous variance. This article examines if either of this component has any predictive power in forecasting future returns of synthetic variance swaps. We work with Indian options market, a market that is dominated by retail participants. We first establish the presence of variance risk premium using both model-based and model-free approaches. Our result is robust to alternate specifications of volatility, sampling frequencies and sample periods. We then split realized variance into its two components. We find that only past continuous variance is significant in forecasting short-term synthetic variance swap returns; realized jumps do not have any predictive power. These results suggest that the continuous component of the quadratic variation in returns is a key determinant of the variance risk premium.

Keywords: Variance risk premium, Model free implied volatility, Synthetic variance swaps, Jumps, Realized variance

JEL Classification: G12, G13, G14

1. Introduction

The price of an option is a non-linear function of the underlying asset price; hence, option writers are exposed to changes in the price of the underlying and its higher moments, specifically increase in volatility. Under the Black and Scholes (1973, BS hereafter) option pricing framework, volatility is assumed to be non-stochastic; hence, an option seller is compensated only for bearing the price risk. In reality, volatility is stochastic; long positions in options that are market-neutral benefit from an unexpected increase in volatility, which typically coincides with market declines. Since such positions act as a hedge against volatility, the seller will demand a premium for bearing the underlying variance risk. This is often referred to as variance risk premium (VRP).

Bollerslev et al. (2009) document that VRP explains a non-trivial fraction of the time-series variation in aggregate stock market returns. What drives VRP? While the presence of VRP is well documented, literature on the drivers of VRP is fairly limited. Realized variance has two components: unexpected market jumps and continuous variance. Recent developments in financial econometrics have permitted researchers to disentangle these components using high frequency data. Still, little is known about the impact that these components have on variance risk premium. In the current paper, we address this question using data from an emerging market.

Our central contributions are two-fold. First, we establish the presence of VRP in an emerging market dominated by retail participants using recently advanced econometric techniques. While prior studies on mature markets find overwhelming evidence in favour of variance risk premium, Yoon and Byun (2009) find that variance risk is not priced in Kospi 200 options traded in Korea Stock Exchange. They attribute this unique finding to lack of hedging demand from retail participants. We use data from Indian market, which is similar to the Korean market in terms of retail participation[77], and examine if variance risk is indeed priced in this market.

Prior investigations on variance risk premium use techniques that can be classified as either model-dependent or model-free. The first approach examines returns from market-neutral option positions (Coval and Shumway, 2001; Bakshi and Kapadia, 2003a; Bakshi and Kapadia, 2003b; Low and Zhang, 2005; Goltz and Lai, 2009). A common feature of these studies is that they employ hedge ratios based on a model such as BS; hence, they are joint tests on the assumptions of the underlying model and the presence of risk premium. To address this, recent studies employ a model-free estimate of variance risk premium; this is defined as the difference between the expectation of future return variance under physical and risk-neutral measures (Bondarenko, 2004; Bollerslev et.al, 2009; Carr and Wu, 2009; Driessen et.al, 2009; Bollerslev et.al, 2011). These studies confirm that variance risk is priced in index options traded in US markets. In the current paper, we employ both model-dependent and model-free techniques to examine the pricing of variance risk.

[77] Globally, Korea Stock Exchange and National Stock Exchange of India are ranked first and second respectively in terms of number of stock index options contracts traded (Source: World Federation of Exchanges, 2010). The retail holding of the shares underlying the index for the period (2004-2012) is about 14% (Indian Securities Market, A Review published by NSE). The retail participation in Indian index options market for the period 2004-10 is 58.4% (Source: nse-india.com). NSE started disseminating data on the extent of shorting by participant category from 2012; the average contribution of retail to short index option positions for the period Jan 2012 - Dec 2012 is about 50%. These participation numbers are comparable to the Korean markets: 50% in options and 20% in the underlying market.

Our second contribution is to examine the drivers of variance risk premium. Specifically, realized variance can be split into two components: unexpected market jumps and continuous variance. We examine if either of these components has an impact on variance risk premium. Jacquier and Okou (2013) find that these two components have different predictive powers on future long-term excess market returns. While continuous variance is found to be a key driver of medium to long-term excess returns, jumps have little predictive power. They conclude that realized jumps are not a state variable driving the market risk premium dynamics. We extend their framework to the context of variance swap returns and examine if either of these components are a determinant of variance risk premium.

In addressing this question, we add to the nascent literature on the drivers of variance risk premium. Bakshi and Madan (2006) posit a linkage between variance risk premium and factors such as investor risk aversion and higher order return moments. Carr and Wu (2009) document the presence of variance risk premium in a large number of indices and stocks. They then examine if such premium can be explained by classical risk factors. Todorov (2010) identifies two sources of variance risk: (i) stochastic volatility, or the continuous change in volatility and (ii) unexpected jumps in market returns. He then examines the pricing of these two components in a general semi-parametric framework.

Our main findings can be briefly summarized as follows. We find that market-neutral straddles yield excess returns that are negative and statistically significant; this suggests that higher order risks are priced. Our results are robust to alternate specifications of volatility, different sampling frequencies and sample periods. To validate this central finding, we compare realized variance estimated from intraday data with model-free implied variance estimated from traded options. This difference is found to be negative and significant. Also, long positions in synthetic variance swaps on Nifty are found to yield statistically significant negative returns. We conclude that variance risk is indeed priced in the Indian options market. Further, we find that while past continuous variance has significant predictive power in forecasting short-term synthetic variance swap returns, realized jumps do not have any such power. These results suggest that the continuous component of the quadratic variation in returns is a key determinant of the variance risk premium.

The rest of the paper is organized as follows. The second section details the underlying theoretical framework and the testable hypothesis that emerge out of it. The third section provides relevant information about National Stock Exchange and our sample data. In the fourth section, we analyze returns of zero-beta straddles and document the findings of our model-free tests. In the same section, we also examine the predictive power of the two components of realized variance. We conclude in the last section.

2. Theoretical framework and testable hypothesis

Examining VRP: Model-dependent tests

Since option positions are highly leveraged bets, the price of an option should reflect the leverage implicit in the option position. Black and Scholes (1973) formalize this linkage through the beta of the option. If the underlying asset has a positive beta, beta of a call option would be positive and that of a put option would be negative. Hence, as illustrated by Coval and Shumway (2001), calls and puts can be combined in an appropriate ratio such that the beta of the resulting position is zero. This combination is referred to as a zero-beta straddle. If only market risk is priced, expected excess returns from such a beta-neutral position should be zero. This yields our first hypothesis: Expected excess returns from zero-beta straddles is zero.

Straddles benefit from large movements in the price of the underlying; hence, their premiums reflect the probabilities assigned to such extreme moves. If ex-ante, high probabilities are assigned to jumps and if these are not subsequently realized, returns from straddles would be low. To address any bias that might arise due to non-realization of such high probabilities, we study returns of crash-resistant straddles. These are constructed by augmenting an At-the-money (ATM) straddle with a short position in an Out-of-The-Money (OTM) put option. These straddles yield a fixed payoff irrespective of the magnitude of crash. If returns from such positions are still negative and statistically significant, it can be concluded that low returns from straddles cannot be explained by the mispricing of crash risk. This yields our second testable hypothesis: Expected excess returns from crash-resistant zero-beta straddles is zero.

Examining VRP: Model-free tests

The tests discussed in the earlier section use BS hedge ratios; hence, they are joint tests of both the underlying assumptions of the BS framework and the presence of risk premium. To address this, we validate our findings with model-free tests of variance risk premium. In accordance with extant literature, we define variance risk premium as the difference between the expectation of future realized variance under the physical (P) and risk-neutral (Q) measures (Bondarenko, 2004; Bollerslev et.al, 2009; Driessen et.al, 2009; Bollerslev et.al, 2011). Formally, if we denote by $RV_{t,T}$ the return variance between time t and T, Variance Risk Premium is defined as

$$VRP_t = \mathrm{E}_t^P RV_{t,T} - \mathrm{E}_t^Q RV_{t,T} \tag{1}$$

As shown by Britten-Jones and Neuberger (2000), the expectation of future return variance under risk-neutral measure can be estimated from prices of European call options. Since this estimate of $\mathrm{E}_t^Q(RV_{t,T})$ doesn't assume any option-pricing model, it is commonly referred to as Model-Free Implied Variance (MFIV). Applying law of iterated expectations to (1), a model-free estimate of variance risk premium can be obtained as the difference between the unconditional means of RV and MFIV.

To obtain the unconditional mean of Realized Variance, we use high frequency data. Prior studies (Andersen and Bollerslev, 1998; Andersen et.al., 2001; Barndoff-Nielsen and Shephard, 2002) establish that realized variance measures estimated using intraday data yield a superior estimate of actual return variance than those estimated using daily returns. If the period $[t, T]$ is divided into N equally spaced intervals and if the intra-day log return over an interval i is denoted by r_i, then the Realized Variance for this period is computed as

$$RV_{t,T} = \sum_{i=1}^{N} r_i^2 \tag{2}$$

Realized Variance as defined above is a consistent estimator of quadratic variation of r_i in the limit $N \to \infty$. In reality, it is not possible to sample continuously. Further, (2) assumes that intraday returns form an iid sequence with mean zero and finite variance; hence, at larger N, microstructure effects such as bid-ask bounce would create a bias in the estimate of volatility. The optimal interval for sampling intraday returns has attracted much attention; Jian and Tiang (2005) provide a good summary of this literature. We adapt their approach and sample returns at five-minute intervals and correct for first-order autocorrelation (using the formulae provided therein). If variance risk is priced in the options market, then the model-free estimate of variance risk premium as specified in (1) should be non-zero; this yields our third hypothesis: the model-free estimate of variance risk premium is zero.

Prior studies have also tested the above hypothesis in its ratio form, i.e. the average of the ratio of Realized Variance and MFIV equals one (Bondarenko, 2004; Driessen et.al, 2009). This is motivated by the fact that this ratio measures the returns from going long a variance swap, which is a forward contract on future variance that pays the difference between Realized Variance and a contractually agreed fixed rate. A long position in variance swap pays off when there is an unexpected increase in variance; hence, it acts as a hedge against the variance risk. If this risk is priced in a market, expected returns of such swaps would be negative. Even if such a product were not traded in the market, it can be synthetically replicated using a portfolio of traded options (Bondarenko, 2004; Carr and Wu, 2009; Trolle and Schwartz, 2010). Hence, the third hypothesis (stated earlier) is equivalent to the following: average return of synthetic variance swaps is zero.

Determinants of VRP

Realized variance as defined above has two components: jumps and continuous variance. We next examine if either of these components can predict VRP or specifically, future returns of synthetic variance swaps. We start by assuming that the logarithmic stock price follows a jump-diffusion process:

$$\delta s_t = \mu_t^s \delta t + \sigma_t^s \delta W_t + J_t^s \delta q_t \qquad (3)$$

where s_t is the logarithmic stock price, μ is the drift, σ is the volatility, and J_t is the jump size. The price process is driven by two stochastic components: W_t, a Brownian motion and q_t, a Poisson process with intensity λ^s. J_t, the size of jumps in log stock prices, in turn is assumed to follow a normal distribution with mean μ_J^s and σ_J^s. A number of techniques have been proposed in recent years to detect jumps from high frequency data; Barndoff-Nielsen and Shephard (2004, BNS hereafter) is an early reference. Zhang et.al. (2009) use BNS' approach with certain modifications; we adapt their approach in the current study. We briefly summarize the main intuition here; implementation details are provided in Appendix A.

In the absence of jumps, RV (realized variance) is a consistent estimator of quadratic variation of returns. In the presence of jumps, $RV_{t,T}$ converges to $\int_t^T \sigma^2 \delta s + \sum_{i=1}^{N} J_i^{\,2}$ where as before N refers to the number of intervals between t and T (here, one day). BNS also define another measure called Realized Bipower Variation (denoted by $BV_{t,T}$), which is computed as

$$BV_{t,T} = \frac{\pi}{2} \sum_{i=2}^{N} | r_i \; || r_{i-1} | \qquad (4)$$

$BV_{t,T}$ in turn converges to $\int_t^T \sigma^2 \delta s$. This is a consistent estimator of the continuous component of the quadratic variation. This is referred to as continuous variance or $C_{t,T.}$ If for any period [*t*,*T*] the underlying process has a jump, the asymptotic difference between $RV_{t,T}$ and $BV_{t,T}$ is strictly positive. The contribution of jumps to total realized variance on any given day can be computed as

$$CJ_{t,T} = \frac{RV_{t,T} - BV_{t,T}}{RV_{t,T}} \qquad (5)$$

If $CJ_{t,T}$ is statistically significant, we infer that a jump has occurred during that period. To test for statistical significance, we need asymptotic results. If we denote by $V(CJ_{t,T})$ the asymptotic variance of $CJ_{t,T}$, then the scaled contribution of jumps is given by $z_{t,T} = \frac{CJ_{t,T}}{\sqrt{V(CJ_{t,T})}}$. This measure converges to a standard normal distribution. If $z_{t,T} > \Phi_{\alpha}^{-1}$, then a "significant jump" is said to have occurred. Here α is some high confidence level (taken here as 99%) and Φ is the cumulative density function for normal distribution. We next compute the distributional characteristics of jump amplitude, namely the mean and volatility of jump sizes. For estimating this, we need to make two further assumptions.

First, we assume that there is at most one jump during the given period. The size of jump can then be estimated as $RV_{t,T} - BV_{t,T}$.If we further assume that on jump periods jump size dominates the return, then the sign of the jump can be obtained as the sign of the return, i.e., *sign($r_{t,T}$)*. We can then compute realized jump for period [t,T], $J_{t,T}$ as

$$J_{t,T} = \text{sign}(r_{t,T}) \times RV_{t,T} - BV_{t,T} \times I\left(z_{t,T} > \Phi_{\alpha}^{-1}\right) \tag{6}$$

The continuous component of quadratic variation can then be estimated as

$$C_{t,T} = I\left(z < \Phi_{\alpha}^{-1}\right) \times RV_{t,T} + I\left(z_{t,T} > \Phi_{\alpha}^{-1}\right) \times BV_{t,T} \tag{7}$$

Jacquier and Okou (2013) show that these two components - $J_{t,T}$ and $C_{t,T}$ - have different predictive powers on future long-term excess market returns. While continuous variance is found to be a key driver of medium to long-term excess returns, jumps have little predictive power. They conclude that realized jumps are not a state variable driving the risk premium dynamics. We adapt their regression framework to the context of variance swap returns. To be specific, we estimate the following regression:

$$\ln\left(\frac{RV_{t,T}}{MFIV_{t,T}}\right) = \alpha + \beta C_{t-h,t} + \lambda J_{t-h,t} + \varepsilon_t \tag{8}$$

where h (=T-t) is the horizon over which the different volatility measures are computed. For our analysis on predictability, we use a horizon of month. To obtain the estimate of multi-step (or the monthly) variance from that of daily variance, we follow Andersen et al. (2007) and aggregate daily estimates of various variance measures. Specifically, $RV_{t,t+h}$ is measured as $RV_{t,t+h} = RV_{t,t+1} + RV_{t+1,t+2} + \cdots + RV_{t+h-1,t+h}$. We obtain similar estimates for other variance measures such as jumps and continuous variance.

We use the above regression framework to examine if either past realized jumps or continuous variance has any predictive power for short-term variance risk premium. We also extend the framework to include contemporaneous market returns. Carr and Wu (2009) use a similar regression framework to examine if variance risk premium can be explained by classical risk factors such as market returns, firm size, book-to-market value.

3. Data

Trading of equities and equity derivatives in India is concentrated on two exchanges: National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). While the latter is the oldest stock exchange in Asia and has the largest number of listed companies in the world, its share of derivatives trading is negligible. NSE accounts for about 98% of derivatives turnover in Indian bourses[78]. S&P CNX Nifty is NSE's key benchmark index; it is a market-capitalization weighted index that is adjusted for free-float. It is a well-diversified index that contains 50 stocks and accounts for 24 sectors of the economy.

Futures and options on Nifty were introduced in June of 2000 and 2001 respectively. These markets have recorded impressive growth since their inception; currently, index options on Nifty are ranked second globally in terms of contracts traded and sixth in terms of notional value. Further evidence on liquidity of Nifty options is provided by Grover and Thomas (2012) who compare bid-ask spreads of options on Nifty and S&P 500 options. They find that cross-sectional variation of liquidity is smaller for Nifty options; they conclude that on this metric, Indian markets are more liquid than US markets.

Our data source is the high frequency database obtained from NSE[79]. It contains time-stamped intraday prices of all transactions in the spot and derivatives segment. Additionally, snapshots of the entire order book at five different time points - 11:00, 12:00, 13:00, 14:00, and 15:00 - are also provided. Index options on Nifty were introduced in June 2001; since the market might have gone through a learning phase in its initial years, we use data for the period January 2004 to August 2010. For each day in our sample period, we identify the best buy and sell prices at 14:00 hours and use their mid-price for our analysis. This ensures that straddles are not constructed using asynchronous prices.

Our options sample is constructed as follows. First, we remove contracts whose prices violate model-free arbitrage bounds. Specifically, we remove call option prices that are outside the range $(Se^{-qT} - Ke^{-rT}, Se^{-qT})$ and put option prices which are outside the range $(Ke^{-rT} - Se^{-qT}, Ke^{-rT})$ where S is the spot price, K is the strike price of the option, T is the time to maturity, r is the risk-free rate and q is the annual dividend yield. We use 30-day T-Bill yield published jointly by FIMMDA[80] and Reuters as a proxy for risk-free interest rate. Nifty is not a total performance index; we build a time series of historical dividend yields by calculating the difference between returns on Nifty and Nifty Total Returns Index, a total performance index published by NSE.

Second, to reduce the impact of illiquidity and stochastic interest rates, we focus on short-maturity options. On any given day, for a given strike price, contracts of three different maturities are available for trading - ones that expire in the same month, ones that expire in the subsequent month and the ones that expire in the month after[81]. The average quoted spread (QS) for ATM same-month call options is about INR 1.65; the relative quoted spread (RQS) is about 2.3% of the call premium. The average QS and RQS for the next-month ATM call options are about INR 5.3 and 3.5%. Hence, near month options are relatively more liquid. Hence, for our model-based tests, we include only contracts that expire in the same month. We make one exception: to control for the expiration week effects (see Vipul, 2005) for early evidence from

[78] Source: Website of Securities and Exchange Board of India

[79] Regular trading on the exchange takes place between 09:15 hours and 15:30 hours. NSE additionally conducts a pre-open session between 09:00 hours and 09:08 hours. These revised timings came to effect from December 2009, prior to which the exchange was open between 09:55 hours and 15:30 hours.

[80] Fixed Income Money Market and Derivatives Association of India (FIMMDA) is an association of commercial banks, financial institutions and primary dealers.

[81] In March 2008, NSE also introduced contracts of fixed maturity: three that expire in the quarterly cycle (March, June, September and December) and five that expire in the subsequent semi-annual cycle (June and December). However, these contracts are highly illiquid

Indian markets), during the expiry week, we consider only contracts that mature during the subsequent month. All index options on Nifty can be exercised only on the expiry date.

For our analysis, we classify options based on their moneyness - ratio of the strike price of the option to the dividend-adjusted value of the underlying index $\frac{K}{Se^{-qT}}$. We identify options that have a moneyness ratio between 0.9875 and 1.0125 as At-the-Money options. We also consider four other bins for our analysis; these are defined in Table 1. For each bin, we choose the contract with the lowest spread between best buy and best sell prices[82].

4. Empirical evidence

To better anchor our empirical findings, we first present some stylized facts about returns of equity indices from Indian (Nifty) and US (S&P500) markets. During the period under study, Nifty earned an average daily return of 0.07%. For the same period, S&P 500 yielded a daily return of -0.01%. Emerging markets such as India are likely to witness more volatility than developed markets; this is affirmed by the higher standard deviation of Nifty daily returns (1.8%, as against 1.4% for S&P 500)[83].

Examining VRP: Model-dependent tests

We document the statistical properties of daily returns on call and put options on Nifty in Table 1. As can be seen from Panel A of Table 1, average excess returns are positive for call options across all bins; for ATM and mildly ITM and OTM options, these are statistically significant and increasing in strike price. The statistical properties of put option returns are reported in Panel B of Table 1. Average excess return is negative, statistically significant and increasing in strike price for all moneyness bins. On average, ATM put options lose 2.4% per day and deep OTM options lose 3.3% per day; a strategy of selling deep OTM options has an attractive Sharpe ratio of 1.94 (annualized). These results are comparable to those from more advanced markets; using data from 1987 - 2005, Broadie et.al (2009) find that OTM put options yield a monthly return of -57% and a Sharpe ratio of 0.27 (monthly).

To examine if risk of higher order moments is priced, we next study returns to zero-beta straddles. For each moneyness bin, we choose that strike price for which the combined bid-ask spread of call and put options is the lowest. We combine these options in a ratio that renders the overall beta of straddle to be zero. For computing option betas, we use BS implied volatility. These positions are held for a day; daily returns are computed based on prices observed on the next trading day.

We record descriptive statistics of zero-beta straddle returns in Panel A of Table 2. Straddles of all moneyness are found to earn returns that are negative and statistically significant. Specifically, ATM zero-beta straddles lose on average 0.74% per day. This implies that sellers of market-neutral straddles can earn returns that are both statistically and economically significant by selling calls and puts in an appropriate ratio. These figures are marginally higher than those reported for US markets; Coval and Shumway (2001) find that beta-neutral ATM straddles on S&P100 lose 0.5% per day and those on S&P 500 lose about 3.15% per week.

[82] During the period 2004 – 2007, options were introduced with a strike price difference of INR 10. Subsequently, options were introduced with a difference of INR 50. Further, new contracts are introduced if index falls or rises beyond a particular threshold during the previous trading day.
[83] Also, based on unreported results, Nifty returns have lower skewness and kurtosis than S&P500 returns.

Table 1: Unhedged option returns

The following table presents descriptive statistics for daily excess returns from Nifty call and put options for the period January 2004 to August 2010. t-statistics are reported in brackets. All returns are reported in daily percentage terms. Contracts have been split into five different bins based on their moneyness. We define moneyness as $\frac{K}{Se^{-qT}}$ where K is the strike price, S is the index level, q is the annual dividend yield and T is the time to maturity. Each bin has contracts whose moneyness is +/- 0.125 from the center. For example the bin 0.95 has contracts with moneyness between 0.9375 and 0.9625. S.R. refers to annualized Sharpe ratios.

Moneyness	0.95	0.975	1	1.025	1.05
Panel A: Call options					
Count	1412	1565	1604	1551	1178
Mean	0.77%	1.28%	1.60%	1.68%	0.71%
	(1.45)	(2.23)	(2.40)	(2.10)	(0.64)
Std. dev.	0.19	0.22	0.26	0.31	0.37
Skew	1.04	0.99	1.11	1.48	3.16
Kurt	7.88	8.19	7.38	9.35	31.63
S.R.	0.74	1.07	1.14	1.02	0.36
Panel B: Put options					
Count	1525	1581	1589	1485	1052
Mean	-3.36%	-2.74%	-2.46%	-2.12%	-1.58%
	(-3.98)	(-3.46)	(-3.61)	(-3.39)	(-2.22)
Std. dev.	0.33	0.32	0.27	0.24	0.23
Skew	1.85	2.19	1.38	0.97	0.84
Kurt	7.30	12.76	5.34	2.70	2.26
S.R.	-1.94	-1.66	-1.73	-1.68	-1.30

Table 2: Zero-beta straddle returns

Panel A of the following table presents descriptive statistics for daily excess returns from zero-beta straddles on Nifty for the period January 2004 to August 2010. Panel B presents the corresponding statistics for Crash-Resistant (CR) zero-beta straddles on Nifty for the same period. Unlike regular straddles, these contracts pay a fixed amount even if the market crashes by a large extent; this is achieved by augmenting ATM straddle with a short OTM put position. Both these straddles are constructed such that they have zero beta at inception; they are further rebalanced on a daily basis. In both panels, t-statistics are reported in brackets. All returns are reported in daily percentage terms. S.R. refers to annualized Sharpe ratios.

Panel A						Panel B
	Zero-beta straddles					CR zero-beta straddles
Moneyness	0.95	0.975	1	1.025	1.05	
Count	1380	1554	1589	1485	1051	1518
Mean	-0.83%	-0.70%	-0.74%	-0.82%	-0.81%	-0.53%
	(-3.83)	(-3.39)	(-3.74)	(-3.74)	(-2.60)	(-2.29)
Std. dev.	0.08	0.08	0.08	0.08	0.10	0.09
Skew	2.43	4.00	3.42	2.85	1.70	4.95
Kurt	16.17	33.17	27.52	22.05	7.93	78.40
S.R.	-1.97	-1.64	-1.79	-1.85	-1.53	-1.12

To verify if high probabilities assigned ex-ante to crashes drive the low returns of straddles, we adapt Coval and Shumway's (2001) approach and construct crash-resistant market-neutral straddles. We combine an ATM straddle with the most liquid OTM put with moneyness between 0.95 and 0.90. The average moneyness of the OTM puts used in our analysis is 0.93. This portfolio is also constructed such that it is beta-neutral. We document statistical properties of returns earned by these strategies in Panel B of Table 2. ATM crash-resistant zero-beta straddles lose 0.53% per day; these returns are statistically significant. Hence, ATM zero-beta straddles earn negative returns even after selling the associated crash insurance. It can be inferred that low returns from zero-beta straddles cannot be explained by mispricing of crash risk.

We next undertake a series of tests to verify the robustness of our results. First, we analyze the sensitivity of our results to the holding period by examining weekly returns. Rebalancing at lower frequencies increases the chances of the position not being market-neutral at the end of the period; however, it translates to lower transaction costs, which could be critical from a trader's perspective. For sake of brevity, we haven't tabulated the results. We find that with weekly rebalancing, ATM straddles lose 3.59% per week; associated t-statistics suggest that these returns are statistically significant even at 1% level. ATM crash-resistant straddles that are rebalanced weekly lose 2.84% per week; these results are again significant. Hence, our findings are robust to the frequency at which the underlying straddles are rebalanced.

Second, we gauge the robustness of our findings to alternate measures of volatility forecasts. Our earlier computations use BS implied volatility of an option as an input for computing its beta. To verify the robustness of our results, we build a time series of model-free implied volatility from traded options; returns from market-neutral straddles whose betas are computed using these model-free estimates of volatility are studied. Based on untabulated results, we find that that our earlier results are largely invariant to the specification used in forecasting volatility; returns are still negative and statistically significant.

Third, we analyze if our findings are consistent over different sample periods. We observe a dramatic increase in both Realized Variance and Model-free Implied Variance during the second half of 2008 (Refer Figure 1); hence, we split our sample into two periods: Jan 2004-June 2008 and July 2008-Aug 2010. Table 3 reports the statistical properties of daily excess returns for zero-beta straddles of our subsample analysis. With minor exceptions, returns from straddles are negative, economically and statistically significant across all sub-samples. Average returns are more negative during the second sub-sample. This could perhaps be due to a higher variance risk premium demanded by market participants following the onset of the 2008 global financial crisis.

Finally, we check the robustness to the underlying model itself. The results reported hitherto are based on Coval and Shumway's (2001) approach of constructing zero-beta straddles. To ensure that our results are robust to the approach used, we also adapt the Bakshi and Kapadia (2003a)'s methodology. Under this, we examine returns from delta-hedged option positions. Net gains from such positions should be zero if (a) variance is constant or (b) if variance is stochastic, but volatility risk is not priced risk in options. We report the results for calls and puts in Table 4**.** We find that in both cases the delta-hedged option positions underperform zero. Hence, we conclude that our result is not sensitivity to the underlying model used for computing the hedge ratios. We conclude that our results are robust to alternate specifications of volatility, different sampling frequencies and sample periods. These findings provide initial evidence for pricing of higher order moments in Indian index options market

Table 3: Robustness tests for model-based tests: Subsample Analysis

The following table presents results for different sub-sample periods. Descriptive statistics for daily returns of market-neutral straddles on Nifty is presented. t-statistics are reported in brackets. Daily returns are reported in daily percentage terms.

	Daily returns	
Moneyness	Jan 2004 - Jun 2008	Jul 2008- Aug 2010
0.95	-0.80%	-0.90%
	(2.92)	(2.54)
0.975	-0.62%	-0.87%
	(2.44)	(2.47)
1.0	-0.58%	-1.11%
	(2.41)	(3.10)
1.025	-0.74%	-0.98%
	(2.73)	(2.62)
1.05	-0.68%	-0.98%
	(1.50)	(2.40)

Examining VRP: Model-free tests

To further validate our findings, we compute model-free estimates of variance risk premium. This is obtained as the difference between the average of Realized Variance and Model-free Implied Variance (MFIV). We adapt the numerical procedure suggested by Jiang and Tian (2005) for estimating MFIV. If call options with strike prices between $K_{\min}$ and $K_{\max}$ are available, then the MFIV can be approximated using the following summation

$$MFIV_{t,T} = 2\sum_{K_{\min}}^{K_{\max}} \frac{C_t\left(T, \frac{K}{B(t,T)}\right) - C_t\left(t, \frac{K}{B(t,t)}\right)}{K^2} \Delta K \qquad (9)$$

where $C_t(T,K)$ denotes the time-t price of a European call option maturing at time T with strike price K, B(t,T) denotes the time-t price of a zero-coupon bond maturing at time T and ΔK refers to the distance between successive strike prices.

Table 4: Delta Hedged Option Position

We report the results of delta hedged gains for both calls and puts on Nifty. Panels A and B report the returns from delta hedged calls and puts respectively. t-statistics are reported in brackets. The hedged portfolio is rebalanced on a daily basis from expiry to expiry. The results reported here are gains as a percentage of the option premium.

Moneyness	0.95	0.975	1	1.025	1.05
	Panel A – Calls				
Count	90	98	120	158	77
Mean	-0.09	-0.07	-0.11	-0.15	0.01
	(-4.17)	(-1.96)	(-2.48)	(-2.13)	(0.06)
Stdev	0.20	0.33	0.48	0.91	1.27
Skew	-0.94	-2.12	-1.60	-1.47	-1.74
Kurt	1.78	9.71	4.09	2.40	7.37
	Panel B – Puts				
Count	75	77	78	76	59
Mean	-0.28	-0.26	-0.19	-0.11	-0.10

	(-3.23)	(-4.54)	(-3.86)	(-2.65)	(-2.61)
Stdev	0.74	0.50	0.44	0.38	0.28
Skew	1.12	1.35	2.38	2.08	1.30
Kurt	7.48	2.80	8.10	5.16	3.10

Lower ΔK, lower is the discretization error introduced by the numerical approximation. However, in reality, lower bound for ΔK is determined by the granularity of strike prices listed in the exchange. To overcome this, option prices are interpolated between available strike prices. To be precise, since option prices are highly non-linear functions of strike prices, the interpolation is done on implied volatilities instead of option prices; these interpolated volatilities are then translated to option prices.

If on day t, options expiring on T_1 are used as inputs to the above procedure, we obtain an estimate of MFIV for maturity $T_1 - t$. To build a time-series of fixed-maturity variance, we adapt Carr and Wu (2009)'s two-step procedure. First, we compute MFIVs based on options expiring on T_1 (the nearest maturity) and T_2 (the next available maturity). Next, we linearly interpolate the variance between these two dates to obtain a one-month MFIV. Specifically, if we denote by T the trading day that is a month ahead, then the fixed-maturity $MFIV_{t,T}$ is given by

$$MFIV_{t,T} = \frac{1}{T-t}\left(\frac{MFIV_{t,T_1}(T_1-t)(T_2-T)+MFIV_{t,T_2}(T_2-t)(T-T_1)}{T_2-T_1}\right) \qquad (10)$$

This procedure yields a MFIV forecast corresponding to the option maturity. To clarify, if on day t, options expiring on T_1 are used as inputs, we obtain an estimate of MFIV for maturity $T_1 - t$. However, we need to build a time-series of fixed-maturity variance. For this, we adapt Carr and Wu's (2009) two-step procedure. First, we compute MFIVs based on options expiring on T_1 (the nearest maturity) and T_2 (the next available maturity). Next, we linearly interpolate the variance between these two dates to obtain a one-month MFIV. We use both options expiring in the current month and the next month for purposes of interpolation.
Next, we estimate the one-month Realized Variance for each trading day using intraday data sampled at 5-minute intervals. These returns display a first order autocorrelation of -0.048. The higher order autocorrelations are much smaller; for instance, the second-order autocorrelation is 0.004. Hence, we correct only for the first-order autocorrelation using the procedure detailed in Jiang and Tian (2005).

Figure 1 plots the Realized Variance and MFIV over our sample period. MFIV is greater than RV for most part of our sample; this stylized fact is in accordance with results from other markets (Jackwerth and Rubinstein, 1996, is an early reference). We report results of our model-free tests in Panel A of Table 5. The sample mean of RV for the period is 0.0470 on annualized basis; this is lower than that of MFIV, which is 0.0955. The corresponding volatility numbers are 21.67% and 30.9%. Hence, the model free estimate of variance risk premium is -0.0484. This estimate is found to be statistically significant based on standard errors corrected using Newey-West (1987) with 22 lags. In Panel B of Table 5, we report the descriptive statistics of log returns from a long position in synthetic variance swaps on Nifty. In using log returns, we follow the recommendation of Carr and Wu (2009) who state that such a transformation renders the return distribution closer to normality. These positions on average yield a monthly log return of -94.19%; the null that the mean of log returns is zero is rejected at the 1% confidence level.

Table 5: Model-free estimates of variance risk premium

Panel A reports our model-free estimate of variance risk premium based on data for the period Jan 2004-Aug 2010. Realized variance is computed from five-minute intra-day returns over a one-month period; the estimate is adjusted for first order autocorrelation. The one-month model-free implied variance is obtained from traded Nifty options using the methodology of Jiang and Tian (2005) described in Section 2. Variances are expressed in annual terms. Panel B reports descriptive statistics of monthly log returns $\left(\ln\left[\frac{RV}{MFIV}\right]\right)$ from long positions in synthetic variance swaps. t-statistics based on standard errors corrected using Newey-West (1987) with 22 lags are reported in brackets. Skew and kurtosis denote skewness and excess kurtosis respectively.

Panel A: Model free estimate of Variance Risk Premium	
Count	1600
Mean RV, $\overline{RV}$	0.0470
Mean MFIV, $\overline{MFIV}$	0.0955
VRP: $\overline{RV} - \overline{MFIV}$	-0.0484
	(-8.39)
Panel B: Descriptive statistics of Variance Swap Returns	
Mean	-94.19%
	(-15.52)
Std. dev.	2.42
Skew	1.26
Kurtosis	2.47

Our estimates of swap returns are higher than those estimated for US markets by Carr and Wu (2009); they find that the mean log return of variance swaps on S&P 500 is -66%. We next examine the robustness of these results. First, we gauge the sensitivity of our results to inclusion of overnight returns. In our earlier analysis, we follow Andersen et.al. (2001) and Wu (2011) and consider only intraday returns for estimating realized variance. However, Bollerslev et.al. (2009) compute realized variance as sum of squared intra-day and overnight returns. The additional information in these overnight returns would result in an increase in the estimate of realized variance and decrease in that of variance risk premium.

For sake of brevity, we do not tabulate these results. We find that realized variance increases to 0.0788 and variance risk premium falls to -0.0167. However, the null of zero variance risk premium continues to be rejected at 5% confidence interval. Returns of synthetic variance swaps continue to be negative and statistically significant. Second, we examine if our results hold across different sample periods. We consider the same sub-samples as before; the results are reported in Table 6. We find that results are consistent across both periods; variance risk premium is negative and significant. Our estimate of variance risk premium increases during the second sub-sample; this partially explains the higher negative returns of market-neutral straddles during this period.

Table 6: Robustness checks for model-free tests: Subsample analysis

The following table presents results for different sub-sample periods: Jan 2004 - Jun 2008 & Jul 2008-Aug 2010. Panel A reports our model-free estimate of variance risk premium. Realized variance is computed as sum of squared five-minute intra-day and overnight returns over a one-month period; this estimate is adjusted for first order autocorrelation. The one-month model-free implied variance is obtained from traded Nifty options using the methodology of Jiang and Tian (2005) described in Section 2. Variances are expressed in annual terms. Panel B reports descriptive statistics of monthly log returns $\left(\ln\left[\frac{RV}{MFIV}\right]\right)$ from long position in synthetic variance swaps. t-statistics based on standard errors corrected using Newey-West (1987) with 22 lags, are reported in brackets. Skew and kurtosis denote skewness and excess kurtosis respectively.

	Jan 2004 - Jun 2008	Jul 2008 - Aug 2010
Panel A: Model free estimates of Variance Risk Premium		
Count	1085	515
Mean RV, $\overline{RV}$	0.0412	0.0594
Mean MFIV, $\overline{MFIV}$	0.0786	0.1311
VRP: $\overline{RV}-\overline{MFIV}$	-0.0374	-0.0717
	(-6.11)	(-6.56)
Panel B: Descriptive Statistics of Variance Swap Returns		
Mean	-88.69%	-105.79%
	(-11.72)	(-11.05)
Std. dev	2.49	2.17
Skew	1.36	0.69
Kurt	2.45	0.83

We conclude that results of our model-free tests are robust to alternate specifications of realized variance and are consistent across different sample periods. These findings affirm our conclusion that variance risk is priced in Indian markets; this is in agreement with most studies from mature markets. It is interesting to note that the result holds despite the fact that the participation of retail investors in options market is higher than that in the underlying market.

We can only speculate on why the results for Indian markets are different from those for Korean markets (Yoon and Byun, 2009, YB hereafter). This could be attributed to differences in either the market structure or the methodology employed. While both markets are similar in terms of retail participation, the nature of these participants could be different. For instance, it is possible that participants in one market are more sophisticated than the others. However, such differences are difficult to quantify or measure. On the methodology front, there are three considerable differences. First, YB employ only the model-dependent approach advanced by Bakshi and Kapadia (2003a). We additionally use the recently developed model-free techniques that enable us to examine variance risk premium in isolation of any specific modeling framework.

Second, YB use standard deviation of returns and forecasts from GARCH (1, 1) models to obtain volatilities required for computing BS hedge ratios. Jiang and Tian (2005) establish that a model-free estimate of implied volatility estimated from traded options subsumes all information contained in BS implied volatility and historical volatility; such an estimate also provides a more efficient forecast of future realized volatility. We use both BS and model-free estimates of implied volatility for computing option hedge ratios. Third, YB use daily data to estimate volatility. As is well documented (Andersen and Bollerslev, 1998; Andersen et.al., 2001; Barndoff-Nielsen and Shephard, 2002) realized variance measures computed using intraday data yield a superior estimate of actual return variance than those that use daily returns.

In our study, we use intraday returns sampled at five minute intervals for estimating volatility. To summarize, the divergence in these results could be attributed to differences in either the market structure or the methodology employed. A detailed empirical comparison of these two markets could provide more insights; however, it is beyond the scope of the current study.

Determinants of VRP

Realized variance has two components: continuous variance and jumps. We use the methodology characterized by Equations 6-7 to disentangle these components from realized variance. All these estimates are computed for a daily horizon. Basic descriptive statistics of jump parameters are presented in Table 7. A plot of daily RV and contribution of daily realized jumps to total realized variance is plotted in Figure 2.

Table 7: Descriptive statistics of Jumps and Continuous Variance

The following table reports the descriptive statistics of jump and continuous variance. We use the methodology outlined in Equations (6) - (7) to disentangle jump (J) and continuous variance (C) from realized variance. All these estimates are computed over daily horizon. Mean and other statistics for jumps are computed based on the days on which such jumps occur. We use a confidence level of 99%ile to detect jumps.

	J	C
Mean	4.72E-05	1.92E-04
Median	2.42E-05	9.64E-05
Stdev	8.76E-05	5.13E-04
Skew	6.3145	19.2821
Kurt	47.8503	507.3861

We detect jumps on 83 out of the 1648 days considered in our analysis. The implied jump intensity of 5% is slightly lower than that reported for US indices by Jacquier and Okou (2013). They use daily prices to identify monthly jumps in NYSE/AMEX value-weighted index. The jump intensity is estimated as 8.3% for the period January 1952 to December 2009. Using daily data for the period January 2001 to December 2003, Zhang et al (2009) estimate jump intensity to be 8.48% for firms that are traded in US and are rated between A and AAA.We further find that on the days that jumps occur, their mean contribution to total realized variance is 34.3%. This is again comparable to figures reported by Zhang et al. (2009). For all firms in their sample (which is not limited to just AAA-A firms), they find that the mean contribution of jumps is 52.3%.

To determine if either of jumps or continuous variance has any predictive power for future variance swap returns, we adapt the regression framework of Jacquier and Okou (2013). We estimate the regression equation (8) with daily time-series of monthly returns. To address the issue of overlapping data, we follow Carr and Wu (2009) and report t-statistics computed according to Newey and West (1987) with 30 lags for the overlapping daily series. To be specific, we use 30-day returns sampled at a daily interval. We report the results in Panel A of Table 8.

We find that lagged continuous variance is positive and statistically significant. This suggests that an increase in the continuous component of variance leads to higher variance risk premium in future periods. Hence, persistent changes in volatility, as captured by the continuous component, appear to be a priced risk factor. The contribution of past realized jumps to the predictable part of variance swap returns appears to be insignificant. This is perhaps due to the non-persistent nature of jumps. We conclude that the jump component, which captures the transient changes in volatility, is not a state variable driving the time-varying variance premium.

Table 8: Predictive power for future variance swap returns

The following table reports result of regression outlined in (8) which determines the factors that can predict the future variance swap returns. Panel A reports the results of regression with lagged values of Jump and Continuous variance. Panel B reports the extended regression which includes contemporaneous market return as an explanatory variable. We report the coefficient followed by the t-stat for each factor along with adjusted R-Sq for each regression. We use 30-day swap returns sampled at daily frequency; as we use overlapping data, we report t-statistics computed according to Newey and West (1987) with 30 lags.

Variance Swap	Panel A	Panel B
J_{t-1}	-321.29	-270.92
	(-1.05)	(-0.93)
C_{t-1}	90.01	75.88
	(6.15)	(5.00)
Excess Nifty Returns		-2.80
		(-3.01)
Intercept	-1.28	-1.29
	(-19.73)	(-19.47)
Adj. R-Sq	0.4605	0.5034

We extend the above regression by adding contemporaneous market return as an explanatory variable. We use return on Nifty as a proxy for the market return. Market return is significant and has the correct sign (Carr and Wu, 2009); however, the inclusion of this variable doesn't qualitatively impact our earlier findings. These results are also presented in Table 8.

We next undertake two robustness tests. First, in Equation (6), we use a confidence level of 99% to detect jumps. We verify the robustness of our main finding to two alternate confidence levels: 95% and 99.9% (used respectively in Jacquier and Okou, 2013 and Zhang et al, 2009). While we observe a difference in jump intensity, our earlier finding that only lagged continuous variance is significant survives. Second, in our earlier estimation, we did not include overnight returns. Adding these returns yields higher jump intensity; however, this doesn't have any impact on our central finding. Hence, we conclude that our findings on predictability are robust to various alternate specifications. For sake of brevity, we do not report these results.

5. Conclusion

We use both model-based and model-free techniques to examine if variance risk is priced in Indian equity index options market. Market-neutral straddles on Nifty are found to earn negative returns that are economically and statistically significant. Selling the embedded crash insurance by augmenting the straddles with a short position in OTM put doesn't significantly alter our results. These findings are robust to alternate volatility forecasts and are consistent across different subsamples and holding horizons. We infer that factors beyond market risk are priced in these options.

To validate our findings, we undertake model-free tests. Variance risk premium, measured as the difference between unconditional means of realized variance and model-free implied variance, is found to be negative and statistically significant. Equivalently, long positions in synthetic variance swaps on indices are found to yield significant negative returns. We conclude that variance risk premium is priced in Indian index options market.

We next separate realized variance into two components: jumps and continuous variance. We find that only the persistent component, as captured by past continuous variance, has predictive power for future returns of synthetic variance swaps. Jumps do not have any significant predictive power; this is perhaps due to their low persistence or high mean reversion. These results suggest that realized jumps are not a priced risk factor governing the dynamics of the variance risk premium.

Appendix A
Estimation of jumps from high frequency data
The contribution of jumps to the total realized variance for period $[t,T]$ is given by

$$CJ_{t,T} = \frac{RV_{t,T} - BV_{t,T}}{RV_{t,T}}$$

where $RV_{t,T}$ is the realized variance given by equation (2) and $BV_{t,T}$ is the realized bipower variation given by equation (4). When scaled by its asymptotic variance $V(CJ_{t,T})$, this ratio converges to a standard normal distribution. The asymptotic variance $V(CJ_{t,T})$ is given by

$$V(CJ_{t,T}) = \left(\left(\pi/2\right)^2 + \pi - 5\right).\Delta.\max\left(1, \frac{TP_{t,T}}{BV_{t,T}^2}\right)$$

where Δ refers to the time interval between consecutive data points and $TP_{t,T}$ is the Tri-Power Quarticity measure computed as

$$TP_{t,T} = \frac{1}{4\Delta\left[\Gamma(7/6).\Gamma(1/2)^{-1}\right]^3}.\sum_{i=3}^{1/\Delta}|r_i|^{4/3}.|r_{i-1}|^{4/3}.|r_{i-2}|^{4/3}$$

The scaled contribution of jumps, $z_{t,T}$ then converges to a standard normal distribution

$$z_{t,T} = \frac{CJ_{t,T}}{V(CJ_{t,T})} \xrightarrow{d} N(0,1)$$

This allows us to detect the presence of jumps; to filter "significant jumps", we set the confidence level as 99%. Using Monte Carlo tests, Huang and Tauchen (2005) demonstrate that this test of detecting jumps is quite accurate and has excellent size and power properties.
High frequency returns can be subject to microstructure noise such as bid-ask bounce. To correct for such serial correlation in adjacent returns, we follow Zhang et.al. (2009) and use staggered returns ($j = 1$) for computing $BV_{t,T}$ and $TP_{t,T}$

$$BV_{t,T} = \frac{\pi}{2}\sum_{i=2+j}^{1/\Delta}|r_i||r_{i-(1+j)}|$$

$$TP_{t,T} = \frac{1}{4\Delta\left[\Gamma(7/6).\Gamma(1/2)^{-1}\right]^3}.\sum_{i=1+2(1+j)}^{1/\Delta}|r_i|^{4/3}.|r_{i-(1+j)}|^{4/3}.|r_{i-2(1+j)}|^{4/3}$$

References

Andersen, T.G., and T. Bollerslev (1998), Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, International Economic Review, 39, 885-905.

Andersen, T. G., T. Bollerslev, and F.X. Diebold (2007), Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, Review of Economics and Statistics, 89, 701-20.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001) The Distribution of Realized Stock Return Volatility, Journal of Financial Economics, 61, 43-76.

Bakshi, G., and N. Kapadia (2003a) Delta-Hedged Gains and the Negative Market Volatility Risk Premium, Review of Financial Studies, 16(2), 527-66.

Bakshi, G., and N. Kapadia (2003b) Volatility Risk Premium Embedded in Individual Equity Options: Some New Insights, Journal of Derivatives, 11(1), 45-54.

Bakshi, G. and D. Madan (2006) A Theory of Volatility Spreads, Management
Science, 52, 1945-56.

Barndorff-Nielsen, O., and N. Shephard (2002) Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models, Journal of Royal Statistical Society, 64(B), 253-80.

Black, F., and M. Scholes (1973) The pricing of options and corporate liabilities, Journal of Political Economy, 81(3), 637-59.

Bollerslev, T., G. Tauchen, and H. Zhou (2009) Expected Stock Returns and Variance Risk Premia, Review of Financial Studies, 22(11), 4463-92.

Bollerslev, T., M. Gibson, and H. Zhou (2011) Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities, Journal of Econometrics, 160(1), 235-45

Bondarenko, O (2004) Market Price of Variance Risk and Performance of Hedge Funds, Working Paper, University of Illinois at Chicago.

Britten-Jones, M., and A. Neuberger (2000) Option Prices, Implied Price Processes, and Stochastic Volatility, Journal of Finance, 55(2), 839-66.

Broadie, M., M.Chernov, and M. Johannes (2009) Understanding Index Option Returns, Review of Financial Studies, 22(11), 4493-4529.

Carr, P., and L. Wu (2009) Variance Risk Premiums, Review of Financial Studies, 22(3), 1311-41.

Coval, J. D., and T. Shumway (2001) Expected Option Return, Journal of Finance, 56(3), 983-1009.

Driessen, J., P.Maenhout, and G. Vilkov (2009) The Price of Correlation Risk: Evidence from Equity Options, Journal of Finance, 64(3), 1377-1406.

Goltz, F., and W.N.Lai (2009) Empirical Properties of Straddle Returns, Journal of Derivatives, 17(1), 38-48.

Grover, R. and S.Thomas (2012), Liquidity Considerations in Estimating Implied Volatility, Journal of Futures Markets, 32(8), 714–741.

Jackwerth, J., and M. Rubinstein (1996) Recovering Probability Distributions from Option Prices, Journal of Finance, 51, 1611-31.

Jacquier, E., and C. Okou (2013) Disentangling Continuous Volatility From Jumps in Long-Run Risk-Return Relationships, Journal of Financial Econometrics, 10, 1093.

Jiang, G. J., and Y. S. Tian (2005) The Model-Free Implied Volatility and Its Information Content, Review of Financial Studies, 18(4), 1305-42.

Kim, I. J. and S. Kim (2005) Is It Important to Consider the Jump Component for Pricing and Hedging Short-tem Options? Journal of Futures Markets, 25, 989-1009.

Low, B.S. and S. Zhang (2005) The Volatility Risk Premium Embedded in Currency Options, The Journal of Financial and Quantitative Analysis, 40(4), 803-832.

Newey, W.K., West, K.D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica, 55, 703-708.

Pan, J (2002) The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, Journal of Financial Economics, 63, 3-50.

Todorov, V. (2010) Variance Risk-Premium Dynamics: The Role of Jumps, Review of Financial Studies, 23, 345-383.

Trolle, A. B., and E.S.Schwartz (2010) Variance Risk Premia in Energy Commodities, Journal of Derivatives, 17(3):15-32.

Vipul (2005) Futures and Options Expiration Day Effects: The Indian Evidence, Journal of Futures Markets, 25(11), 1045-65.

Wu (2011) Variance Dynamics: Joint Evidence from Options and High Frequency Data, Journal of Econometrics, 160(1), 280-287.

Yoon S.J., and S.J.Byun (2009) Is Volatility Risk Priced in the KOSPI 200 Index Options Market? Journal of Futures Markets, 29 (9), 797-825.

Oil price and Gulf Corporation Council Stock Indices: New Evidence From Time-Varying Copula Models

Jean-Pierre Fenech, Monash University, Department of Banking and Finance, Melbourne, Australia

Hamed Vosgha, Monash University, Department of Banking and Finance, Melbourne, Australia

ABSTRACT

Using both constant and time-varying copula approaches, we determine the conditional dependence of Saudi, Qatar, Oman, Dubai, Abu Dhabi and Kuwait stock indices on oil price between 2007 and 2016. We find strong empirical evidence of co-movement between the two variables. The time-varying approach reports negative associations for Saudi Arabia and Dubai, with the latter being more negative and more prevalenrtrtt when oil price drops. Finally, we predict the co-movement with an average accuracy rate of 69.43% and 80.69% for the Saudi and Dubai indices respectively. Such findings have implications for equity traders seeking portfolio diversification strategies.

JEL classification: C1, C6, E3, G1
Keywords: Crude oil prices, Copulas, Tail dependence, Co-movement

1. Introduction

The correlation of the Gulf Cooperation Council (GCC) stock markets with oil price is well documented, with empirical studies reporting evidence of feedback loops (Khandelwal, Miyajima and Santos, 2016; Salisu and Isah, 2017). Overall, the main findings depend on the data range selected and the methodology employed. The magnitude and direction of the interdependences also vary over time, with the main body of research remaining inconclusive and sometimes contradictory. Smyth and Narayan (2015) survey the Energy Economics literature and find that mixed findings reflect differences in econometric approaches and model specifications, amongst other things. For instance, Hammoudeh and Choi (2006) claim oil price movements do not directly affect the GCC stock markets. Where according to Noguera-Santaella (2016) not even geo-political events post 2000 should impact oil price. This is in contrast to Maghyereh and Al-Kandari (2007) and Zarour (2006), wherein Saudi and Omani stock markets do appear to be affected by oil price movements. Aloui, Nguyen and Njeh (2012) find that oil price risk is priced across the oil exporting countries. Such differences in findings are further exacerbated as different jurisdictions within the region advance their respective financial liberalization/privatization processes on different timelines.

Furthermore, from a price perspective, structural changes are an ongoing process within the oil sector. The sustained and worldwide increase in shale gas production is often claimed to push oil prices down (Chapman, 2014; Gevorkyan and Semmler, 2016). Moreover, the significant drop in oil price in June 2008 was followed by an upward trend, only to be reversed by April 2011. Although the West Texas Intermediate (WTI) spot price was $30 a barrel in February 2016 (monthly average data), pricing of future contracts suggests a recovery of only 5 to 10 percent over the coming two years. Additionally, it is not clear whether the current attempts to curtail oil supply result in an immediate impact on the stock indices, if any. Therefore, empirical studies investigating the consistent drop in oil price is required to investigate how it co-moves with the GCC stock markets. In the interim, Awartani and Maghyereh (2013) report evidence of oil consistently playing a dominant role in the information transmission mechanism affecting equities in the region (period of study 2004 to 2012). Aloui and Hkiri (2014) note frequent changes in the pattern of GCC stock co-movements, particularly after 2007. However, given the current oil price, it is even more important to consider how a steady drop in oil price affects such stock markets. In addition, albeit in a European context, Arouri *et al.* (2012) argue that stock returns are more sensitive to negative oil shocks than to positive. On a global perspective, Martin-Barragan, Ramos and Veiga (2015) find that co-movements tend to be stable in non-shock periods, with a breakdown in correlation during oil shocks and stock market crashes. Finally, Smyth and Narayan (2018) survey the oil price and stock returns literature across financial markets and highlight its complexity both in terms of breadth of coverage and econometric methods employed.

In this paper, we therefore attempt to address this gap within the literature by initially applying a constant copula followed by a time-varying approach, determining the conditional dependence between the oil price and stock markets in their respective GCC economies across Saudi, Qatar, Oman, Dubai, Abu Dhabi and Kuwait stock indices respectively from 2007 to 2016. We observe the correlations without holding any parameter constant and finally apply a time-varying approach.

Overall, copulas are a more accurate way of measuring associations. They are far more flexible to asymmetric dependence structures than correlation coefficients. Most copulas address the tail risk or extreme event scenario by modelling the joint distribution of random variables by separating the marginal distribution from its dependence structure to be modelled separately and independently (Ding, Kim and Park, 2016). However, selecting the right marginal is crucial for adequately modelling the tail risk, as we are particularly interested in extreme or tail dependence events above a certain threshold. This gives us a better estimate of tail risk associated with oil price and the respective stock index.

Furthermore, the consistent downside oil price trend has not yet been extensively investigated in the literature. Its implications for retail investors and portfolio managers seeking equity diversification opportunities require further investigation. Due to the valuable insight and consequences for portfolio management strategies within this oil dependent region, we also attempt to accurately forecast the co-movement of the two variables across the major indices in a downside market. We capture the fat tail distributions and asymmetrical behaviour linked to the sensitivity of their respective co-movements. In addition, our approach makes no assumptions about the marginal distributions, as we do not assume normality, allowing us to relate non-normal marginals to a dependence structure calculating probability distributions. Unlike Li (2000), we do not restrict the analysis to a Gaussian approach; we empirically demonstrate the best-fit Archimedean copula to model the tail dependencies, illustrating the impact of tail dependency on their respective non-linear parameters. This is a significant improvement on the Gaussian approach, where co-association is captured by a single scalar quantity keeping certain parameters constant.

Hence, we start by using constant Archimedean copulas, where we find the GCC stock markets are more correlated with oil prices on the left tail, meaning the pattern is more prevalent on the lower side. This is supported by Kendall Tau association measurements. Such outcomes are predominant in Abu Dhabi, Bahrain, and Dubai. However, it appears that the Kuwait, Oman, Qatar and Saudi markets do not follow the reverse correlation pattern exhibited in the other GCC markets. We also note that stock returns are more responsive to decreases in oil price drops than to increases. This asymmetric behaviour is seen across all the GCC markets, highlighting the importance of understanding how stock markets react to a consistent downturn in oil price.

However, constant copula models fail to exhibit the sensitivity of price changes and neglect the time factor in the correlation patterns. This was evident during the global financial crisis, wherein Li's (2000) model was employed to measure default correlation. Hence, we employ a similar approach to Reboredo (2011) and adopt time-varying copulas. We therefore investigate the asymmetric behaviour in view of the oil price fluctuations, particularly downturn. Contrary to Hammoudeh and Choi (2006) and Arouri and Fouquau (2009), we find evidence of bi-directional behaviour, which is more in line with Awartani and Maghyereh (2013) and Jouini and Harrathi (2014). Unlike Naifar and Al-Dohaiman (2013), our time-varying approach reports asymmetric dependence structures towards the lower side for both variables.

Therefore, we report new evidence on this bivariate correlation with respect to Saudi Arabia and Dubai, the former being more oil-dependent than the latter. Contrary to economic rationale and intuition, we report the two variables move in opposite directions. Historically, studies suggest higher oil prices should be good for stock prices and vice-versa. However, this study captures a unique phase within the oil market, and evidence of oil/stock price moving in opposite directions appears to represent a paradigm shift. We observe this new phenomenon by employing a correlation measure that is dynamic and time varying.

Furthermore, we forecast the co-movement of the two variables with back-testing techniques, determining our overall accuracy level via GARCH-type processes. We also empirically report oil/stock prices moving in similar and opposite directions. Such a phenomenon presents portfolio managers with the opportunity for a buy low/sell high equity strategy. This pattern is more evident in the Dubai market than its Saudi counterpart. Furthermore, using the two larger indices in the region, i.e., Saudi Arabia (highly dependent on oil) and the Dubai Financial Market (the least oil-dependent and more diversified equity market), we are also able to forecast the co-movement in the next period, with accuracy levels of 69.43% and 80.69% respectively.

This paper provides a number of contributions to the literature. The time-varying copula approach, compared to its constant counterpart study by Naifar and Al-Dohaiman (2013), clearly suggests that the co-movement varies across time. In line with the Awartani and Maghyereh (2013) and Ajmi, EI-montasser, Hammoudeh and Nguyen (2014), we provide further detail of the non-linear co-movement through this unique period of low oil price compared to the past decade. Clearly, although the GCC region is considered to be homogenous, there are instances where idiosyncrasies emerge, presenting opportunities for fund managers and their interested stakeholders to investigate further. Hence, in contrast to Naifar and Al-Dohaiman (2013), we separate the UAE market into Abu Dhabi and Dubai to provide further granularity. By calculating time-varying country-specific dynamic correlations on a monthly basis, we demonstrate the ongoing change to the bivariate relationship. In line with Jouini (2013) where evidence of non-linear long run relationships was reported, our analysis does not constrict the data and does not make any assumptions or set any pre-conditions. Given the positive and negative correlations recorded over time, equity diversification is clearly a key investment strategy within the GCC equity markets.

From an economic perspective, correlations lead to changes in fiscal and external positions. Higher equity market returns are expected from more oil-dependent jurisdictions (e.g., Saudi Arabia compared to Dubai) as investors anticipate a positive impact on the corporate sector as oil price increases. With more government spending, credit growth is increased, causing higher asset prices and a positive overall wealth effect. Conversely, an oil price downturn blended with other factors may have a reverse effect. The bursting of the real estate bubble in the UAE's 2009 financial crisis together with defaults in 2008 by two of the largest investment companies in Kuwait is evidence of this phenomenon. Therefore, diversifying equity exposures to non-oil sector investments may have its own limitations as systemic risk affects the whole economy.

Finally, despite the low forecasting accuracy levels for stock returns, the techniques employed in this paper provide forecasters with some opportunities. As this critical but unique market remains under-researched, more intensive analysis is expected as more data becomes available within the GCC. Our study benefits from taking a long horizon use of disaggregated data (different indices) and employs non-linear modelling techniques to model such co-movements. We also engage with out-of-sample testing and by assessing a month-to-month rolling window, we are able to predict such parameters with reasonable accuracy. Based on this correlation predictability, potential equity portfolio strategies may be designed.

The remainder of the paper is structured as follows. Section 2 reviews the literature with respect to the co-movement of oil price and stock indices, particularly in the GCC region. We also elaborate on the constant/time-varying conditional copula empirical models in Sections 3 and 4 respectively. Section 5 describes the data, methodology employed and the results. Section 6 concludes the study.

2. Related literature

Zarour (2006) investigates the relationship between oil prices and five stock markets in the Gulf countries over the period 2001–2005, using a VAR method. As this period represents high oil prices, the markets reacted positively to such oil price shocks. Similarly, Mohanty, Nandha, Turkistani and Alaitani (2011) find a similar result with the exception of Kuwait. Conversely, Hammoudeh and Choi (2006) used a series of co-integration tests with vector error correction approaches, looking into the same markets as Zarour (2006), and reported no direct effects on any GCC stock market. The study period commenced from 1994. Furthermore, Malik and Hammoudeh (2007) examine the volatility and shock transmission mechanisms among US stocks, Gulf equities and oil price. They apply a multivariate GARCH from 1994 till 2001 and highlight the importance the Saudi market plays within the oil market. They find that the Gulf equity markets are affected by oil market volatility. However, contrary to all the other Gulf countries, only the Saudi stock index provides a significant spillover effect to the oil market. Hammoudeh *et al*. (2009) capture a different time period, i.e., 2001–2007, and employ a VAR-GARCH method to report moderate volatility spillovers between the sectors amongst the individual countries, with the exception of Qatar. In a more recent study, Arouri *et al*. (2011), looking into the period 2005–2010 and employing a VAR-GARCH approach, find that the recent 2007–08 global financial crisis led to an increase in volatility spillovers between oil and Gulf equity markets. Awartani and Maghyereh (2013) extend the data set to 2012, using a multivariate GARCH and find return and volatility transmissions are bi-directional and asymmetric.

Maghyereh and Al-Kandari (2007) address the issue of non-linearity, where nonparametric rank testing for non-linear co-integration is executed on the long-run linkages between oil and GCC markets. The results support the phenomenon where oil price affects stock markets in a nonlinear way. Their study period is from 1996 to 2003. Arouri and Rault (2012) study the long-term links in the period 2005–2010 using a panel co-integration and seemingly unrelated regression. They showed that the casual relationship is bi-directional for Saudi Arabia. For other GCC countries, strong statistical evidence is reported showing oil price disturbances cause stock price changes. At a short-run level, Arouri and Fouquan (2009) show that stock markets in Qatar, Oman and the UAE exhibit asymmetric behaviour with changes to oil price. They also employ a non-parametric method; their study period is 1981–2007. More recent studies capture similar patterns to the results reported in prior work. Fyyad and Daly (2011) use a VAR approach over 2005–2010 and report Qatar and UAE show more responsiveness to oil shocks than Kuwait, Bahrain and Oman. Mensi, Hammoudeh, Yoon and Balcilar (2017) study the non-linear relationship between stock markets in the GCC and oil price together with other major macroeconomic factors and find oil price increased the performance of the GCC stock markets.

Naifar and Al-Dohaiman (2013) cover 2004 till 2011 and use an Archimedean copula to study the relationship between oil and stock returns in the GCC. They find that the dependence structure is asymmetric and leans towards the upper side during the recent financial crisis. The use of copulas, particularly an Archimedean approach, allows non-Gaussian approaches to model the relationship between oil price and the stock market. This methodology is an improvement as compared to VAR, co-integration tests, VEC and multivariate-GARCH approaches, as there are no data assumptions and pre-conditions on the distributions.

Further to the body of work executed, copulas are alternative tools for dealing with multivariate extremes, and they have recently become popular in finance due to the inability of previous models to handle extreme values (Cherubini, Luciano and Vecchiato, 2004). Copulas became a mainstream tool in finance measuring complex risk, pricing collateral debt obligations consisting of mortgages with brokers quoting prices for bond tranches based on their correlations. Li's formula, known as a Gaussian copula function (Li, 2000) made no allowance for unpredictability, assuming correlation is held constant. Hence the underlying assumption of a Gaussian copula where a single scalar quantity describes the relationship between two assets is questionable. Risk managers should have realised that small adjustments to the underlying assumptions could result in very large changes in correlation figures.

Overall, copula functions permit flexible modelling of the dependence structure between random variables by allowing the construction of multivariate densities consistent with the univariate marginal densities. The advantage of using copulas relies on the separation of the marginal distributions and their dependence structures. Such separation enables the construction of multivariate distribution functions, avoiding the common assumption of normality. Copula functions start by determining their marginal distributions. Sklar developed the theorem stating that any joint distribution can be written in terms of a copula and marginal distribution functions. Sklar (1959) showed that for *n*-dimensional continuous random variables $(X_1,\ldots,X_p)$ with marginal cumulative functions (CDF) $u_i = F_i(x_i)$, i=1,…,p, there exists one unique n-copula C such that $H_{X_1,\ldots,X_p}(x_1,x_2,\ldots x_p) = C_{U_1,\ldots,U_n}(F_1(x_1),\ldots.F_n(x_p))$ where U_i is the *i*th marginal and $H_{X_1,\ldots,X_p}$ is the joint- CDF of $(X_1,\ldots,X_p)$. We observe that copulas represent the multivariate dependence structures. Alternatively, the density representation is given by:

$$h_{X_1,\ldots,X_p}(x_1,\ldots,x_p) = \prod_{i=1}^{p} f_{X_i} \times c(F_1(x_1),\ldots.F_p(x_p))$$

(1)

where $c(F_1(x_1),\ldots.F_p(x_p)) = (\partial^n / \partial u_1 \ldots \partial u_p)C(F_1(x_1),\ldots.F_p(x_p))$ is the copula density, and f_{X_i} are marginal density functions. The density of *H* has been expressed as the product of the copula density and the univariate marginal densities. It is in this sense that copulas can be considered a powerful tool for identifying and modelling dependence structures. An important property of copulas is that they are invariant under strictly increasing transformations of the variables. This invariance property guarantees that variables $(X_1,\ldots,X_p)$ and their logarithms have the same copula. There are a number of copulas, and in our survey we concentrate on Archimedean copulas where $A:[0,1] \to [1/2,1]$ is a convex function such that max (t, 1-t) < $A(t)$ <1 for all $t \in [0,1]$. The function A(*t*) is called the dependence function. Many extreme value copulas are introduced in the literature. In our study, extreme dependence between each pair of markets is modelled by the more commonly applied copulas. The most appropriate one is finally chosen based on its distance being minimal.

3. Constant copula concepts and theory

3.1 Models for marginal distributions

There are different types of copulas that risk practitioners may employ in different scenarios. Depending upon the situation, we can opt for the copula matching the scenario. For example, to model the dependency of a bivariate data series on one side of the tails, a Gumbel is used for modelling tail dependency on the right side, Clayton for left and Frank for symmetrical dependency. In general, Archimedean copulas are mathematically expressed as follows:

$$C(U_1, U_2, \ldots, U_n) = \Phi^{-1}(\Phi(U_1) + \ldots + \Phi(U_n)) \quad (2)$$

Where $U_1, U_2, \ldots, U_n \in I$ and Φ is a function that must satisfy:

- $\Phi(1) = 0$
- For all $t \in I, \Phi(t) < 0$ (decreasing function)
- For all $t \in I, \Phi(t)' \geq 0$ (Convex)

Theorem 1 (Kimberling, 1974): Let Φ be a generator. The function $C: [0,1]^n \rightarrow [0,1]$ Defined by: $C(U_1, U_2, \ldots, U_n) = \Phi^{-1}(\Phi(U_1) + \ldots + \Phi(U_n))$ is a copula if Φ^{-1} is completely monotonic on $[0, \infty]$.

Definition 1 A function f with domain $(0, \infty)$ is completely monotonic if it possesses derivatives $f^n(x)$ for all n= 0,1,2,3,…. and if $(-1)f^n(x) \geq 0$.

Theorem 2 (Feller, 1971)**:** A function φ on $[0, \infty)$ is the Laplace transform of a Cumulative Distribution Function (c.d.f.'s) F if and only if φ is completely monotonic and $\varphi(0) = 1$. This theorem is also known as Bernsteins Theorem.

Corollary 1 An important source of generators for Archimedean *N*-copulas consists of the inverse of the Laplace transforms of c.d.f's.

Definition 2 (Laplace transform**):** Let *f* be a function of time, with value *f*(*t*) at time *t*, the Laplace transform of *f* is denoted ˜ *f* and it gives an average value of *f* taken over all positive values of *t* such that the value *f* ˜(*s*) represents an average of *f* taken over all possible time intervals of length *s*.

$$L\,[f(t)] = f\tilde{}(s) = \int_0^\infty e^{-st} f(t)dt, for\ s > 0 \qquad (3)$$

As mentioned earlier, if Φ is a convex decreasing function with domain (0,1) and range [0, ∞] such that $\Phi(1) = 0$, using inverse function of Φ, then function *C* is represented as:

$$C(U_1, U_2, \ldots, U_n) = \Phi^{-1}(\Phi(U_1) + \ldots + \Phi(U_n)) \qquad (4)$$

However, different choices of generator will result in different Archimedean copulas such as Clayton and Frank copula provide a tight correlation at the low end of each random variable. The generator is given by $\Phi(u) = u^\alpha$-1 so $\Phi^{-1}(t) = (t+1)^{-\frac{1}{\alpha}}$; the inverse generator is completely monotonic if $\alpha > 0$. Then, the Clayton *n*-copula is asymmetrical

$$C(U_1, U_2, \ldots, U_n) = [\sum_{i=1}^{n} U_i^{-\alpha} - n + 1]^{\frac{-1}{\alpha}} \quad \text{with } \alpha > 0 \qquad (5)$$

This is an asymmetric Archimedean copula demonstrating greater dependency in the negative tail than the positive. This copula has a heavy concentration of probability near (0,0), i.e., the intersection of axes X and Y, where it correlates well with the small returns.

A Frank copula is a radially symmetric Archimedean in dimension when *n* = 2, producing the correlation across the range of variables. Its generator is given by $\Phi(u) = Ln\,(\frac{\exp(-\alpha u)-1}{\exp(-\alpha)-1})$, so $\Phi^{-1}(t) = -\frac{1}{\alpha}$ Ln $(1+e^t(e^{-\alpha}$-1)); if $\alpha > 0$ the inverse of generator is completely monotonic and then the Frank copula is given by:

$$C(U_1, U_2, \ldots, U_n) = -\frac{1}{\alpha} Ln\,\{1 + \frac{\prod_{i=1}^{n}(e^{-\alpha u_i} - 1)}{(e^{-\alpha}-1)^{n-1}}\} \qquad (6)$$

The Frank copula is even lighter in the right tail compared with the Gaussian copula and does not show tail dependency in its limits. In this regard, a Frank copula is similar to a Gaussian. It does not generate a strong relationship between large losses. Therefore, it is less likely to be a suitable candidate for the modelling of tail estimation.

The Gumbel copula is another member of the family of non-linear copulas. It is more likely to capture a bivariate data series in the tails than the Frank and Gaussian copulas. As a Gumbel copula assigns more probability to tails, it is the better model for recording the dependency of joint extreme events resulting in unexpected returns or losses. In other words, Gumbel copulas are more asymmetric, i.e., having more weight in the right tail. Its generator is given by:

$\Phi(\mathrm{u}) = (-\mathrm{Ln}\,(\mathrm{u}))^{\alpha}$, so $\Phi^{-1}(\mathrm{t}) = \exp\ (-t^{\frac{1}{\alpha}})$; it is completely monotonic if $\alpha > 0$. The Gumbel copula is represented as:

$$\mathrm{C}(\mathrm{U}_1, \mathrm{U}_2, \ldots, \mathrm{U}_\mathrm{n}) = \exp\{-[\sum_{i=1}^{n}(-LnU_i)^{\alpha}]^{\frac{1}{\alpha}}\}; \text{ with } \alpha > 0. \qquad (7)$$

There is more correlation at the two extremes of the correlated distributions in the Gumbel copula function, as it keeps a strong relationship even for the large returns. The consequences of the least probable outcomes exist in the tails. Archimedean copulas are clearly preferable to the Gaussian copula functions in modelling non- linear correlations. These copulas may be adapted to different circumstances and since they do not follow normal patterns, they provide a more accurate way for constructing the modelling of joint distributions, providing great variety of dependence structure.

The Clayton copula has lower tail dependence, but lacks upper tail dependence. The lower tail dependence of the Clayton copula increases as the degree of dependence (θ) increases. In fact, the Clayton copula is an asymmetric Archimedean copula, exhibiting greater dependence in the negative tail than in the positive one. Conversely, the Frank copula is a symmetric Archimedean copula exhibiting neither upper nor lower tail dependencies.

Archimedean copulas better handle the correlation structure amongst the marginal distribution related to τ. The distribution of this statistic has better statistical properties as it forms Archimedean copulas with simple parametric copulas. The sample version of the measure of association, i.e., τ, is defined in terms of concordance, defined as a pair of random variables concordant if large values of one tend to be associated with large values of the other and small values of one with small values of the other. Let (x_i, y_i) and (x_j, y_j) denote two observations from a vector (X,Y) of continuous random variables. (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$, and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. However, (x_i, y_i) and (x_j, y_j) are discordant if $x_i < x_j$ and $y_i > y_j$ or if $x_i > x_j$ and $y_i < y_j$. Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ denote a random sample of *n* observations from a vector (X,Y) of continuous random variables. There are $\binom{n}{2}$ distinct pairs of (x_i, y_i) and (x_j, y_j) of observations in the sample, and each pair is either concordant or discordant. Let *c* denote the number of concordant pairs and *d* the number of discordant pairs. Then τ for the sample is defined as:

$$\tau = \frac{\mathbf{c-d}}{\mathbf{c+d}} = \frac{\mathbf{c-d}}{\binom{n}{2}} \qquad (8)$$

Equivalently, τ is the probability of concordance minus the probability of discordance for a pair of observations (x_i, y_i) and (x_j, y_j) chosen randomly from the sample. τ is invariant under strictly increasing transformations of the underlying random variables. In fact, τ is only dependent on the copula, whereas the linear correlation in the Gaussian copula is a variant under strictly increasing transformations of the underlying random variables.

Archimedean copulas are in a better position to model the dependency, compared to Gaussian approaches. Archimedean copulas measure dependency solely on the copula itself, while the Gaussian's correlation coefficient depends not only on the copula function but also on its marginal distributions. Thus, this measure is affected by changes of scale in its marginal variables as the correlation acts as an invariant measure of dependency with the Gaussian using linear coefficient correlation.

Table 1: The relationship between Kendall tau and Archimedean copula function parameter (α).

Family	Range of α	Generation Function Φ(u)	Kendall tau (τ)
Gumbel (1960)	[1,∞)	$-(Ln(u))^{\alpha}$	$1-\alpha^{-1}$
Clayton (1978)	[0,∞)	$u^{-\alpha}$ -1	$\alpha/(\alpha+2)$
Frank (1979)	(-∞,+∞)	-Ln $\frac{e^{-\alpha u}-1}{e^{-\alpha}-1}$	$1+4[D_1(\alpha)-1]/\alpha$

The copula family used in our work includes commonly used copulas in the Archimedean copula family, such as the Clayton, Frank and Gumbel. The class of Archimedean copulas was named by Ling (1965), recognized by Schweizer and Sklar (1961) in the study of *t*-norms. Their non-elliptical characteristics allow us to model a large variety of different dependence structures. We therefore consider in particular the one-parameter Archimedean copula.

4. Time varying (conditional) copulas

We employ conditional copula models using GARCH theory, where some of the parameters are potentially time varying, conditional on the set of past information. Patton (2011) extends Sklar's theorem to the conditional case and studies the attributes of this new model class. Applications of copula/GARCH models in finance can be found throughout the literature (Panchenko and Diks, 2006; Serban *et al*., 2007; and Wang *et al*., 2012).

GARCH models have become important in the analysis of time series data, particularly in financial applications when the goal is to analyse and forecast volatility. Engle (1982) noted that although many financial time series such as stock returns and rates are unpredictable, there is an apparent clustering in the variability or volatility. This is often referred to as conditional heteroscedasticity, since it is assumed that overall the series is stationary, but the conditional expected value of the variance may be time-dependent. Our marginal model is built on the classical GARCH model and the GJR model, in which the standard innovation is to follow the normal distribution and Student-t distribution respectively.

4.1. Model specification – An overview

The estimation procedure used in our work employs a GARCH model to filter the original data sets; we then check to make sure that the marginal models are correctly specified before the standardised residuals are transformed into i.i.d. Uniform (0,1). Finally the probability integral transforms of the standardised residuals are plugged from the marginal models to the chosen copula. This estimation method is called the inference functions of margins (IFM) method, proposed by Joe and Xu (1996). For the purpose of goodness of fit, we have examined the performance of the chosen copula by the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). There are two potential models to be employed. The marginal model, using the marginal distribution of the time series datasets using GARCH type procedure filters the raw data with a AR (k)-GARCH (p,q) type of models. This model has been used by Patton (2011), amongst others. The marginal model is specified as follows:

$$r_{i,t} = c_i + \Sigma_k AR_{i,k} \text{ X } r_{i,t-k} + \varepsilon_{i,t} \quad (9)$$
$$\sigma_{i,t}^2 = Arch0_i + \Sigma_p Garch(p)_i \text{ X } \sigma_{t-p}^2 + + \Sigma_q Arch(q)_i \text{ X } \varepsilon_{i,t-p}^2$$

Where $r_{i,t}$ is—as an example—the return for a stock for company I as time t, $\sigma_{i,t}^2$ is the varaiance of the $\varepsilon_{i,t}$term in the mean equation. The estimation results of the marginal models in our work are depicted in Section 5. To examine the time vaying dependnce structure in the return series, we apply a dynamic copula model as described by Patton (2011), defined as follows for TVC- Clayton copula:

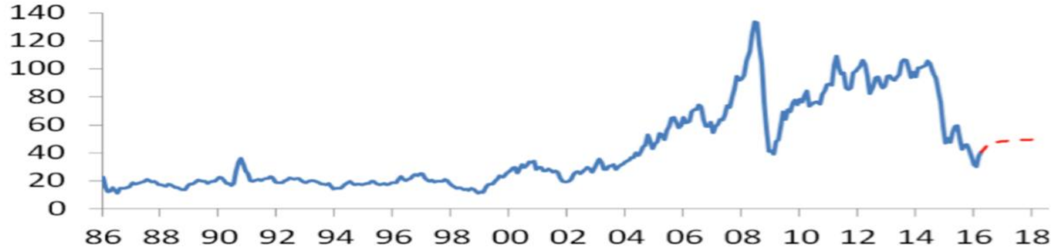
$$\lambda_t = \Lambda\left(\omega + \beta\lambda_{t-1} + \alpha.\frac{1}{10}\Sigma_{i=1}^{10} \mid \upsilon_{i-t} - \nu_{i-t} \mid\right) \quad (10)$$

where Λ denotes the logistics transformation to keep the parameters of the TVC Clayton copula in (0,1). It is defined as $\Lambda(x) = (1 + e^{-x})^{-1}$. The dynamic copula model contains an autoregressive term designed to capture persistence in dependence and a forcing variable which is the mean absolute difference between uniform margins in the form of bivariate (υ, ν).

5. Data and empirical results

The daily stock prices with respect to Saudi Arabia, Bahrain, Qatar, Oman, Dubai, Abu Dhabi and Kuwait stock indices from 2007 to the 1st quarter of 2016 are downloaded from Bloomberg. We initially use daily OPEC oil spot prices, and later replace with both West Texas Intermediate and Brent spot prices as robustness tests and obtain similar results. We commence our study from 2007 to capture the start of the extreme events emanating from the global financial crisis. Figure 1 shows the consistency in oil price over the 80s and 90s, followed by more volatile and extreme values. Maximum prices reached \$130 a barrel during 2008, only to drop down to \$40 the following year. Clearly, this irregular price behaviour is a sign of uncertainty within the oil market, with unknown impact on the respective stock price.

Figure 1: West Texas Intermediate Oil Price



West Texas Intermediate Oil Prices (U\$ a barrel, monthly average). Source: Haver and IMF staff calculations. Note: Actual through end-April 2016 with the broken line based on 3-, 6-, 12- and 24-month future prices.

5.1. Constant copula results

As the main research objective of this study is to measure the correlation between oil price and its relative stock index, the next stage is to calculate such association. Most empirical models consisting of a number of macroeconomic variables relating stock prices to oil prices are potentially wrongly specified. It is almost impossible to capture every variable of interest whilst trying to examine the causal relationship between oil and equity index. This phenomenon is well established in the literature, and such models attempting to capture such a relationship have certain limitations. (Conrad, Loch and Rittler, 2014). Therefore, by initially using constant copula methodology, we measure the dependency structures and report the association patterns in Figure 2. The seven graphs (Saudi Arabia, Bahrain, Qatar, Oman, Dubai, Abu Dhabi and Kuwait) capture the appropriate copula with the best goodness of fit for each index, which is a function of its dependency structure.

Kendall tau is initially measured. It is a degree of concordance, calculating the strength of association between -1 and 1. Subsequently, the best copula is fitted based on its optimal goodness-of-fit. Each graph may be seen as a quadrant with concentrations of observations observed within either upper or lower tails. Lower tail dependencies to the left suggest a low oil price with a bearish stock market and therefore a Clayton approach may be more appropriate in this context. Conversely, with concentration in the middle, compared to the Gaussian approach, a Frank is more appropriate, as it takes into consideration the Kendal tau measure. Upper tail dependencies to the right suggest that as oil price increases, stock prices also increase. Upper tail dependencies to the left suggest high oil prices with low stock index prices.

Abu Dhabi and Oman have upper tail dependencies, best described by a Gumbel approach. Dubai, Saudi Arabia and Qatar have a lower tail dependency, hence a Clayton approach is applied. Bahrain and Kuwait follow a lighter tail dependency, with a Frank approach best describing the co-movement with oil prices.

Further to the graphs reported in Figure 2, Table 2 summarises the results of the dependency modelling between the two selected variables. The Kendall tau correlation coefficient, which is a nonparametric measure of the strength and direction of association between two variables measured at least on an ordinal scale, is recorded. It is considered a nonparametric alternative to the Pearson's product-moment correlation and the nonparametric Spearman rank-order correlation coefficient.

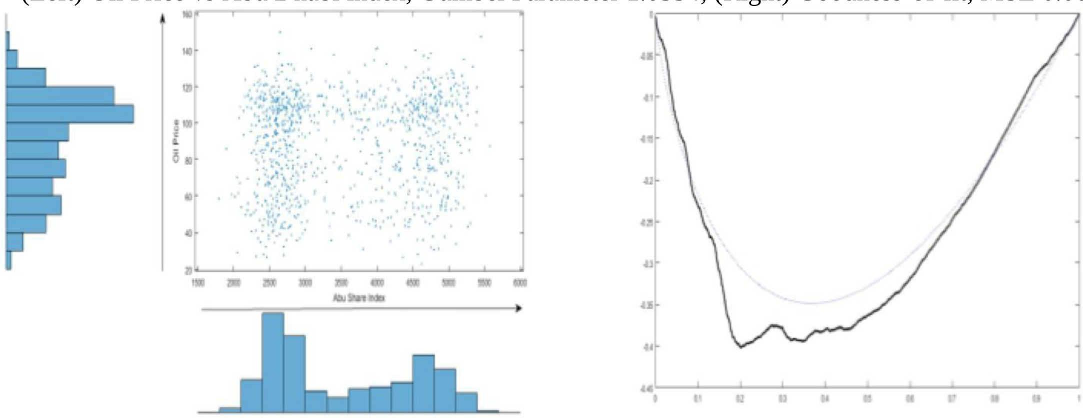
Table 2: **Dependence modelling GCC indices/oil prices using constant Archimedean copulas.**

	Abu Dhabi	Bahrain	Dubai	Kuwait	Oman	Qatar	Saudi Arabia
Kendall Tau	-0.0331	-0.1985	-0.0178	0.1688	0.1688	0.1586	0.2081
Constant Copula Parameter	1.0554	-2.2065	$1.45e^{-06}$	1.7184	1.247	0.4085	0.6234
AIC	-35.6085	-462.8878	-0.4598	-43.5591	-154.6427	-755.36	-655.8759
MSE	0.0012	$5.65e^{-04}$	0.0110	0.0440	0.0024	0.0019	$9.49e^{-04}$
Copula selected	Gumbel	Frank	Clayton	Frank	Gumbel	Clayton	Clayton

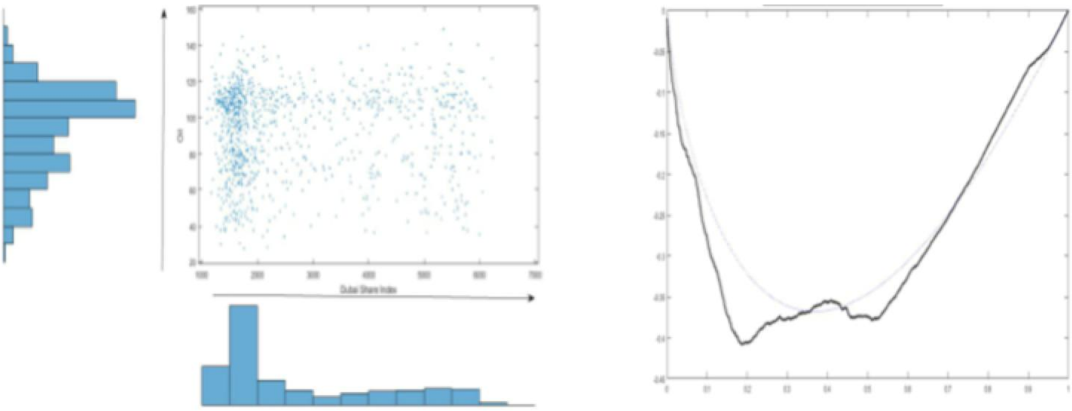
The Akaike Information Criterion (AIC) has been employed to determine the best copula fit along with the Mean Squared Error (MSE). The optimal goodness-of-fit scenario is represented by its lowest measure where the function used is different for each type of copula taking the underlying copula parameter as the main input.

Figure 2: Constant copula dependency bivariate structure representing oil price vs. stock index and Goodness-of-fit measures.

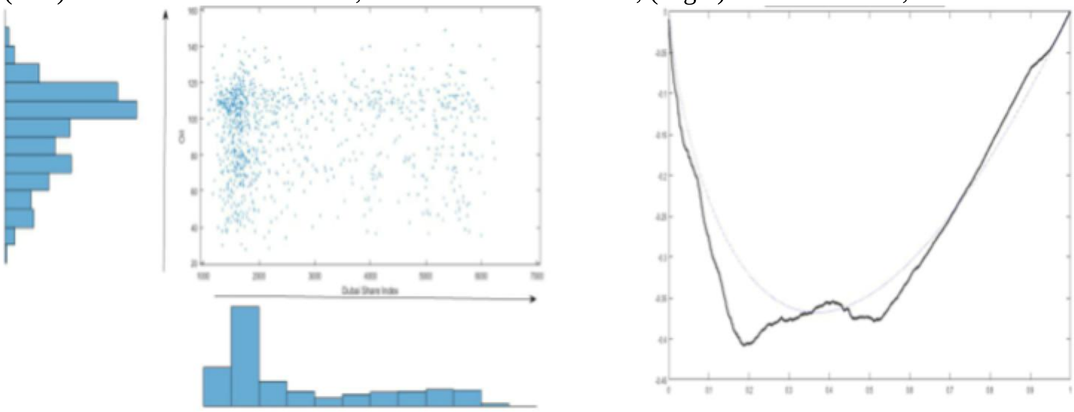
(Left) Oil Price vs Abu Dhabi index, Gumbel Parameter 1.0554; (Right) Goodness-of-fit, MSE 0.0012



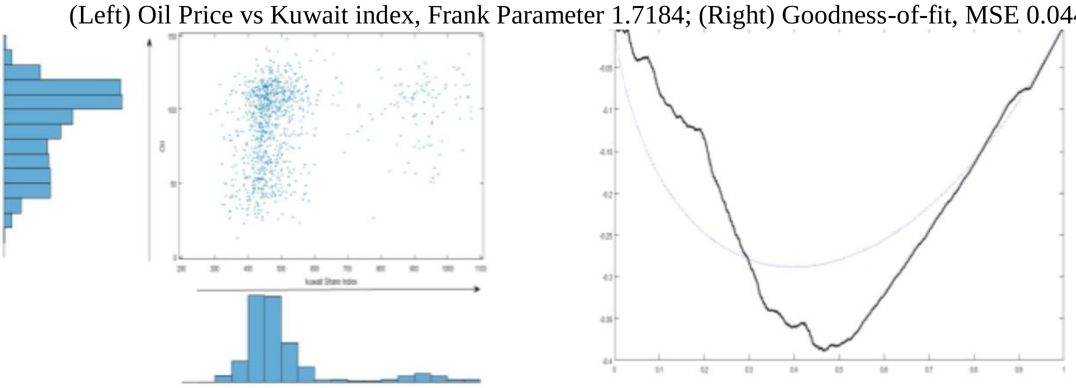
(Left) Oil Price vs Dubai index, Clayton Parameter $1.45e^{-06}$; (Right) Goodness-of-fit, MSE 0.0110



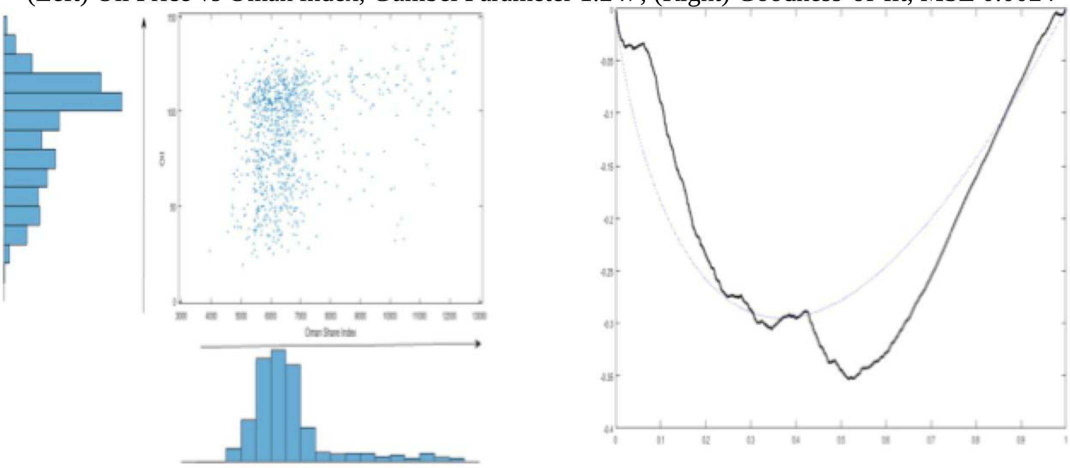
(Left) Oil Price vs Bahrain index, Frank Parameter -2.2065; (Right) Goodness-of-fit, MSE $5.65e^{-04}$



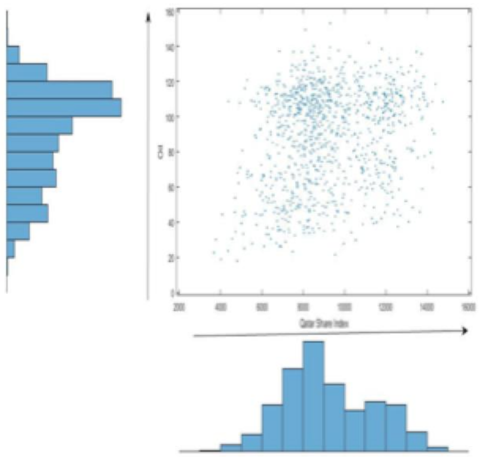
(Left) Oil Price vs Kuwait index, Frank Parameter 1.7184; (Right) Goodness-of-fit, MSE 0.0440

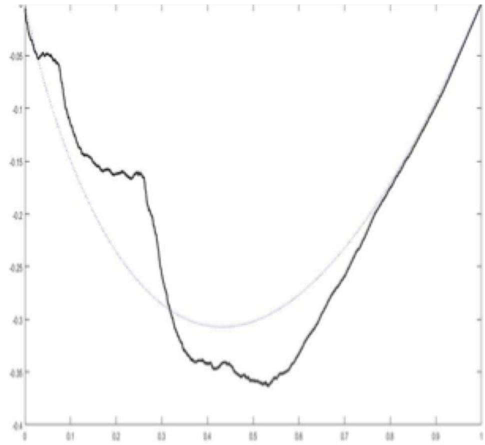


(Left) Oil Price vs Oman index, Gumbel Parameter 1.247; (Right) Goodness-of-fit, MSE 0.0024

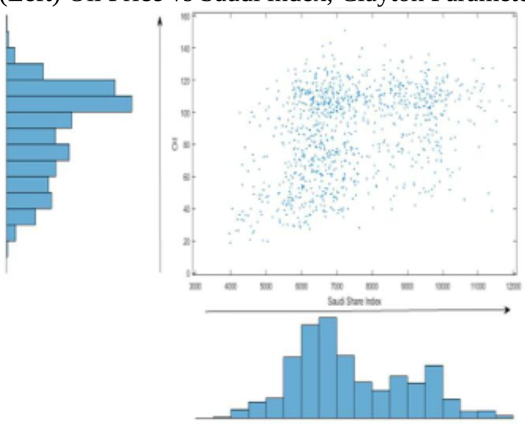


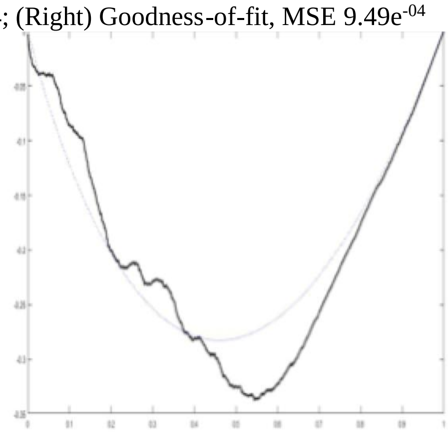
(Left) Oil Price vs Qatar index, Clayton Parameter 0.4085; (Right) Goodness-of-fit, MSE 0.0019





(Left) Oil Price vs Saudi index, Clayton Parameter 0.6234; (Right) Goodness-of-fit, MSE 9.49e^{-04}





Constant copula dependency bivariate structure representing oil price vs. stock index and Goodness-of-fit measures. Order of graphs: Abu Dhabi, Dubai, Bahrain, Kuwait, Oman, Qatar and Saudi. One thousand simulated points from the copula fitted on an index and oil price. The bivariate copula is reported within the graph, with marginal being represented along the x and y-axes. The copula function is constructed based on the Kendall tau of two datasets. Based on the theory described in Section 3, the calculation of the copula parameters for GCC stock Indices vs. Oil price undertaken using nonlinear dependence measure: Kendall's tau. The estimated Kendall's tau and the associated copula parameters are illustrated in Table 2. In summary, Gumbel copula best fits Abu Dhabi and Oman stock indices vs. Oil price, characterized by strong upper tail dependency. Whereas, Clayton copula matches Dubai, Saudi, and Qatar stock indices vs. oil price showing lower tail dependency structure. The estimated Frank copula for Bahrain and Kuwait equity markets vs. oil price characterizes the weak tail dependency.

Abu Dhabi, Bahrain and Dubai report a negative correlation, whilst Kuwait, Oman, Qatar and Saudi Arabia are positive. The inconsistency in correlation sign and size demonstrates how differently the stock indices react to movements in oil price over time. In view of a downward oil price trajectory, this is an indication that such equity markets may not be as highly correlated as previously thought. Additionally, the association of such two variables (oil vs. stock) is key for pairs-trading strategies, where long positions are matched with short, based on their correlation levels, creating a market-neutral trading strategy.

The statistical arbitrage approach used by hedge funds (Connor and Lasarte, 2004) is primarily based on their correlation behaviour over time. Therefore, the stock/oil commodity pair constructed portfolio requires active management to achieve maximum efficiency. In doing so, copulas may provide traders and portfolio managers with insight.

The copula functions have their parameters summarised with the characteristics of the dependence structures of two variables. The information contained in the dependence structures expose the two data points distancing from each other. Furthermore, it also demonstrates when one of the variables move higher or lower than expected, given the historical relationship between the two data points. This information is critical, as sometimes different financial markets across the region evolve independently from others. Once again, the variety in the parameters recorded in Table 2 corroborates the same trend with Kendall tau.

Akaike Information Criterion (AIC) and Mean Squared Error (MSE) are measures of the relative quality of statistical models for a given set of data. The optimal goodness-of-fit scenario is represented by its lowest measure where the function used is different for each type of copula taking the underlying copula parameter as the main input. Hence the copula selected is based on such measures. However, notwithstanding the benefits of such an approach over other more conventional and linear approaches, constant copulas could be misleading. They incorporate constant parameters to their historical data and therefore restrict the analysis conditional to the pre-conceived assumptions embedded within the parameters. Studies suggest correlation structures in financial markets significantly change over time, impacting asset pricing and risk management areas. This phenomenon has been further exacerbated during financial crisis. Consequently, placing distribution assumptions on the oil/stock data correlation may result in inaccurate outcomes.

5.2 Time-varying copula empirical results

Therefore, selecting the two largest and most diverse indices by market capitalization within the region, we apply a copula-GARCH model on the Saudi and Dubai indices respectively with respect to oil prices. This approach is different to the constant method and is a significant improvement over Li's (2000) Gaussian copula. It imposes no pre-conceived distribution assumptions, allowing the data to determine the outcome of the co-movement amongst the two variables. We use a two-step procedure for estimating the model parameters. Firstly, we use AR(q) - GARCH(1,1) for the data set with the Gaussian distribution for the residuals. The distribution of the residuals defines the log likelihood function of the margins and the method that transforms the standardized, i.i.d. residuals from the filtration, to uniform (IFM). This transformation is achieved by the probability integral transform: Let $\varepsilon_{i,t}$ t = 1, … ,T be a time series of i.i.d. variables, where we assume that: ε_i~ F,i = 1, … , T. Then the series:

$$\upsilon_t = F(\varepsilon_t) \qquad (11)$$

is the probability integral transform of ε_t and it holds that ui ~U [0; 1] ; i = 1,…, T using the IFM method, where transformation to U [0; 1] is being made parametrically, by using the distribution assumed for the residuals. With the GARCH parameters defined, we express the time-varying copula parameters.

The second step demonstrates how copula parameters evolve through time using Patton's (2011) approach on the Saudi and Dubai dataset and determine the dependency parameter Kendall's tau for the Clayton copula in line with the following equation:

$$\tau_t = \Lambda\,(\omega + \beta\tau_{t-1} + \alpha.\frac{1}{10}\,\Sigma_{i=1}^{10} \mid \upsilon_{i-t} - \nu_{i-t} \mid) \quad (12)$$

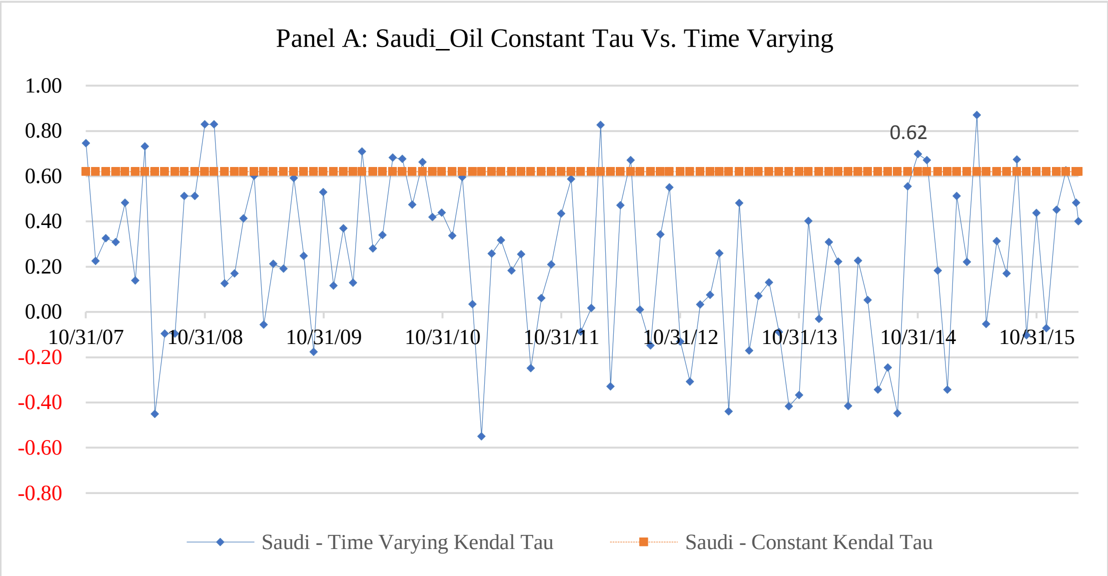
In Figure 3, we compare and contrast the constant with the time-varying co-movement for Saudi Arabia and Dubai respectively. It is evident the time-varying approach shows more granularity amongst the two variables over the selected time period. Panel A in Figure 3 represents a constant tau of 0.62 for the Saudi market. Compared to the time-varying tau, the constant co-movement statistic is an over estimation of the correlation once time is controlled for. Panel B compares the Dubai market to oil price and the constant tau of -0.178 is an under-estimation of its co-movement compared to the time-varying approach. Both graphs depicting the time-varying co-movements have significant implications for pairs-trading strategies. It is a more robust tool to develop a market-neutral position between equity indices and oil prices over time. The absence of pre-conditions on the data parameterisation reflects a more realistic co-movement, even during periods of highly volatility.

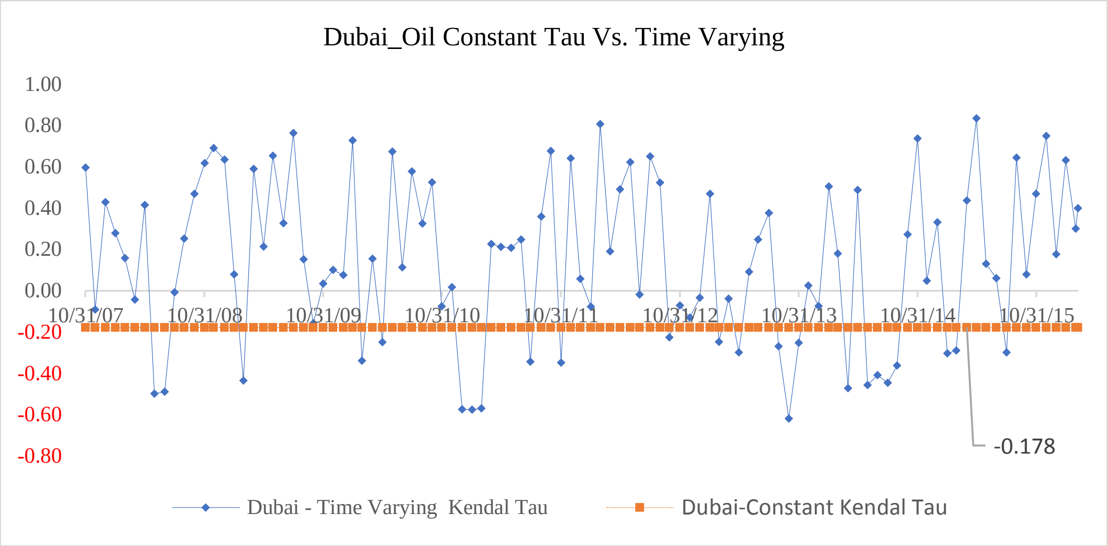
Another benefit of time-varying copulas is the ability to predict the correlation structure over the next period. A body of literature has developed around whether oil price may predict stock returns, with Narayan and Gupta (2015) oil price predicts US stock returns. In this paper we examine such possible dependence structures over time, we specifically extend our analysis over two indices, i.e., Saudi and Dubai. We follow the two-step procedure method as defined earlier and compute the required parameters for the two indices as listed in Table 3. The correlation structure is measured for oil prices and Saudi/Dubai respectively for the 1st quarter of 2016 and is represented in Table 3.

Table 3: Clayton TVC Estimation output for Saudi vs. Oil Prices.

SAUDI	Parameter	St. Error	t-stats
ω	0.9089	0.193	4.7201
α	-3.0428	0.644	-4.7280
β	0.7612	0.051	14.9155
AIC: -894.1068; BIC: -877.0401; Log Likelihood: 450.053			
DUBAI	Parameter	St. Error	t-stats
ω	2.50882	0.193	4.7201
α	-11.4452	0.644	-4.7280
β	0.0612	0.051	14.9155
AIC: -2229.4537; BIC: -2212.4229; Log Likelihood: 1117.727			

Figure 3: Comparison of constant and time varying tau for Saudi Arabia and Dubai to oil price



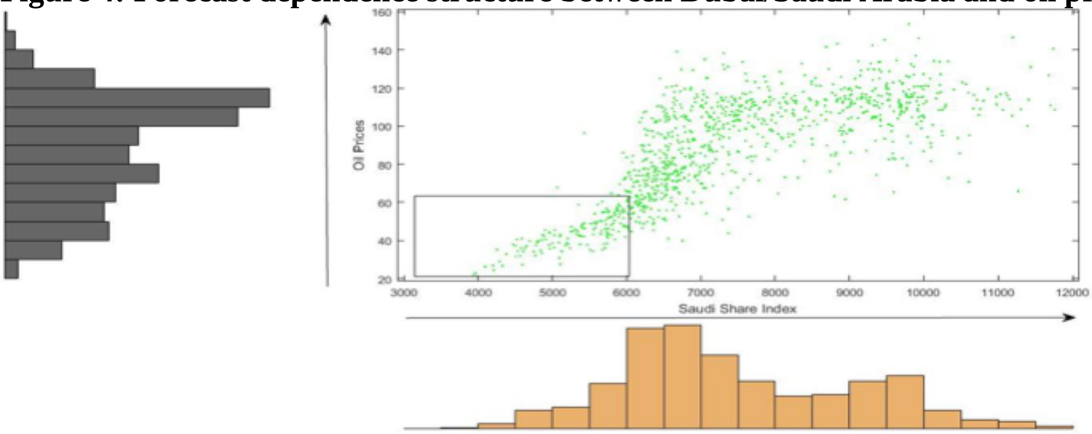


Comparison of constant and time varying tau for Saudi Arabia (Panel A) and Dubai to oil price (Panel B). The top figure represents a constant tau of 0.62 for the Saudi market. Compared to the time-varying approach, the constant approach is an over-estimation of the co-movement once time is controlled for. The bottom figure is the comparison of the Dubai market to oil price. The constant tau of -0.178 is an under-estimation of its co-movement compared to the time-varying approach.

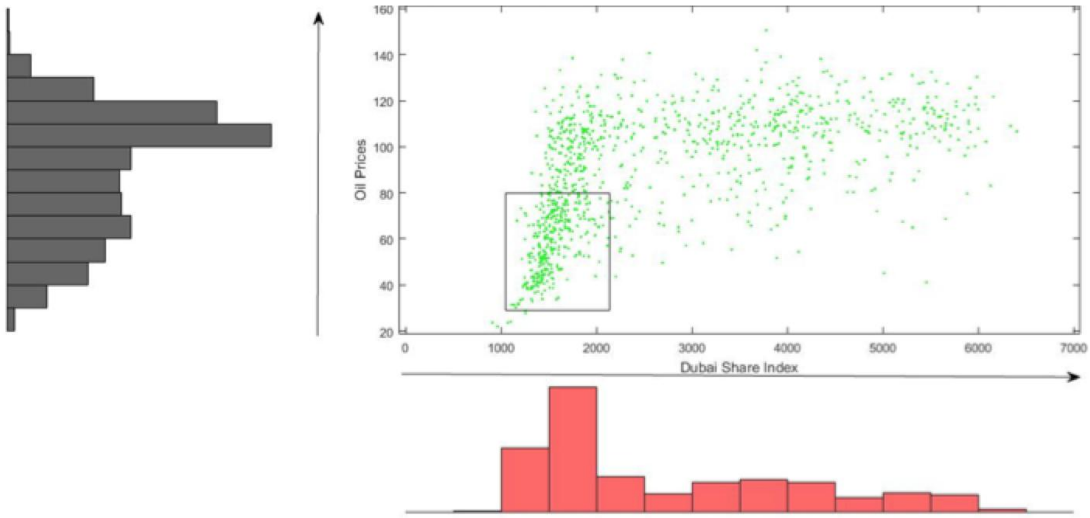
Both Saudi Arabia and Dubai show a concentration within the lower left-hand side quadrant, suggesting that low oil prices correspond with low equity price. This is in contrast to Naifar and Al-Dohaiman (2013), wherein upper side concentrations were reported, highlighting the importance of a time-varying approach.

Figure 4 also shows evidence of negative correlation behaviour, which is in line with Awartani and Maghyereh (2013), where bi-directionality is reported. We demonstrate the extent of such a phenomenon within the new paradigm of low oil prices, capturing the non-linear co-movement over time.

Figure 4: Forecast dependence structure between Dubai/Saudi Arabia and oil price



Forecast dependence structure between Dubai and oil for Q1, 2016, TVC Clayton parameter = 1.7820.



Forecast dependence structure between Saudi Arabia and oil for Q1, 2016, TVC Clayton parameter = 2.7619.

We have back-tested the predicted results against the actuals for their respective month. The test results are presented in Table 4. By using the two-step process described above, we extract the forecasted Kendal tau for the subsequent month based on the current historical information. Moreover, we then compare the forecasted with the actual Kendal tau measure. Finally, we extract the accuracy levels on a monthly basis. Using the TV copula methodology, we predict the results for Saudi Arabia and Dubai with 69.43% and 80.69% average accuracy levels respectively.

Table 4: Back-testing results for Saudi and Dubai markets Kendal tau.

SAUDI	Actual	Forecast	Acc. level (%)
Jan 2016	0.63	0.58	91.60
Feb 2016	0.48	0.23	47.99
Mar 2016	0.40	0.58	68.71
		Avg. accuracy	69.43
DUBAI	Actual	Forecast	Acc. level (%)
Jan 2016	0.63	0.99	63.64
Feb 2016	0.30	0.28	93.33
Mar 2016	0.40	0.47	85.11
		Avg. accuracy	80.69

6. Conclusion

Prior correlation modelling techniques employed a constant dependence structure to calculate the co-movement amongst the two variables (Naifar and Al-Dohaiman, 2013). In this study, we have applied Patton's (2011) approach and employed a time-varying dependence structure from 2007 till 2016, representing significant volatility in both oil and their corresponding equity markets. Hence, we are able to address the limitations surrounding Li's (2000) model and capture the extreme co-movements, finding significant dependency in the left tail. Compared to its constant counterpart, we find negative co-movement across the different indices. Overall, the analysis demonstrates that as the region develops further, different markets are less homogenous, developing their own idiosyncrasies. However, systemic risk remains applicable to all markets.

We capture both positive and negative co-movements over time, suggesting diversification benefits may still be achieved by investing within the GCC markets. With every oil price fluctuation, portfolios may be rebalanced in line with expected changes in oil price. However, the expected equity diversification benefits may become less evident with systemic risk.

However, from an economic perspective, it is evident that high oil prices lead to an increase in government revenues, which in turn leads to stronger fiscal and external positions anticipating a positive impact on the corporate sector. Nevertheless, more government spending may also lead to higher non-oil output growth, increasing credit growth, higher asset prices with positive overall wealth effects. Conversely, an oil price downturn may have a reverse effect. This study demonstrates that even in a downside oil price market, equity markets may respond positively to such events, presenting opportunities for equity rebalancing. We empirically demonstrate this phenomenon, with Dubai being more prone to negative co-movements than Saudi Arabia, probably because of its lesser reliance to the oil sector; its index is not predominantly conditional on oil price.

Finally, the techniques employed in this paper provide forecasters with reasonable accuracy levels. As this critical but unique market remains under-researched, the GCC market requires more intensive analysis based on more data being made available. We benefit from taking a longer horizon, using disaggregated data (different indices), and employing non-linear modelling techniques to model such co-movements. We also engage with out-of-sample testing, and by assessing a month-to-month rolling window, we are able to predict such parameters with reasonable accuracy. Based on this correlation predictability, better-paired equities may be designed.

This paper is not without limitations. To provide profitable strategies, the forecast returns are required to be corrected for risk levels and also for transaction costs. Furthermore, due to the direct influence of governments on the region's liquidity and available credit, it is not clear to what extent governments influence such markets. In terms of future research, the GCC remains under-researched and as more databases become available it would be insightful to build causal models to capture the determinants for the extreme values in the GCC equity indices.

References

Ajmi, A.N., EI-montasser, G., Hammoudeh, S. & Nguyen, D.K. 2014. Oil prices and MENA stock markets: New evidence from non-linear and asymmetric casualities during and after the crises period. Applied Economics 34, 1370-1379.

Aloui, C., and Hkiri, B. 2014. Co-movements of GCC emerging markets: New evidence from wavelet coherence analysis. 36, 421-431.

Aloui, C. Nguyen, D.K. & Njeh, H. 2012. Assessing the impact of oil price fluctuations on stock returns in emerging markets. Economic Modelling 29, 2686-2695.

Arouri, M.E.H. & Rault, C. 2012. Oil prices and stock markets in GCC countries: Empirical evidence from panel analysis. International Journal of Finance and Economics 17 (3), 242–253.

Arouri, M., Fouquau, J., 2009. On the short-term influence of oil price changes on stock markets in GCC countries: linear and nonlinear analyses. Economics Bulletin, Economics Bulletin 29 (2), 795-804.

Arouri, M.E.H., Jouini, J., Nguyen, D.K., 2011. Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management. Journal of International Money and Finance 30, 1387-1405.

Arouri, M., Jouini, J., Nguyen, D., 2012. On the impacts of oil price fluctuations on European equity markets: Volatility spillover and hedging effectiveness. Energy Economics 34, 611-617.

Awartani, B., Maghyereh, A.I., 2013. Dynamic spillovers between oil and stock markets in the Gulf Cooperation Council Countries. Energy Economics 36, 28-42.

Chapman, I., 2014. The end of Peak Oil? Why this topic is still relevant despite recent denials. Energy Policy 64, 93-101.

Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. Wiley Finance.

Connor, G., Lasarte, T. 2004. An Introduction to Hedge Fund Strategies. Research Report. Hedge Fund Research Programme. Financial Markets Group. London School of Economics.

Conrad, C., Loch, K., Rittler, D., 2014. On the macroeconomic determinants of long-term volatilities and correlations in US stock and crude oil markets. Journal of Empirical Finance, 29, 26-40.

Ding, H., Kim, H.G. and S. Park, 2016. Crude oil and stock markets: Casual relationships in tails?

Engle, R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50(4), 987-1007.

Fayyad, A., Daly, K.J., 2011. The impact of oil price shocks on stock market returns: comparing GCC countries with the UK and USA. Emerging Markets Review 12(1), 61-78.

Gevorkyan, A., and Semmler, W., 2016. Oil price, overleveraging and shakeout in the shale energy sector – Game changers in the oil industry, Economic Modelling, 54, 244-259.

Hammoudeh, S., Choi, K., 2006. Behaviour of GCC Stock Markets and Impacts of US Oil and Financial Markets. Research in International Business and Finance 20, 22-44.

Hammoudeh, S.M, Yuan, Y., McAleer, M., 2009. Shock and volatility spillovers among equity sectors of the Gulf Arab stock markets. The Quarterly Review of Economics and Finance 49, 829-842.

Joe, H. and Xu, J.J. (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jouini, J. 2013. Stock markets in GCC countries and global factors: a further investigation. Economic Modelling 31, 80-86.
Jouini, J. and Harrathi, N., 2014. Revisiting the shock and volatility transmissions among GCC stock and oil markets: A further investigation. 38, 486-494.
Khandelwal, P., Miyajima, K., Santos, A., 2016. The Impact of Oil Prices on the Banking System in the GCC. IMF Working Paper 16/161.
Li, D., 2000. On default correlation: A copula function approach. Journal of Fixed Income Vol. 9, No. 4, pp. 43-54.
Ling, C. M., 1965. Representation of associative functions. Publ. Math. Debrecen, 12, 189-212.
Maghyereh, A., Al-Kandari, A., 2007. Oil prices and stock markets in GCC countries: new evidence from nonlinear cointegration analysis. Managerial Finance 33, 449-460.
Malik, F., Hammoudeh, S., 2007. Shock and volatility transmission in the oil, US and Gulf equity markets. International Review of Economics & Finance 16(3), 357-368.
Martin-Barragan, B., Ramos, S. and Veiga, H., 2015. Correlations between oil and stock markets: A wavelet-based approach. 50, 212-227.
Mensi, W., Hammoudeh, S., Yoon, S.M. and Balcilar, M. 2017. Impact of macroeconomic factors and country risk ratings on GCC stock markets: evidence from a dynamic panel threshold model with regime switching. Applied Economics 49 (13), 1255-1272.
Mohanty, S., Nandha, M., Turkistani, A.Q., and Alaitani, M.Y. 2011. Oil price movements and stock market returns: Evidence from Gulf Cooperation Council (GCC) countries. Global Finance Journal vol. 22 ,1, 42-55.
Noguera-Santaella, J., 2016. Geopolitics and oil price. Economic Modelling 52, 301-309.
Naifar, N., Al-Dohaiman, M., 2013. Nonlinear analysis among crude oil prices, stock markets' return and macroeconomic variables. International Review of Economics & Finance 27, 416-431.
Narayan, P.K., & Gupta, R. 2015. Has oil price predicted stock returns for over a century? Energy Economics, 48, 18-23.
Panchenko, V., Diks, C., 2006. A new statistic and practical guidelines for nonparametric Granger causality testing. Journal of Economic Dynamics and Control 30, 1647-1669.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246-256.
Reboredo, J., 2011. How do crude oil prices co-move? A Copula Approach. Energy Economics 33(5), 948-955.
Salisu, A. and Isah, K., 2017. Revisiting the oil price and stock market nexus: A non-linear Panel ARDL approach. Economic Modelling, 66, 258-271.
Schweizer, B., Sklar, A., 1961. Mathematische Annalen 143, 440-447.
Serban, M. Brockwell, A. Lehoczky, J. Srivastava, S., 2007. Modelling the Dynamic Dependence Structure in Multivariate Financial Time Series. Journal of time-series analysis. 28(5), 763–782.
Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges", Publ. Inst. Statist. Univ. Paris (in French), **8**: 229–231.
Smyth, R. and Narayan, P.K., 2015. Applied Econometrics and Implications for Energy Economics Research, 50, 351-358.
Smyth, R. and Narayan, P.K., 2018. What do we know about oil prices and stock returns? Working paper.

Wang, Y., Wu, C., 2012. Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models? Energy Economics 34(6), 2167-2181.

Zarour, B.A., 2006. Wild oil prices, but brave stock markets! The case of GCC stock markets. Operational Research: An International Journal 6, 145-162.

Central Bank Interventions and Limit Order Behavior in the Foreign Exchange Market

Masayuki Susai, Nagasaki University

Yushi Yoshida, Shiga University

ABSTRACT

We investigate the intra-day effect of interventions in both the post- global crisis and pre-crisis periods by the Bank of Japan (BOJ) in foreign exchange markets using limit order data at intra-day high frequency. First, we find that the relationship between order flow and market return in dollar/yen exchange markets breaks down following unexpected and very high volumes of offer/sell orders by BOJ interventions. Then, a simple methodology of using large recursive residual is proposed to detect the exact timing of interventions. Second, the dataset allows measuring how long an individual limit order stays in the market. With the measured lifetime of limit orders, we find interventions, detected by the proposed methodology, significantly reduce the life-time of limit order in the market. By applying the same methodology on non-intervention days, we find no such evidence on the life-time of limit orders although large recursive residuals are also pervasive in non-intervention days.

JEL classification: F31, G12, G14, G15, E58.
Keywords: *the Bank of Japan; Central bank interventions; Foreign exchange market; Life time of limit order; Order flow.*

1. Introduction

The foreign exchange market interventions by the Bank of Japan (BOJ) were both voluminous and frequent in 2003 and 2004. US dollars equivalent to 20,246.5 (2003) and 14,831.3 (2004) billion Japanese yen, were purchased on the yen/dollar exchange market over 82 (2003) and 47 (2004) trading days. However, after the last day of the BOJ intervention on March 16, 2004, the Ministry of Finance of Japan and the BOJ quietly observed the movements of the Japanese yen on foreign exchange markets despite the fact that the Japanese yen was experiencing its historically highest level of appreciation since the World War II.

At the beginning of 2007, the sub-prime housing market in the U.S. started plummeting, and the consequent financial turmoil spread to the rest of the world. The US dollar and Euro depreciated against the other major currencies, especially against the Japanese yen. After six years of inactivity, the BOJ intervened in the yen/dollar exchange market on September 15, 2010, to the surprise of many market participants[84]. The size of the intervention transaction per day was unprecedentedly high, at 2,124.9 billion yen[85]. This is equivalent of 25,601 (24,999) million US dollars, calculated at the rate of 83.0 (85.0) Japanese yen per US dollar.

In this paper, we investigate the effects of the BOJ intervention on September 15, 2010 on trading activities on the yen/dollar market of the Electronic Broking System (EBS). To determine whether the findings are specific to this event, i.e., the first time in six years and the first time after the global crisis, we also apply the same methodology to five intervention days during the pre-crisis period. The major two contributions of this investigation are the following: First, we find that the relationship between order flow and market return in dollar/yen exchange markets breaks down following unexpected and very high volumes of offer/sell orders by BOJ interventions. Then, a simple methodology of using large recursive residual is proposed to detect the exact timing of interventions. Second, the dataset allows measuring how long an individual limit order stays in the market. With the measured lifetime of limit orders, we find interventions, detected by the proposed methodology, significantly reduce the life-time of limit order in the market. We find interventions significantly reduce the life-time of limit order by about 27 to 44 seconds. By applying the same methodology on non-intervention days, we find no such evidence on the life-time of limit orders although large recursive residuals are also pervasive in non-intervention days.

The remainder of this paper is structured as follows. The next section discusses the key concepts used in this paper and reviews the relevant studies in the literature. Section 3 describes the structure of the EBS dataset. Section 4 summarizes distinctive characteristics of the yen/dollar foreign exchange market on the day that the BOJ intervened for the first time in six years and reports preliminary investigations of the EBS dataset. Section 5 provides the empirical results on the relationship between order flows and exchange rate returns and proposes the simple method of detecting the exact timing of interventions (and after-effects)

---

[84] Prior to September 15, 2010, Mr. Noda, then Minister of Finance, repeatedly spoke to the media saying that the MOF and the BOJ would take necessary actions, including interventions, to halt further appreciation of the Japanese yen against the US dollar. However, the market participants, as reported frequently in the media, did not believe that the MOF and the BOJ would intervene in the yen-dollar exchange market on this particular day. Mr. Noda revealed in a morning interview with the press on the same day that the MOF requested the intervention of the BOJ at 10:30 AM (1:30 in GMT).

[85] After this intervention, the BOJ intervened in the foreign exchange market to the amount of 692.5 billion Japanese yen on March 18, 2011 and (the historically highest amount per day) 4,512.9 billion Japanese yen on August 4, 2011. By the end of November, 2011, the Ministry of Finance reports that the total value of interventions between October 28 and November 28, 2011 was 9,916 billion Japanese yen.

by using recursive errors. Section 6 investigates the impact of intra-day interventions on limit order behavior of foreign exchange market participants. By limit order behavior in this paper, we focus on the life-time (how long an individual limit order stays in the market) of limit orders. Section 7 provides robustness checks on the effect of interventions on the life-time of limit orders by applying the same method on non-intervention days. The final section discusses the findings and reports the conclusions.

2. Order flow, limit order, and intervention

In this section, we discuss the existing literature, with an emphasis on the relationship between three important features of this study: the use of limit orders as order flows, measuring the lifetime of limit orders, and an investigation of intervention at intra-day frequency. First, we argue that further investigation of the possible information dissemination role of limit orders is necessary. The current definition of order flow is based on actual transactions, i.e., observationally equivalent to market orders[86]. Second, the investigation of limit orders requires a new approach because, unlike market orders, most limit orders are canceled (or revised with a new price). We propose to investigate the effect of possible determinants of the life (i.e., the length of time that they remain in the order book) of limit orders. Third, using intra-day high frequency data for limit orders on the foreign exchange market, detection of the exact timing of intervention becomes an unavoidable issue[87]. Using an unusual, isolated incident of a publicized timing of intervention by the BOJ, we test the accuracy of an intervention-timing candidate found using a proposed approach and intend to apply the same approach to other interventions if the approach is proven valid. In the following, we discuss the three key concepts used in this paper in turn: the relationship of order flow to market and limit orders, the life-time of limit orders, and interventions at intra-day frequency.

2-1. The relationship of order flow to market and limit orders

The microstructural approach to the foreign exchange market (e.g., Lyons, 1997) emphasizes the role of order flow as a determinant of the exchange rate. Order flow is defined as the net result of buyers' initiated transactions minus sellers' initiated transactions (Evans and Lyons, 2002b). Because customer transactions are private dealer information, order flow in interbank transactions disseminates this information and affects the market price. Acquiring proprietary order flow data from one of the largest market makers, Cerrato et al. (2011) and Marsh (2011) investigate the effect of order flows of various customer types on exchange rate. Order flow is found to affect exchange rate by reflecting macroeconomic information (Rime et al., 2010 and Frömmel et al. 2011) and commodity price (King et al., 2010). In this context, market orders are treated as the only tool conveying private information throughout the market, whereas the role of limit orders is considered to be only passive, at best providing liquidity to the market.

However, current foreign exchange markets, which are dominated by electronic brokering platforms such as EBS and Reuters, are limit order markets. Cumulative limit orders constitute the order book with best bid-ask quotes, and submitted market orders are matched with existing limit orders at the best quotes. The number of limit orders submitted exceeds that of market orders in various financial markets. The theoretical framework in which limit order traders act only as liquidity providers is not suitable for explaining the current limit order markets.

[86] Market orders are orders matching the existing best quote in the market, and limit orders are orders set at specific prices, which may not be the same as the best quotes.

[87] See Menkhoff (2010) for a current survey on the high-frequency analysis of interventions and Vitale (2011) for a theoretical model of interventions in a market microstructure model.

Theoretical models have been developed to allow a trader to choose between market and limit orders, e.g., Cohen et al. (1981), Foucault (1999) and Bloomfield et al. (2005), among others. Unlike market orders, limit orders face the risk of non-execution. Foucault (1999) examines the sub-game perfect equilibrium in a dynamic limit order market in which a trader chooses to submit either a market or limit order, with explicit consideration for non-execution risk and the risk of being picked off (or free-option risk). The results show that high volatility leads to more limit orders than market orders being placed and to a lower fill rate, i.e., the probability of being hit by a market order, for a limit order. Considering the two types of traders in a model, Bloomfield et al. (2005) show that informed traders and liquidity traders use both market and limit orders. Informed traders, in particular, use market orders to realize profit at the opening of the market and switch to limit orders as the market price approaches true value at the close of the market.

Based on theoretical developments regarding limit orders, we argue that the limit order has a more active role in disseminating private information to the market than market orders[88]. In this paper, therefore, we define order flow as the bid limit orders minus the offer limit orders. We investigate whether the order flow defined by limit orders has a significant effect on the market price on the foreign exchange market.

2-2. Life-time of limit orders

Limit orders by their nature need not be executed instantly and are frequently canceled without any transaction taking place. Many studies document high levels of cancellations in various limit order markets. Biais et al. (1995) is the first to investigate the order book of the limit order market at the Paris Bourse, which provides traders with the best five quotes and the corresponding volumes each time a new order or cancellation occurs. They document that approximately 20 percent of orders (at best five quotes) are canceled[89]. Harris and Hasbrouck (1996) document that 56.2 percent of limit orders on the New York Stock Exchange remain unfilled. This figure should not be interpreted as active cancellation, as some limit orders simply remain unmatched at the close of the market. Using the complete tick data for a company on the Stockholm Stock Exchange, Hollifield and Miller (2004) report that the execution probability for two days is 68, 33, and 12 percent for limit orders that are, respectively, 1, 2, and 3 ticks away from the best quote. Eventually, 88 percent of limit orders with prices 3 ticks away from the best quote are canceled. Yeo (2005) reports that the ratio of cancellations to submitted limit orders on the New York Stock Exchange has recently increased to 40 percent[90]. Hasbrouck and Saar (2002) document that roughly 25 (40) percent of limit orders are canceled after two (ten) seconds on the Island ECN, which constitutes 11 % of the trades on the Nasdaq exchange in 1999.

[88] The effects of limit orders on market characteristics are also investigated empirically and theoretically. As an empirical work, Biais et al. (1995) find that the conditional probability of placing limit orders rather than market orders is larger when the bid-ask spread is large or the order book is thin. In the model of Foucault et al. (2007), in which limit order traders possess asymmetric information about future volatility, the bid-ask spread signals the size of future volatility.

[89] This percentage is calculated by the ratio between unconditional new orders and cancellations shown in Table III (p. 1670, Biais et al., 1995).

[90] Yeo (2005) compares the percentage of cancellations in all submitted requests, which include market orders, limit orders, and cancelations. Note that this percentage has the highest limit of 50 percent for cancellations because the number of cancellations cannot exceed the number of limit orders. Approximately 20 percent of orders were cancellations in 2001, compared with 5 percent in 1990.

This large number of canceled limit orders can be attributed to the order splitting strategy and undercutting, according to Yeo (2005). Traders split orders in multiple submissions when they intend not to disseminate their private information. This strategy results in multiple cancellations when traders revise their orders. On the other hand, traders who compete to undercut other traders need to revise their prices frequently. The dynamic limit order market model of Foucault (1999) indicates that higher volatility leads to a lower fill rate. A lower fill rate, then, can be interpreted as a higher probability of cancellation in the foreign exchange market because no specific closing time exists. Foucault et al. (2005) theoretically show that the average time to a transaction increases with the size of the spread. This result in turn can be interpreted as indicating a lower fill rate at a fixed time interval during sporadic incoming orders.

To investigate why a significant proportion of order activities consists of cancelations and revisions, Fong and Liu (2010) consider the effect of non-execution risk, free-option risk, and monitoring cost in the order strategy. They find that both a closer submitted price to the best bid-ask quotes and a larger order volume increase the likelihood of cancellation or order revision. Using probit analysis, Yeo (2005) finds that a move in the market quote away from the submitted price induces cancellations. Both a larger volume in limit order and a larger volume at the best quote, i.e., the depth at quotes, deter cancellations.

Susai and Yoshida (2012) investigates the determinants of life-time of limit orders in the JPY/USD foreign exchange market. They measure how long each limit order stays in the market by calculating the length of time between the time stamp of order submission and that of either cancelation or transaction execution. The full discussions on how the possible determinants, conditional on market conditions and other participants behaviors, may influence the limit order behaviors in foreign exchange market.

2-3. Intervention at intra-day frequency

The availability of intervention data at intra-day level for the Swiss National Bank (hereafter SNB) motivates the research by Fischer and Zurlinden (1999), Payne and Vitale (2003), and Pasquariello (2007), among other studies. Using tick by tick data directly, Fischer and Zurlinden (1999) find that intervention, especially the first transaction, affects the exchange rate movement. Payne and Vitale (2003) also examine the SNB interventions by aggregating tick data at 15-minute intervals, whereas Fischer and Zurlinden (1999) used irregularly spaced tick series. They find that the SNB intervention has a stronger impact when it is leaning with the wind and concerted with other central banks. Pasquariello (2007) further aggregates tick by tick data into daily variables.

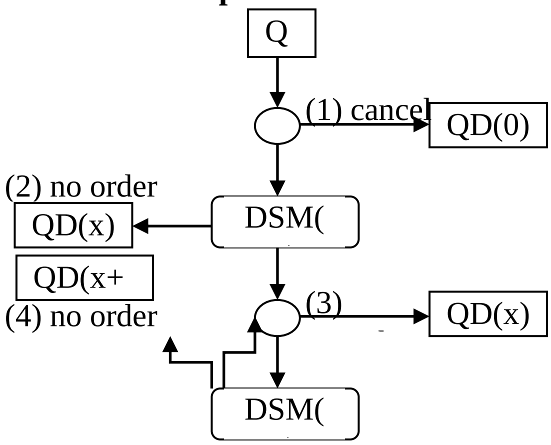
Chari (2007) combines the news reports of interventions with tick by tick quotes from Reuters and finds that the BOJ and the FRB interventions lead to increased volatility and a widening of bid-ask spreads. Using hourly aggregates of tick data for the Czech krouna-euro, Scalia (2008) finds that interventions (news) by the Czech National Bank increase the impact of order flow on the exchange rate. In contrast, by aggregating tick data for the Russian rouble-US dollar into 30-second intervals, Melvin et al. (2009) find that the price impact of order flow is smaller on intervention days.

The exact timing of interventions can be traced back to headline news reports (e.g., Chari, 2007); however, the inaccuracy of news reports regarding interventions is well documented in Klein (1993), Osterberg and Wetmore Humes (1993), and Fischer (2006). Among the central banks making their daily intervention data available to the public, only the SNB reveals to her counterpart that transactions are carried out for the purpose of intervention, see Fischer (2006) for detail. The exact time in minutes can be confirmed only for SNB interventions. Regarding the interventions of other central banks, researchers may make educated guesses about the exact timing of particular intervention episode by gathering newswire reports and scrutinizing the tick data but are never able to confirm whether their guesses are correct. Notwithstanding this vagueness about the timing of interventions, on a single occasion (September 15, 2010), the Minister of Finance publicly revealed the exact timing of a BOJ intervention, see footnote 1. This incident provides a great advantage to BOJ intervention research, and we use this fact to its full extent to check the validity of our proposed methodology to detect the timing of interventions.

3. The EBS data structure

Traders can either initiate a quote (i.e., submit a limit order) or match a posted quote (i.e., submit a market order). In the EBS dataset, all data entries are assigned one of five indicators: QS, QD, HS, HAD, and DSM. A quote begins with QS and a specific 20-digit ID and ends with QD. A hit begins with HS and ends with HAD. When two parties are matched in a transaction, DSM records the information for the transaction. The life of a quote can be described by the four cases shown in Figure 1: (1) a quote is deleted by cancellation; (2) a quote is filled either by another quote or by a market order; (3) a quote is canceled after part of the order is executed; (4) a quote is filled by multiple transactions. In the EBS dataset, which we purchased with a limited contract, all data cannot be made public unless aggregated to conceal the characteristics of individual transactions[91].

Figure 1. Records of quotes on the EBS Spot market



Note: QS indicates the start of the quote. QD indicates the end of the quote. DSM indicates a transaction. There are four cases for the life of quotes. (1) a quote is deleted by cancellation; (2) a quote is filled with either another quote or a market order; (3) a quote is canceled after a part of the order is executed; (4) a quote is filled by multiple transactions.

4. Summary of trading activities on September 15, 2010

The dataset is all limit orders in the EBS JPY/USD spot market with the sample covering the period between 21:00:00 (GMT) on September 14, 2010 and 20:59:59 (GMT) on September 15, 2010. The number of all limit orders is 625,725. A large portion of orders is submitted literally within a split second after the last order has been placed in the market. The orders submitted within a second after the last order constitutes 97.3 percent (608,793) of all orders for the day. In addition, approximately 0.7 percent of all orders are submitted simultaneously, measured in terms of milliseconds, with another order. This extremely fast speed of orders is

[91] For a more detailed description of this EBS dataset, see Susai and Yoshida (2012).

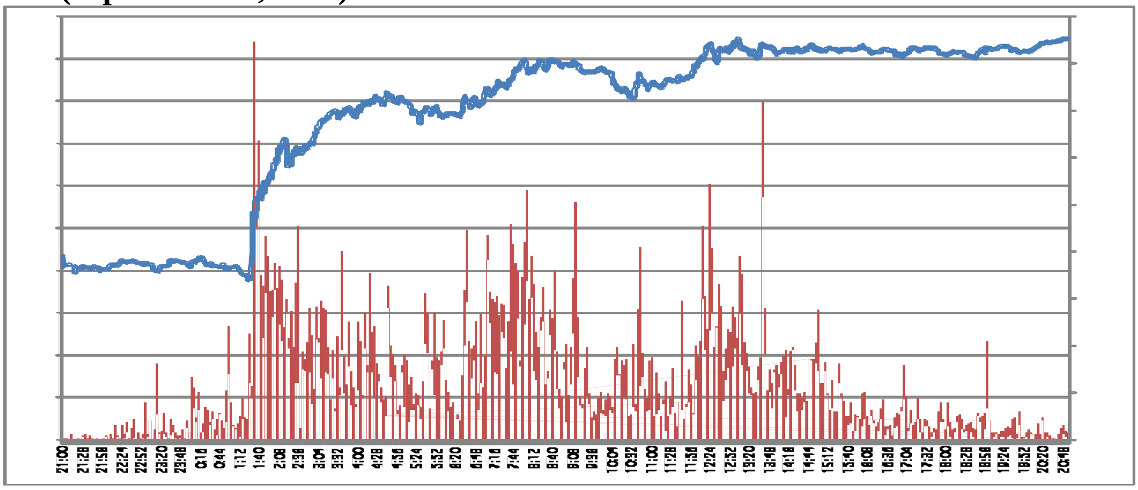
explained in part by the pervasive use of algorithm trading using computers[92]. The asymmetric information models described in Easley and O'Hara (1987, 1992) suggest that large orders and short durations are evidence of trading by informed traders. Manganelli (2005) find supporting evidence for the link between short durations and trading by informed traders on the NYSE.

4-1. Order volume and rate by minute intervals

In the sample period of 24 hours, the number of limit orders is 625,725, time-stamped in milliseconds. Due to the irregular time spans of quote submissions, interpreting a series of raw data requires special care, even for a simple graph. We choose to convert these raw quote data to minute intervals. The minute interval sums all volumes for limit orders submitted within one minute. For example, the volume for 23:02 covers all transactions processed between 23:02:00 and 23:02:59. The minute interval rate is the rate for the earliest limit order in the next minute. If no limit order is submitted in the current minute, the previous minute interval rate is maintained.

Figure 2 provides the simultaneous limit order volume and JPY/USD rate. The beginning of swift intervention by the BOJ near 1:00 (GMT) is obvious in the figure, although a smaller BOJ intervention before this time cannot be ruled out. Limit order volume for the 01:32 minute interval increased suddenly to 4,235 million US dollars from 167 million US dollars in the previous minute interval. In approximately ten minutes, the JPY/USD rate increased from 82.915 (mid-rate) at 01:31 (minute interval) to 83.835 (mid-rate) at 01:40 (minute interval). The Japanese yen continued depreciating and remained at approximately 85.70 at the end of the day.

Figure 2. Exchange rate and limit-order volume plotted against minute interval (September 15, 2010)



Note: The minute interval sums all limit order volumes in a one-minute period. For example, the volume at 23:02 covers all transactions that occur between 23:02:00 and 23:02:59. The minute interval exchange rate is the mid-rate of the latest best bid-ask rates during the current minute. The exchange rate is scaled on the left vertical axis and the limit-order volume (bar) is scaled on the right vertical axis.

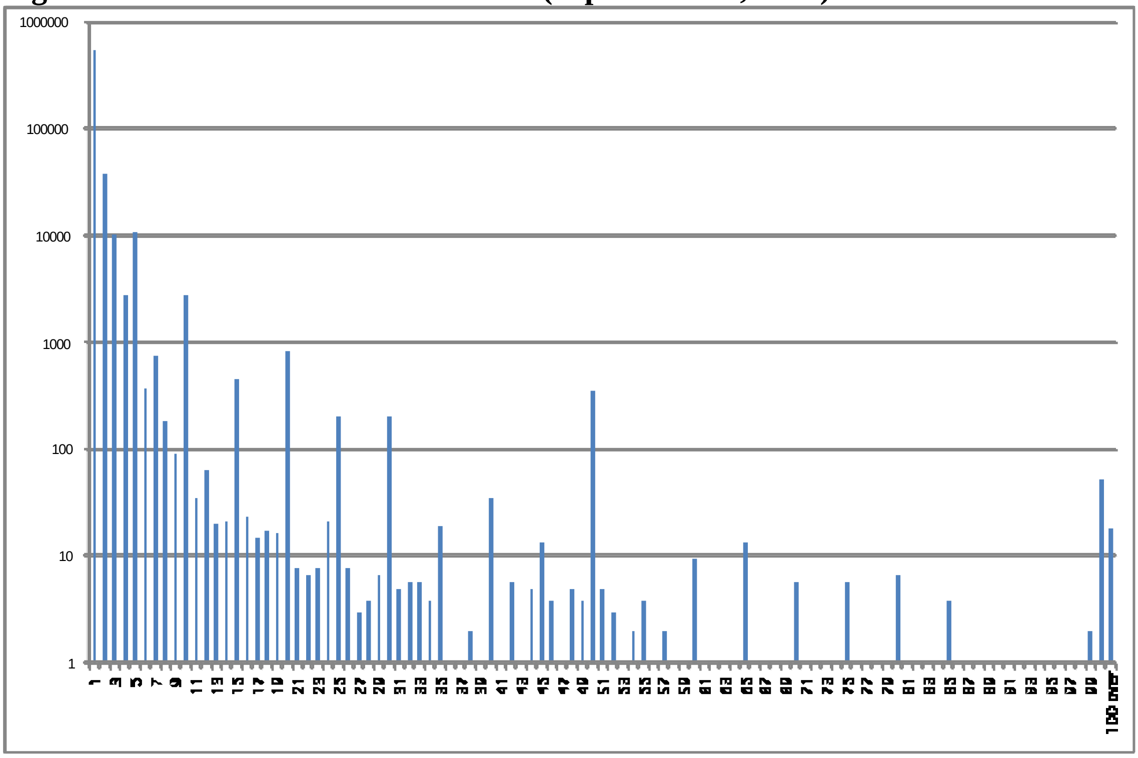
4-2. Hourly order size by order type

The order size in the dataset ranges from a minimum of one million US dollars to a maximum of 430 million US dollars (Figure 3). The size of orders used most is the minimum requirement

[92] Corwin and Lipson (2011) distinguish program (algorithm) traders, institutional traders, retail traders, and member traders in their empirical analysis of NYSE-listed securities. See Section 2 of their paper for the significant presence of program traders.

of one million US dollars, and the use generally declines with increasing size of orders except for some focal numbers such as 10 million. The clustering in small orders is consistent with the limit of open positions for traders (Cheung and Chinn, 2001). The intraday position limit of most dealers in the US does not exceed 50 million US dollars. Cheung and Chinn (2001) report in their survey that 54 % (74 %) of dealers are authorized to have a maximum open position of less than 25 (over 50) million US dollars. Orders exceeding 50 million US dollars are exceptionally high against the limit of open positions for the most trading institutions. For the intervention days in the pre-crisis period, the maximum volumes of limit orders are 100, 520, 500, 500, and 300 million US dollars, respectively, for September 12, 2003; September 30, 2003; December 10, 2003; January 9, 2004; and March 5, 2004.

Figure 3. The distribution of order size (September 15, 2010)



Note: All limit orders on the JPN/USD spot market from 21:00:00 (GMT) on September 14 to 20:59:59 (GMT) on September 15, 2010. The number of data points is 625,725. The vertical axis is the number of orders, shown on a log scale. The size of orders on the horizontal axis is marked in US million dollars. Over 85 percent of all orders have a minimum value of 1 million US dollars.

Based on the limit data reported in Figure 2, there is substantial variation in the activity of the foreign exchange market, especially on the intervention day. In this subsection, we break down the market orders by hour and compare the characteristics of orders between limit and market orders and between offer (sell) and bid (buy) orders. Table 1 reports the number of orders submitted in a particular one-hour period. Consistent with the transaction volume reported in Figure 2, panel (A) in Table 1 indicates that the number of orders exceeds 47,000, i.e., there were approximately 13 orders per second after 01:00, and the trading continues to be highly active until 15:00. In terms of dollar purchase interventions by the BOJ, the frequency of order type does not hint at any peculiarity possibly caused by the official intervention.

Table 1. Order size by hour (September 15, 2010)

	panel (A)				panel (B)		panel (C)					
							10over			50over		
hour	all	bid	offer	bid-offer	average	max	all	bid	offer	all	bid	offer
21	689	374	315	59	2.09	60	23	22	1	4	4	0
22	2,823	1,299	1,524	-225	1.33	75	30	19	11	4	2	2
23	5,921	2,982	2,939	43	1.16	51	35	24	11	2	1	1
0	14,612	7,396	7,216	180	1.27	100	136	61	75	13	7	6
1	47,334	22,087	25,247	-3160	1.42	300	707	439	268	63	42	21
2	50,029	23,287	26,742	-3455	1.41	250	668	398	270	87	52	35
3	36,076	17,119	18,957	-1838	1.38	250	503	207	296	33	22	11
4	30,433	14,641	15,792	-1151	1.57	250	505	261	244	85	68	17
5	35,088	16,629	18,459	-1830	1.29	250	307	215	92	25	20	5
6	33,452	16,095	17,357	-1262	1.30	95	226	135	91	10	2	8
7	51,683	25,423	26,260	-837	1.45	100	624	308	316	44	25	19
8	41,633	20,326	21,307	-981	1.47	411	348	169	179	35	22	13
9	23,053	11,409	11,644	-235	1.41	100	141	70	71	13	9	4
10	31,100	14,864	16,236	-1372	1.37	150	196	120	76	11	9	2
11	19,922	9,898	10,024	-126	1.43	65	93	52	41	15	13	2
12	56,145	27,780	28,365	-585	1.34	430	467	263	204	27	22	5
13	42,465	20,630	21,835	-1205	1.29	100	239	129	110	14	7	7
14	39,020	19,064	19,956	-892	1.22	100	141	84	57	12	8	4
15	20,537	10,209	10,328	-119	1.26	80	66	36	30	8	2	6
16	12,911	6,460	6,451	9	1.21	60	49	11	38	3	1	2
17	9,813	4,901	4,912	-11	1.14	50	27	15	12	2	2	0
18	11,085	5,664	5,421	243	1.17	199	41	14	27	7	4	3
19	5,955	3,107	2,848	259	1.17	50	17	1	16	1	0	1
20	3,946	2,054	1,892	162	1.35	40	36	18	18	0	0	0
Total	625,725	303,698	322,027				5,625			518		

Note: The figures in panels (A) and (C) are the number of orders submitted within one hour. In panel (B), the unit is one million US dollars for the average and maximum values. In panel (C), 10 over and 50 over represent the number of orders greater than or equal to 10 and 50 million US dollars, respectively.

Panel (B) in Table 1 provides the hourly breakdown of order size on the foreign exchange market. The average size of orders varies from 1.14 million US dollars at 17:00 to 2.09 million US dollars at 21:00. It is noteworthy that the average volume of 1:00 (the start of the sharp depreciation of the Japanese yen) takes place within the high part of the range at 1.49 million US dollars. The maximum volumes per orders are 430, 411, and 300 million US dollars, respectively at 12:00, 8:00, and 1:00.

We further investigate order activity by examining the hourly breakdown of order size by market order types in Panel (C) in Table 1, which provides the number of orders greater than or equal to 10 and 50 million US dollars for bid, offer, and both. Focusing on the three most active orders in each category, the orders at 1:00, 2:00, and 7:00 show quote activities that are active in high-volume orders. In these high-volume order categories, the number of bids exceeding offers is observed at more hours.

5. Empirical analysis by using aggregate data at minute intervals

In this section we investigate the relationship between order flows and exchange rate returns by using aggregated data at minute intervals. Our choice of aggregating at minute intervals is the best compromise between taking advantages of data availability at the tick level and the aggregate concept of order flows by definition. The next subsection provides on the estimation results on the relationship between order flows and exchange rate returns and the following subsections discuss the method to detect the exact timing (and possible after-effects) of intra-day interventions.

5-1. Exchange rate return and order flow

The effect of order flow on price change is well documented (Jones et al., 1994, Evans and Lyons, 2002). In the literature, the net order flow is the difference between purchase orders and sales orders. Distinguishing intervention days from non-intervention days, Marsh (2011) estimates the effect of net order flow on the change in the exchange rate. Interestingly, the significant effect of order flow on non-intervention days disappears on intervention days.

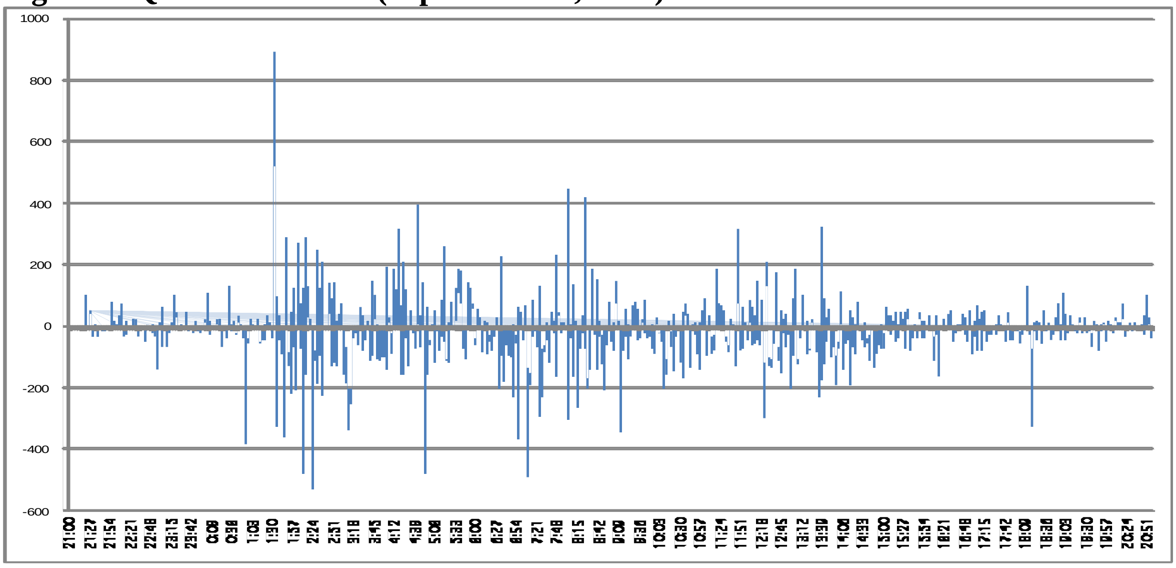
The net order flow is constructed for minute intervals between 21:00 on September 14, 2010 and 20:59 on September 15, 2010. As discussed in detail in Section 2, we define the net quote order flow as the bid limit order (dollar purchase) minus the offer limit order (dollar sales), and a positive value indicates a net purchase order for dollars. Figure 4 shows the net order flows of yen/dollar spot foreign exchange market on the EBS by minute. The US dollar purchase order by the BOJ intended to depreciate the Japanese yen should appear as a positive value in the figure.

Plotting the net order flow against the change in exchange rate in Figure 5, we observe a positive relationship between the net dollar purchase and a positive return of the yen/dollar exchange rate. More formally, following the simple regression approach in Marsh (2011), we estimate the following equation (1):

$$R_t = \alpha_0 + \alpha_1 OF_t + \varepsilon_t, \qquad (1)$$

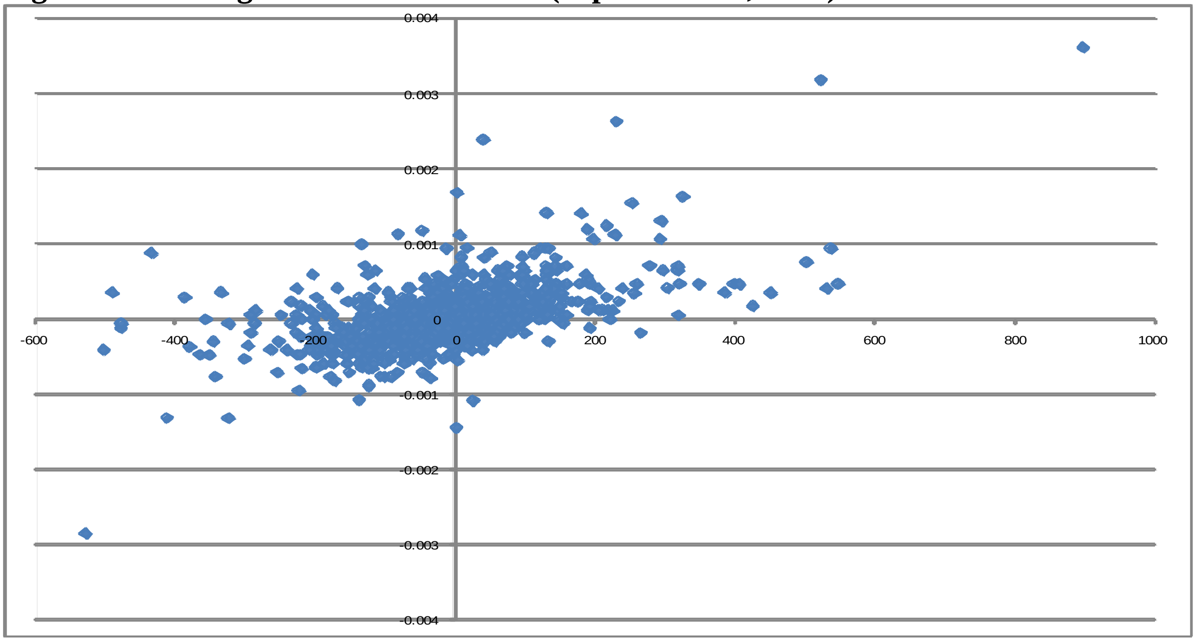
where R_t is the change in the log of the exchange rate at minute intervals (used in Figure 2), OF_t is the net order flow by minute and error terms, ε_t, are independent and normally distributed with means zero and variance, σ^2.

Figure 4. Quote order flow (September 15, 2010)



Note: The net quote order flow is defined as the bid (dollar purchase) minus offer (dollar sales). The minute interval sums all quote volumes for the one minute.

Figure 5. Exchange rate and order flow (September 15, 2010)



Note: The one-minute change measured in log yen/dollar is plotted on the vertical axis, and the net (quote) order flow is plotted on the horizontal axis.

Table 2. The per minute change in yen-dollar on net order flow (September 15, 2010)

	post-crisis
	2010Sep15
constant	0.00004006***
	(0.00000792)
OF	0.00000180***
	(0.00000019)
Adj.R2	0.31
NOB	1439

	pre-crisis				
	2003Sep12	2003Sep30	2003Dec10	2004Jan9	2004Mar5
constant	0.00000142	0.00001117	0.00000805	-0.00000618	0.00000283
	(0.00000335)	(0.00000811)	(0.00000508)	(0.00000567)	(0.00000386)
OF	0.00000072***	0.00000186***	0.00000121***	0.00000181**	0.00000037**
	(0.00000024)	(0.00000059)	(0.00000034)	(0.00000084)	(0.00000015)
Adj.R2	0.07	0.13	0.10	0.29	0.04
NOB	1439	1439	1439	1439	1439

Note: The dependent variable is the per minute change in the log of the yen/dollar exchange rate, and a positive value indicates dollar appreciation. The net quote order flow is defined as the bid (dollar purchase) minus offer (dollar sales), and a positive value indicates the net purchase order measured in dollars. The standard errors in parenthesis are robust to heteroskedasticity. *, **, and *** denote statistical significance at the ten-, five-, and one-percent levels, respectively.

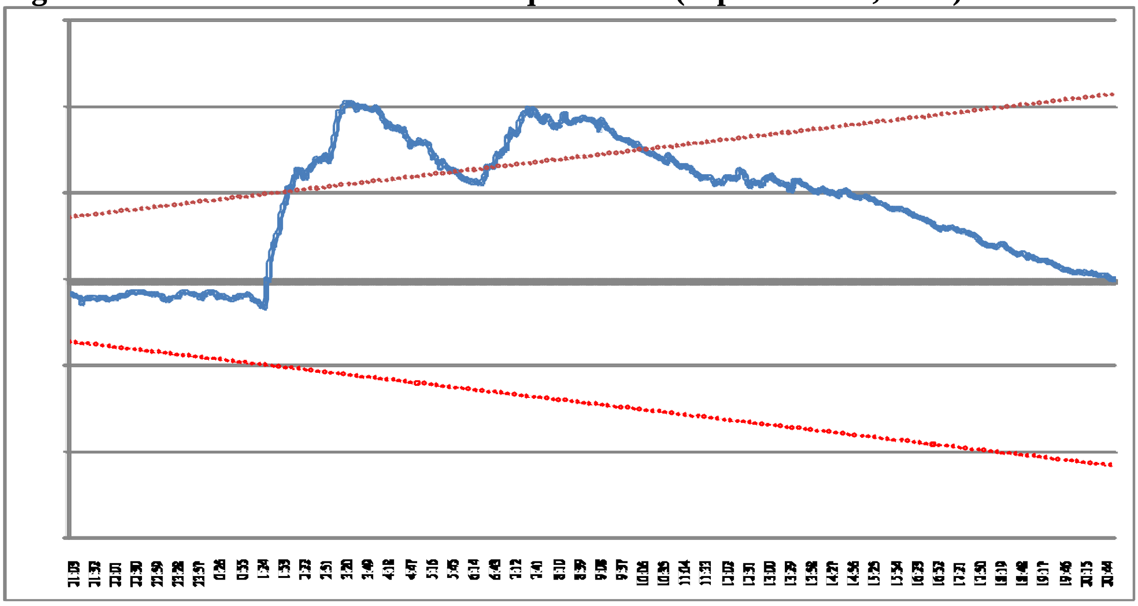
The upper panel of Table 2 reports the estimation results for September 15, 2010. The coefficient of net order flow is correctly signed and statistically significant at the one-percent level. The fitness of regression in terms of adjusted R^2 is 0.31 and is relatively high compared to the order flow literature. It is noteworthy that order flow in previous studies is defined using market orders (or observationally equivalent transactions).

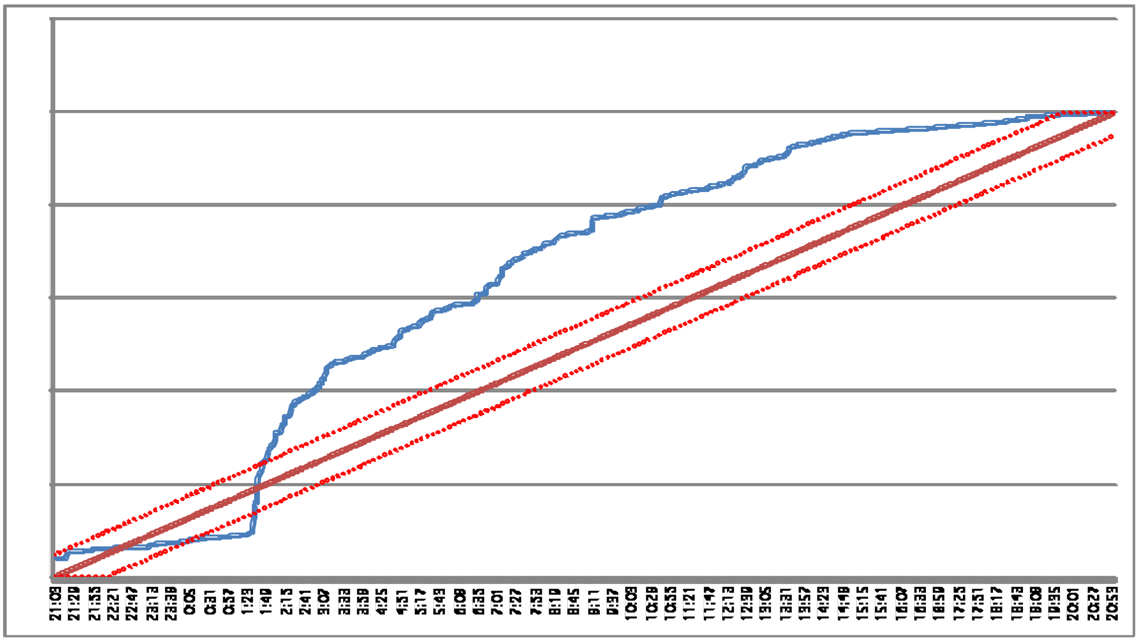
We also estimated equation (1) for the five intervention days in the pre-crisis period. The estimation results are shown in the lower panel of Table 2. The coefficients of order flow are all statistically significant at the five-percent level, but the degrees of fitness of the regression are smaller than that for the post-crisis priod. In this study, we obtain supporting evidence that the limit order version of order flow also affects the price.
This result supports the notion that the submission of a limit order carries important private information to be disseminated through the market.

5-2. Stability of the relationship between order flow and exchange rate

Now, reflecting the fact that previous studies using daily exchange rate data and daily order flow data can distinguish intervention days from non-intervention days as in Marsh (2011), we test whether the parameters of the empirical model in equation (1) is stable throughout the entire period including pre-intervention hours, intervening hours, and post-intervention hours. For this examination, we implement the CUSUM test for the residual from the regression of equation (1). The cumulated sum of the residuals is plotted along the 95 percent upper-bound and lower-bound lines in Figure 6. The structural change is first detected at the 95 percent level at 02:03 on September 15, 2010. It is noteworthy that the cumulated sum of the residuals begins a sharp increase near 01:30. We have evidence that the relationship between the order flow and exchange rate returns is affected by the BOJ interventions.

Figure 6. The CUSUM and CUSUM square tests (September 15, 2010)





Note: The solid line represents the accumulated sum of the forecast residuals (upper panel) and the squared forecast residuals (lower panel), and the dotted lines represent the 95 percent upper and lower bounds.

The cumulated sums of the residuals are similarly plotted for five intervention days in the pre-crisis period in Figure 7. Strikingly, the standard CUSUM tests do not reject the null hypothesis of no structural change occurring during the pre-crisis period except for September 30, 2003 in Panel B of Figure 7. On September 30, 2003, the cumulative sum of the residuals exceeds the 95 percent upper bound at 23:25 (GMT).

What is striking in these plots of the cumulative sum of residuals and the cumulative sum of squared residuals is that sudden increases are observed, almost vertical increases, at several points in the sample. Three occur on September 15, 2010, one on September 12, 2003, three on September 30, 2003, two on December 10, a less obvious one on January 9, 2004, and at least two on March 5, 2004. These plots indicate that unusually high volume orders by the BOJ interventions can be detected at the minute or at least hour level by plotting the cumulative sum of the residuals. This reflects the fact that the relationship between limit order flow and exchange rate drastically changes when the BOJ is intervening in the foreign exchange market.

5-3. **intra-day intervention detection**

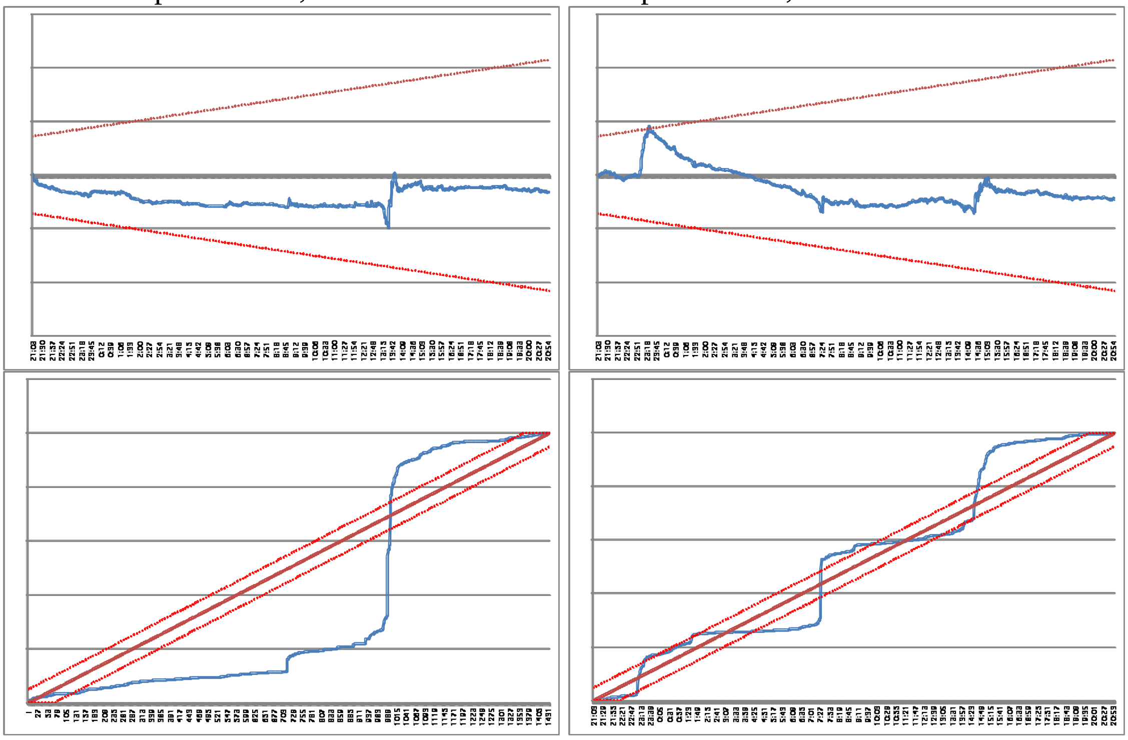
Given the strong detection power of recursive residuals for intervention activities, we formally introduce our proposed method in this subsection. Let x_t be (*1*, *OF$_t$*) , $X_{t-1}^{'}=[x_1,...,x_{t-1}]$, y_t be R_t and $Y_{t-1}^{'}=[y_1,...,y_{t-1}]$. Let $a_t^{'}$ be the least-squares estimates of $[\alpha_0,\alpha_1]$ based on the first t observations. Under the hypothesis of constant parameters and constant variance of error terms, recursive residual, w_t, is proved to follow $N(0,\sigma^2)$, see Brown et al. (1975).

$$w_t=\frac{y_t-x_{t-1}^{'}a_{t-1}}{\sqrt{1+x_t^{'}(X_{t-1}^{'}X_{t-1})^{-1}x_t}} \qquad (2)$$

By using the estimated standard deviation determined by $\hat{\sigma}^2=S_T/(T-2)$ where S_T is the residual sum of squares after fitting the model to the entire observations, $w_t/\hat{\sigma}^2$ is asymptotically normal with zero mean and unity variance.

Figure 7. The CUSUM tests for pre-crisis intervention days

Panel A: September 12, 2003 Panel B: September 30, 2003



Panel C: December 10, 2003 Panel D: January 9, 2004

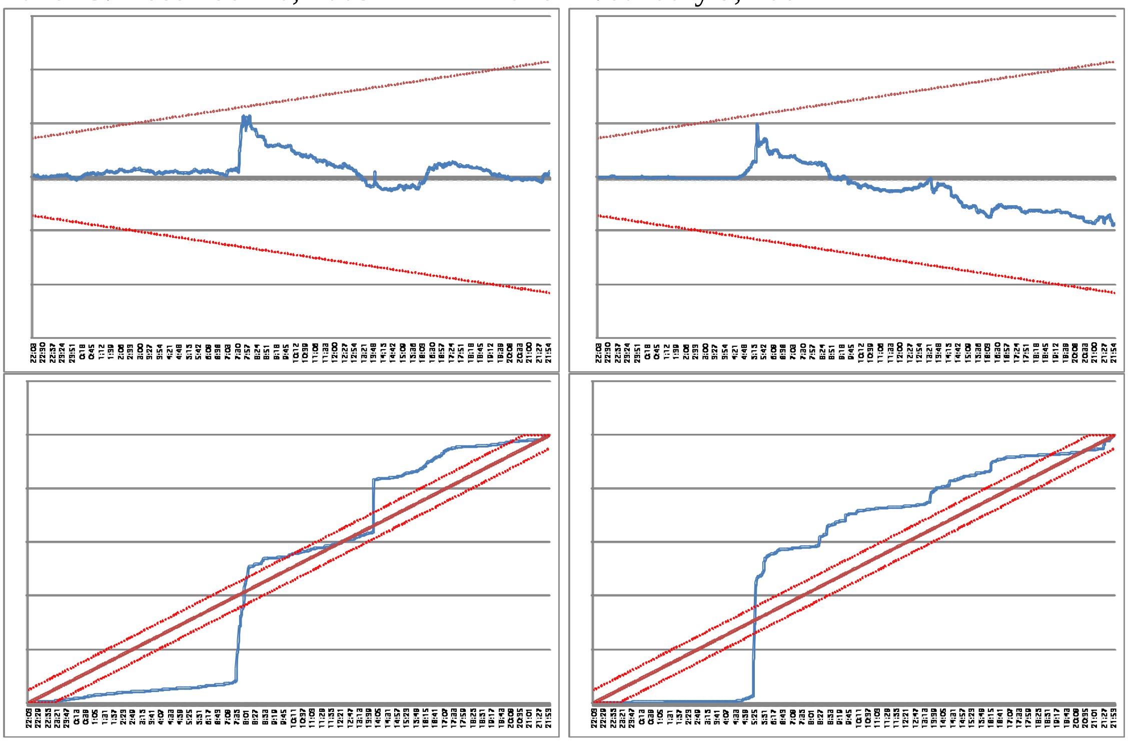
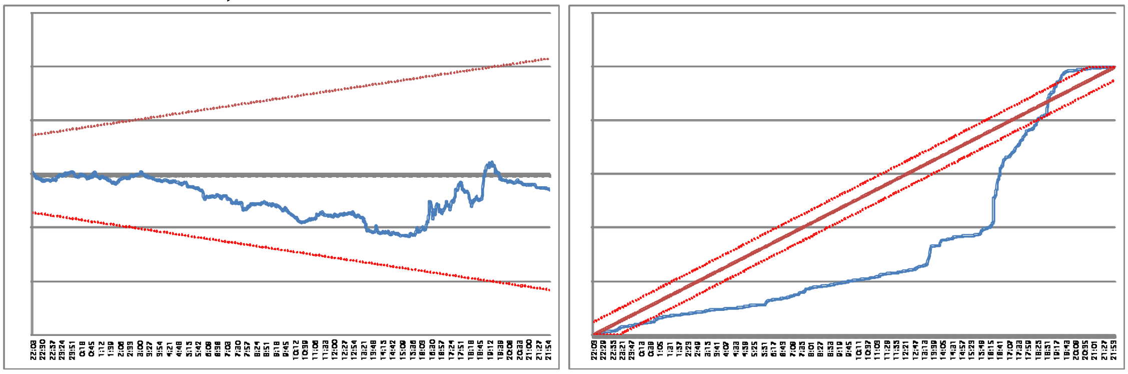


Figure 7. (Continued)
Panel E: March 5, 2004



Note: The solid line represents the accumulated sum of the forecast residuals (upper figure) and the squared forecast residuals (lower figure), and the dotted lines represent the 95 percent upper and lower bounds.

Instead of using pre-specified statistical significance level, we choose the threshold levels which vary from 0.5 to 4.0 by 0.5 steps. With these threshold values, we define binary intervention variable, $INTERV_t$, which takes the value of one when recursive residual exceeds the corresponding threshold values and zero otherwise. For an illustrative purpose, we report that the number of observations exceeding the threshold value of 4 is 14 out of 1436 observations, in September 15, 2010. In terms of percentage, this 1 percent (14/1436) exceeds far beyond the level of 0.02 percent for the standard normal distribution at four standard deviations. Note that these extreme values may originate from the well know fact that error terms in financial assets follow non-normal distribution, especially with fat tails. We see in the following sections that these extreme values are whether mere statistical outliers or meaningful indicators of intra-day interventions.

6. The effect of intra-day interventions on limit order behavior

In this section we investigate whether intervention variables, constructed in a manner of section 5-3, have significant effect on the financial behaviors of participants in the JPY/USD foreign exchange market. First, we introduce the life-time estimation model, closely following Susai and Yoshida (2012), and apply the model to six intervention days in subsection 6-1. Then we add *INTERV* variables in the life-time estimation model to see whether the BOJ intervention activities (or just extreme values of recursive residuals) may affect the life-time of limit orders in subsection 6-2.

6-1. Life-time estimation model

For each limit order i, clock times are measured at the start of order, $t^s(i)$, and at the end of order, $t^e(i)$. The volume and quote are recorded as v_i and q_i. The best bid and ask (offer) are time-varying and are $b(t)$ and $o(t)$, respectively. I_i is an indicator function, taking the value of one for bid orders and zero for offer orders. The order book is kept as the sum of the volume at the rate by each tick, by 0.01 yen, on the bid and offer sides, $bv(t, rate)$ and $ov(t, rate)$, respectively.

$$d_i^{se} = \beta_1 Vol_i + \beta_2 Gap_i + \beta_3 Depth_i + \beta_4 Calm_i + \varepsilon_{t^s(i)} \qquad (3)$$

where $d_i^{se} = t^e(i) - t^s(i)$, $d_i^{ss} = t^s(i) - t^s(i-1)$, and ε_t is independently and identically distributed. Four independent variables are defined as follows: $Vol_i = v_i$,

$Gap_i \equiv I_i \left| b(t^s(i)) - q_i \right| + (1 - I_i) \left| o(t^s(i)) - q_i \right|$,

$$Depth_i \equiv \sum_{j=b_i-0.04}^{b_i} bv(t^s(i), j) + \sum_{j=o_i}^{o_i+0.04} ov(t^s(i), j) \text{, and}$$

$$Calm_i \equiv \sum_{j=i-19}^{i} d_j^{ss} .$$

The estimation results for equation (3) are shown in Table 4. The lifetimes of all limit orders are calculated, except for the first 29 instances due to the reconstruction of the order book to recover the best bid and ask quotes. The lifetime of these limit orders are regressed on four independent variables. All estimated coefficients are statistically significant at the one percent level. First, larger individual order volumes affect the order lifetime positively. As discussed in Section 4-2 and shown in Figure 4, most limit orders are submitted at a minimum size of one million US dollars. The estimated coefficient indicates, when other effects are controlled for, that minimum volume orders leave the order book within ten seconds (9.898 seconds) on average. Second, the larger the difference between the quote and market prices, the longer a limit order stays in the order book. If a quote price is 0.01 points away from the market price, the additional lifetime of a limit order is approximately 26 seconds. Third, market depth hastens the exit of a limit order from the market. The more orders are stocked in the order book, the more rapidly a limit order disappears from the market. Note that an infinitesimal increase, i.e., 1 million US dollars, in the order book does not have a large effect. An additional 100 million US dollars shortens the lifetime of limit orders by approximately 18 seconds. Fourth, the calmness (the reciprocal of volatility) of the market allows a limit order to stay longer in the order book. At the mean value of the Calm variable (2.15 from Table 3), a limit order stays approximately 11 seconds in the order book on average.

Table 3. Descriptive statistics (September 15, 2010)

	mean	s.d.	min	max
d_i^{se} (lifetime of limit order)	41.9	526	0.001	85716
Vol (volume)	1.36	2.70	1	430
Gap (difference between the quote and market price)	0.0225	0.0696	-0.3	33.5
Depth (the volume sum in the order book)	178	86.2	4	845
Calm (previous durations for consecutive orders)	2.15	5.60	0.009	508

Note: The number of observations is 625,573, excluding the first 29 observations.

The estimation may be biased if both hit orders and canceled orders are included in the same sample because the former is more likely to represent orders with quotes close to a contemporaneous market price and thus has a shorter life time. We also investigate the empirical model of equation (3) using only canceled limit orders, and the results are shown in the third and fourth columns of Table 4.

The qualitative results are very similar to those of the sample including all limit orders. To shed light on the empirical question of whether the global financial crisis affects the behaviors of foreign exchange market, we repeat the same exercise for the five different days on which the BOJ interventions took place.

We note that both the changes in the overall market behavior due the global financial crisis and the different effects of long-forgotten interventions may cause different response patterns in the estimation results between samples recorded during the pre- and post-crisis periods.

Table 4. The lifetime of limit orders (September 15, 2010)

	all orders		canceled orders	
	coef.	s.d.	coef.	s.d.
Vol	9.898	(0.226)***	13.316	(0.325)***
Gap	2610.5	(8.83)***	2498.4	(9.04)***
Depth	-0.182	(0.0036)***	-0.217	(0.004)***
Calm	5.763	(0.108)***	5.211	(0.119)***
adj. R^2	0.13		0.13	
NOB	625,573		553,732	

Note: All orders includes realized and canceled orders. The first 29 observations are dropped to construct the best bid-ask quotes. S.d. is the standard deviation robust to heteroskedasticity. *, **, and *** denote statistical significance at the ten-, five-, and one-percent levels, respectively.

In Table 5, the qualitative results of the sample recorded before the crisis in terms of the degree of fitness of the regressions, statistical significance, and the signs of the coefficients are quite consistent with those of the sample after the crisis, except for the positive coefficient of $Depth_i$ on Apr 9, 2004.

By comparing the range of the estimated coefficients in the pre-crisis period with those in the post-crisis period, we find that the coefficients of Gap and Calm in the post-crisis period fall outside of those in the pre-crisis period. For Vol, the estimated coefficients range from 4 to 31 in the pre-crisis period, and 9.898 (all orders) and 13.316 (canceled orders) in the post-crisis period lie within the range. For Depth, the estimated coefficients range from -0.7 to 0.14, and -0.182 (all) and -0.217 (canceled) also lie within the range.

However, for Gap, the effect of quote prices deviating from the market price on the life of limit orders is approximately half of the minimum value of the range in the pre-crisis period. This implies that limit orders on Sep 15, 2010 were withdrawn from the order book much faster than during the pre-crisis period. No-execution risk and foregone opportunity cost are much higher in the market after the crisis.

Comparison of the estimated coefficients of Calm shows that a slower-paced market condition during the post-crisis period allows limit orders to enjoy positions in the order book for longer periods. In other words, the relative exiting decision time from the order book is shorter in the market during a volatile period than a calm period after the crisis.

6-2. The effect of intra-day interventions on life-time

Given the significant change in the magnitude of recursive residuals plotted as cumulated sum of residuals in Figure 6 at the same timing as the announced official intervention by the BOJ on September 15, 2010, we propose to construct an intervention proxy variable at minute intervals. This (intermediate) intervention variable takes the value of one when a recursive residual from the regression in equation (1) exceeds the threshold value and zero otherwise. We only account for large positive errors in this study because all interventions are US dollar purchase, but for a general case of interventions in both directions recursive residuals in absolute terms should be applied. The threshold values are chosen to at least cover the obvious intervention timing of around 1:30 (GMT) and not include too much portion of the entire sample.

Table 5. The lifetime of limit orders (prior to the subprime financial crisis)

all limit orders

	2003Sep12	2003Sep30	2003Dec10	2004Jan09	2004Mar05
Vol	31.28***	3.67***	5.83***	9.64***	11.73***
	(1.48)	(0.45)	(0.56)	(0.47)	(0.37)
Gap	10,895.80***	4,913.44***	7,232.77***	5,678.54***	7,855.36***
	(165.96)	(44.74)	(68.92)	(59.70)	(78.00)
Depth	-0.26***	-0.42***	-0.18***	0.09***	-0.05***
	(0.04)	(0.05)	(0.02)	(0.01)	(0.00)
Calm	1.91***	1.12***	1.01***	1.35***	0.74***
	(0.10)	(0.07)	(0.06)	(0.08)	(0.07)
Adj.R2	0.21	0.19	0.20	0.21	0.20
NOB	21,674	53,085	45,303	42,825	48,912

canceled limit orders

	2003Sep12	2003Sep30	2003Dec10	2004Jan09	2004Mar05
Vol	7.80**	3.52***	8.57***	22.42***	16.94***
	(3.15)	(0.96)	(1.07)	(0.92)	(0.66)
Gap	11,022.60***	5,142.72***	7,502.63***	5,439.98***	7,280.17***
	(224.20)	(67.28)	(92.84)	(78.09)	(102.59)
Depth	-0.35***	-0.70***	-0.30***	0.14***	-0.07***
	(0.06)	(0.10)	(0.03)	(0.02)	(0.01)
Calm	2.61***	1.34***	0.89***	0.95***	0.47***
	(0.16)	(0.12)	(0.10)	(0.13)	(0.12)
Adj.R2	0.21	0.19	0.21	0.23	0.19
NOB	10,647	25,242	26,240	23,793	28,279

Note: All limit orders include realized and canceled orders. The first 29 observations are dropped to construct the best bid-ask quotes. S.d. is the standard deviation robust to heteroskedasticity. *, **, and *** denote statistical significance at the ten-, five-, and one-percent levels, respectively.

These values are 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0. By this way, each minute interval is assigned to either one or zero. The number (the percentage) of minutes assigned the value of one is 260 minutes (18.1%), 134 (9.3%), 85 (5.9%), 49 (3.4%), 32 (2.2%), and 23 (1.6%), respectively for the threshold values of recursive residuals being 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0.

These binary classification based on minute interval is then applied to tick-base limit orders at the submission time. If a limit order is submitted in a minute interval at which the recursive residual from order flow regression exceed the threshold value, $INTERV_i$ takes the value of one and zero otherwise. The equation (4) is estimated with the tick data on September 15, 2010.

$$d_i^{se} = \beta_1 Vol_i + \beta_2 Gap_i + \beta_3 Depth_i + \beta_4 Calm_i + \beta_5 INTERV_i + \varepsilon_{t^s(i)} \qquad (4)$$

In Table 6 the estimated results of equation (4) is presented. The result of equation (3), without $INTERV_i$ variable in equation (4), is provided as specification (i) for comparison. Specification (ii) through (vii) include an intervention dummy variable, *INTERV*, which takes value of one when a recursive residual exceeds respectively the threshold value of 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0.First, *INTERV* variable is statistically significant for all specifications. The magnitude of impact, when interventions (and lingering post-intervention effect) are observed in the market, is the reduction of about 27 to 44 seconds in the life-time of limit orders. This is consistent with the result in Fong and Liu (2010) that limit order cancellations and revisions increase with the market volatility. Second, the magnitude of impact on reducing the life-time is monotonically larger when the threshold for recursive residuals becomes greater. The greater volatility in exchange rate caused by interventions affects existing orders to be canceled or revised.

Table 6. The effect of intra-day interventions on the life-time of limit order

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Vol	9.90***	10.25***	10.24***	10.23***	10.17***	10.11***	10.08***
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Gap	2,608.26***	2,620.07***	2,620.02***	2,619.11***	2,617.12***	2,615.12***	2,614.08**
	(8.83)	(8.85)	(8.84)	(8.84)	(8.84)	(8.84)	(8.84)
Depth	-0.19***	-0.15***	-0.16***	-0.17***	-0.17***	-0.18***	-0.18***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Calm	4.83***	4.77***	4.74***	4.75***	4.77***	4.78***	4.78***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
INTERV		-26.89***	-32.73***	-36.83***	-39.91***	-41.77***	-44.09***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Adj.R2	0.13	0.13	0.13	0.13	0.13	0.13	0.13
SBIC	4763180	4762950	4762930	4762940	4762980	4763020	4763040
NOB	625571	625571	625571	625571	625571	625571	625571

Note: The first 31 (not 29) observations are dropped for all specifications to have equal number of observations. The first specification is the same as in Table 4 except estimators are a little different due to the difference in the number of observations. Specification (ii) through (vii) include an intervention dummy variable, *INTERV*, which takes value of one when a recursive residual exceeds respectively the threshold value of 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0.

Third, a relatively moderate size of threshold, i.e., specification (iii) is chosen best for our model by the Schwarz BIC. For this specification, only 9.3 percent of the entire sample is designated as interventions or post-intervention effects. Finally, we note, however, the increase in terms of overall fitness of regression is only marginal.

In Table 7 the estimated results of equation (4) on the other intervention days in 2003 and 2004 are provided. For the estimations in Table 7, the comparison with 2010 intervention should be made with specification (vii) in Table 6 because the threshold value of 3 is used.

Two findings are noteworthy. First, all intervention variables are consistent with expected negative sign and statistically significant at one percent leve. Second, the most distinct feature appears between post-crisis period and pre-crisis period interventions. The difference in the magnitude of intervention impact on reducing the life-time of limit orders is much greater in the pre-crisis period. The impact of intervention is about five to ten times greater in the pre-crisis period.

Table 7. The effect of intra-day interventions (pre-crisis periods)

	2003Sep12	2003Sep30	2003Dec10	2004Jan09	2004Mar05
Vol	33.59***	4.30***	7.42***	10.48***	12.03***
Gap	11,007.18***	4,980.00***	7,461.90***	5,744.69***	7,901.93***
Depth	-0.23***	-0.22***	-0.15***	0.10***	-0.05***
Calm	1.82***	1.01***	0.95***	1.31***	0.74***
INTERV	-314.08***	-172.20***	-292.88***	-293.63***	-168.19***
Adj.R2	0.21	0.19	0.21	0.21	0.20
NOB	21672	53083	45301	42823	48910

Note: An intervention dummy variable, *INTERV*, which takes value of one when a recursive residual exceeds 3.

7. Robustness checks

In this section, we check how reliable our results are by estimating the same regressions for non-intervention days as well as using other criteria for constructing intervention variables.

7-1. Non-intervention days

So far our analysis is solely based on the days intervention actually occurred. This selection bias may bring spurious results of the effectiveness of intervention on reducing the life-time of limit orders. The reduction of the life-time may be caused by factors other than intervention activities and the analysis of non-intervention days may generate the similar results in which large recursive residuals simply reduce the life-time of limit orders without central bank involvements. For non-intervention days, we selected the following four days: September 8, 2010 (a week before the intervention); September 14 and 16, 2010 (one day before and after the intervention); and September 22, 2010 (a week after the intervention).

The estimated results are presented in Table 8. Similarly in Table 7, we chose the threshold value of 3. First, unlike intervention days, not all control/auxiliary variables are statistically significant. *Gap* and *Depth* are not statistically significant while *Vol* and *Calm* are consistent with expected sign and statistically significant. Second, the fitness of regressions on non-intervention days are much lower than intervention days. Therefore we have some evidence that intervention activities by the central bank may strengthen financial behavior relationship besides the effectiveness of intervention variable itself. Third, *INTERV* variables become insignificant for two days. This is as expected because large recursive residual is just mere statistical phenomena in non-intervention days. However, the effectiveness of *INTERV* variables on September 8, 2010 and September 14, 2010 are disturbing because these days are strictly prior to intervention activities and difficult to be explained by intervention after-effects. We return to this issue in the next subsection by using different *INTERV* variables.

Table 8. Robustness check (non-intervention days)

	2010Sep8	2010Sep14	2010Sep16	2010Sep22
Vol	66.62***	33.11***	33.75***	20.22***
Gap	0.00	0.00	0.01	0.00
Depth	-0.21***	0.02	0.03	0.09***
Calm	3.54***	3.11***	4.19***	3.39***
INTERV	-19.88***	-21.10***	17.86	-11.33
Adj.R2	0.02	0.01	0.01	0.01
NOB	307184	315216	328310	290775

Note: An intervention dummy variable, *INTERV*, which takes value of one when a recursive residual exceeds 3.

7-2. Different criteria for constructing INTERV variables

In Table 7 and 8 results only by threshold value of 3 for *INTERV* are provided to make comparison easier among different dates. Regression model (4) is estimated for all dates (both intervention days and non-intervention days) by using different threshold values for *INTERV*. P-values for *INTERV* are reported in Table 9 and different threshold values for recursive residuals are denoted by RR with corresponding two digit number.

For all intervention days, intervention binary variables, constructed by using thresholds of large recursive residuals, are statistically significant at one percent level regardless of threshold values. In contrast, for non-intervention days, the results are unstable at the best. For each threshold value, there are always two dates (out of four) in which *INTERV* variables are not statistically significant even at ten percent level. For each date, the statistical significance of *INTERV* varies widely by threshold values. We confirmed the results presented in both Table 7 and 8 remain qualitatively the same for other threshold values of *INTERV*.

One possible critic of using recursive residuals is a possible bias introduced in recursive residuals for the later observations after interventions if a structural change actually occurred in the relationship between exchange rate return and order flow. In addition to using recursive residual method in which estimation sample extends for later observations, we also used forecast errors with rolling windows of sample periods. Different size of windows (60, 120, 180, and 240 minutes) and different size of threshold values (2.0, 2.5, 3.0, 3.5, and 4.0) are used to construct corresponding *INTERV* variables. These results are also shown in Table 9 for each criteria denoted by FE plus corresponding windows (in terms of hours) and two digits indicating threshold values.

The comparison between results in intervention days and non-intervention days is quite contrast that statistical significance of *INTERV* seems at random at the best for non-intervention days whereas *INTERV* variables are statistically significant at one percent level regardless of window size and threshold values. With these robustness checks, this study find the strong supporting evidence that intra-day interventions significantly affect the limit order behaviors of financial institutions in the foreign exchange markets.

Table 9. Robustness check (different criteria): p-values for *INTERV* variable

Large error	Intervention Days						Non-intervention Days			
criteria	2003Sep12	2003Sep30	2003Dec10	2004Jan09	2004Mar05	2010Sep15	2010Sep8	2010Sep14	2010Sep16	2010Sep22
RR05	0.000	0.000	0.000	0.000	0.000	0.010	0.738	0.000	0.003	0.670
RR10	0.000	0.000	0.000	0.000	0.000	0.004	0.225	0.328	0.000	0.210
RR15	0.000	0.000	0.000	0.000	0.000	0.002	0.008	0.122	0.000	0.759
RR20	0.000	0.000	0.000	0.000	0.000	0.001	0.016	0.832	0.000	0.109
RR25	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.384	0.095	0.269
RR30	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.111	0.498
RR35	0.000	0.000	0.000	0.000	0.000	0.000	0.163	0.008	0.602	0.000
RR40	0.000	0.000	0.000	0.000	0.000	0.003	0.198	0.002	0.594	0.000
FE1_20	0.000	0.000	0.000	0.000	0.000	0.002	0.615	0.001	0.000	0.024
FE1_25	0.000	0.000	0.000	0.000	0.000	0.000	0.506	0.704	0.000	0.033
FE1_30	0.000	0.000	0.000	0.000	0.000	0.000	0.295	0.518	0.002	0.003
FE1_35	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.631	0.002	0.023
FE1_40	0.000	0.000	0.000	0.000	0.000	0.000	0.103	0.574	0.000	0.028
FE2_20	0.000	0.000	0.000	0.000	0.000	0.001	0.205	0.076	0.185	0.273
FE2_25	0.000	0.000	0.000	0.000	0.000	0.001	0.155	0.288	0.361	0.391
FE2_30	0.000	0.000	0.000	0.000	0.000	0.000	0.258	0.922	0.527	0.078
FE2_35	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.831	0.473	0.234
FE2_40	0.000	0.000	0.000	0.000	0.000	0.000	0.099	0.122	0.070	0.234
FE3_20	0.000	0.000	0.000	0.000	0.000	0.000	0.133	0.071	0.507	0.734
FE3_25	0.000	0.000	0.000	0.000	0.000	0.000	0.208	0.880	0.312	0.460
FE3_30	0.000	0.000	0.000	0.000	0.000	0.000	0.721	0.037	0.546	0.062
FE3_35	0.000	0.000	0.000	0.000	0.000	0.000	0.168	0.026	0.188	0.411
FE3_40	0.000	0.000	0.000	0.000	0.000	0.000	0.171	0.006	0.874	0.000
FE4_20	0.000	0.000	0.000	0.000	0.000	0.001	0.011	0.253	0.049	0.238
FE4_25	0.000	0.000	0.000	0.000	0.000	0.000	0.210	0.521	0.034	0.496
FE4_30	0.000	0.000	0.000	0.000	0.000	0.000	0.150	0.033	0.036	0.785
FE4_35	0.000	0.000	0.000	0.000	0.000	0.001	0.313	0.000	0.333	0.730
FE4_40	0.000	0.000	0.000	0.000	0.000	0.000	0.313	0.000	0.897	0.000
over 10%	0	0	0	0	0	0	20	15	14	17

Note: Recursive residual is used for RR variables and forecast errors with rolling windows for FE variables. The figures are p-values for *INTERV* variable in life-time equations.

8. Conclusions

In this paper, we investigate the impact of the BOJ intervention on September 15, 2010 on trading activities on the yen/dollar market of the EBS. The BOJ had refrained from interventions for more than six years and the world is hit severely by the global financial crisis during this period. We also investigate five intervention days that occurred in 2003 and 2004.

Given the major role of limit orders on the EBS, it is imperative to investigate the effect of limit orders on the foreign exchange market. Contrary to previous studies, the net order flow in this study is constructed from limit orders rather than from market orders, which are observationally equivalent to transaction data. The main contributions of the investigation on the relationship between order flow and exchange rate are the following two points. First, we find that the order flow of limit orders has a positive impact on the exchange rate, i.e., an excess of bid over offer orders appreciate the value of the US dollar against the Japanese yen. This is consistent with many previous studies which use the market order (transaction) definition of order flow. Second, we find that the relationship between order flow and market return on the dollar/yen exchange market experiences a break down following the unexpected and very high volume of offer/sell orders following the BOJ interventions. The recursive residuals detect the timing of the BOJ interventions with striking clarity. A simple methodology is proposed to detect the exact timing of interventions. We propose to construct an intervention proxy variable which takes the value of one when a recursive residual from the order-flow-exchange-rate regression exceeds the threshold value.

Using the EBS data provided by the ICAP, we are able to track the termination of limit orders (by either transaction or cancellation) and we measured the lifetime of limit orders. We find interventions, detected by the proposed methodology, significantly reduce the life-time of limit order in the market by about 27 to 44 seconds while controlling for the volume, the gap between quote and market price, the slower pace of the market, and large outstanding orders in the order book.

As robustness checks, we also applied the same methodology to non-intervention days and could not find the similar relationship between large recursive residuals (from exchange rate return regressions) and the life-time of limit orders. These results stand robust to regardless of various ways to construct intervention variables.

References

Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. Journal of Finance, 50(5), 1655-1689.

Bloomfield, R., O'Hara, M., Saar, G., 2005. The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. Journal of Financial Economics, 75, 165-199.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society, Series B, 37(2), 149-192.

Cerrato, M., Sarantis, N., Saunders, A., 2011. An investigation of customer order flow in the foreign exchange market. Journal of Banking & Finance, 35, 1892-1906.

Chari, A. 2007. Heterogeneous market-making in foreign exchange markets: Evidence from individual bank response to central bank interventions, Journal of Money, Credit and Banking, 39(5), 1131-1162.

Cheung, Y.W., Chinn, M.D., 2001. Currency traders and exchange rate dynamics: A survey of the US market. Journal of International Money and Finance, 20, 439-471.

Cohen, K.J., Maier, S.F., Schwartz, R.A., and Whitcomb, D.K., 1981. Transaction costs, order placement strategy, and existence of the bid-ask spread. Journal of Political Economy, 89(2), 287-305.

Corwin, S. A., Lipson, M. L., 2011. Order characteristics and the sources of commonality in prices and liquidity. Journal of Financial Markets, 14(1), 47-81.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics, 19, 69-90.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance, 47, 577-606.

Evans, M.D.D., Lyons, R.K., 2002a. Information integration and FX trading. Journal of International Money and Finance, 21, 807-831.

Evans, M.D.D., Lyons, R.K., 2002b. Order flow and exchange rate dynamics. Journal of Political Economy, 110(1), 170-180.

Fischer, A.M., Zurlinden, M., 1999. Exchange rate effects of central bank interventions: An analysis of transaction prices, The Economic Journal, 109, 662-676.

Fischer, A.M., 2006. On the inadequacy of newswire reports for empirical research on foreign exchange interventions. Journal of International Money and Finance, 25, 1226-1240.

Fong, K.Y.L., Liu, W.M., 2010. Limit order revisions. Journal of Banking and Finance, 34, 1873-1885.

Foucault, T., 1999. Order flow composition and trading costs in a dynamic limit order market. Journal of Financial Markets, 2, 99-134.

Foucault, T., Kadan, O., Kandel, E., 2005. Limit order book as a market for liquidity. Review of Financial Studies, 18(4), 1171-1217.

Foucault, T., Moinas, S., Theissen, E., 2007. Does anonymity matter in electronic limit order markets? The Review of Financial Studies, 20(5), 1707-1747.

Frömmel, M., Kiss M, N., Pintér, K., 2011. Macroeconomic announcements, communication and order flow on the Hungarian foreign exchange market. International Journal of Finance and Economics, 16, 172-188.

Harris, L., Hasbouck, J., 1996. Market vs. limit orders: The superDOT evidence on order submission strategy. Journal of financial and quantitative analysis, 31(2), 213-231.

Hasbrouck, J., Saar, G., 2002. Limit orders and volatility in a hybrid market: The Island ECN. Working paper, New York University.

Hollifield, B., Miller, R.A., Sandås, P., 2004. Empirical analysis of limit order markets, Review of Economic Studies, 71, 1027-1063.

Jones, C.M., Kaul, G., Lipson, M.L., 1994. Transactions, volume, and volatility. Review of Financial Studies, 7(4), 631-651.

King, M., Sarno, L., Sojli, E., 2010. Timing exchange rates using order flow: The case of the Loonie. Journal of Banking & Finance, 34, 2917-2928.

Klein, M.W., 1993, The accuracy of reports of foreign exchange intervention. Journal of International Money and Finance, 12, 644-653.

Lyons, R.K., 1997. A simultaneous trade model of the foreign exchange hot potato. Journal of International Economics, 42, 275-298.

Manganelli, S., 2005. Duration, volume and volatility impact of trades. Journal of Financial Markets, 8(4), 377-399.

Marsh, I.W., 2011. Order flow and central bank intervention: An empirical analysis of recent Bank of Japan actions in the foreign exchange market. Journal of International Money and Finance, 30, 377-392.

Melvin, M., Menkhoff, L., Schmeling, M., 2009. Exchange rate management in emerging markets: Intervention via an electronic limit order book. Journal of International Economics, 79, 54-63.

Menkhoff, L., 2010. High-frequency analysis of foreign exchange interventions: what do we learn? Journal of Economic Surveys, 24(1), 85-112.

Osterberg, W.P., Wetmore Humes, R., 1993. The inaccuracy of newspaper reports of U.S. foreign exchange intervention. Federal Reserve Bank of Cleveland, Economic Review, 29(4), 25-34.

Pasquariello, P., 2007. Informative trading or just costly noise? An analysis of Central Bank interventions. Journal of Financial Markets, 10(2), 107-143.

Payne, R., Vitale, P., 2003. A transaction level study of the effects of central bank intervention on exchange rates, Journal of International Economics, 61, 331-352.

Rime, D., Sarno, L., Sojli, E., 2010. Exchange rate forecasting, order flow and macroeconomic information. Journal of International Economics, 80, 72-88.

Scalia, A., 2008. Is foreign exchange intervention effective? Some microanalysis evidence from the Czech Republic. Journal of International Money and Finance, 27, 529-546.

Susai, M., Yoshida, Y., 2012. Life-time of limit orders in the EBS foreign exchange spot market. Mimeo.

Vitale, P., 2011. The impact of FX intervention on FX markets: A market microstructure analysis. International Journal of Finance and Economics, 16, 41-62.

Yeo, W.Y., 2005. Cancellations of limit order. Working paper, National University of Singapore.

Corporate Derivatives Use, Leverage, and the Cost of Equity: New Insights from Indian Non-financial Firms

Praveen Bhagawan. M, Institute for Financial Management and Research, Sri City, India

Chinmoy Ghosh, University of Connecticut, USA

Jijo Lukose. P. J, Indian Institute of Management, Kozhikode, India

Shankar. R. L, Great Lakes Institute of Management, Chennai, India

ABSTRACT

This paper investigates the relationship between a firm's derivatives use, leverage, and the cost of equity of S&P CNX 500 non-financial constituents in a simultaneous equations framework. Using a hand-collected data on firm's currency derivatives, we find no evidence of reduction in firm's cost of equity due to the usage of currency derivatives. This result is plausible since the firm's currency risk in relation to total risk is negligible as argued by Copeland and Joshi (1996). Further, we find a positive association between firm's leverage and its decision to hedge currency risk, suggesting that firms with higher financial distress costs are more likely to hedge. Finally, we do not find evidence of increase in leverage ratio due to usage of currency derivatives. These results are consistent with the theory and also remain robust even after controlling for potential endogeneity using various econometric techniques.

JEL Classification: G3, G32.
Key words: Corporate risk management; Currency derivatives; Cost of Equity

1. Introduction

Risk management is considered one of the important financial policies for firms (Rawls and Smithson, 1990).[93] Firms' financial policies are often dependent on the macroeconomic environment in which firms operate. For example, during the financial crisis in 2008, the US, the UK, and Europe followed quantitative easing policy, which resulted in huge variation in foreign currency inflows and outflows in Asian emerging countries which, in turn, increased volatilities in equity and currency markets in these countries (Farhi and Borghi, 2009). Consequently, firms' revenue and cash flows were adversely affected. The severity of the problem triggered heavy use of derivatives by firms to protect themselves from the ensuing future currency and equity risks. Indeed, a 2009 survey conducted by the International Swaps and Derivatives Association Derivatives (ISDA) reports that 94 per cent of the Fortune Global 500 firms use derivatives to manage their exposure to fluctuations in exchange rate, interest rate and value of equity.[94]

Our objective in this paper is to investigate the relationship between hedging, financing, and the cost of equity for Indian non-financial firms. Our research design, especially the focus on India, constitutes a significant extension to the existing literature. Most of the studies on the impact of hedging on firm value and efficient use of data from the U.S. or other developed economies where the regulatory, legal and institutional structure are well established and the derivatives market has been in existence for years, and managers and investors are fully cognizant of the advantages and risks associated with these complex instruments. In comparison, the derivatives market in India is relatively young and the regulation relating to the use of these instruments is still evolving. The combination of the financial collapse, the fast growing capital market in India, and the relatively recent start of the use of derivatives provides us an unique opportunity to provide insight on issues that are unique to emerging economies.

Apart from our focus on one of the fastest growing economies, our analyses make at least three important contributions to the extant literature. First, empirical studies on the relationship between the firms' usage of derivatives and cost of equity employ either survey method (Ameer et al. 2011) or electronic databases (Gay et al. 2011) to collect the data on currency derivatives. In this paper, we hand-collect the data on currency derivatives from firms' annual reports to mitigate the problem of non-availability of data on currency derivatives in electronic databases for an emerging country like India. As such, our dataset is more comprehensive. Second, as noted in the related literature, often a firm's decision to hedge, and how much to hedge, while controlling for leverage, is potentially endogenous (Graham and Rogers (2002); Gay et al. (2011)). The notable studies that directly examine the relationship between firms' usage of derivatives and cost of equity include Ameer et al. (2011) and Gay et al. (2011). Ameer et al. (2011) do not address the endogeneity problem. Gay et al. (2011) use Instrumental Variable Two-Stage Least Squares (IV2SLS) regression to address the endogeneity that arises due to correlated omitted variable bias (COVB). However, they do not consider a firm's decision to hedge as a linear predictor while estimating firm's cost of equity in the second-stage regression using system approach.[95] Similar to Gay et al. (2011), we employ IV2SLS regression to address the endogeneity problem. However, we employ linear regression for firm's decision to hedge

[93] Rawls and Smithson (1990) review the survey studies of Davis (1989) and Millar (1989) and document the importance of risk management for firms.
[94] http://www.isda.org/press/press042309der.pdf
[95] Angrist and Kruger (2001) suggest to use linear regression in the first of IV2SLS even if the endogenous variable is binary, which gives consistent estimates in the second stage regression.

in the first-stage, which provides consistent estimates in the second-stage regression (Angrist and Krugeur, 2001). Finally, to correct for self-selection bias in the hedging decision, we employ treatment-effects regression. Neither of the two studies addressed this type of endogeneity.

Our empirical analyses develop in several steps. First, we examine whether currency hedging through the usage of derivatives helps firms to reduce their cost of equity. Our analyses reveal that compared to firms that do not use derivatives, firms that use derivatives have lower cost of equity, but the difference is not significant. Specifically, using the Fama-French three-factor model, we find that the market beta, SMB beta, and HML beta are not statistically different from zero for hedgers or non-hedgers, which implies that hedging does not benefit firms in reducing their cost of equity. Our finding also resonates with Copeland and Joshi (1998) who argued that the contribution of foreign currency risk in a firm's total risk is not significant. Hence, hedging currency risk should not result in reduction in firm's cost of equity. Our results are robust while controlling for endogeneity of a firm's derivatives use and leverage.

Next, we examine the relationship between a firm's financing policy, measured by leverage ratio, and its hedging policy in a simultaneous equations framework. We find a positive association between the two policies, which implies that firms with higher financial distress costs are more likely to hedge currency risk. This result is justifiable since the leverage ratio of Indian firms are higher than their counterparts in developed countries aftermath of global financial crisis as argued by Guimaraes-Filho et al. (2014). Therefore, firms in India are more likely to hedge their exchange rate exposure. We further find that firms with lower liquidity and higher foreign exchange exposure are more likely to hedge. These results are robust to controlling for endogeneity. Finally, we investigate the impact of a firm's foreign currency hedging policy on its debt ratio in a simultaneous equation setting. We find no evidence of increase in debt ratio of firms that use currency derivatives or hedging. These results are similar in spirit to the findings of Geczy et al. (1997), and Graham and Rogers (2002). We further examine whether our results are sensitive to an alternative measure of hedging, specifically the extent of hedging. We find a positive relationship between a firm's decision on how much to hedge and its leverage. However, it loses its significance after controlling for endogeneity. Overall, the results indicate that firm's hedging policy does not have any impact on its capital structure.

The remainder of the paper is organized as follows. Section 2 reviews the literature that examines the relationship between a firm's usage of derivatives, leverage, and its cost of equity. Section 3 and 4 describe data sources and the methods to estimate system of simultaneous equations. Section 5 reports econometric techniques applied in our work. Section 6 presents the empirical results. Section 7 discusses the robustness checks. Section 8 forms the summary and conclusion.

2. Review of Literature

2.1 Hedging and the Cost of Capital

There are survey based studies that document reduction in funding costs by firms as the one of the main reasons for using derivatives (Bodnar et al, 2013). Bodnar et al. (2013) examine 464 Italian firms that have sales revenue of at least 25 million Euros for 2007-08 and document that reduction in funding costs as one of important reasons why their sample firms prefer to use derivatives.

Coutinho et al. (2012) empirically examine the relationship between a firm's usage of currency derivatives and the weighted average of cost of capital (WACC). They document positive relationship between a firm's foreign currency derivatives and its WACC for 47 large Brazilian firms for the period between 2004 and 2010. Further, they examine whether this relationship has changed after 2008 financial crisis by allowing a separate dummy variable to differentiate the data before and after financial crisis 2008, and find this dummy variable to be negative and statistically significant; suggesting that the usage of derivatives helps firms to reduce their cost of capital post 2008 financial crisis.

2.2 Hedging and the cost of equity

In the existing literature, there are three contrasting theories that explain the relationship between a firm's usage of derivatives and the cost of equity. First, the usage of derivatives can attenuate financial risks, and hence can result in lower cost of equity[96]. Gay et al. (2011) argue that hedging helps firms to reduce their financial distress costs thereby reduces cost of equity. This relationship becomes more robust in an emerging country like India due to regulatory changes in Reserve Bank of India (RBI) guidelines. A recently issued RBI guideline recommends banks to maintain higher provisions if they lend to non-hedged firms[97]. This results in increase in cost of bank borrowing for non-hedged firms. Since these firms' future cash flows are discounted at a higher rate, and the duration of equity is higher than bank borrowing, equity shareholders would expect higher returns from non-hedged firms than hedged firms.

Second, the use of derivatives by firms might entail greater complexities in the analysis of financial statements. This might prompt analysts to devalue the stock price which may lead to increase in the cost of capital so much so that it outweighs perceived benefits of hedging[98]. Third, the usage of derivatives by firms would have no impact on its cost of equity based on Modigliani and Miller argument. This is justifiable based on the proposition of Modigliani and Miller (1958) that hedging by firms is irrelevant if these firms operate in an environment where markets are perfect with no taxes, no transaction costs, and fixed investment policies[99].

2.3 Financing and the Hedging

While reviewing the papers that examine the effect of the usage of derivatives on shareholders' value, Aretz and Bartram (2010) interpret the findings of Morellec and Smith (2007) that firms would benefit more if they negotiate the debt contracts and hedging strategies jointly with their lenders. This reduces information asymmetries between managers and debt holders. As a result, firms may get loans at a lower rate. Consistent with aforementioned argument, firms also prefer to have credit lines and hedge their exposure to exchange rates with the same bank. This helps

[96] See the post titled "Should Companies Hedge Currency Risk?" published in Knowledge@Wharton on Jun 12, 2013 available at http://knowledge.wharton.upenn.edu/article/should-companies-hedge-currency-risk/ accessed on Aug 5th 2014)

[97] Srinivas, RBI proposes extra norms for Unhedged forex exposure for firms, THE HINDU, July 3, 2013 available at http://www.thehindu.com/business/Industry/rbi-proposes-extra-norms-for-unhedged-forex-exposure-of-firms/article4874070.ece; RBI moots extra provisioning for unhedged forex exposure, Business Standard, July 3, 2013 available at http://www.business-standard.com/article/finance/rbi-moots-extra-provisioning-for-unhedged-forex-exposure-113070300039_1.html

[98] See the post titled "Should Companies Hedge Currency Risk?" published in Knowledge@Wharton on Jun 12, 2013 available at http://knowledge.wharton.upenn.edu/article/should-companies-hedge-currency-risk/ accessed on Aug 5th 2014)

[99] On the contrary, DeMarzo and Duffie (1995) argue that since managers tend to have more information than the investors, and hence the former can take better decisions than the latter related to risk management thereby increases firm value.

these firms to borrow in more favourable terms. Therefore, we argue that there is an endogenous relationship between a firm's debt ratio and its hedging decision. Apparently, there are a few notable studies that empirically examine the relationship between a firm's leverage and its hedging decision in a simultaneous equations framework (Geczy et al. (2007); Graham and Rogers (2002); Judge (2006))[100]. To examine this relationship is more robust in an emerging country like India. Guimaraes-Filho et al. (2014) find the evidence that leverage ratio in emerging Asian firms, including India, has significantly increased aftermath of global financial crisis. However, this ratio has declined or remains stable in advanced countries. Further, they argue that the most leveraged firms in Asia, including India, not only have lower profitability but also suffer from both lower liquidity and lower solvency. For example, in case of India about 20 per cent of corporate debt is owed by firms with an interest coverage ratio (ICR) less than one. They also find that firms that have lower ICR tend to have more debt in their capital structures. This result implies that firms in these countries are more likely to have higher financial distress costs as compared to their counterparts in developed countries. Therefore, we investigate the relationship between a firm's leverage ratio and its hedging policy in this paper.

2.4 Hedging and the leverage

Theories in finance are usually unclear about whether hedging helps firms to increase their debt capacity. Two related strands of literature exist in this regard. The first strand of literature suggests that hedging may not help firms in increasing their debt capacity based on Modigliani and Miller (1958) argument that markets are perfect.

A second strand of literature advocates that risk management through the use of derivatives would help firms to increase their debt capacity (Graham and Rogers (2002); Gay et al. (2011)). Graham and Rogers (2002) examine the relationship between a firm's debt ratio and its usage of derivatives in a simultaneous equations framework and find the evidence of increase in leverage ratio due to the usage of derivatives. This result is comparable to the findings of Gay et al (2011). Dionne and Triki (2013) also document a positive relationship between a firm's hedging and its leverage ratio. However, this significant effect disappears after controlling for endogeneity. They further argue that controlling for endogeneity is essential while examining the relationship between a firm's hedging and its leverage ratio.

Overall, to our knowledge, no study has examined the interactions between a firm's cost of equity, financing decisions and hedging decision in an emerging country like India. Our goal is to fill this gap in the literature.

3. Data and Summary Statistics

For our empirical analysis, we use S&P CNX 500 firms in the year 2009. During the middle of the financial year 2009-2010 (September 30, 2009) these firms represent 92.57 per cent of total market capitalization and 91.17 per cent of the total turnover on the National Stock Exchange of India Limited[101]. We exclude banking and financial services firms as they use some or all of

[100] Geczy et al. (1997) examine the relationship between a firm's capital structure and its hedging decision simulataneously in the context of the U.S. and find no statistical relationship between the two decisions. This result implies that a firm's decision on capital structure and its hedging decision can be determined independently. A corroborating evidence is provided by Judge (2006) in the context of U.K. On the contrary, Graham and Rogers (2002) find the evidence that firms with higher debt ratios are more likely to hedge since the former is more likely to have higher financial distress costs.

[101] Retrieved from www.nseindia.com. The figures are collected from the National Stock Exchange of India Limited.

their derivatives for trading and not for hedging. For the 437 non-financial firms, we hand-collect data on financial derivatives from annual reports. Our final sample for analysis consists of 332 non-financial firms which report their usage and/or non-usage of derivatives in their annual reports for 2009.

We obtain the data from five sources. Firms' share prices and firm specific accounting variables are obtained from Centre for Monitoring Indian Economy's (CMIE) PROWESS database. The data on foreign currency derivatives is manually collected from firms' annual reports for 2009. The Electronic copies of these annual reports are obtained from Capitaline database, maintained by Capital Market Publishers India Limited. We search the annual reports with the key words such as derivative, forward, option, call, put, swap, hedge, foreign, and currency to identify the information on currency derivatives. The list of search terms are finalized by manually analyzing a subsample of 50 annual reports. Then, we shortlist the expressions that help us to find the information on currency derivatives. A similar method has been employed by other researchers examining the information on derivatives (i.e. Bartram et al., (2009); Lievenbruck and Schmid (2014)). A firm is considered a currency derivative user (hedger) if its annual report mentions the usage of derivatives to hedge its foreign exchange exposure. A firm is considered a currency derivative non-user (non-hedger) if its annual report clearly states that it does not use currency derivatives to hedge its foreign exchange exposure and/or discloses only unhedged details on foreign exchange exposure. The information on a firm's notional amount of outstanding derivative instruments is collected from annual reports.

The notional amounts of exposure expressed in terms of foreign currency are converted into Indian Rupee by using the exchange rate from Thomson Reuters' database. We restrict our analysis to those firms which report hedging details on export receivables, import payables, and foreign currency loans in their annual reports. We winsorize all accounting ratios used in our analysis at one and ninety-nine percentile levels respectively to eliminate potential data errors or outliers. Finally, the data on Fama and French three-factor returns for Indian Market are obtained from Data Library maintained by Indian Institute of Management Ahmedabad (IIMA)[102].

3.1 Summary Statistics

Table 1 provides the definitions of the variables used in examining the relationship between a firm's use of derivatives and its cost of equity. Table 2 reports the summary statistics for firm characteristics of all (332) sample firms, hedged firms, and non-hedged firms . The last two columns in this table also report the difference in means and the difference in medians between hedged firms and non-hedged firms. The difference in means and difference in medians are tested using t-test with unequal variance and Mann Whitney Wilcoxon test respectively. We find that 84.3 per cent of our sample firms report that they use derivatives to hedge their currency exposure. Anand and Kaushik (2008) also report that 83.6 per cent of their sample firms use derivatives to hedge exchange rate risk. When we examine the firm's extent of hedging, as measured by the amount of currency derivatives used by firms scaled by total assets, we find the mean (median) for the pooled sample across categories is 0.153 (0.050). We also find that the difference in mean and median of extent of hedging for hedgers and non-hedgers is highly significant at one percent of significance.

[102] http://www.iimahd.ernet.in/~jrvarma/Indian-Fama-French-Momentum/ . See Agarwalla et al. (2013) for more details.

Table 1: Definition of variables for equations relating to firm's cost of equity, the usage of derivatives by firms, and the leverage ratio.

Variables	Definition of the variables
HEDDUM	A dummy variable, which is equal to one if the firm reports the use of foreign currency derivatives to hedge their exchange rate exposure, and zero otherwise.
EX_HED	The amount of currency derivatives used by the firm scaled by total assets.
LEV	Ratio between long term debt and the sum of long term debt and net worth.
LOG_TA	Natural logarithm of total assets
BP	Book value of equity to closing market price of equity ratio
LOG_SH.TR	Natural logarithm of share trading volume in National Stock Exchange as of 31st March, 2009.
QR	Ratio between current assets minus inventory and current liabilities
FR	Ratio of foreign revenue to total revenue
DEPN_TA	Depreciation scaled by total assets
FA_TA	Ratio of fixed assets to total assets.
ROA	Ratio of PBDITA to average total assets
SGA_N.SALES	Ratio of the sum of selling, general, and administration expenses scaled to net sales.

The difference in leverage ratios for hedgers and non-hedgers is also significant, which supports the hypothesis that financially distress firms, as measured by leverage, prefer to use more derivatives to hedge their currency risk. The difference in means of book-to-market ratios between the two groups is statistically insignificant, while the difference in medians between the two groups is statistically significant; suggesting that firms with higher book-to-market ratio prefer to hedge more.

It is evident from the data that the mean size of firms that our sample contains more larger firms than smaller firms. However, there is no significant difference in mean size between hedged and non-hedged group, while hedgers are significantly larger in size than non-hedgers when we consider the median values. The hedged firms maintain lower liquidity than non-hedged firms; which implies that maintaining higher liquidity is a substitute for firms' decision to hedge.

Table 2: Mean and median for All firms, Hedged firms and Non-hedged firms

Table 2 presents the values of mean and median for All firms (332), Hedged firms (N = 280) and Non-hedged firms (N = 52). We winsorize all the variables (except the natural logarithm of shares traded, the natural logarithm of book value of total assets, and firm's hedge dummy) at one and 99 percentiles to eliminate some apparent data errors or the impact of outliers. The last two columns present the difference in mean and median values between hedged and non-hedged firms. The difference in means is tested by the t-test with unequal variances. The difference in medians is tested by Wilcoxon rank sum (Mann Whitney) test. The details of each variable are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level.

	All firms		Hedged firms		Non-hedged firms		Difference (Non-hedged firms-Hedged firms)	
	Mean	Median	Mean	Median	Mean	Median	In means	In medians
HED_DUM	0.843	1.000						
EX_HED	0.153	0.050	0.193	0.112	0.000	0.000	-0.193***	-0.112***
LEV	0.467	0.491	0.480	0.511	0.398	0.417	-0.082**	-0.094**
BP	1.178	0.881	1.201	0.898	1.049	0.666	-0.152	-0.233**
LOG_SH.TR	11.395	11.608	11.497	11.612	10.828	11.117	-0.669	-0.496
LOGTA	7.789	7.706	7.835	7.787	7.537	7.133	-0.299	-0.654*
QR	1.853	1.187	1.553	1.133	3.469	1.647	1.916**	0.514***
FR_TR	0.245	0.135	0.265	0.149	0.123	0.047	-0.142***	-0.102***
DEPN_TA	0.027	0.023	0.027	0.024	0.025	0.023	-0.002	-0.001
FA_TA	0.464	0.458	0.479	0.471	0.386	0.309	-0.093**	-0.162**
ROA	0.173	0.155	0.171	0.156	0.184	0.146	0.013	-0.010
SGA_N.SALES	0.111	0.091	0.108	0.091	0.126	0.091	0.017	-0.001

The mean and median values of foreign revenue to total revenue ratio imply that larger firms tend to have higher foreign revenue relative to its total revenue. The ratio between foreign revenue and total revenue is higher for hedged firms as compared to non-hedged firms; suggesting that geographically diversified firms prefer to hedge more. The mean (median) value of hedged firms compared to non-hedged firms suggests that hedged firms have higher tangible assets than non-hedged firms. There are no significant differences between hedged and non-hedgers with respect to other firm characteristics.

4. Computing a Firm's Cost of Equity

We estimate the cost of equity with the Fama-French three-factor model which takes into account market beta, Small Minus Big (SMB) beta, and High Minus Low (HML) beta. The sensitivity of these factor betas are computed using daily return data for the period between 1st April 2008 and 31st March 2009. Using daily returns data is appropriate since a firm's usage of derivative is unlikely to change in the short run, although a firm may switch from being a derivative user to a non-user and vice-versa over a longer period of time. The coefficients that are required to estimate the cost of equity for each firm is obtained from the following time-series regression model:

$$R_{i,t} - R_{f,t} = a_i + \sum_{k=-1}^{k=1} b_{i,k} \left(R_{m,t+k} - R_{f,t+k}\right) + \sum_{k=-1}^{k=1} s_{i,k}\, SMB_{t+k} + \sum_{k=-1}^{k=1} h_{i,k}\, HML_{t+k} + e_{i,t} \qquad (1)$$

$R_{i,t}$ is the return of firm *i* in period *t*; $R_{f,t}$is the return on the 91-day Treasury bill in period *t*; $R_{m.t}$ is the return in period *t* of the value-weighted portfolio of all listed firms in BSE that are covered in CMIE Prowess database; SMB is the difference in returns between small and large stock portfolios; and HML is the difference in returns between high and low book-to-market portfolios. We account for infrequent trading using contemporaneous daily returns for each factor, as well as one lag and one lead daily returns for each factor (Dimson (1979); Fowler and Rorke (1983); Gay et al. (2011)). As suggested by Gay et al. (2011), we get market, SMB, and HML beta by adding the coefficient estimates of the contemporaneous, lead, and lagged values of the corresponding risk premiums.

Follwing D'Mello and Shroff (2000), we estimate the cost of equity for each firm using eqn. (2) below:

$$COE_{i,t} = E\left(R_{i,t}\right) = R_f + B_{i,t}\left[E(R_m) - R_f\right] + S_{i,t}E(SMB) + H_{i,t}E(HML) \qquad (2)$$

$COE_{i,t}$is the cost of equity of firm *i* in period *t*; $B_{i,t}$, $S_{i,t}$, and $H_{i,t}$ are the estimated market beta, SMB size beta, and HML growth betas from equation (1), respectively. R_f is 91 day Treasury Bill as on 31st March 2009. $E(R_m)$ is the expected market return. $E(R_m)$, SMB, and HML are the arithmetic average daily factor returns computed for the period between 4th January 1993 and 31st March 2009, since the Fama and French three-factor returns for India are available only from 4th January 1993[103]. As in (Gay et al, 2011), we calculate the annualized the cost of equity for each firm by multiplying the cost of equity estimated from equation (2) by 252 days.

It is worth noting that in a similar study, Coutinho et al. (2012) find the cost of equity to be less than the cost of debt for a few firms, and for these firms they consider the cost of debt for the cost of equity to ensure that the cost of equity is not lower than the cost of debt. In the absence of an active bond market in India, computation of the cost of debt is an arduous exercise. Hence, we consider a firm's cost of equity as the maximum of the firm's cost of equity and the risk free rate of interest to ensure that a firm's cost of equity is never lower than the risk free rate.

Table 3 presents the differences in mean and median values of three factors, namely market beta, SMB beta, and HML betas from the Fama and French (1993) model for hedgers and non-hedgers. We also report the differences in mean and median values of firm's cost of equity between these groups of firms. The difference between the mean (median) values of market beta for non-hedgers and hedgers is -0.098 (-0.109), which are not statistically significant at conventional levels. In general, these results are comparable to those of Gay et al. (2011),

[103] http://www.iimahd.ernet.in/~jrvarma/Indian-Fama-French-Momentum/

although they find some evidence to support the contention that hedging currency risk would reduce market beta and thereby may result in lower cost of equity. We further find that the difference in mean and median SMB factors between non-hedgers and hedgers category is not statistically significant. According to Gay et al., hedgers should have significantly lower SMB betas than non-hedgers, which would reduce their cost of equity. Our results suggest that SMB betas have no impact on firm's cost of equity.

We also find no significant differences in the mean and median betas for the HML factor for hedgers vis-à-vis non-hedgers. Contrary to our results, Gay et al. (2011) find the differences in mean and median values of the HML factor to be significant in two out of three samples. Finally, we examine the difference between the firm's cost of equity for hedgers and non-hedgers. Our results imply that hedgers face a higher cost of equity by 20 basis points and 10 basis points based on mean and median values, respectively. But these differences are not statistically different from zero. On the contrary, Gay et al. (2011) find that hedged firms tend to have lower cost of equity than non-hedged firms. However, our results are intuitive given that we find no significant differences in market beta, SMB beta, and HML beta between hedgers and non-hedgers. However, these results are based on univariate tests, we must control for other factors that may influence a firm's cost of equity. We turn to this issue next.

Table 3: the mean and median values of market beta, SMB beta, HML beta, and the cost of equity for hedgers and non-hedgers

Table 3 reports the mean and median values of market beta, SMB beta, HML beta, and the cost of equity for hedgers and non-hedgers. Market beta, SMB beta, and HML betas are computed using equation (1) for 2009, and firm's cost of equity is computed using equation (2) for 2009. Market beta is market risk premium; SMB and HML betas are the beta risk factors. A firm's cost of equity is the maximum of firm's cost of equity and the risk-free rate of interest. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level.

	Hedgers		Non-hedgers		Difference (Non-hedgers - Hedgers)	
Variable	Mean	Median	Mean	Median	In means	In medians
Market beta	1.055	1.020	0.956	0.911	-0.098	-0.109
SMB beta	0.591	0.570	0.503	0.490	-0.088	-0.080
HML beta	0.164	0.167	0.167	0.203	0.003	0.036
Cost of equity	0.098	0.091	0.096	0.090	-0.002	-0.001

5. Relationship between firm's derivatives use and cost of equity

5.1 Method

We examine the relationship between a firm's derivatives use and its cost of equity using OLS regression, treatment-effects regression, and IV2SLS regression. Following Gay et al. (2011), we estimate the simultaneous equations as follows:

$$COE_i = \beta_0 + \beta_1\, HED\_DUM_i + \beta_2 LEV_i + \beta_3 Log\_TA_i + \beta_4\, BP_i + \beta_5 Log\_SH.TR_i + \varepsilon_i \qquad (3)$$

$$COE_i = \alpha_0 + \alpha_1\, EX\_HED_i + \alpha_2 LEV_i + \alpha_3 Log\_TA_i + \alpha_4\, BP_i + \alpha_5 Log\_SH.TR_i + v_i \qquad (3.1)$$

COE_i is the cost of equity for firm *i* estimated using equation (2). The definition of the variables is provided in Table 1. Equation (3.1) is similar to equation (3) except that the hedging dummy in equation (3) is replaced with the extent of hedging in equation (3.1). The above equations are similar to Gay et al. (2011). However, unlike Gay et al. (2011), we do not include the number of business segments and the number of analysts mainly because number of business segments is not available in CMIE Prowess database for a majority of firms, and as noted by (Lee et al. (2003)), the data on analysts' forecasts is unavailable for a large number of firms in emerging markets. The inclusion of these two variables in the analysis would significantly reduce our sample size, and hence we have not considered them.

A firm's decision to hedge and the extent of hedging currency risk are modeled in equations (4) and (4.1). These are similar to Gay et al. (2011) except that we have not considered Tax Loss Carry Forwards in equation (4) and Investment Tax Credit in equation (5).

$$HED\_DUM_i = \gamma_0 + \gamma_1\, LEV_i + \gamma_2\, BP_i + \gamma_3\, Log\_TA_i + \gamma_4 QR_i + \gamma_5 FR_i + \eta_i \quad (4)$$

$$EX\_HED_i = \delta_0 + \delta_1\, LEV_i + \delta_2\, BP_i + \delta_3\, Log\_TA_i + \delta_4 QR_i + \delta_5 FR_i + u_i \quad (4.1)$$

Equation (4.1) is similar to equation (4) except that firm's decision to hedge is the dependent variable in equation (4), and firm's extent of hedging is the dependent variable in equation (4.1). η_i and u_i are the error terms.

Finally, a firm's leverage ratio is modeled as follows:

$$LEV_i = \theta_0 + \theta_1\, HED\_DUM_i + \theta_2\, BP_i + \theta_3\, Log\_TA_i + \theta_4 DEPN_i + \theta_5 FA\_TA_i + \theta_6 ROA_i + \theta_6 SGA_i + \nu_i \quad (5)$$

$$LEV_i = \pi_0 + \pi_1\, EX\_HED_i + \pi_2\, BP_i + \pi_3\, Log\_TA_i + \pi_4 DEPN_i + \pi_5 FA\_TA_i + \pi_6 ROA_i + \pi_6 SGA_i + \xi_i \quad (5.1)$$

5.2 Control Variables

Gay et al. (2011) argue that firms that have higher leverage would have higher financial distress costs, and hence investors would require higher returns which implies a firm's leverage is positively related to its cost of equity. A firm's size is inversely related to its cost of equity due to the fact that larger firms are more likely to have higher tangible assets, lower variability in cash flows, and better-established operations than smaller firms. Firms with high book-to-market ratio tend to have not only lower earnings, but also higher leverage and higher earnings uncertainty as compared to the firms with low book-to-market ratios (Fama and French (1995); Chen and Zhang (1998)), suggesting that these firms would have higher financial distress costs (Fama and French (1996); Vassalou and Xing (2004)), more sensitivity to business cycle (Vassalou and Xing, 2004) such that investors would require higher returns (Fama and French (1993); Fama and French (1995); Lewellen (1999); Chen and Zhang (1998)). Therefore, the relationship between book-to-market ratio and a firm's cost of equity is positive (Gay et al., (2011)). On the contrary, Chen and Zhang (1998) argue that book-to-market ratio does not contribute to the cross-sectional variation in returns Thailand and Taiwan stocks.[104]

Trading volume is an important control variable in examining the relationship between a firm's derivative use and its cost of equity. Luez and Verrecchia (2000) argue that the trading volume is inversely related to information asymmetries as the former is a proxy for market liquidity. Myers and Majluf (1984) argue that firms with more information asymmetries between managers and shareholders incur high cost to obtain external finance. Hence, it can be argued that the trading volume is negatively associated with its cost of equity. On the contrary, Gervais et al. (2001) argue that high trading volume stocks outperform normal trading volume stocks due to high visibility, which induces investors to pay more for these stocks, which contradicts

[104] Chen and Zhang (1998) conclude that the association between a firm's book-to-market ratio and its stock returns is positive only in well-established markets like the US, lesser impact in growth markets like Japan, Hong Kong, and Malaysia, and no impact in Thailand and Taiwan, which are considered to be high-growth markets.

that trading volume is positively related to firm's cost of equity. Overall, the evidence between a firm's trading volume and its cost of equity is mixed.

The relationship between size and hedging is ambiguous. Larger firms tend to use more derivatives to hedge currency risk than smaller firms for two reasons. First, larger firms benefit from the transactional and informational economies of scale more than smaller firms in implementing the risk management programme. Second, the fixed cost in setting up of treasury desk dealing with derivatives is high, and larger firms find it more affordable (Mian (1996); Geczy et al. (1997); Bodnar et al. (1998); Bartram et al. (2009)). On the contrary, Warner (1977) documents an inverse relationship between costs of bankruptcy and the size of the firm. Moreover, firms facing liquidity problem tend to have high financial distress costs and thus are more likely to use derivatives to hedge their currency risk (Geczy, et al. (1997); Bartram et al. (2009)).

We expect a positive association between a firm's usage of derivatives and its leverage since the usage of derivatives help firms to reduce their financial distress costs (Leland (1998)), which in turn motivates these firms to increase leverage. Myers (1977) argues that firms with high growth and investment opportunities prefer to have lesser leverage. Warner (1977) argues that larger firms tend to have lower bankruptcy costs than smaller firms. Hence, larger firms can afford higher levels of debt. Firms with high non-debt tax shield, measured by depreciation relative to its total assets, is a substitute for tax benefits of debt financing. Therefore, these firms tend to have lower debt (DeAngelo and Masulis (1980)). Firms with more tangible assets, as measured by fixed assets, prefer to have high leverage since these firms have higher collateral value (Myers and Majluf (1984)). In addition, firms prefer, in the order of priority, retained earnings, long term debt, and new equity, mainly due to high transaction cost associated with new equity (Myers and Majluf (1984)). This line of reasoning suggests that past profitability and retained earnings are negatively related to its current debt levels. Hence, firms with higher profitability prefer to have lesser leverage (Titman and Wessels (1988)).
Finally, in examining the relationship between uniqueness of a firm and its leverage, we consider its non-financial stakeholders, such as customers, suppliers, and employees, of a firm. Customers may incur high switching costs when the firm is liquidated since it may be difficult to find alternative servicing for their products; similarly, employees may have acquire job specific skills that make it difficult for them to find alternative jobs; and, suppliers might have made firm-specific investments (Titman and Wessels (1988); Maksimovic and Titman (1991); Grinblatt and Titman (2002); Kale and Shahrur (2007); Banerjee et al. (2008); Bae et al. (2011); Agrawal and Matsa (2013)). Therefore, a firm's uniqueness is negatively related to its leverage. In most previous studies, the authors consider spending on selling, general and administrative expenses as a proxy for firm's uniqueness. The usage of proxy is justifiable as firms that produce unique products may have to advertise and spend more in promoting and selling their products.

6. Econometric Methods

Larcker and Rusticus (2010) document that it is a common practice that researchers first assume all explanatory variables to be exogenous, and later they employ various techniques to control for endogeneity in order to check whether results vary after controlling for endogeneity. Since the sources of endogeneity are not known, we employ techniques such as Treatment-effects regression and IV2SLS regression techniques to examine whether our results are sensitive to endogeneity.

6.1 Ordinary Least Squares (OLS) Regression

Following Gay et al. (2011), we employ OLS regression to examine the effect of firm's decision to hedge, as measured by hedging dummy in equation (3), and also the extent of hedging, as measured by the amount of currency derivatives scaled by total assets in equation (3.1), on its cost of equity. We first employ OLS regression by treating all explanatory variables as exogenous for equation (3) and (3.1).

6.2. Treatment-effects Two-step Regression approach

Tucker (2011) suggests that Propensity Score Matching (PSM) is a popular technique used to correct for self-selection bias due to observable factors, and Heckman's Treatment-effects Regression is the one of the best techniques to control for self-selection bias that arises from firm's unobservable factors. However, we do not employ PSM due to two reasons. First, Zhao (2004) argues that in order to get reasonable match for PSM, a large sample is essential. Since we use cross-sectional data on firms' usage of derivatives, it may not be appropriate to employ PSM. Second, Tucker (2011) argues that most of the firm's decisions are at the discretion of managers and multiple parties are involved in decision-making process, which cannot be observed by researchers. Hence, Tucker believes that researchers in accounting and finance assign more weights to Treatment-effects regression over PSM.

We apply Heckman's two-step approach of Treatment-effects regression to correct for self-selection bias. Treatment-effects regression is a widely used technique in related literature since it controls for self-selection bias due to differences in unobservable factors. In a notable study, Chen and King (2014) argue that firms that benefit from hedging are most likely to hedge their currency, interest rate, and commodity risks, and hence the standard assumption of OLS that sample should be random is violated. Allayannis et al. (2012) also employ Treatment-effects regression while examining the effect of hedging on Firm value. They justify the usage of technique as firms with high value, as measured by Tobin's Q, prefer to use more of derivatives. Heckman's Two-step Treatment-effects regression is estimated using equations (3) and (4). The parameter $\beta_1$in equation (3) measures the average treatment effect of firm's decision to hedge on its cost of equity. To perform two-step approach of Heckman's Treatment-effects regression procedure, we follow two steps. First, Equation (4), which is a firm's self-selected decision to use currency derivatives, is estimated as a reduced-form probit by including all exogenous variables in the system i.e., equations (3) and (4)[105]. Then, we compute IMR (Inverse Mills ratio). Second, we include IMR, which is error correction variable estimated from the first-stage, as one of the additional regressors in equation (5) to examine the effect of firm's decision to hedge on its cost of equity. Later, we test whether the self-selection bias is present by examining the statistical significance of IMR in the outcome equation, i.e., equation (3). If the IMR is statistically significant, which implies that self-selection bias is present in the model, and the treatment-effects regression overcomes this and provides consistent estimates. However, an insignificant IMR suggests that the model does not suffer from self-selection bias, and hence the OLS yields consistent estimates.

6.3. Instrumental Variable Two-Stage Least Squares (IV2SLS) regression

Roberts and Whited (2012) argue that endogeneity can also arise if researchers have omitted explanatory variables (OEVs) in the regression model, and if these OEVs are correlated with any of the included explanatory variables. This bias is usually referred as COVB. As a result of this, OLS provides inconsistent estimates for all the coefficients of the regressors. In order to overcome the problem with COVB, the authors suggest IV regression. The IV regression is

[105] As Hamilton and Nickerson (2003) argue that estimating selection equation in reduced-form probit in the first-stage yields robust estimates in the outcome equation.

widely used in accounting and finance literature since it mitigates COVB and also measurement error in the explanatory variables (Larcker and Rusticus (2010)).

As noted earlier, we measure a firm's hedging activity by hedging dummy as well as by the extent of hedging. We examine the impact of firm's decision to hedge on its cost of equity in a multivariate framework using simultaneous equation modeling using IV2SLS within a system. Following Gay et al. (2011), we estimate the structural equations, i.e., equations (3), (4), and (5) to examine this relationship. Similarly to investigate the effect of extent of hedging on firm's cost of equity, we use equations (3.1), (4.1), and (5.1). To perform IV2SLS regression, we check for the validity and relevance of the instruments, and we also examine whether the proposed model suffers from endogeneity using Hansen-Sargan test. The null hypothesis of Hansen-Sargan test is that the instruments are valid. Failure to reject for null hypothesis; suggests that the instruments are valid in addressing the endogeneity. If the instruments are valid, then we examine whether there is any endogeneity between a firm's cost of equity, its derivatives use, and leverage using Durbin-Wu-Hausman (DWH) test. We check for the relevance of the instruments using Anderson Canonical Correlation Likelihood ratio test. The null hypothesis of this test is that instruments are not relevant. Failure to reject the null hypothesis implies that the instruments are not relevant. If we find that derivatives use and leverage are endogenous variables[106], then we estimate the equations as a system using IV2SLS regression.

While examining the effect of financing policy by firms, as measured by leverage, on its hedging decisions and also on the extent of hedging, we first estimate equations (4) and (4.1) using probit and Tobit regressions treating all explanatory variables as exogenous. We do not employ Treatment-effects regression for equations (4) and (4.1) since the IMR of Treatment-effects regression does not correct for the self-selection bias if the second-stage model is non-linear (Tucker, 2011). In order to correct for COVB, we estimate equation (4), firm's decision to hedge, using IVprobit regression in a system approach. Similarly, the extent of hedging by firms, equation (4.1), is estimated using IVtobit regression in a system approach. In both equations (4) and (4.1), we get the instruments from the system. We also check for the relevance and validity of the instruments.

While examining the effect of firm's decision to hedge on its capital structure, we estimate equation (5) using OLS regression. Further, we employ Treatment-effects regression using equations (5) and (4). We also use IV2SLS regression to correct for COVB in equation (5). For this purpose, we first measure firm's hedging dummy variable to hedge using linear regression. In the second stage, we plug the fitted value of hedging dummy variable into the leverage equation in equation (5) using linear regression.

Finally, to examine the effect of firm's extent of hedging on its capital structure, as measured by leverage, we first estimate equation (5.1) using OLS regression. Further, we employ IV2SLS regression to control for COVB in equation (5.1).

7. Results and Interpretation

7.1. Impact of hedging on firm's cost of equity

Table 4 reports the results of firm's cost of equity, i.e., equations (3) and (3.1), using OLS regression. The estimated coefficient of firm's hedging dummy is statistically insignificant in specifications 1 and 2. Hence, we conclude that hedging currency risk with currency derivatives

[106] Also if the instruments that are used to address the endogeneity are valid and relevant

has no effect for our sample firms in reducing their cost of equity. This result is in line with the findings of Ameer et al. (2011), who also find statistically insignificant relationship between a firm's extent of hedging and its cost of equity in Malaysia.

We find a positive relationship between a firm's leverage and the cost of equity in specifications 1 and 2 of table 4, which implies that firms with higher leverage tend to have higher financial distress costs, and therefore investors in these firms demand higher returns (Gay et al (2011)). The relationship between a firm's book-to-market ratio and its cost of equity is positive and statistically significant in both models. This may be due to three reasons. First, firms with higher book-to-market ratio tend to have lower earnings and higher uncertainty in future earnings compared to firms with lower book-to-market ratios (Fama and French (1995); Chen and Zhang (1998)). Second, firms with higher book-to-market ratio are more likely to have higher financial distress costs (Fama and French (1996); Vassalou and Xing (2004))

Table 4: Empirical Results: equations (3) and (3.1)

Table 4 reports the results of equations (3) and (3.1) using OLS regression. The dependent variable under specification 1 and 2 is firm's cost of equity. Robust t statistics are reported in parentheses. Firm's measure of hedging is its decision to hedge in specification 1 and it is the extent of hedging in specification 2. The details of each variable are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level

Covariates	OLS	OLS
	(1)	(2)
HEDDUM	-0.004	
	(-0.71)	
EX_HED		-0.006
		(-0.62)
LEV	0.040***	0.039***
	(3.58)	(2.92)
BP	0.006**	0.008**
	(1.99)	(2.36)
LOG_SH.TR	0.002*	0.001
	(1.77)	(0.69)
LOGTA	0.001	0.002
	(0.47)	(0.53)
Intercept	0.044***	0.050***
	(3.10)	(3.10)
No. of observations	330	245
Adjusted R^2	0.077	0.079
F-stat (5,324)	7.30***	
F-stat (5,239)		5.11***

Third, firms with higher book-to-market ratio are more sensitive to business cycles (Vassalou and Xing (2004)). The coefficient of log of shares traded is positive in models 1 and 2, but it is significant only in model 1. This result is consistent with the idea that high trading volume shares outperform normal trading volume stocks due to its higher visibility. This result is similar to the findings reported by Gay et al. (2011). The results reported in table 4 might change if we treat firm's cost of equity, derivatives use, and its leverage all as endogenous leading to endogeneity bias in simple OLS. Since the sources of endogeneity can be due to self-selection bias and COVB. Hence, it is important to control for aforementioned biases. Therefore, we employ suitable econometric techniques, such as treatment-effects regression and IV2SLS regression, to address each of the bias as reported above. Table 5 reports the regression results of equations (3) and (3.1) after controlling for endogeneity using aforementioned econometric techniques. It also reports the regression results of IV2SLS in specifications 2 and 3. In specification 2, firm's measure of hedging is its decision to hedge currency risk and it is extent of hedging in specification 3.

The inverse mills ratio in specification 1 is not statistically different from zero; suggesting that the model does not suffer from self-selection bias. The Hansen J statistic as reported in specification 2 is 4.774 and it is 2.650 in specification 3. These statistics are statistically insignificant in both specifications. These results do not reject the null hypothesis that the instruments are valid, and therefore the instruments that are generated from the system of equations are valid. Having ensured the validity of the instruments, it is important to examine whether firm's derivatives use and leverage are endogenous using Durbin-Wu-Hausman (DWH) Chi-square test. The DWH Chi-Square test statistic (4.7957) is significant at ten per cent level of significance in specification 2, and this statistic (9.170) is statistically significant at five per cent of level of significance in specification 3. These results do not accept the null hypothesis that firm's hedging dummy and leverage are exogenous, and it implies that hedging dummy and leverage are endogenous. In order to check for relevance of the instruments, the study employs Anderson Canonical Correlation Likelihood ratio test.

The Anderson Canonical Correlation Likelihood ratio statistic is 35.797 in specification 2, and it is 56.550 in specification 3. These statistics statistically significant at one percent level of significance. Therefore, the instruments considered are relevant.

Table 5: Treatment-effects regression of equation (3): empirical results

Table 5 reports the results from Treatment-effects regression of equation (3) under specification 1. The table also reports IV2SLS regression for equations (3) and (3.1) under specification 2 and 3. The dependent variable is firm's cost of equity. The firm's hedging measure is its decision to hedge in specifications 1 and 2 and it is the extent of hedging in specification 3. Robust t statistics are reported in parentheses for specification 1. Robust z statistics are reported in parentheses for specifications 2 and 3. The details of each explanatory variables are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level.

Explanatory Variables	Treatment effects regression	IV2SLS regression	
	(1)	(2)	(3)
HEDDUM	-0.0004	-0.008	
	(-0.03)	(-0.43)	
EX_HED			-0.034
			(-1.48)
LEV	0.043***	0.080***	0.080***
	(3.86)	(3.90)	(3.43)
BP	0.005*	0.004	0.006*
	(1.65)	(1.22)	(1.71)
LOG_Sh.Tr	0.002	0.002	0.001
	(1.49)	(1.40)	(0.84)

LOGTA	0.001	0.0004	-0.001
	(0.44)	(0.15)	(-0.24)
Intercept	0.043**	0.040**	0.052***
	(2.48)	(2.26)	(3.12)
No. of observations	308	308	233
Adjusted R^2	0.075		
F-stat(6,301)	5.66***		
F-stat(5,302)		7.15***	
F-stat(5,227)			6.25***
IMR	-0.0003		
Hansen J statistic		4.774	2.650
Durbin-Wu-Hausman(Chi sq)		4.7957*	9.17**
Anderson LR statistic		35.797***	56.550***

In table 5, the coefficient of hedging dummy is negative in specifications 1 and 2. Our results are line with the results as reported in OLS regression as reported under specification 1 in table 4. The estimated coefficient of the extent of hedging in specification 3 is also negative and statistically insignificant in table 5. This result strongly rejects the hypothesis that the use of derivatives by firms helps them to reduce their cost of equity.

The leverage coefficient is consistently positive, and it is statistically significant at one per cent level in all the three models, which is qualitatively similar to the results estimated by OLS regression under specifications 1 and 2 in table 4. Moreover, the results imply that the positive association between a firm's leverage and its cost of equity is robust even after controlling for endogeneity. These results are comparable to the findings of Gay et al. (2011).

The coefficient of book-to-market ratio is positive and statistically significant in specifications 1 and 3 in table 5. This finding can be interpreted as the relationship between a firm's book-to-market ratio and its cost of equity, which is positive and statistically significant, is robust even after controlling for endogeneity. However, the estimated coefficient is statistically insignificant in specification 2. The sign with respect to other two exogenous variables, such as the log of shares traded and firm's size, are in line with the theory and also qualitatively similar to that of those reported by OLS regression under specifications 1 and 2 in table 5. However, the aforementioned two exogenous variables are not statistically significant at ten percent level of significance.

7.2 Examining the relationship between debt ratio and firm's hedging policy

Table 6 reports the regression results of both firm's decision to hedge (equation 4) and firm's extent of hedging (equation 4.1). The dependent variable in specification 1 and 3 in table 6 is a firm's decision to hedge. Specification 1 in table 6 reports the results of probit regression, and the results of probit regression using instruments (IVprobit) to control for endogeneity are reported in specification 3 of table 6. The coefficient of leverage in specifications 1 and 3 is positive, but it is statistically significant at five per cent level in specification 3. These results of specification 3 can be interpreted as firms with higher leverage tend to have higher financial distress costs, and hence they are most likely to hedge their currency exposure. This result is in line with the findings of Bartram et al. (2009).

Table 6: Probit and Tobit Results

Table 6 reports the results of equations (4) and (4.1). The dependent variable under specifications 1 and 3 is firm's decision to hedge its currency risk. The dependent variable under specifications 2 and 4 is firm's extent of hedging. The above table reports the results from probit and Tobit regressions under specifications 1 and 2 respectively. Instrumental Variable (IV) probit and IV Tobit regression results are reported under specifications 3 and 4 respectively. Robust t statistics are reported in parentheses for specification 3. Robust z statistics are reported in parentheses for specifications 1, 2, and 4. The details of each variable are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level.

Independent variables	Probit (1)	Tobit (2)	IVProbit (3)	IVTobit (4)
LEV	0.401	0.056	1.707**	0.098
	(0.81)	(0.70)	(1.99)	(0.52)
BP	0.169	0.037**	0.096	0.036**
	(1.48)	(2.12)	(0.92)	(2.06)
LOGTA	0.136*	-0.008	0.096	-0.009
	(1.67)	(-0.78)	(1.13)	(-0.83)
QR	-0.200***	-0.032***	-0.163***	-0.031***
	(-3.88)	(-3.16)	(-2.89)	(-2.76)
FR	2.071***	0.515***	2.118***	0.520***
	(3.44)	(5.83)	(3.70)	(6.08)
Intercept	-0.319	0.031	-0.659	0.021
	(-0.47)	(0.42)	(-0.94)	(0.23)
No of observations	308	233	308	233
Wald chi-square (5)	22.74***		27.25***	42.41***
F (5,228)		8.17***		
Pseudo R^2	0.176	0.673		

The significantly positive coefficient of firm's size in specification 1 indicates that firms that are larger in size are more likely to hedge. These results suggest that larger firms enjoy economies of scale in terms of transactions and information in accessing the risk management expertise than smaller firms. Thus, the positive coefficient of a firm's size is justifiable. This result is in line with the findings of Mian (1996), Geczy et al. (1997), and Bartram et al. (2009). However, this coefficient turns out to be statistically insignificant after controlling for endogeneity in specification 3. This implies that probit regression results reported under specification 1 is biased and inconsistent due to the presence of COVB. The coefficient of quick ratio is negative and statistically significant in specifications 1 and 3; suggesting that firms that have higher liquid assets are less likely to hedge. This result is comparable to the findings of Geczy et al. (1997). The study finds a significant positive relationship between a firm's foreign exchange exposure and its decision to hedge, suggests that geographically diversified firms are more likely to hedge. This result is consistent with the findings of Geczy et al. (1997). Therefore, the conclusion follows that the relationship between liquidity, measured by quick ratio, and a firm's decision to hedge remain robust after controlling for COVB. It is true for the relationship between a firm's exchange exposure and its decision to hedge.

The dependent variable under specifications 2 and 4 of table 6 is the firm's extent of hedging. Specifications 2 and 4 of table 6 report the results of Tobit and Tobit regressions using instruments (IVTobit). The positive coefficient of book-to-market ratio under specifications 2 and 4 in table 6 is counter-intuitive from the perspective of underinvestment problem. A plausible explanation for this contradictory result is that firms with lower growth opportunities prefer to hedge more since these firms are more likely to have free cash problems (Aretz and Bartram (2010)). The coefficient of firm's liquidity is negative and statistically significant at one per cent level of significance under specifications 2 and 4. This result is justifiable since the firms with higher liquid assets prefer to hedge more since these firms tend to have higher financial distress costs. Foreign exchange exposure coefficient is positive and significant at one per cent level of significance in specifications 2 and 4, which suggests that firms with higher foreign currency exposure prefer to hedge more.

7.3 Examining the impact of hedging policy on firm's debt ratio

Table 7 presents the estimated results of firm's leverage ratio i.e., equations (5) and (5.1) using OLS regression. The firm's decision to hedge and the extent of hedging are positively associated with its leverage as reported in specifications 1 and 2, but it is statistically significant only in specification 2. This result implies that it is the decision on how much to hedge matters, instead of decision to hedge, for firms to increase debt in their capital structure. This result is tenable for the following reason. Hedging reduces firm's financial distress costs by reducing cash flow volatility of firms as reported by Smith and Stulz (1985), and hence firms can afford to have higher leverage. These results are also consistent with the results reported by Leland (1988), who argues that hedging by firms would help them to reduce their financial distress costs, and thereby they can enjoy more debt in their capital structure. The estimated coefficient of book-to-market ratio is negative, and is statistically significant in both specifications 1 and 2; suggesting that the firms who face higher uncertainty in future cash flows prefer to have lesser leverage. The relationship between a firm's size and its leverage is positive and is statistically significant in specifications 1 and 2. This implies that firms with higher tangible assets, as measured by its total assets, would have higher leverage.

With respect to other control variables, the coefficient of non-debt tax shield, as measured by depreciation scaled by total assets, is negative, and significant in both the specifications. This result indicates that non-debt tax shield acts as a substitute for debt financing firms. This result is in line with the results of DeAngelo and Masulis (1980) who find that firms with higher non-debt tax shields tend to have lower debt in their capital structure. Further, the coefficient of the ratio between fixed assets and total assets is positive and is statistically significant, suggesting that firms with higher tangible assets prefer to have higher debt in their capital structure since these firms enjoy higher collateral value of their assets. This result can be compared with those of Myers and Majluf (1984). The negative coefficient of profitability in both the specifications suggests that firms with higher profitability are more likely to have higher retained earnings, which is one of the possible sources of financing their projects by firms. According to the pecking order hypothesis, firms would prefer retained earnings most followed by debt and equity. Therefore, the results are consistent with this hypothesis that firms with higher profitability would prefer to reduce debt in capital structure. This result is consistent with the results of Titman and Wessels (1988).

Table 7: Empirical results relating to Firm's leverage ratio

Table 7 reports the regression results of firm's leverage ratio i.e., equations (5) and (5.1) using OLS regression. The dependent variable under specifications 1 and 2 is firm's leverage ratio. Firm's hedging measure is its decision to hedge in specification 1, and it is the extent of hedging in specification 2. Robust t statistics are reported in parentheses. The details of each variable are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level

Covariates	OLS	OLS
	(1)	(2)
HEDDUM	0.048	
	(1.46)	
EX_HED		0.093*
		(1.89)
BP	-0.028**	-0.031**
	(-2.28)	(-2.33)
LOGTA	0.016**	0.020**
	(1.97)	(2.11)
DEPN_TA	-3.053***	-3.791***
	(-3.96)	(-4.47)
FA_TA	0.386***	0.433***
	(7.47)	(7.31)
ROA	-0.859***	-0.855***
	(-7.11)	(-6.55)
SGA_N.SALES	-0.228	-0.251
	(-1.49)	(-1.35)
Intercept	0.411***	0.401***
	(4.59)	(3.95)
No of observations	329	245
Adjusted R^2	0.294	0.308
F-stat (7,321)	25.19***	
F-stat (7,237)		25.13***

The coefficient of selling, general, and administration expenses to net sales ratio, which is a proxy for firm's uniqueness of the product, is negative, but it is not statistically significant. Hence, the results reported in this study suggest that uniqueness of a firm's product does not matter for leverage.

The results reported in table 7 using OLS regression have not used controls for self-selection bias and COVB. Hence, the suitable econometric techniques are employed to address each bias separately in table 8. Specifications 1 and 2 in table 8 report the results of how firm's decision to hedge affects its leverage ratio using Treatment-effects regression and IV2SLS regression respectively. The results reported under specifications 1 and 2 in table 8 can be compared with results reported under specification 1 in table 7 to examine whether the results are sensitive to any of the previously mentioned biases. The coefficient of firm's hedging dummy is positive, but is statistically insignificant in specifications 1 and 2; suggesting that no association between a firm's hedging and its capital structure.

These results are consistent with Geczy et al. (1997), who do not find statistically significant association between a firm's decision to use currency derivatives and its capital structure in simultaneous logit-OLS framework. In a similar study in simultaneous probit-OLS framework by Graham and Rogers (2002), who also do not find the effect of foreign currency hedging by firms on debt ratio. The results as reported in table 8 are similar to the results reported in table 7, except for logta. The coefficient of logta is positive in specifications 1 and 2 in table 8. However, it is statistically insignificant only at the significance level of ten percent. This implies that OLS results reported under specification 1 in table 7 are biased and inconsistent due to the presence of the abovementioned biases. Therefore, it can be concluded that the size of the firm may not be a determining factor for firm's leverage.

Table 8: Simultaneous equation framework

Table 8 reports the results of equation (5) for specifications 1 and 2, and the results of equation (5.1) for specification 3 in a simultaneous equation framework. The dependent variable under specifications 1, 2, and 3 is firm's leverage ratio. The above table reports the results from Treatment-effects regression under specification 1, IV2SLS regression from equation (5) under specification 2, and IV2SLS from equation (5.1) from specification 3. Robust t statistics are reported in parentheses for specification 1. Robust z statistics are reported in parentheses for specifications 2 and 3. The details of each variable are presented in table 1. *** Significant at the 0.01 level, ** Significant at the 0.05 level, and * Significant at the 0.10 level.

Covariates	Treatment Effects regression	IV2SLS regression	
	(1)	(2)	(3)
HEDDUM	0.095	0.165	
	(0.99)	(1.48)	
EX_HED			0.141
			(1.10)
BP	-0.025*	-0.026**	-0.034**
	(-1.93)	(-2.08)	(-2.29)
LOGTA	0.012	0.011	0.017*
	(1.47)	(1.30)	(1.80)
DEPN_TA	-2.974***	-2.926***	-3.812***
	(-3.79)	(-3.71)	(-4.40)
FA_TA	0.346***	0.338***	0.410***
	(6.41)	(6.23)	(6.22)
ROA	-0.951***	-0.929***	-0.968***
	(-7.38)	(-6.94)	(-7.28)
SGA_N.SALES	-0.248	-0.222	-0.276
	(-1.63)	(-1.54)	(-1.51)
Intercept	0.436***	0.381***	0.460***
	(3.92)	(3.06)	(4.49)
No. of observations	308	308	233
Adjusted R2	0.32		

F-stat (8,299)	21.23***		
F-stat (7,300)		21.56***	
F-stat (7,225)			23.04***
IMR	-0.0512		
Hansen J statistic		2.38	2.48
Durbin-Wu-Hausman (Chi sq)		3.99**	0.445
Anderson LR statistic		44.06***	54.95***

Specification 3 in table 8 examines how the extent of hedging can affect a firm's leverage after controlling for endogeneity using IV2SLS regression. The results reported in table 8 are qualitatively similar to the results reported in specification 2 of table 7 i.e., firm's extent of hedging. This coefficient is positive, but it is not statistically significant at ten percent level of significance. This result suggests that hedging currency risk by firms does not matter for the characteristics of debt contracts. However, this coefficient is not only positive but also statistically significant in OLS regression results reported under specification 2 in table 7. This result implies that OLS results are biased and inconsistent due to the presence of COVB between a firm's hedging and its leverage. These results are comparable to the findings of Dionne and Triki (2013), who also document a positive relationship between a firm's hedging and its leverage using Tobit regression. However, this relationship turns out to be statistically insignificant after controlling for endogeneity. The results relating to other control variables such as depreciation to total assets, fixed assets to total assets, return on total assets, and selling and distribution expenses to net sales under specifications 1, 2, and 3 of table 8 are qualitatively similar to the results reported under specifications 1 and 2 in table 7.

8. Robustness checks

Addressing endogeneity is inevitable in corporate finance since most of the firm's decisions are endogenous. The source of the endogeneity is not known, and hence researchers should employ relevant econometric techniques to control for endogeneity. In addition to a battery of econometric techniques in addressing endogeneity in empirical results section, we further employ two more techniques to control for endogeneity as robustness checks. First, while examining the relationship between a firm's extent of hedging and its total risk, Hentschel and Kothari (2001) argue that if all the independent variables are endogenous, then the researcher has to find out the instruments for all independent variables, which may not be an easy task. They suggest using portfolio ranks as instruments. For firm's extent of hedging variable, they assign portfolio rank as zero to all non-hedged firms. They further split firms' extent of hedging, as measured by the amount of currency derivatives scaled by total assets, into two groups, namely above-median and below-median groups. The firms that belong to below-median and above-median groups are assigned portfolio ranks as 1 and 2, respectively. They use these portfolio ranks as instruments. They follow a similar methodology to compute portfolio ranks for all other independent variables other than firm's extent of hedging, except that portfolio break points to compute ranks are different. They divide each independent variable into three sub-groups. The first sub-group contains firms upto the 33rd percentile, second group contains firms between 33rd and 67th percentile, and third group contains firms above the 67th percentile. They assign portfolio ranks 1, 2, and 3 for the first, second, and third sub-groups, respectively. Finally, they use portfolio ranks so generated for each of their independent variables as instruments, and then employ IV2SLS regression.

The instruments generated based on portfolio ranks must satisfy two conditions, namely validity and relevance. The validity of the instruments can be justified on the grounds that portfolio ranks are unlikely to be correlated with the error term in the structural equation. The instruments might be highly correlated with endogenous variables, since the higher values of independent variable would contain higher portfolio rank and lower values of independent variables would contain lower portfolio rank. However, we check for validity and relevance econometrically using relevant techniques. Larcker and Rusticus (2010), while reviewing the usage of IV in accounting research, have also suggested this approach as one of ways to control for endogeneity. Therefore, in our paper we employ a similar methodology as suggested by Hentschel and Kothari (2001) by computing portfolio ranks as instruments for all independent variables, and then use these computed ranks as instruments in IV2SLS regression. Our results are qualitatively similar to the base case regressions as reported in empirical results section.

Second, we include an industry-adjusted long term debt ratio[107], instead of actual long term debt ratio, to measure firm's financial distress risks as suggested by Geczy et al. (1997) and Gay et al. (2011). Geczy et al. (1997) argue that firms would always target for industry median long term debt ratio. They further argue that firms with debt ratio higher than its industry-adjusted would be considered as financially distressed, and firms with debt ratio lower than its industry-adjusted would come under lower financially distressed firms. While examining the relationship between a firm's cost of equity, derivatives use, and its debt ratio, Gay et al. (2011) employ the cost of equity and debt ratio after adjusting for industry effect, instead of actual cost of equity and debt ratio, in their regression analysis. Moreover, we investigate the inter-relationship between a firm's cost of equity, derivatives use, and its debt ratios using industry-adjusted cost of equity and industry-adjusted debt ratios. We compute industry-adjusted leverage as the difference between a firm's leverage and the median industry leverage based on NIC two-digit code for 2009. While examining the relationship between a firm's cost of equity and its derivatives usage, we find the coefficient of leverage ratio is statistically insignificant. However, the coefficients of other variables are qualitatively similar as compared to the base case results. While examining the relationship between a firm's leverage ratio and its hedging decision and vice-versa, the results are qualitatively similar to those reported in base case results.

Titman and Wessels (1988) employ different measures of leverage for examining the determinants of firm's leverage. Hence, in order to cross-check whether our results are sensitive to the usage of different measures, we measure leverage alternatively as: long term debt to net worth, long term debt to market value of equity, long term debt is scaled by the sum of long term debt and market value of equity. Our results are qualitatively similar under different measures.

9. Conclusion

To our knowledge, this is the first study to examine the inter-relationship between a firm's usage of derivatives, leverage, and its cost of equity in a simultaneous equations framework in the context of an emerging country like India. Our findings suggest that reduction in firm's cost of equity for hedged firms is statistically insignificant as compared to non-hedged firms. This result implies that a firm's decision to hedge and the extent of hedging currency risks do not affect its financing costs. This result is tenable since the proportion of firm's currency risk as compared to its total risk is negligible as argued by Copeland and Joshi (1996), and hence hedging currency risk may not result in reduction in firm's cost of equity. Our findings also support the argument that more financially distressed firms, as measured by its leverage, tend

[107]We compute industry-adjusted leverage as the difference between firm's leverage and the median industry leverage based on NIC two-digit code for 2009. This methodology is similar to those of Geczy et al. (1997) and Gay et al. (2011).

to have higher cost of equity. These results are robust even after controlling for different sources of endogeneity using relevant econometrics techniques such as Heckman Treatment-effects regression and IV2SLS regression.

We also examine the relationship between a firm's leverage and its different measures of hedging, such as decision to hedge and the extent of hedging in a simultaneous equations framework. We find a positive and significant association between a firm's leverage and its decision to hedge currency risk; suggesting that financial distress firms are more likely to hedge. However this is not true for the relationship between a firm's debt ratio and its extent of hedging. Firms with lesser liquid assets and higher foreign exchange exposure are more likely to hedge. These results are consistent even after controlling for endogeneity using IV2SLS estimation. Larger firms are more likely to hedge currency risk than smaller firms since the former enjoys the transactional and informational economies of scale in assessing the risk management expertise. However, this significant relationship disappears after controlling for endogeneity. This result implies that firm's size may not matter for firm's decision to hedge and for the extent of hedging.

Finally, we investigate whether a firm's decision to hedge would enhance debt capacity of firms in a simultaneous equations framework. We find no effect on firm's debt ratio due to usage of currency hedging. This result is consistent with those of Geczy et al. (1997) and Graham and Rogers (2002). We further extend our analysis to test whether the extent of hedging affects firm's leverage, and we find that the extent of hedging is positively associated with firm's leverage. However, this significant effect disappears when we control for endogeneity. From the discussions so far, it can be argued that firm's decision to hedge and the extent of hedging have no role to play in determining its capital structure. Firms with higher book-to-market ratio, higher non-debt substitutes, and higher profitability prefer to have lesser debt in capital structure. Firms with higher tangible assets prefer to have higher leverage. These results are consistent with the theory and remain consistent even after controlling for potential endogeneity using various econometrics techniques. Overall, we conclude that risk management through the use of derivatives is an important decision to consider as it can create value by reducing potential financial distress costs.

Appendix

Asset-pricing model comparison between FF (1993) and the CAPM is reported below.

Table 9: Asset-pricing model comparison between FF (1993) and the CAPM

	FF 1993	CAPM
Chi-square statistics	43.298	57.229
P-value for Chi-square statistic	0.416	0.059
Average absolute value of intercepts	0.0008	0.0014
Adjusted R^2	0.621	0.573

Note: The table 9 reports Gibbons, Ross, and Shaken (GRS) (1989) test is applied at the industry level. The second and third column of the table indicates Chi-square statistics, P-value for Chi-Square statistic, the average absolute value of intercepts, and adjusted R^2 for FF (1993) and the CAPM.

We examine whether the three-factor model of Fama and French (FF) (1993) is superior to the Capital Asset Pricing Model (CAPM) in estimating firm's cost of equity. To assess this, we adopt three alternative measures as available in the related literature. First, GRS (1989) test is applied at the industry level[108]. We find that chi-square statistics (p values) are 43.298 (0.416) and 57.229 (0.059) for FF (1993) and for the CAPM respectively. Therefore, we conclude that FF (1993) model fits empirical data better than the CAPM since the former has lower chi-square statistic and also insignificant statistic[109] but which is not true for the CAPM. Second, the average absolute pricing errors, as measured by average absolute value of intercepts, for FF (1993) and for the CAPM the corresponding figures are 0.0008 and 0.0014, respectively. This further confirms our previous findings that FF (1993) is a better performing model than the CAPM[110]. Third, the mean Adjusted R^2 across all industry portfolios is 0.621 for FF (1993), and that of the CAPM is 0.573; suggesting that the explanatory power of the former is higher than the latter. Hence, we measure firm's cost of equity using FF (1993) rather than the CAPM.

[108] Since we have more number of stocks (332) than the number of time-series observations (243), we cannot apply GRS (1989) test at the firm level. To circumvent this problem, we group them into industry portfolios based on two-digit National Industry Classification (NIC) code, which is similar to SIC code in the context of US.

[109] Insignificance suggests that the intercepts of the model in question are jointly statistically zero.

[110] Connor and Sehgal (2001); and Mehta and Chander (2010)

References

Agarwalla. S.K., Jacob, J.J., and Varma.J.R (2013), "Four Factor Model in Indian Equities Market", Indian Institute of Management Ahmedabad, Working Paper (2013-09-05), 1-22.

Agrawal. A.K., and Matsa, D.A. (2013), "Labor Unemployment Risk and Corporate Financing Decisions", Journal of Financial Economics, 108, 449-470

Allayannis.G., Lel, U., and Miller. D.P. (2012), "The Use of Foreign Currency Derivatives, Corporate Governance, and Firm Value around the World", Journal of International Economics, 87 (1), 65-79.

Ameer, R., Isa, R.B.M., and Abdullah, A.B. (2011). "A Survey on the Usage of Derivatives and Their Effect on Cost of Equity Capital", The Journal of Derivatives, 19 (1), 56-71

Anand, M. and Kaushik, K.P. (2008), "Currency Derivatives: A Survey of Indian Firms". IIMB Management Review, 20 (3), 324 – 339

Angrist. J.D, and Krueger, A.B. (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", Journal of Economic Perspectives, 15 (4), 69-85.

Aretz, K. and Bartram, S.M. (2010). "Corporate Hedging and Shareholder Value", The Journal of Financial Research, 33 (4), 317 – 371.

Bae, K.H., Kang, J.K., and Wang, J., (2011), "Employee Treatment and Firm Leverage: a Test of the Stakeholder Theory of Capital Structure", Journal of Financial Economics, 100, 130-153.

Banerjee, S., Dasgupta, S., and Kim, Y., (2008), "Buyer-supplier relationships and the Stakeholder Theory of Capital Structure", The Journal of Finance, 63 (5), 2507–2552

Bartram, S.M., Brown, G.W., and Fehle, F.R. (2009), "International Evidence on Financial Derivatives Usage". Financial Management, 38 (1), 185-206

Bodnar.G.M., Hayt, G.S., Marston, R.C., and Smithson, C.W. (1995), "Wharton Survey of Derivatives Usage by U.S. Non-Financial Firms", Financial Management, 24 (2), 104-114.

Bodnar, G.M., Hayt, G.S. and Marston, R.C. (1998), "1998 Wharton Survey of Financial Risk Management by US Non – Financial Firms", Financial Management, 27 (4), 70 – 91.

Bodnar. G.M., Consolandi, C., Gabbi G., and Jaiswal-Dale A. (2013), "Risk Management for Italian Non-Financial Firms: Currency and Interest Rate Exposure", European Financial Management, 19 (5), 887-910

Bound, J., Jaeger, D,A., and Baker, R.M. (1995), "Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogeneous Variable is Weak", Journal of the American Statistical Association, 90 (430), 443-450.

Chava.S., and Purnandam.A. (2010), "Is Default Risk Negatively related to Stock Returns?" Review of Financial Studies, 23 (6), 2523-2559

Chen.J. and King T.D. (2014), "Corporate Hedging and the Cost of Debt", Journal of Corporate Finance, 29, 221-245.

Chen, Nai-Fu. and Zhang.F. (1998), "Risk and Return of Value Stocks", The Journal of Business, 71 (4), 501-535.

Copeland, T.E. and Joshi, Y. (1996), "Why derivatives don't reduce FX risk", The McKinsey Quarterly, 1, 66-79

Connor, G. and Sehgal, S. (2001), "Tests of the Fama and French Model in India", Financial Market Group, London School of Economics, London, UK, Discussion Paper 379.

Coutinho. J.R.R., Sheng H.H., and Lora M.I. (2012), "The Use of Fx Derivatives and the Cost of Capital: Evidence of Brazilian Companies", Emerging Markets Review, 13 (4), 411-423

DeAngelo, H., and Masulis, R.W. (1980). "Optimal Capital Structure under Corporate and Personal Taxation", Journal of Financial Economics, 8 (1), 3-29.

DeMarzo P.M. and Duffie. D. (1995), "Corporate Incentives for Hedging and Hedge Accounting", The Review of Financial Studies, 8 (3), 743-771.

D'Mello R. and Sheroff. P.K. (2000). "Equity Undervaluation and Decisions Related to Repurchase Tender Offers: An Empirical Investigation", The Journal of Finance, 55 (5), 2399-2424.

Dimson, E. (1979), "Risk Measurement when Shares are subject to Infrequent Trading", Journal of Financial Economics, 7 (2), 197-226

Dionne. G. and Triki. T. (2013), "On Risk Management Determinants: What Really Matters?", The European Journal of Finance, 19 (2), 145 – 164.

Fama, E.F. and French, K.R. (1993), "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics, 33 (1), 3-56

Fama, E.F. and French, K.R. (1995), "Size and Book-to-Market Factors in Earnings and Returns", The Journal of Finance, 50 (1), 131-155.

Fama, E.F. and French, K.R. (1996), "Multifactor Explanations of Asset Pricing Anomalies", The Journal of Finance, 51 (1), 55-84.

Farhi. M. and Borghi, R.A.Z. (2009), "Operations with Financial Derivatives of Corporations from Emerging Economies", Estudos Avancados, 23, 66, 169-188

Fowler, D., and Rorke, C.H. (1983), "Risk Measurement when Shares are subject to Infrequent Trading: Comment". Journal of Financial Economics, 12 (2), 279-283.

Gay, G.D. and Nam, J. (1998), "The Underinvestment Problem and Corporate Derivatives Use", Financial Management, 27 (4), 53 – 69.

Gay, G.D., Lin, C., and Smith, S.D. (2011), "Corporate Derivatives Use and the Cost of Equity", Journal of Banking & Finance, 35, 1491-1506.

Geczy, C., Minton, B.A., and Schrand, C. (1997). "Why Firms Use Currency Derivatives". The Journal of Finance, 52 (4), 1323-1354.

Gibbons. M. R., Ross. S. A, and Shanken.J, (1989), "A Test of the Efficiency of a Given Portfolio", Econometrica, 57 (5), 1121-1152.

Graham J.R. and Rogers, D.A. (2002), "Do Firms Hedge in Response to Tax Incentive?", The Journal of Finance, 57 (2), 815-839.

Greene, W. H. (2000), Econometric Analysis, Prentice Hall International, Inc, Fourth Edition, Upper Saddle River, New Jersey

Grinblatt, M. and Titman. S, (2002), "Financial Markets and Corporate Strategy", The McGraw Hill Companies, Second Edition

Godfrey, L.G., and Hutton,J.P., (1994), "Discriminating between Error-in-variables/Simultaneity and Misspecification in Linear Regression Models", Economics Letters, 44 (4), 359-364

Gervais, S., Kaniel. R., and Mingelgrin. D. H. (2001), "The High-Volume Return Premium", The Journal of Finance, 56 (3), 877-919.

Guimaraes-Filho. R., Piao. S, and Zhang. L. (2014), "Corporate Leverage in Asia: A Fault Line?, Regional Economic Outlook: Asia and Pacific, 33-46.

Heckman. J.J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", Annals of Economic and Social Measurement, 5 (4), 475 – 492.

Hentschel, L. and Kothari, S.P. (2001), "Are Corporations Reducing or Taking Risks with Derivatives?" The Journal of Financial and Quantitative Analysis, 36 (1), 93-118.

International Swaps and Derivatives Association (ISDA) Research Notes. (2009), 2009 ISDA Derivatives Usage Survey,
(http://www.isda.org/researchnotes/pdf/ISDA-Research-Notes2.pdf)

Judge, A. (2006). The Determinants of Foreign Currency Hedging by UK Non-Financial Firms, Multinational Finance Journal, 10 (1), 1 – 41.

Kale, J., and Shahrur, H., (2007), "Corporate Capital Structure and the Characteristics of Supplier and Customer Markets", Journal of Financial Economics, 83, 321–365

Larcker. D.F. and Rusticus T.O. (2010), "On the use of Instrumental Variables in Accounting Research", Journal of Accounting and Economics, 49 (3), 186-205.

Lee, C., Ng. D., and Swaminathan. B. (2003), "The Cross Section of International Cost of Capital", Cornell University, Working Paper

Leland, H. E. (1998), "Agency Costs, Risk Management, and Capital Structure", The Journal of Finance, 53 (4), 1213-1243.

Lewellen, J. (1999), "The Time-series relations among Expected Return and Book-to-market", Journal of Financial Economics, 54 (1), 5-43.

Lievenbruck, M., and Schmid, T. (2014), "Why do Firms (not) Hedge? – Novel Evidence on Cultural Influence", Journal of Corporate Finance, 25, 92-106.

Luez. C. and Verrecchia. R.E. (2000), "The Economic Consequences of Increased Disclosure", Journal of Accounting Research, 38, 91-124

MacMinn, R, D. (1987), "Forward Markets, Stock Markets, and the Theory of the Firm", The Journal of Finance, 42 (5), 1167-1185

Maksimovic, V. and Titman, S. (1991), "Financial Policy and Reputation for Product Quality", The Review of Financial Studies, 4 (1), 175–200.

Mian, S.L. (1996), "Evidence on Corporate Hedging Policy", The Journal of Financial and Quantitative Analysis, 31 (3), 419 – 439

Modigliani. F. and Miller. M. H. (1958), "The Cost of Capital, Corporation Finance, and the Theory of Investment", The American Economic Review, 48 (3), 261-297.

Morellec, E. and Smith.Jr. C.W. (2007), "Agency Conflicts and Risk Management", Review of Finance, 11, 1-23.

Myers, S.C. (1977), "The Determinants of Corporate Borrowing", Journal of Financial Economics, 5 (2), 147-175

Myers. S. C. and Majluf, N. (1984). "Corporate Financing and Investment Decisions when Firms have information that Investors do not have". Journal of Financial Economics, 13(2), 187-221.

Rawls III, S.W. and Smithson, C.W. (1990), "Strategic Risk Management", Journal of Applied Corporate Finance, 2.4, 6 – 18.

Roberts, M.R. and Whited T.M. (2012), "Endogeneity in Empirical Corporate Finance", Simon School Working Paper No. FR 11-29, Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1748604

Smith, C.W. and Stulz, R.M. (1985), "The Determinants of Firms' Hedging Policies", The Journal of Financial and Quantitative Analysis, 20 (4), 391 – 405.

Stulz, R.M. (1996), "Rethinking Risk Management", Journal of Applied Corporate Finance, 9 (3), 8 – 24.

Titman, S and Wessels, R. (1988), "The Determinants of Capital Structure Choice", The Journal of Finance, 43 (1), 1-19.

Tucker J.W. (2011). Selection Bias and Econometric Remedies in Accounting and Finance Research. Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1756911

Varma, J.R., and Barua, S.K. (2006), "A first cut estimate of the equity risk premium in India", Indian Institute of Management Ahmedabad, Working Paper (2006-06-04)

Vassalou.M. and Xing Y. (2004), "Default Risk in Equity Returns", The Journal of Finance, 59 (2), 831- 868.

Warner, J. B. (1977), “Bankruptcy Costs: Some Evidence”, The Journal of Finance, 32 (2), 337-347.

Zhao, Z. (2004), “Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence”, The Review of Economics and Statistics, 86 (1), 91-107.

Debt Externality in Equity Markets: Leveraged Portfolios and Islamic Indices

Salman Khan, Lahore University of Management Sciences, Lahore, Pakistan

Saad Azmat, Lahore University of Management Sciences, Lahore, Pakistan

ABSTRACT

This paper tests for the externality of debt in the equity markets. Adopting an exogenous view of the business cycles and assuming myopia amongst borrowers and lenders, the paper argues that when the markets are going up, portfolios and indices with high debt should perform better than those with low debt, while during the downward phase, low debt portfolios and indices would perform better. We use firm as well as index level data to compare performance based on high and low debt in the up and down market and conduct a series of robustness tests. We use monthly data from 4131 listed nonfinancial US firms from 1982 to 2016 to create low and high debt portfolios. At the index level, we use Islamic indices as a proxy for low debt indices. Our results show that low debt portfolio and Islamic indices outperform the high debt portfolio and conventional indices in the down market and underperform in the up market, respectively. The paper contributes to the literature on debt externality by extending the idea to the equity markets. The paper also contributes to the Islamic finance literature by identifying their better performance in the down markets. We theorize that the low debt of Islamic equity indices could be the moderating cause of their better performance.

JEL classification: C34 G11 G29 Z12
Keywords: Debt externality; Islamic equities; Risk-return characteristics; Islamic finance; Systematic risk; Liquidity risk, Financial crisis

1. Introduction

The 2008 financial crisis might be responsible for the renewed interest amongst researchers in understanding the externality of debt. A surge in the financial literature, however, had started well before the crisis which questioned the extent to which investors can rationally foresee risks occurring in the future and their ability to choose financially optimal contracts (see Kahneman and Tversky, 1979; Berntazi and Thaler, 1995). The financial crisis of 2008 acted as a reminder of how the micro level misjudgments in our financial decisions can have a macro impact (Gennaioli, Shliefer, and Vishny, 2012). Another stream of literature suggests that financial crisis and the fragility of our financial system might not simply be a behavioral issue but could be an inherent feature of debt structures (see Ebrahim et al., 2016). Given the cyclical nature of our macro economy and the individual tendency to neglect long-term risk, debt contracts might be more prone to misjudgment. In this paper, we refer to this phenomenon as the externality of debt.

The debt externality can be best explained by taking an exogenous view of the business cycle along with assuming myopia amongst borrowers and lenders. (see Sufi and Mian, 2010, 2011, and 2015; Mian, Sufi and Verner, 2017). It can be argued that during the boom period, when the economy is doing well, the debt contracts should seem more optimal for both the lenders and the borrowers. This is because during an upturn the defaults are low resulting in a relatively safe return for the lenders while the borrower (particularly the borrowing firm) can enjoy the significant upside, which the high growth period offers in the form of greater profits. During the downturn, when the economy underperforms, debt contracts should be less optimal as the possibility of defaults can end up imposing a cost on all parties. Ignoring the possibility of a downturn, when making decisions during an upturn, could be a possible cause of the debt externality. In this paper, we try to extend these ideas to test for the externality of debt in the equity markets.

In the equity markets, the presence of debt externality implies that portfolios or indices with high debt stocks might underperform those with lower debt when the markets are going down and outperform them when the markets are going up. More importantly, we argue that the benefit that the low debt portfolios or indices experience during their downturn is greater than the cost they have to bear during the upturn. We theorize that this externality might emanate from the investor's myopic behavior; hence, they would tend to ignore it.

We test for the debt externality by using both firm and index level data. For firm level analysis, we use monthly data from 4131 listed nonfinancial US firms from 1982 to 2016. Using debt-to-asset ratio, we create high and low leverage portfolios from our sample firms. This allows us to explicitly analyze the externality of debt by comparing the performance of high and low debt portfolios while controlling for other factors. For the index level analysis, we examine Islamic indices from 2001 to 2015, which serve multiple purposes. Firstly, the financial filters that are required to ensure that a stock qualifies for the Islamic equity index results in a low debt ratio for the index. This allows us to use it as a proxy for the low debt index. Secondly, it has allowed us to borrow from and contribute to the emerging literature on Islamic finance. This stream of literature on Islamic finance suggests that Islamic equity indices performed better than conventional indices during the global financial crisis; however, the literature seems to offer conflicting evidence on the difference in their performance for the non-crisis period (see Nainggolan, How and Verhoeven, 2015; El Alaoui et al., 2016). We use the Dow Jones family of indices for both the Islamic indices and conventional indices.

For our firm level analysis, we use Fama and French (1993) and Carhart (1997) factors. We divide the sample into high leverage portfolio (HLP) and low leverage portfolio (LLP). The factors for each portfolio are further divided into the up and down market using excess market return greater or lower than zero. We use Fama Macbeth (1973) two step regression with the first step entailing time series regression, and in the second step we estimate the cross sectional regressions. The results are further tested using bivariate sorted portfolios based on size and leverage. We also test for the impact of the financial crises of 1998 and 2008. We also vary the debt-to-asset ratio when constructing the high and low leverage portfolio. We particularly test for the 33% threshold ratio which is used for screening Islamic equities.

For the index analysis, the data are divided into pre-crisis, crisis and post-crisis phases. Following Pettengill, Sundaram and Mathur (1995), the data are divided into up and down markets. We examine the risk-adjusted performance of Islamic and conventional indices using Sharpe and Treynor ratios respectively. These performance benchmarks are validated by employing the 30 and 120 days rolling analyses. The comparative performance is further tested for robustness by controlling for four liquidity risk-channels in line with Acharya and Pedersen (2005).

Our results, both at the firm and the index level, confirm the presence of debt externality in the equity market. The firm level analysis suggests that high leverage portfolios (HLP) outperform low leverage portfolios (LLP) in the up market while the LLP outperforms the HLP in the down market after controlling for different factors such as excess market return, size, value, momentum, and leverage. Also, the results are robust to the changes such as bivariate sorting (size and leverage), the financial crisis of 1998 and 2007, and change in debt-to-asset ratio constraints. Similar findings are observed in the case of index level analysis. Using Sharpe and Treynor ratios, we find that Islamic equity index performs better than conventional equity index in all down markets during the pre-crisis, crisis or post-crisis phases. The results also indicate that the conventional equity index outperforms the Islamic equity index in the up markets. After controlling for multiple liquidity-risk channels as well as size effect, the results lend further support to the behavior demonstrated by these indices. We argue that this behavior is an indication of the low debt nature of Islamic indices. The results confirm the findings of some of the previous studies suggesting that Islamic equities outperformed conventional equities during the financial crisis of 2008. However, our results indicate that better performance of Islamic indices during the global financial crisis was not an anomaly. Overall, the results corroborate the theory that debt externality might be prevalent in the equity markets.

The study contributes to two streams of literature. Firstly, it contributes to the literature on the externality of debt by offering empirical support that this externality is present in the equity markets, where indices with a higher level of debt are exposed to significant downside risk compared to those with low debt. The argument regarding the externality of debt, which is the main focus of our paper, is inspired by Mian and Sufi (2010, 2011, and 2015) and Mian, Sufi and Verner (2017). They have shown the existence of debt externality at the level of the household and the macro economy. They remain cautious in extending their implications to the equity markets. Our paper is perhaps the first paper that makes a case that debt externality may also exist in the equity markets. The presence of behavioral biases, particularly myopia amongst investors and the tendency to neglected risk (see Gennaioli, Shliefer, and Vishny, 2012) may be responsible for the persistence of this externality**.** Secondly, this paper contributes to the literature on Islamic finance (see Nainggolan, How and Verhoeven, 2015; El Alaoui et al., 2016). We suggest that the performance of Islamic indices during the global financial crisis was not an anomaly. The low debt ratio of Islamic indices, which is an outcome of their screening criteria, ensures that Islamic indices perform well in all down markets.

This paper has important implications for investors and portfolio managers. Our results should help them examine the costs and benefits of investing in high and low debt portfolios and indices. It would also enable them to appreciate better the benefits of using high and low debt portfolios as well as Islamic indices as a hedge.

The rest of the article is organized as follows: Section 2 highlights the key differences between Islamic and conventional indices, it particularly focuses on their low debt feature. The literature review and relevant hypotheses are discussed in Section 3. Section 4 explains the data. The methodology is analyzed in Section 5. Section 6 examines the empirical results. Section 7 concludes the paper.

2. Islamic Indices and Debt

Islamic indices by design have lower debt. For an equity to be classified as Islamic, it has to follow multiple Shariah screening criteria. These include both financial and non-financial filters. Most of these filters have their origin in Islamic law. The foremost amongst these is the Islamic prohibition of interest. In Islamic law, the charging of excess interest in a debt contract is considered exploitative and hence prohibited (Ayub, 2009). Extending the argument to equities, Shariah scholars and regulators have prescribed financial filters to ensure that the percentage of the prohibited debt should be below a certain threshold. For the Dow Jones Islamic indices, debt to market capitalization should be less 33%. The debt constraint at 33% level essentially reduces the universe of stocks for Islamic indices but also makes them a good proxy for low debt indices (Derigs et. al, 2009).

The adherence to Islamic law also means that certain industries are excluded from the list, foremost amongst them is the conventional finance industry. Conventional banking stocks whose core business is to provide loans are excluded from Islamic indices. Apart from debt, other Islamic financial filters restrict the company's investment in debt securities, their receivables and other factors that would reflect different financial aspects of the stock[111].
It can be concluded from the discussion above that Islamic indices, in the presence of certain financial filters, should have less debt by design. They would be immune to the adverse impact of high debt and enjoy the benefit of low debt. The next section discusses the impact of low debt on the financial performance of equities and their indices.

[111] Two category of screens/filters are applied before a company is included in the Dow Jones family of Islamic indices (www.djindexes.com)

- **Sector Based Screens:** Based on the Shariah Supervisory Board established parameters, the businesses listed below are inconsistent with Shariah law. Income from the following sources cannot exceed 5% of revenue: i) Alcohol ii) Tobacco iii) Pork-related products iv) Conventional financial services (banking, insurance, etc.) v) Weapons and defense vi) Entertainment (hotels, casinos/gambling, cinema, pornography, music, etc.)
- **Accounting Based Screens:** All of the following must be less than 33%: i) Total debt as a percentage of trailing 24-month average market capitalization ii) The sum of a company's cash and interest-bearing securities as a percentage of trailing 24-month average market capitalization iii) Accounts receivables as a percentage of trailing 24-month average market capitalization.

3. Literature Review

This section discusses the empirical literature on Islamic finance in order to explain why Islamic equity indices might perform better or worse than conventional indices. The section first highlights the impact of debt on the firm value, equity performance, and portfolio returns. This helps us to develop Hypotheses H1 and H2. Then we examine empirical studies that offer conflicting evidence regarding the performance of Islamic indices in the non-crisis and crisis periods. This helps us develop Hypotheses H3, H4, and H5.

3.1. The Debt Externality

The literature on capital structure theories suggests that higher debt compared to equity should have a positive impact on firm value. Two theories in particular favor this conclusion. These include the trade-off theory and the pecking order theory (see Fama and French, 2002). The trade-off theory suggests that firms balance the benefits of tax exemption of debt with the cost of the potential financial distress. In the presence of tax benefits and low financial distress cost, the use of debt would be favored over equity. The pecking order theory argues that in the presence of asymmetric information, the capital structure acts as a signaling mechanism. Issuance of equity sends the signal that the firm has a lower prospect for profitable projects, while the use of retained earnings and debt issuance confirms the firm's belief in profitable future prospects (Myers and Majluf, 1984). Extending these ideas, individual stocks and indices with higher debt should outperform those with lower debt. Particularly, when the economy is booming, according to the trade-off theory, the expected bankruptcy cost is likely to be low, hence, increasing debt would be associated with higher firm profitability. The same would be true for the pecking order theory, where increasing debt may send signals to the industry regarding the firm's better profitability potential. When the economy is doing well, higher debt firms, owing to their capital structure may experience increased profitability and better returns.[112] These capital structure theories ignore the externality of debt arising from the misjudgment of the investors regarding a possible downturn. The favorability of debt over equity, indicated by these theories, might hold true during a boom period when the financial markets are doing well but not so when the markets are experiencing a downturn. Mian and Sufi (2010, 2011, and 2015) and Mian, Sufi and Verner (2017) argue that during an upturn when the economy is doing well, households and firms might undermine the prospect of a downturn and hence might end up over borrowing. The real cost of this over borrowing would be manifested when the economy takes a downturn. Assuming exogenous business cycles, the externality of debt becomes inevitable. A somewhat similar conclusion but adopting a completely different approach, Gennaioli et al., (2012) make a case for the externality of debt securities by suggesting that the investor tends to neglect extreme risks. This neglect along with the investor's preference for relatively fixed returns implies an over-investment in debt securities. The investor eventually realizes his neglect of extreme risk, thereby resulting in a financial crisis. Extending the literature on capital structure theories and debt externality, indices with high debt stocks should perform better in an upturn and perform worse in a downturn.

H1: High leverage portfolios perform better than low leverage portfolios in the up market.
H2: Low leverage portfolios perform better than high leverage portfolios in the down market.

The empirical evidence related to the performance of Islamic indices which is a proxy for low debt indices is discussed next.

3.2. Performance of Islamic Equity Indices (IEI) versus Conventional Equity Indices (CEI)

The literature on the comparative performance of Islamic and conventional indices seems to offer conflicting evidence. Some studies suggest that Islamic indices outperform conventional indices while others find no significant difference between them. Some studies compare the performance by focusing on the crisis period; however, the results are again conflicting with some favoring Islamic indices over conventional indices during the crisis period while others find no significant difference. Some of these studies are examined below.

Ho et al., (2014) compared the global Islamic indices with the conventional indices using standard Sharpe and Treynor ratios. Their results indicate that IEI outperformed their CEI during the crisis. The results were found to be inconclusive for the non-crisis periods. They attribute this particular behavior to the 'conservative nature' of Islamic investments. Similarly, Ashraf et al., (2014) investigate the claim that global and regional IEI perform better compared to CEI. They find that overall IEI exhibit lower systematic risk, compared to their conventional benchmark, during the declining phase of capital markets.

Charles et al., (2015) study the impact of Shari'ah filtering criteria on the risk of Islamic indices. They also examine the effects of the (Global Financial Crisis, 2008) GFC using various performance measures with and without risk adjustment. Their findings suggest that the IEI outperform the counterpart CEI on a risk-adjusted basis over the full period (1996–2013). The results, in the case of sub-period samples, indicate that Islamic indices have a higher risk compared to conventional indices. They also found that in most cases either the Islamic indices outperform the conventional counterparts, or there is no significant difference in performance between them. They attribute this to lack of diversification of Islamic indices.

Arouri and et al., (2013) examines Islamic finance innovations may help investors to escape from a financial downturn. They find that the impact of the 2008 crisis on the Islamic finance industry is less significant than on conventional finance. Jawadi et al., (2014) analyzes the financial performance of Islamic and conventional indices and suggests that conventional funds perform better before the crisis and during periods of calmness; however, Islamic funds outperform them during the crisis. Hayat et al., (2011) examine the risk-return characteristics of Islamic equity funds. They find that Islamic equity funds underperform their conventional equity benchmarks. They report that the underperformance increases during the GFC. In sharp contrast, Hoepner et al., (2011) find that Islamic funds outperform international conventional equity indices.

Narayan et al., (2016) find that the market risk factors—namely, excess market returns, value, size, and betting-against-beta factors—and macroeconomic risk factors are appropriately priced in Islamic indices. They conclude that the profitability of Islamic stocks is merely a compensation for risks and is not due to mispricing. Similarly, Albaity and Mudor (2012) do not find any significant difference in mean returns between the Islamic and non-Islamic indices. They argue that the stock screening criteria in principle should eliminate bad stocks from the Islamic index, resulting in a possible reduction in the return's volatility during the crisis. Their empirical results, however, do not find this to be the case.

El Alaoui et al., (2016) examine the connection between debt and Islamic equities' performance. They investigate the relationship between the firm leverage and systematic risk while controlling for the Shari'ah stock screening rules. Their results suggest that the Islamic equities carry lower systematic risk because of their low debt. Nainggolan, How and

Verhoeven, (2015) study the link between the ethical screening and portfolio performance of Islamic equity funds They found that Islamic equity funds outperform conventional equity funds only during the 2008 financial crisis.

The conflicting results reported in the literature seem to be at odds with both the theories of capital structure and the externality of debt. If traditional theories of capital structure are to be extended, Islamic indices owing to their lower debt should perform worse than conventional indices. On the contrary, the literature focusing on the externality of debt would imply that the Islamic indices should outperform their conventional counterparts. One way to reconcile these theories with the conflicting empirical evidence is to decouple the up and the down market. The debt becomes profound when the markets are going down. It is during these times that Islamic indices with low debt stocks should outperform conventional indices with higher debt. When the markets are going up, then one would favor the traditional capital structure theories where Islamic indices having low debt stocks would underperform conventional indices. Similarly, one would expect Islamic indices to perform during a financial crisis.[113] The discussion motivates the following hypotheses.

H3: Islamic equity indices outperform conventional equity indices in all down markets.
H4: Conventional equity indices outperform Islamic equity indices in all up markets.
H5: Islamic equity indices outperform conventional equity indices during the crisis period.

4. Data and Descriptive Statistics

4.1 Firm Level Data

For the purpose of H1 and H2, we use monthly equity-level data for firms listed on the NYSE and NASDAQ from 1982 to 2016. There are 4131 non-financial listed firms (out of 5294 firms), out of which 2044 firms belong to NYSE and 2283 to NASDAQ. The data is acquired from Datastream. The paper employs 1-month US TBill return provided by Ibbotson and Associates, Inc. The data consists of firms' total stock returns including dividends (frequency: monthly), the market capitalization (frequency: monthly), the book equity to market equity ratio (frequency: monthly), and the debt-to-asset ratio (frequency: annual).
The paper uses total return index. The index uses adjusted closing prices, which takes into account dividends, splits, and repurchases as given in Eq. (1).

$$RI_t = RI_{t-1} \times \frac{P_t + D_t}{P_{t-1}} \tag{1}$$

Where RI is the return index, P is the share price, D the dividend paid, and t is the time period. The firm market equity is defined as the number of ordinary shares outstanding per share class in the issue multiplied by the share price. The paper employs the firm leverage, which is given in Eq. (2):

$$Leverage\ (\%) = \frac{Total\ Debt}{Total\ Assets} \tag{2}$$

Where *Total Debt= Long Term Debt +Short Term Debt and Current Portion of Long Term Debt*. As discussed in Fama and French (1992), the paper excludes financial firms since the leverage for financials firms does not have the same meaning for nonfinancial firms, in which high leverage more likely indicates financial distress.
Finally, the sample data are split into two groups based on high and low leverage:

[113] We argue in this paper that the low debt nature of Islamic equities would cause them to outperform other indices, in all down markets, holding all other factors constant. Our argument is based on conditions of ceteris paribus. We are not in any way suggesting that Islamic indices would outperform other indices in all states of the world.

i) High Leverage Portfolios (HLP): The portfolio includes the *top* 40% firms based on leverage. In terms of debt-to-asset ratio, the threshold translates to an average of 40% with a maximum of 70%.
ii) Low Leverage Portfolios (LLP): The portfolio includes the *bottom* 40% firms based on leverage. In terms of debt-to-asset ratio, the threshold translates to an average of 11% and a maximum of 18%.

The proportion of 40% helps us to have a balanced and normalized data set, however, a robustness test is added in order to test the stability of results with respect to variation in debt-to-asset ratio. Particularly, we have tested the low leverage portfolios at the 33% debt-to-asset ratio, which is the threshold for Islamic equity screening.

4.2 Index Level Data

In order to test Hypothesis H3 to H5, we use the Dow Jones family of indices for both the IEIs and CEIs. Selecting Islamic and conventional indices from Dow Jones minimizes the biases that may arise due to differences in index methodology or Shari'ah screening methods. Within the Dow Jones family of indices, the Dow Jones Islamic Market (DJIM) Index remains the premier benchmark of investment performance for the global universe of Shari'ah compliant equities. On the conventional front, we select Dow Jones Global Index (DJGI)[114]. The study employs the Dow Jones Global Titan 50 Index (DJGT) as benchmark index.

Table 1.1 shows the information on the selected Islamic, conventional and benchmark indices. The DJIM and DJGI closely follow each other in terms of mean and median market capitalization and therefore comparable indices. The benchmark index DJGT comprises highly liquid and large firms, in order to account for size effect, we use index market capitalization as an additional factor. Other characteristics are found to be similar across all sample indices.

Table 1.1: Market Indices Description

The DJIM and DJGI are float-adjusted market capitalization (CAP) indexes. Data is in USD as of the end of May 2016. Excl. Neg. stands for excluding negative price to earnings ratio.

	Dow Jones IEI Index (DJIM)	*Dow Jones Global Index(DJGI)*	*Dow Jones Global Titans 50 Index (DJGT)*
Data (Daily)	1/1/2001-12/31/2015	1/1/2001-12/31/2015	1/1/2001-12/31/2015
Mean (CAP)	7.2 Billion	5.7 Billion	151.4 Billion
Median: (CAP)	1.2 Billion	1.1 Billion	133.9 Billion
Trailing PE (Excl. Neg.)	21.00	16.95	18.59
Price to Book	3.16	1.92	2.24
Dividend Yield	2.20	2.59	2.91
Price to Sales	2.01	1.41	2.24
Price to Cash Flow	13.01	10.72	10.27
Country Allocation (Greater than 2%)			
United States	60.34%	51.24%	76.39%
Japan	6.54%	8.95%	2.42%
United Kingdom	5.80%	6.42%	6.66%
Canada	2.16%	3.15%	-
France	2.39%	3.08%	2.37%
Switzerland	4.82%	2.89%	7.03%
Germany	2.32%	2.79%	-
China	2.61%	2.74%	-
Australia	-	2.27%	-
Total	**86.98%**	**83.53%**	**94.87%**

[114] As of July 2016, the Dow Jones Islamic Market World Index has 2586 components compared to 7279 components in the Dow Jones Global Index.

To examine the risk-adjusted return performance of the respective indices in line with H3, H4, and H5, the data has been divided into two phases. The pre-crisis phases stretch between 1/1/2001 –31/12/2007 (7 years), the crisis phase ranges between 1/1/2008 –31/12/2011 (4 years) and the post-crisis phase is between 1/1/2012-1/1/2015(4 years). While testing the robustness of the results, we employ liquidity based controlled variables for which the available data is limited and extends between 1/1/2008 –31/12/2012 (5 years). We reconstitute the sample including the crisis and non-crisis periods ranging from 1/1/2008 –31/12/2012 (5 years) and 1/1/2009-31/12/2012(4 years) respectively. We take US 3-month T-bill rate as the risk-free rate. The data have been obtained from Bloomberg[115].

The descriptive statistics of the sample data are shown in Table 1.2. The total sample exhibits little difference between the Islamic and conventional indices in terms of risk, return, skewness as well as kurtosis. The IEI appears to be riskier compared to CEI. Similar results prevail during pre-crisis. However, during the crisis, the IEI exhibits lower losses at a lower risk compared to CEI, essentially displaying resilience against the overall market fall. In post-crisis, it appears that the risk-return equilibrium is restored given that the IEI being riskier offers higher returns and vice versa in case of CEI.

Looking at the bottom panel, Liquidity and Return Data (2008-2012), the IEI appears more liquid compared to CEIs. Given that the lower liquidity number shows higher liquidity level the IEI appears more liquid and offers higher returns (lower losses) at a lower risk compared to CEI during 2008-2012. The particular liquidity statistics may be pointing towards the debt component which is lower in the case of IEI compared to CEI.

5. Methodology

In order to test the Hypotheses H1 and H2, we construct the factors (control variables) which include the excess market return, size, value, momentum, and leverage. The impact of these factors is assessed on dependent variables which include 25 single portfolios sorted based on leverage and 5X5 double sorted quantiles (based on size and leverage).

These portfolios are divided into HLP (High Leveraged Portfolio) and LLP (Low Leveraged Portfolios) groups. Using Fama Macbeth (1973) two step regressions (time series and cross-sectional), we estimate the risk exposures (betas) for each group. Next, we compare the performance of HLP and LLP by categorizing the factors into up and down market based on whether the excess market return for a given month is positive or negative.

In order to test the hypotheses H3, H4 and H5 we divide the sample for Islamic equity indices into the up and down markets. Next, we compare the risk-adjusted returns using Sharpe and Treynor ratio to test hypotheses H3-H5 while employing rolling analysis to test the stability of the resulting parameters. Finally, we run robustness tests based on liquidity-adjusted CAPM to control for different channels of liquidity risk.

[115] The volume data for DJGI is available until December 2012 which is critical in estimating index liquidity. Other indices such as MSCI and FTSE as well as country indices have limited data on Islamic indices mostly missing partially or completely the crisis-period especially in case of volume data. This may limit the possible robustness tests.

Table 1.2: Descriptive Statistics

The index returns are calculated using $\ln\left({P_t}/{P_{t-1}}\right)$. The returns are multiplied by 100. The market level liquidity is calculated using the Eq. (1) i.e., $ILLIQ_{i,t} = \frac{1}{D_{i,t}}\sum_{d=1}^{D_{i,t}}\frac{|R_{i,t,d}|}{V_{i,t,d}}$ where $D_{i,t} = 260\ days$. The liquidity is standardized around mean and then using Eq. (2), we fit an AR (2) model to eliminate any autocorrelation. The residual of the AR (2) model serves as Liquidity. The daily data ranges from 1/1/2001-12/31/2015 for all samples except liquidity for which the daily data ranges from 1/1/2008-12/31/2012. The US 3-month T-bill is used as a risk free rate. The returns are tested for normality using Jarque-Bera test. * represents the parameters significance at 5% confidence level.

Index	Log Returns (x100)	Std.Div.	Skew	Kurtosis
Total Sample: 2001-2015				
DJIM-Mkt	0.010*	1.045	-0.326	10.949
DJGI-Mkt	0.010*	1.023	-0.389	10.590
DJGT-Mkt	0.000*	1.073	-0.177	11.240
3MTBill	0.006*	0.007	0.949	2.492
Pre-Crisis: 2001-2007				
DJIM-Mkt	0.013*	0.911	-0.074	5.137
DJGI-Mkt	0.020*	0.841	-0.095	5.118
DJGT-Mkt	0.002*	0.952	0.056	6.722
Crisis: 2008-2011				
DJIM-Mkt	-0.015*	1.467	-0.337	8.953
DJGI-Mkt	-0.027*	1.486	-0.350	7.693
DJGT-Mkt	-0.031*	1.492	-0.196	9.032
Post-Crisis: 2012-2015				
DJIM-Mkt	0.029*	0.702	-0.330	5.392
DJGI-Mkt	0.028*	0.688	-0.421	5.660
DJGT-Mkt	0.028*	0.712	-0.264	5.596
Liquidity and Return Data: 2008-2012				
DJIM-Mkt	-0.007*	0.014	-0.353	9.894
DJGI-Mkt	-0.015*	0.014	-0.378	8.537
DJGT-Mkt	-0.019*	0.014	-0.224	10.194
DJIM-Liq	0.177*	0.769	1.164	5.849
DJGI -Liq	0.284*	0.849	3.086	30.130
DJGT-Liq	0.357*	0.759	0.936	4.093

5.1 Firm Level

5.1.1 Factor Portfolios

The CAPM, which is a single factor model is calculated for each month by taking NYSE index return and subtracting the corresponding risk-free rate, to get the excess return of the market ("ERM") as given in Eq. (3),

$$ER_{it} = \alpha_i + \beta_{i,ERM}ERM_t + \varepsilon_{it} \tag{3}$$

Where ER_{it} is the monthly stock return in excess of the risk-free rate for portfolio i in month t. Alpha (α_i) is the risk-adjusted abnormal return (pricing error). ERM_t is the excess return on the market portfolio. The ε_{it} denotes the error term. In addition to market factor, we construct four additional factors, with size and value related factors corresponding to Fama and French (1993), a momentum factor related to Carhart (1997) and finally a leverage factor which is constructed similar to value factor except that value is replaced with leverage.

We construct the factors using six equally-weighted portfolios formed on size (ME) and book-to-market (BE/ME). The firms included in the sample are ranked for each year at the end of June based on their ME and BE/ME and then formed into portfolios. If a firm during a given year does not have any ME or BE/ME ranking it is not included in the factor relevant for that year. Similarly, the firm is only included if it has all the 12 monthly returns in the subsequent holding period. The median stock size is used to split the firms into two groups, Small S and Big B. The sample is then split into three groups based on BE/ME, where the bottom percentile is 30% (Low L or Growth), the middle percentile is 40% (Neutral M), and the top percentile is 70% (High H or Value). The six portfolios that are formed are; S/L, S/M, S/H, B/L, B/M, and B/H. For example, the portfolio S/L contains firms with small market values and low book-to-market ratios. The BE/ME is used to sort the portfolios in June of each year. The size ranking is carried out using a similar procedure.

5.1.1.1 Size Factor

The portfolio SMB is the difference each month between the simple average of the returns on the three small stock portfolios, S/L, S/M, and S/H, and the simple average of the returns on the three big stock portfolios, B/L, B/M, and B/H. The SMB portfolio, which is meant to mimic the risk factor related to size, is thus the difference between the returns on small and big stock portfolios with about the same weighted average book-to-market ratios. The Small-minus-Big, SMB factor is calculated using Eq. (4):

$$SMB = \frac{1}{3}(small\ value + small\ Neutral + small\ growth) - \frac{1}{3}(big\ value + big\ Neutral + big\ growth) \tag{4}$$

5.1.1.2 HML Factor

The portfolio HML is defined similarly and is thus the difference each month between the simple average of the returns on the two high BE/ME portfolios, S/H and B/H, and the simple average of returns on the two low BE/ME portfolios, S/L and B/L. HML is meant to mimic the risk factor in return related to book-to-market equity, and should largely be free of the size factor in returns. The High-minus-Low, HML factor is calculated using Eq. (5):

$$HML = \frac{1}{2}(small\ value + big\ value) - \frac{1}{2}(small\ growth + big\ growth) \tag{5}$$

5.1.1.3 Momentum Factor

In order to construct a factor mimicking portfolio for momentum in stock returns, we employ the same method as Carhart (1997). The factor WML is defined as the equal-weighted average return of firms with the highest 30% eleven month returns lagged one month minus the equal weighted average return of firms with the lowest 30% eleven-month returns lagged one month. The portfolios are rebalanced at the end of June each year.

5.1.1.4 Leverage Factor

For leverage, we form a portfolio to mimic the risk factor related to the leverage of firms. At the end of June each year, all firms are ranked based on their leverage as reported for December t-1. Similar to the above treatment of HML and WML, we group firms based on the breakpoints for the bottom 30% (Low), middle 40% (Neutral), and top 30% (High). The difference each month between the simple average of the high leverage firms' returns and the simple average of the low leverage firms' returns is used to create the High-Leverage-minus-Low-Leverage, HLMLL portfolio which is calculated using Eq. (6):

$$HMLL = \frac{1}{2}(small\ highDA + big\ HighDA) - \frac{1}{2}(small\ lowDA + big\ lowDA) \tag{6}$$

5.1.2 Regression Portfolios

In order to produce the empirical results needed to test our hypotheses, we use equally weighted portfolios[116]. The forming of portfolios is desirable as it reduces the residual variance of the estimated betas and produces more stable betas over time. It also avoids the problem of dealing with individual stock returns that can be very volatile and yield results that cannot reject the proposition that all average returns are equivalent (Cochrane (2005)). Furthermore, the portfolios are divided based on High (HLP) and Low (LLP) leveraged. The portfolios formation takes two forms:

i) ***Single Sorting***: Forming 25 equally weighted portfolios sorted based on leverage, such that portfolio 1 consists of firms with low leverage while portfolio 25 consists of firms with high leverage ratio.

ii) ***Bivariate Sorting***: As a robustness measure, we construct 5X5 quintiles, i.e., the firms are sorted based on size and leverage similar to Fama and French (1993).

5.1.3 Testing Models

5.1.3.1 Time Series Regression

The study employs Fama Macbeth (1973) two step regression method. The first step is given in Eq. (7) below.

$$ER_{it} = \alpha_i + \beta_{i,ERM}ERM_t + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,WML}WML_t + \beta_{i,HMLL}HMLL_t + \varepsilon_{it} \quad (7)$$

Where t=*1980…2016=T* and i represent portfolio i=*1,2…25.* ER_{it} is the monthly stock return in excess of the risk-free rate for portfolio i in month t. Alpha (α_i) is the risk-adjusted abnormal return, known as the pricing error. ERM_t represents the excess return on the market portfolio (market risk premium). The other factors such as SMB_t (small minus big), HML_t (high BE/ME value minus low BE/ME), WML_t (winner minus losers or momentum), $HMLL_t$ (high leverage minus low leverage). The ε_{it} denotes the error term.

In order to estimate the up and down market conditions, we multiply the Eq. (7) with $\delta = 1$ if $(ERM) > 0$ i.e., when the market excess returns are positive, and $\delta = 0$ if $(ERM) < 0$ i.e., when the market excess returns are negative. We use NYSE Composite index as market index.

5.1.3.2 Cross Sectional Regression

The second step is to estimate the regression given in Eq. (8).

$$ER_i = \gamma_0 + \gamma_{ERM}\beta_{i,ERM} + \gamma_{SMB}\beta_{i,SMB} + \gamma_{WML}\beta_{i,WML} + \gamma_{HML}\beta_{i,HML} + \gamma_{HMLL}\beta_{i,HMLL} + \alpha_i \quad (8)$$

Where t=*1980…2016=T* and i represent portfolio i=*1,2…25.* ER_{it} is the average monthly stock return in excess of the risk-free rate for portfolio *i* in month *t*. $\beta_{i,ERM}$ is the market beta. $\beta_{i,SMB}$ is the natural logarithm of the average portfolio size. The $\beta_{i,HML}$ is the average portfolio book-to-market ratio. The $\beta_{i,WML}$ is the momentum beta. The $\beta_{i,HMLL}$ is the average portfolio leverage. The respective betas are estimated from the time series regression i.e., from Eq. (7) The parameter γ is the premium for each respective beta. The γ_0 is the intercept and α_i is the error term. Since Eq. (8) almost always suffers from heteroscedasticity and autocorrelation, we replace the standard t-test with Newey and West (1987) and Hansen and Hodrick (1980) which provide heteroscedasticity and autocorrelation adjusted standard errors.

[116] The results based on value weighted portfolios are not different while using equally weighted portfolios.

5.2 Index Level

5.2.1 Up and Down Market

We employ Pettengill, Sundaram and Mathur (1995), hereafter called the PSM method to categorize up and down markets. It is argued that in CAPM as given in Eq. (3), for the $\beta_{i,ERM}$ to be a useful measure of risk, a systematic relationship should exist between the beta and the returns. The existence of a large number of negative excess returns shows that the positive correlation between beta and realized returns may be biased. In order to rectify the bias, PSM uses a dummy variable to estimate the lambda (λ) for up and down market separately as given in Eq. (9),

$$ER_{it} = \hat{\lambda}_{0t} + \hat{\lambda}_{1t} * \delta * \beta_{i,ERM} + \hat{\lambda}_{2t} * (1 - \delta) * \beta_{i,ERM} + \epsilon_t \tag{9}$$

Where $\hat{\lambda}_{1t}$is the market risk premium estimated when $\delta = 1$ if $ERM_t > 0$ that is when the market excess returns are positive, and $\hat{\lambda}_{2t}$ is the risk factor estimated $\delta = 0$ if $ERM_t < 0$ when the market excess returns are negative.

5.2.2 Market Risk-Adjusted Return Analysis

In order to test the hypotheses, H3-H5, we estimate Sharpe and Treynor performance ratios. Sharpe ratio shows the investment's return per unit of total risk, where total risk is estimated as the standard deviation of returns. A higher Sharpe ratio implies a higher probability of the index return exceeding the risk-free return. The Sharpe ratio is calculated as $SR = \frac{E(R) - rfr}{\sigma}$ where $E(R)$ is the expected return for the index over the period, rfr is average of the risk free rate and σ is standard deviation of index return.

The second performance measure is Treynor ratio which measures the index performance for a given level of market risk. It is computed as $TR = \frac{E(R) - rfr}{\beta}$. It uses beta or systematic risk as a measure of total risk. A higher Treynor ratio indicates superior performance. The average $E(R)$ of each index and the parameters such as σ, β are tested against the mean using standard student's t-test.

In time series, a rolling analysis can be used to validate the model's parameter stability over time. When analyzing financial time series data using a statistical model, a key assumption is that the parameters of the model are constant over time. However, the economic environment often changes considerably, and it may not be reasonable to assume that a model's parameters are constant. A common technique to assess the constancy of a model's parameters is to compute parameter estimates over a rolling window of a fixed size through the sample. If the parameters are truly constant over the entire sample, then the estimates over the rolling windows should not be significantly different. If the parameters change at some point during the sample, then the rolling estimates should capture this instability.

We perform the rolling analysis with respect to mean returns, the standard deviation in returns and market sensitivity (beta) over 30 and 120 days moving the window. We then compute Sharpe and Treynor ratios in order to analyze the change in performance, if any.

5.2.3 Liquidity Risk Adjusted Return Analysis

We compare the performance of Islamic indices with conventional indices during both the upturns and downturns in the presence of liquidity risk. In the standard CAPM model, a key assumption is the presence of frictionless markets which implies that a security can be traded at zero cost. In reality, however, market frictions exist. Liquidity, defined as 'the ease of trading a security', is considered as one of the most important frictions in the market (Amihud et al., 2006). We use Liquidity-adjusted Capital Asset Pricing Model (LCAPM) to examine the liquidity risk premium differential in the presence of four channels through which liquidity risk might affect the market returns (Acharya and Pedersen, 2005). These different channels include co-movement between IEI and CEI liquidities (Cochrane, 2001), co-movement between IEI return and CEI liquidity (Pastor and Stambaugh, 2003), and co-movement between IEI liquidity and CEI returns (Acharya and Pedersen, 2005). Furthermore, we also study these risks at the aggregate level.

We use the Amihud (2002) measure of illiquidity (ILLIQ). It takes into account the impact of trade order on returns. It also qualifies the Kayle (1985) concept of illiquidity[117]. The Amihud (2002) ratio is considered as reliable price impact measure when dealing with a lower frequency of data sets (Goyenko et al., 2009; Hasbrouck, 2009). ILLIQ[A] is computed as follows.

$$ILLIQ_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{|R_{i,t,d}|}{V_{i,t,d}} \tag{10}$$

Where $|R_{i,t,d}|$ is the absolute in returns of stock i in any month $t, V_{i,t,d}$ is the dollar denominated trading volume for stock i on day of month t, and $D_{i,t}$ is the total trading days for stock i in month t. The ratio $|R_{i,t,d}|/V_{i,t,d}$ gives absolute change in return per dollar traded (dollar cost per dollar invested) or daily price impact. Essentially, higher value of the ratio is associated with lower liquidity. Since we are using index, therefore the Eq. (10) is measured using index returns i.

Liquidity is empirically documented to be highly persistent and therefore exhibits strong first-order autocorrelation. Therefore, the study employs the AR (2) as given in Eq. (11) and uses the residuals as ILLIQ. Furthermore, the Eq. (11) is mean-adjusted in line with Amihud, (2002).

$$C_t^i = \alpha_0 + \alpha_1 C_{t-1}^i + \alpha_2 C_{t-2}^i + \cdots + \alpha_x C_{t-x}^i + \nu_t^i \tag{11}$$

Where, C_{t-1}^irepresents the liquidity of index i at time t. In order to capture the liquidity effects, we use the unconditional version of LCAPM. The model is essentially derived under the assumption of constant conditional variance and is given by Eq. (12) below:

$$E\left(r_t^i - r_t^f\right) = E\left(c_t^i\right) + \lambda^1\beta^{1i} + \lambda^2\beta^{2i} - \lambda^3\beta^{3i} - \lambda^4\beta^{4i} \tag{12}$$

where r_t^i is the return of index i at time t, r_t^f is the risk free rate time t, $\lambda^1 = E(\lambda_t) = E(r_t^M - c_t^M - r^f)$ (the factor loading), r_t^M is the market return of the index i at time t, c_t^i is the liquidity cost of index i at time t and c_t^M is the liquidity cost of the market M at time t. The betas represent different channels through which liquidity affects the returns. The betas are defined as follows:

$$\beta^{1i} = \frac{cov(r_t^i, r_t^M - E_{t-1}(r_t^M))}{var(r_t^M - E_{t-1}(r_t^M) - [c_t^M - E_{t-1}(c_t^M)])} \tag{13}$$

$$\beta^{2i} = \frac{cov(c_t^i - E_{t-1}\left(c_t^i\right), c_t^M - E_{t-1}(c_t^M))}{var(r_t^M - E_{t-1}(r_t^M) - [c_t^M - E_{t-1}(c_t^M)])} \tag{14}$$

$$\beta^{3i} = \frac{cov(r_t^i, c_t^M - E_{t-1}(c_t^M))}{var(r_t^M - E_{t-1}(r_t^M) - [c_t^M - E_{t-1}(c_t^M)])} \tag{15}$$

[117] Kyle (1985) proposed that because market makers cannot distinguish between order flow that is generated by informed traders and by liquidity (noise) traders, they set prices that are an increasing function of the imbalance in the order flow which may indicate informed trading. This creates a positive relationship between the order flow or transaction volume and price change, commonly called the price impact.

$$\beta^{4i} = \frac{cov(c_t^i - E_{t-1}(c_t^i), r_t^M - E_{t-1}(r_t^M))}{var(r_t^M - E_{t-1}(r_t^M) - [c_t^M - E_{t-1}(c_t^M)])} \tag{16}$$

The Eq. (13) represents the standard CAPM beta-adjusted for trading costs. The β^{2i} in Eq. (14) shows the commonality between index and market liquidity, i.e., the investor is expected to be compensated for holding an illiquid asset when the market as a whole is illiquid leading to a positive sign. The argument further supports the wealth effect (Cochrane, 2001). The β^{3i} in Eq. (15) represents the commonality between the index return and market liquidity, i.e., the investor will accept lower return on an asset that pays high return in the presence of an illiquid market (Pastor and Stambaugh, 2003) resulting in a negative sign. Eq. (16) show the co-movement between the index liquidity and market returns. Acharya and Pederson (2005) explain this relationship by suggesting that when the markets declines, the ability to sell easily and quickly becomes more valuable. Hence, an investor is willing to accept a discounted return on index with low illiquidity costs when the market returns are low. Hence, the expected sign of β^{4i} is negative. The $\lambda^1, \lambda^2, \lambda^3$ and λ^4 represents the respective factor risk premiums.

The total effect of liquidity risks is given in Eq. (17) leading to an estimation of aggregate liquidity risk using LCAPM given in Eq. (18) respectively.

$$\beta^{5i} = \beta^{2i} - \beta^{3i} - \beta^{4i} \tag{17}$$

$$E(r_t^i - r_t^f) = \alpha + \kappa E(c_t^i) + \lambda^1\beta^{1i} + \lambda^5\beta^{5i} \tag{18}$$

Similarly, the total systematic risk, i.e., market risk plus liquidity risk is given by Eq. (19) leading to an estimation of aggregate systematic risk given in Eq. (20) respectively.

$$\beta^{6i} = \beta^{1i} + \beta^{2i} - \beta^{3\mathrm{i}} - \beta^{4\mathrm{i}} \tag{19}$$

$$E(r_t^i - r_t^f) = \alpha + \kappa E(c_t^i) + +\lambda^6\beta^{6i} \tag{20}$$

In the first step, the monthly coefficients of Eq. (13) to (16) are tested against the respective means using t-statistics. In the second step, we employ time series regression on Eq. (3), (12), (18) and (20) respectively.

5.2.4 Size Effect

In addition to the liquidity channels, we control for the size effect which is based on market capitalization of the index. The size difference among the indices can create a bias since the small firms are generally less traded (less liquid) than the big firms (more liquid). The size effect has a negative relationship with returns (Banz, 1981; Reiganum, 1981; Fama and French, 1992).

The size effect is generally associated with liquidity such that small-cap stocks are less liquid than large-cap stocks and; therefore, provide correspondingly higher returns to offset the higher transactions costs (e.g., Brennan, Chordia, and Subramanian, 1998). The size effect provides another dimension of liquidity, i.e., the trading difference between the small (less traded) versus large (highly traded) firms in the respective index as noted by Amihud and Mendelson (1986) and Berk (1995).

Firm size is thought to proxy for underlying risk factors associated with smaller firms. Observed variations in the size effect can be explained by such underlying factors like market liquidity that change over time (Crain, 2011). We capture the size effect by including the natural log of the market capitalization of the benchmark index and including it as an additional factor in Eq. (3), (12), (18) and (20) respectively.

6. Results and Discussion

We find strong evidence to support Hypotheses H1 and H2. For Hypothesis H1, the firm level results suggest that the high leveraged portfolio (HLP) outperform the low leveraged portfolio (LLP) for the entire data sample, along with the two crises we have examined (1998 and 2008). The results are found to be robust for univariate (25) and bivariate (5X5) portfolios under both equal and value weighting schemes. Additionally, to proxy Islamic screening criteria, the LLPs are constrained to include firms with a debt-to-asset ratio of less than or equal to 33% while HLPs include firms with debt-to-asset greater than 33%. For the Hypotheses H3 to H5, our results support the hypotheses. For Hypothesis H3, the index level results suggest that the IEIs outperform the CEIs during the down market. In the up market, CEI outperforms IEI supporting our Hypothesis H4. The results indicate that for the crisis period, the IEIs outperform the CEIs supporting Hypothesis H5. These results are robust to rolling analysis and hold even after incorporating liquidity and size effects.

6.1 Firm Level Analysis

Initially, we construct the risk factors MKT, SMB, HML, WML, and HMLL as given in section 5.1.1. Using Eqs. (3) to (6), the respective factor summary statistics are given in Table 1.3. The negative SMB reflects that the large cap firms outweigh the small cap firms. Similarly, the positive HML shows the spread in returns between the value and growth stocks with value stocks contributing more than the growth stocks. The negative HMLL (Leverage) shows the spread in returns attributed to low leveraged firms. This may be due to the fact that leverage entails more losses than the widely ascribed gains.

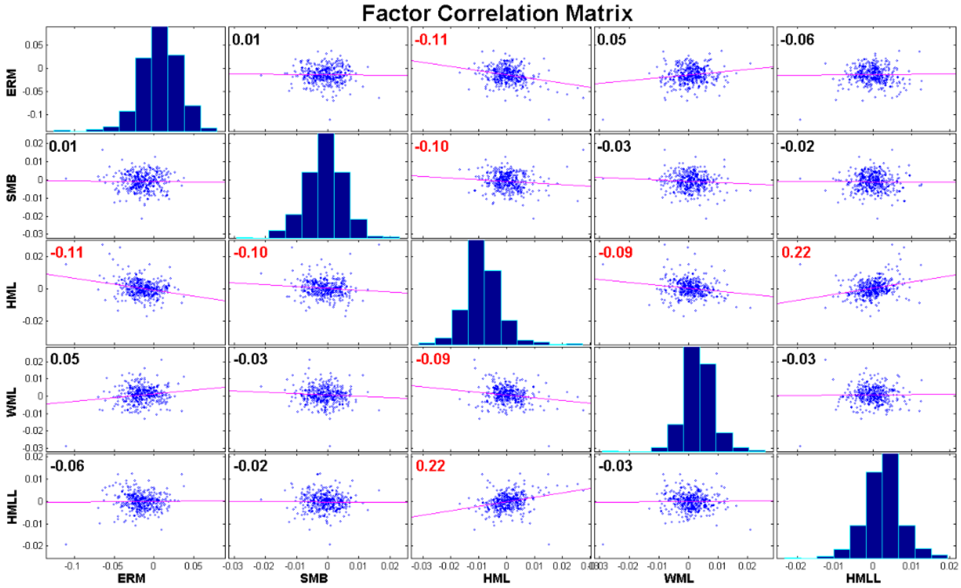
Table 1.3 : Descriptive Statistics (Equity Factors)

The ERM excess market returns obtained by subtracting one month US TBill rate from monthly returns from the NYSE composite index. The SMB and HML are Fama and French, 1993 factors for ‘Small minus Big (market capitalization)’ and ‘High minus Low (book to market ratio)’, WML is the Carhart (1997) factor for ‘Winner minus Loser’, HMLL is the leverage risk factor (High minus Low Leverage). The data consist of 420 months from 1982-2016.

	ERM	SMB	WML	HML	HMLL
Mean (X100)	0.13	-0.09	0.08	0.04	-0.01
S.D (X100)	1.24	0.45	0.47	0.50	0.37
Skewness	-2.33	-0.20	-0.31	0.90	-0.17
Kurtosis	20.75	4.13	7.24	7.27	5.44
Maximum	0.04	0.02	0.02	0.03	0.01
Minimum	-0.11	-0.02	-0.03	-0.02	-0.02

The correlation between different factors is shown in Figure 1.1, which essentially captures certain interesting dimensions of the data. The most pertinent include the i) correlation between the HML and the ERM (-0.25) indicating that the book to market spread has a larger share in ERM compared to other factors. ii) the correlation between the HML and HMLL is 0.28 which is in line with Fama and French (1992) that HML measures the financial distress and; therefore, the leverage need not be explicitly included in the cross sectional regression. In contrast, Peterkort and Nielsen (2005) suggested that though the two share commonality, leverage measure more than what HML captures. This allows us to use leverage as a separate risk factor. The leverage has a negative correlation with SMB suggesting that small size portfolios will have larger leverages and vice versa. In addition, the leverage is positively correlated to ERM as well as the HML. The correlation coefficients highlighted in red indicate which pairs of variables have correlations significantly different from zero. For these time series, all pairs of variables have correlations significantly different from zero.

Figure 1.1: Risk Factors Correlation Matrix



6.1.1 Time Series Regressions

In order to test the Hypotheses H1 and H2, we estimate the Eq. (7) to test the basic relationship between leverage factor and the returns. To find the evidence of a similar relationship in the long term, we will conduct cross sectional regression in section 6.1.2. The time series regression results for the entire sample are given in Table 1.4. The GRS test statistic is used to analyze whether the alphas are jointly zero. The GRS does not reject the null hypothesis that the alphas are jointly zero. This means that the model is well-specified in explaining the returns across the portfolios. The $\beta_{i,HMLL}$ is negative for portfolios 1 to 17 and becomes positive until the 25th portfolio. The trend shows that the $\beta_{i,HMLL}$ for low to medium leveraged portfolios is negative while it is positive for high leveraged firms. This means that firms holding low to medium debt have lower returns while the high leveraged firms tend to have higher returns on account of leverage risk. Interestingly, the key statistics in Table 1.4 show that the positive leverage risk is largely associated with small cap firms with the characteristic of being value stocks. The opposite is true for the negative leverage risk.

Similarly, the $\beta_{i,HMLL}$ results for the Low Leveraged Portfolios (LLPs) sample is given in Table 1.5. The results corroborate the results in Table 1.4 such that the leverage parameters for LLPs are negative and significant at 5% confidence level. This shows that the type of firms that carry lower debt-to-asset ratio are big firms. The average $\beta_{i,HMLL}$ is around -0.32, thus reducing the returns. The result is consistent with the notion that the value firms should be bigger in size and hold lower debts compared to smaller-growth firms.

The $\beta_{i,HMLL}$ results for the High Leveraged Portfolios (HLPs) is given in Table 1.6. The results again corroborate the results in Table 1.4 showing that the leverage parameters for HLPs are largely significant and positive. The result show that the small firms carries higher average debt-to-asset ratios. The average $\beta_{i,HMLL}$ is around 0.42, this higher risk should entail a higher return.

The leverage risk ($\beta_{i,HMLL}$) for the entire sample, LLP and HLP are graphically shown in Figure 1.2. The figure show the portfolio leverage risk for both LLP and HLP and the entire sample in the up and down market conditions[118]. The figure indicates that the down market is influencing the entire sample far more than the up market. This result signifies that the higher number of small firms in an equally weighted portfolio creates a bias in favor of small firms, thus resulting in negative leverage coefficient during the cross sectional regression.

Table 1.4: Time Series Regression (Univariate Sorting, Complete Sample)

The dependent variables consist of excess returns of 25 leveraged-based sorted portfolios. The independent variable consists of ERM, the market risk premium, SMB and HML are Fama and French, 1993 factors for 'Small minus Big (market capitalization)' and 'High minus Low (book to market ratio)', WML is the Carhart (1997) factor for 'Winner minus Loser', HMLL is the leverage risk factor (High minus Low Leverage). The 1-month US TBill return is used as the risk-free rate. The data consist of 420 months from 1982- 2016. The GRS, Gibbons, Ross, Shanken (1989) asset pricing test. The * represents that values are significant (at least) at 5% significance. The key statistics pertaining to each portfolio are given for debt-to-asset ratio (D/A), the absolute amount of debt, market capitalization (Cap) and book to the market ratio (B/M).

	Risk Factors								Key Statistics			
Portfolio	$\alpha_{i \times 100}$	$\beta_{i,ERM}$	$\beta_{i,SMB}$	$\beta_{i,HML}$	$\beta_{i,WML}$	$\beta_{i,HMLL}$	R^2	Adj R^2	D/A	Debt	Cap	B/M
Low Lev	0.11*	1.04*	0.36*	-0.03	0.06	-0.59*	0.89	0.88	0.04	110 M	4 B	0.53
2	0.03	1.00*	0.26*	-0.09	0.18*	-0.53*	0.89	0.88	0.04	285 M	3 B	0.57
3	0.05	0.99*	0.23*	-0.29*	0.07	-0.27*	0.88	0.88	0.06	573 M	7 B	0.57
4	0.04	1.01*	0.16*	-0.05	0.09	-0.42*	0.89	0.89	0.08	962 M	8 B	0.53
5	0.01	1.01*	0.15*	-0.15*	-0.12*	-0.13	0.90	0.90	0.10	778 M	6 B	0.58
6	-0.02	0.98*	-0.17*	-0.18*	-0.09	-0.31*	0.90	0.90	0.12	1253 M	10 B	0.54
7	0.04	1.01*	-0.16*	0.03	-0.08	-0.23*	0.90	0.90	0.15	1273 M	9 B	0.55
8	0.01	1.00*	-0.07	0.10*	-0.01	-0.39*	0.93	0.93	0.16	1702 M	9 B	0.57
9	0.04	1.01*	0.04	0.04	0.06	-0.15*	0.92	0.92	0.17	1963 M	10 B	0.56
10	0.01	1.01*	-0.20*	-0.10	-0.01	-0.14	0.90	0.90	0.19	1843 M	9 B	0.59
11	0.04	1.04*	-0.27*	-0.07	-0.10	-0.26*	0.91	0.91	0.20	1995 M	9 B	0.62
12	0.01	1.04*	-0.08	0.12*	0.00	-0.16*	0.92	0.92	0.22	2116 M	9 B	0.59
13	-0.06	1.00*	-0.10	0.19*	-0.04	0.01	0.90	0.90	0.23	2532 M	8 B	0.56
14	0.00	1.02*	-0.19*	0.04	-0.05	-0.05	0.91	0.91	0.25	2372 M	7 B	0.58
15	-0.01	1.02*	-0.23*	-0.01	0.05	-0.25*	0.93	0.93	0.27	2423 M	6 B	0.63
16	-0.06	0.98*	-0.11*	0.13*	0.13*	-0.14*	0.92	0.92	0.29	2756 M	7 B	0.68

[118] The tables for the LLP and HLP in up and down markets are available on request.

17	-0.02	0.98*	-0.12*	0.08	-0.02	-0.08	0.92	0.92	0.31	2466 M	6 B	0.65
18	0.00	1.00*	-0.15*	0.13*	-0.07	0.12	0.91	0.90	0.31	2767 M	5 B	0.63
19	-0.03	0.97*	-0.13*	0.12*	-0.13*	0.40*	0.92	0.92	0.34	3632 M	6 B	0.65
20	0.03	1.02*	-0.15*	-0.10	-0.09	0.49*	0.92	0.91	0.36	4148 M	6 B	0.62
21	-0.06	0.97*	-0.08	0.18*	0.07	0.50*	0.93	0.92	0.38	3700 M	5 B	0.67
22	0.01	1.01*	-0.14*	0.06	-0.09	0.54*	0.92	0.92	0.41	3807 M	5 B	0.66
23	0.01	1.02*	-0.11	-0.06	0.10	0.68*	0.92	0.92	0.44	5036 M	5 B	0.64
24	0.08*	1.05*	-0.01	-0.06	-0.17*	0.83*	0.91	0.91	0.50	5628 M	4 B	0.67
High Lev	0.04	1.04*	0.09	-0.02	0.18*	0.68*	0.89	0.89	0.61	3612 M	4 B	0.64
GRS	1.25	**p-val**	0.19	**Average**		0.01	0.91	0.91	0.25	2389 M	7 B	0.60

Table 1.5: Time Series Regression (Univariate Sorting, LLP)

The dependent variables consist of excess returns of 25 Highly Leveraged Portfolios (HLP) which represents top 40% of the sample sorted based on leverage. The independent variable consists of ERM, the market risk premium, SMB and HML are Fama and French, 1993 factors for ‘Small minus Big (market capitalization)’ and ‘High minus Low (book to market ratio)’, WML is the Carhart (1997) factor for ‘Winner minus Loser’, HMLL is the leverage risk factor (High minus Low Leverage). The 1-month US TBill return is used as the risk-free rate. The data consist of 420 months from 1982- 2016. The GRS, Gibbons, Ross, Shanken (1989) asset pricing test. The * represents the values are significant (at least) at 5% confidence. The key statistics pertaining to each portfolio are given for debt-to-asset ratio (D/A), the Absolute amount of debt, market capitalization (Cap) and book to market ratio (B/M).

	Risk Factors								**Key Statistics**			
Portfolio	$\alpha_{i\times100}$	$\beta_{i,ERM}$	$\beta_{i,SMB}$	$\beta_{i,HML}$	$\beta_{i,WML}$	$\beta_{i,HMLL}$	**R^2**	**Adj R^2**	**D/A**	**Debt**	**Cap**	**B/M**
Low Lev	0.07	1.03*	0.08	-0.24*	0.15	-0.62*	0.79	0.79	0.03	77 M	6 B	0.52
2	0.13	1.06*	0.66*	0.29*	0.04	-0.76*	0.70	0.69	0.04	159 M	3 B	0.54
3	0.08	0.99*	0.43*	0.06	0.26*	-0.33*	0.69	0.69	0.03	119 M	2 B	0.57
4	0.06	1.02*	0.30*	-0.08	0.04	-0.41*	0.79	0.79	0.04	275 M	3 B	0.59
5	0.04	1.01*	0.18	-0.24*	0.14	-0.69*	0.78	0.78	0.04	363 M	4 B	0.56
6	-0.01	0.97*	0.23*	-0.41*	0.12	-0.04	0.75	0.75	0.06	538 M	9 B	0.58
7	0.16*	1.05*	0.38*	-0.05	0.13	-0.51*	0.75	0.75	0.06	556 M	6 B	0.56
8	-0.08	0.95*	0.07	-0.29*	0.20	-0.17	0.70	0.70	0.09	993 M	7 B	0.55
9	0.08	1.01*	0.22*	-0.15	-0.03	-0.40*	0.80	0.80	0.07	1054 M	7 B	0.58
10	0.04	1.02*	0.13	0.04	0.16	-0.53*	0.79	0.78	0.08	726 M	7 B	0.49
11	-0.06	0.96*	0.30*	-0.28*	0.15	0.15	0.75	0.75	0.10	863 M	5 B	0.58

12	0.06	1.01*	0.18	-0.10	-0.17	-0.27*	0.80	0.80	0.11	774 M	6 B	0.57
13	-0.03	1.03*	-0.17	0.03	-0.37*	-0.47*	0.76	0.76	0.11	717 M	8 B	0.53
14	-0.01	0.99*	-0.21*	-0.42*	-0.06	-0.16	0.81	0.81	0.12	1373 M	10 B	0.52
15	0.00	0.95*	-0.12	-0.07	-0.10	-0.33*	0.68	0.68	0.11	1332 M	13 B	0.57
16	0.01	0.95*	0.00	-0.05	0.12	-0.11	0.80	0.80	0.14	1381 M	11 B	0.55
17	0.05	1.02*	-0.13	0.12	-0.20*	-0.05	0.79	0.78	0.15	1136 M	8 B	0.54
18	0.06	1.02*	-0.40*	-0.09	-0.02	-0.49*	0.80	0.80	0.15	1117 M	8 B	0.56
19	-0.02	1.02*	-0.11	0.10	-0.01	-0.32*	0.86	0.86	0.15	1294 M	9 B	0.59
20	0.03	1.00*	0.01	0.24*	-0.19*	-0.54*	0.79	0.79	0.17	2248 M	10 B	0.57
21	0.00	1.02*	0.06	0.09	0.13	-0.36*	0.80	0.79	0.17	1700 M	10 B	0.55
22	0.11*	1.04*	0.28*	0.27*	0.18*	-0.12	0.82	0.82	0.17	2427 M	9 B	0.56
23	-0.05	0.95*	-0.30*	-0.14	0.09	-0.12	0.78	0.78	0.18	1980 M	11 B	0.56
24	0.01	1.01*	-0.22*	-0.17	-0.01	-0.07	0.79	0.79	0.18	1673 M	9 B	0.60
High Lev	0.04	1.01*	-0.17	0.00	-0.23*	-0.25*	0.80	0.80	0.18	2014 M	8 B	0.59
GRS	1.12	**p-val**	0.32	**Average**		-0.32	0.77	0.47	0.11	1075 M	8 B	0.56

Table 1.6: Time Series Regression (Univariate Sorting, HLP)

The dependent variables consist of excess returns of 25 Highly Leveraged Portfolios (HLP) which represents top 40% of the sample sorted based on leverage. The independent variable consists of ERM, the market risk premium, SMB and HML are Fama and French, 1993 factors for 'Small minus Big (market capitalization)' and 'High minus Low (book to market ratio)', WML is the Carhart (1997) factor for 'Winner minus Loser', HMLL is the leverage risk factor (High minus Low Leverage). The 1-month US TBill return is used as the risk-free rate. The data consist of 420 months from 1982-2016. The GRS, Gibbons, Ross, Shanken (1989) asset pricing test. The * represents that the values are significant (at least) at 5% significance level. The key statistics pertaining to each portfolio are given for debt-to-asset ratio (D/A), the absolute amount of debt, market capitalization (Cap) and book to market ratio (B/M).

	Risk Factors								**Key Statistics**			
Portfolio	$\alpha_{i\times 100}$	$\beta_{i,ERM}$	$\beta_{i,SMB}$	$\beta_{i,HML}$	$\beta_{i,WML}$	$\beta_{i,HMLL}$	R^2	Adj R^2	**D/A**	**Debt**	**Cap**	**B/M**
Low Lev	-0.02	0.97*	-0.29*	0.21*	0.02	-0.22	0.79	0.78	0.28	2528 M	7 B	0.67
2	-0.13*	0.98*	-0.14	0.00	0.30*	-0.24*	0.82	0.82	0.29	3352 M	8 B	0.72
3	-0.07	0.97*	0.04	0.29*	0.18	0.04	0.76	0.76	0.30	2020 M	5 B	0.61
4	-0.01	1.01*	-0.23*	0.07	-0.01	-0.02	0.85	0.85	0.30	2236 M	6 B	0.64
5	-0.03	0.95*	-0.01	0.12	-0.10	-0.09	0.81	0.81	0.31	2861 M	6 B	0.67

6	0.03	1.03*	-0.16	0.24*	-0.15	-0.16	0.79	0.79	0.31	2853 M	6 B	0.66
7	-0.01	0.98*	-0.20*	-0.04	-0.02	0.26*	0.80	0.79	0.31	2617 M	5 B	0.61
8	0.00	0.98*	-0.24*	0.20*	-0.06	0.20	0.81	0.81	0.33	2907 M	6 B	0.67
9	-0.07	0.99*	-0.08	0.02	-0.19*	0.56*	0.79	0.78	0.34	4154 M	7 B	0.62
10	0.00	0.96*	-0.14	0.10	-0.11	0.33*	0.75	0.75	0.34	3530 M	6 B	0.64
11	0.05	1.04*	-0.12	-0.04	-0.17*	0.49*	0.86	0.86	0.34	2961 M	6 B	0.61
12	-0.03	0.98*	-0.21*	-0.14	-0.11	0.51*	0.80	0.80	0.36	4852 M	6 B	0.63
13	0.07	1.02*	-0.14	0.03	-0.15	0.41*	0.75	0.75	0.37	4999 M	6 B	0.63
14	-0.03	0.98*	-0.04	0.08	0.16*	0.55	0.85	0.84	0.37	2563 M	5 B	0.66
15	-0.10*	0.98*	-0.06	0.29*	0.15	0.47*	0.84	0.84	0.40	4248 M	4 B	0.70
16	0.04	1.01*	-0.23*	-0.17*	-0.17	0.54*	0.82	0.82	0.40	3913 M	5 B	0.66
17	-0.07	0.98*	-0.14	0.25*	-0.01	0.47*	0.81	0.81	0.41	3737 M	5 B	0.66
18	0.04	1.03*	0.06	0.13	-0.01	0.59*	0.85	0.84	0.42	4547 M	4 B	0.68
19	0.02	1.02*	-0.29*	-0.02	0.09	0.71*	0.82	0.82	0.44	4989 M	5 B	0.63
20	0.01	1.01*	-0.03	-0.18	0.24*	0.80*	0.80	0.80	0.46	5063 M	6 B	0.63
21	0.14*	1.04*	-0.07	0.05	-0.22*	0.95*	0.79	0.79	0.47	4608 M	4 B	0.72
22	0.04	1.03*	0.02	-0.20*	-0.18	0.80*	0.82	0.81	0.51	5416 M	3 B	0.65
23	-0.01	1.06*	-0.06	0.02	0.16	0.66*	0.83	0.83	0.53	6329 M	5 B	0.60
24	-0.02	1.04*	0.07	-0.10	0.34*	0.73*	0.79	0.79	0.57	3326 M	2 B	0.66
High Lev	0.10	1.04*	0.16	-0.05	-0.05	0.56*	0.72	0.72	0.70	3414 M	3 B	0.61
GRS	1.09	**p-val**	0.35	**Average**		0.40	0.80	-0.08	0.40	3761 M	5 B	0.65

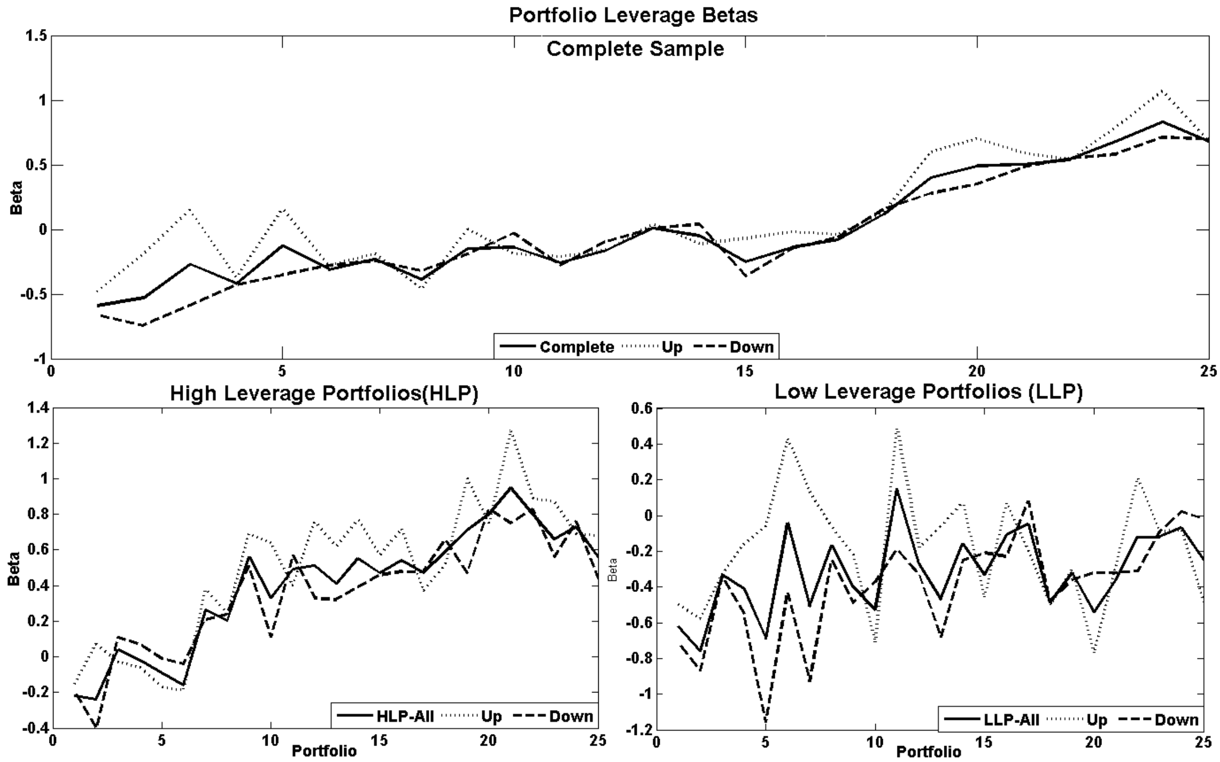


Figure 1.2: Portfolio Leverage Betas

6.1.2 Cross Sectional Regressions

The cross sectional regressions are estimated for i) Complete Sample, ii) Financial Crisis-I (1998-2003), iii) Financial Crisis-II (2006-2010) and iv) Debt Variation. The purpose of using different samples allows us to test the relationship between the leverage risk and the respective return. These results are robust to change in the sample duration. The findings support the Hypotheses H1 and H2 suggesting that the HLP outperforms the LLP in the up market and LLP outperforms the HLP in the down market.

6.1.2.1 Complete Sample (1982-2016)

The cross sectional results for the univariate leverage sorted portfolios which include HLP and LLP in up and down market using Eq. (8) are given in Table 1.7. In the up market, the expected average returns of the portfolio are positive, while in the down market they are negative. The result for leverage risk, γ_{HMLL} implies higher returns for LLP as compared to HLP during the down market. The results are significant at 5% using Newey and West (1987) and Hansen and Hodrick (1980) standard errors. The results support our Hypothesis H2.

In order to test the robustness of the above results, the 5x5 quintile portfolios, which are sorted based on size and leverage, are tested. The results are given in the second panel of Table 1.7. The results for γ_{HMLL} are found significant at 5% in the down market. The results indicate that LLP outperforms HLP in the down market. The expected market returns are negative extending support to Hypothesis H2.

6.1.2.2 Financial Crisis

In order to test the robustness of results, we estimate cross sectional regression on two major financial crises of 1998 and 2008. The first financial crisis covers the period from 1998-2003 which includes direct and indirect as well as lagged effects of the East Asian crisis in mid of 1997, the 1998 Russian default on domestic debt and sovereign debt, and the dot-com bubble

(1997-2001). To test for the impact of the global financial crisis, we use the period 2006 to 2010. For both the crises, we use a time period window that contains the pre and post crises periods. The cross sectional results for the financial crises are given in Table 1.8.

The results suggest that for the period 1998-2003, the γ_{HMLL} is found significant in the case of up market for both HLP and LLP. The results are significant at 5% using Newey and West (1987) and Hansen and Hodrick (1980) standard errors. The results suggest, based on the coefficient size, the HLP outperforms the LLP during the up market.

These results support the Hypothesis H2. For the 2006-2010, the γ_{HMLL} is also found significant in the case of up market for both HLP and LLP. The findings indicate that the HLP outperforms LLP in up market, supporting Hypothesis H1.

Table 1.7: Cross Sectional Regression (Complete Sample)

In the first panel, the dependent variables consist of excess returns of 25 leveraged-based sorted portfolios while in the second panel, the dependent variables consist of excess returns of 5X5 Size and leverage-based sorted portfolios. The 'Complete' sample contains the entire sample while the up or down sample is based on whether the excess market returns are positive $\mathrm{ERM_t} > 0$ or negative $\mathrm{ERM_t} < 0$. The 1-month US TBill return is used as the risk free rate. The 'High Leverage Portfolio' (HLP) represents the returns of the top 40% of the sample leverage-based sorted equity while the 'Low Leverage Portfolio' indicates the bottom 40% of the sample leverage-based sorted equity. The independent variable consists of SMB, HML, WML, Leverage and Lagged Leverage betas obtained from time series regressions in Eq. (7). The inputs are used to capture the cross-sectional effect using Eq. (8). The γ_{ERM} denotes the market risk premium, γ_{SMB} and γ_{HML} are Fama and French, 1993 factors for 'Small minus Big (market capitalization)' and 'High minus Low (book to market ratio)', γ_{WML} is the Carhart (1997) factor for 'Winner minus Loser', γ_{HMLL} is the leverage risk factor (High minus Low Leverage). The data consist of 420 months from 1982- 2016. The * represents that the values are significant (at least) at 5% significance level while ** says significant at 10% based on both or either Newey and West (1987) and Hansen and Hodrick (1980).

	Complete Sample			**High Leverage Portfolio**		**Low Leverage Portfolio**	
	Complete	**Up**	**Down**	**Up**	**Down**	**Up**	**Down**
	Univariate Sorting (Leverage)						
A	-2.048*	-1.736*	-0.736*	-1.372*	-1.189*	-1.414*	-1.271*
γ_{ERM}	0.664*	0.564*	-0.800*	-0.028	-0.288*	-0.038	-0.194
γ_{SMB}	-0.082*	-0.012	0.024	-0.031	-0.027*	0.040	-0.044*
γ_{WML}	-0.030	-0.084*	-0.077*	-0.100*	-0.102*	0.029	0.070*
γ_{HML}	0.109*	-0.001	-0.062*	0.021**	0.018	-0.042	0.077*
γ_{HMLL}	-0.061*	-0.041*	-0.023**	0.011	-0.025*	-0.046*	-0.015*
R^2	0.149	0.450	0.300	0.228	0.275	0.135	0.205
Adj. R^2	0.188	0.568	0.379	0.288	0.347	0.171	0.258
F-Stat	23.07*	100.70*	98.24*	25.37*	54.95*	4.40*	106.46*
	Bivariate Sorting (Size and Leverage)						
A	-1.221*	-1.357*	-0.645*	-0.820*	-0.505*	-1.217*	-0.912*
γ_{ERM}	-0.204**	-0.069	-0.786*	-0.542*	-0.903*	-0.178**	-0.471*
γ_{SMB}	-0.074*	-0.067*	-0.091*	-0.075*	-0.116*	-0.034**	-0.061*
γ_{WML}	0.014	-0.017	0.058	0.128*	-0.009	0.129**	0.020
γ_{HML}	-0.051*	-0.063*	0.026	0.010	0.105*	-0.120*	0.098*
γ_{HMLL}	-0.003*	-0.002*	-0.003*	-0.002	0.003*	-0.007*	-0.008*
R^2	0.710	0.694	0.847	0.569	0.662	0.304	0.399
Adj. R^2	0.896	0.877	1.070	0.718	0.836	0.384	0.505
F-Stat	165.32*	456.42*	1563.9*	120.05*	73.44*	30.16*	180.04*

Table 1.8: Financial Crisis (Leverage Sorted)

The dependent variables consist of excess returns of 25 leveraged-based sorted portfolios. The 'Complete' sample contains the entire sample while the up or down sample is based on whether the excess market returns are positive $ERM_t > 0$ or negative $ERM_t < 0$. The 1-month US TBill return is used as the risk free rate. The 'High Leverage Portfolio' (HLP) represents the returns of the top 40% of the sample leverage-based sorted equity while the 'Low Leverage Portfolio' indicates the bottom 40% of the sample leverage-based sorted equity. The independent variable consists of SMB, HML, WML, Leverage and Lagged Leverage betas obtained from time series regressions in Eq. (7). The inputs are used to capture the cross-sectional effect using Eq. (8). The γ_{ERM} denotes the market risk premium, γ_{SMB} and γ_{HML} are Fama and French, 1993 factors for 'Small minus Big (market capitalization)' and 'High minus Low (book to market ratio)', $\gamma_{WML}$is the Carhart (1997) factor for 'Winner minus Loser', γ_{HMLL} is the leverage risk factor (High minus Low Leverage). The * represents that the values are significant (at least) at 5% significance level while ** says significant at 10% based on both or either Newey and West (1987) and Hansen and Hodrick (1980).

	Complete Sample			*High Leverage Portfolio*		*Low Leverage Portfolio*	
	Complete	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*
	1998-2003						
α	-1.011*	-1.453*	-1.133*	-1.449*	-1.148*	-1.374*	-1.560*
γ_{ERM}	-0.446*	-0.108	-0.557*	-0.035	-0.479*	-0.241	-0.070
γ_{SMB}	-0.036	-0.003	-0.054*	-0.110*	-0.013	-0.053	-0.152
γ_{WML}	0.112*	-0.029	0.145*	0.027	-0.007	0.024	0.251*
γ_{HML}	-0.042*	-0.060*	0.074*	0.094*	0.078	-0.042	0.022
γ_{HMLL}	0.020*	0.051	0.030*	0.030*	-0.078	0.148*	-0.104*
R^2	0.452	0.101	0.404	0.149	0.320	0.468	0.323
Adj. R^2	0.571	0.128	0.511	0.188	0.405	0.591	0.408
F-Stat	493.56*	16.90*	116.22*	25.61*	67.41*	40.33*	12.03*
	2006-2010						
α	-0.786*	-0.715*	-0.922*	-0.889*	-0.838*	0.501*	-1.091*
γ_{ERM}	-0.114	-0.194**	0.000	0.008	-0.005	-1.471*	0.192
γ_{SMB}	-0.056	-0.033	-0.183*	-0.018	0.079*	0.081	-0.033
γ_{WML}	-0.181*	-0.131*	-0.134*	-0.101*	-0.004	-0.091*	-0.163*
γ_{HML}	0.112	0.127*	0.017	0.000	-0.076*	0.134*	-0.031
γ_{HMLL}	-0.070*	-0.062*	-0.082*	-0.075*	-0.014*	-0.053*	0.003
R^2	0.451	0.436	0.388	0.460	0.172	0.597	0.244
Adj. R^2	0.570	0.551	0.490	0.581	0.218	0.754	0.308
F-Stat	95.76*	69.00*	121.97*	60.12*	31.71*	264.84*	14.84*

6.1.2.3 Debt-to-Asset Variation

For much of our analysis, we have sorted the portfolios such that the top 40% of highly leveraged portfolios are considered as HLP and the bottom 40% of portfolios with low leverage are considered as LLP. For the LLP, the average debt-to-asset ratio is 11% (and maximum of 18%).

For HLP the average debt-to-asset ratio is 40% (and maximum of 70%). To check the robustness of our results and also to test the effectiveness of the screening criteria for Islamic equity indices, we sort the portfolio based on the Shariah permissible threshold of 33% for the debt-to-asset ratio. For the LLP, we change the debt-to-asset ratio to a maximum of 33%. For HLP, the debt-to-asset ratio is set to be greater than 33%. With the revised benchmark, 76% of the firms in the sample fall in the LLP and 24% in the HLP, compared to the previous 40% even distribution.

We conduct the time series and cross-sectional regressions as given in Eq. (7) and Eq. (8) respectively. Only cross-sectional results are given in Table 1.9. Given that in the up and down market, the average returns are expected to be positive and negative respectively, the results in the first panel (univariate sorting) indicate that the HLP outperforms the LLP in the up market while the LLP outperforms the HLP in the down market. Similar results are found for bivariate sorting as shown in second panel.

Table 1.9: Debt-to-Asset Ratio Variation

In the first panel, the dependent variables consists of excess returns of 25 leveraged-based sorted portfolios while in the second panel, the dependent variables consist of excess returns of 5X5 Size and Leveraged-based sorted portfolios. The 'Complete' sample contains the entire sample while the up or down sample is based on whether the excess market returns are positive $ERM_t > 0$ or negative $ERM_t < 0$. The 1-month US TBill return is used as the risk free rate. The 'High Leverage Portfolio' (HLP) represents the returns of the sample with leverage threshold of 33% while the 'Low Leverage Portfolio' includes the firms with leverage threshold greater than 33%. The independent variable consists of SMB, HML, WML, Leverage and Lagged Leverage betas obtained from time series regressions in Eq. (7). The inputs are used to capture the cross-sectional effect using Eq. (8). The γ_{ERM} denotes the market risk premium, γ_{SMB} and γ_{HML} are Fama and French, 1993 factors for 'Small minus Big (market capitalization)' and 'High minus Low (book to market ratio)', $\gamma_{WML}$is the Carhart (1997) factor for 'Winner minus Loser', γ_{HMLL} is the leverage risk factor (High minus Low Leverage). The data consists of 420 months from 1982- 2016. The * represents that the values are significant (at least) at 5% significance level while ** says significant at 10% based on both or either Newey and West (1987) and Hansen and Hodrick (1980).

	Complete Sample			*High Leverage Portfolio*		*Low Leverage Portfolio*	
	Complete	*Up*	*Down*	*Up*	*Down*	*Up*	*Down*
	Univariate Sorting (Leverage)						
α	-2.048*	-1.736*	-0.736*	-1.614*	-1.384*	-1.263*	-1.250*
γ_{ERM}	0.664*	0.564*	-0.800*	0.318*	0.054*	-0.190	-0.237
γ_{SMB}	-0.082*	-0.012	0.024	-0.015	0.089	0.060	-0.022*
γ_{WML}	-0.030	-0.084*	-0.077*	-0.137*	-0.085*	-0.046*	0.024*
γ_{HML}	0.109*	-0.001	-0.062*	-0.014	-0.071	-0.045	0.007*
γ_{HMLL}	-0.061*	-0.041*	-0.023**	0.069*	-0.057*	-0.025*	-0.035*
R^2	0.149	0.450	0.300	0.327	0.230	0.224	0.170
Adj. R^2	0.188	0.568	0.379	0.414	0.290	0.283	0.215
F-Stat	23.07*	100.70*	98.24*	120.40*	104.70*	38.34*	28.86*
	Bivariate Sorting (Size and Leverage)						
α	-1.221*	-1.357*	-0.645*	-1.021*	-0.633*	-1.460*	-0.861*
γ_{ERM}	-0.204**	-0.069	-0.786*	-0.327*	-0.747*	0.050	-0.553*
γ_{SMB}	-0.074*	-0.067*	-0.091*	-0.086*	-0.107*	-0.131*	-0.122*
γ_{WML}	0.014	-0.017	0.058	-0.095**	-0.131	-0.043	0.121*
γ_{HML}	-0.051*	-0.063*	0.026	-0.008	0.065	0.049	0.046*
γ_{HMLL}	-0.003*	-0.002*	-0.003*	0.004*	0.006*	-0.002*	-0.001*
R^2	0.710	0.694	0.847	0.346	0.475	0.656	0.656
Adj. R^2	0.896	0.877	1.070	0.437	0.600	0.829	0.829
F-Stat	165.32*	456.42*	1563.9*	18.12*	617.13*	303.11*	111.31*

6.2 Index Level Analysis

This section presents the empirical results to test the hypotheses H3 to H5. It first discusses the results for risk-adjusted return using Sharpe and Treynor ratios. The stability of the different parameters, i.e., mean, standard deviation and market beta using rolling based analysis is then examined. The robustness of the results is then discussed after controlling for the liquidity risk as well as size effect.

6.2.1 Discussion on Risk-Adjusted Return Analysis

Sharpe and Treynor ratios are shown in Table 2.0. The expected signs for the up and down phases are positive and negative respectively. The results show that the Sharpe ratio is higher (lower loss per unit of total risk) for IEI compared to CEI in down markets in all phases, i.e., pre-crisis, crisis, and post-crisis. Similarly, in the case of Treynor ratio (lower loss per unit of market risk), the IEI outperforms the CEI in down markets in all phases. The outperformance of IEI compared to CEI in the down market supports hypothesis H3. The results corroborate the presence of debt externality in equity markets as IEI includes firms with lower debt levels, which enables the index to offer resistance to losses during the down market periods.

In the case of the up market, the results for Sharpe ratio and Treynor ratio show that CEI outperforms the IEI during all phases, i.e., pre-crisis, crisis, and post-crisis. The findings support hypothesis H4 suggesting that relatively higher debt in CEIs enables them to better perform IEI when the markets are going up. Finally, IEI outperforms CEIs during the crisis period, supporting hypothesis H5. The results suggest that Islamic equities owing to their lower debt might be more resilient during the crisis period.

Table 2.0: Risk-Adjusted Return Analysis

In Panel A (based on monthly data), Sharpe Ratio is computed as $SR = \frac{E(R)-rfr}{\sigma}$, in Panel B, Treynor Ratio is computed as $TR = \frac{E(R)-rfr}{\beta}$ where beta is the market risk factor, i.e., $\beta = \frac{Cov(R_i,mkt)}{var(mkt)}$. The IEI represents the Islamic Equity Index while CEI represents the Conventional Equity Index. The mean β has been tested to be significant at 5% level. The δ is a dummy variable such that $\delta = 1\ if\ ERM_t > 0$ reflects up markets and $\delta = 0$ if $ERM_t < 0$ reflects down market. The ERM_t represents the excess return on the market portfolio. The grey color shows the greater value among the two.

	Excess Returns Per Unit of Total Risk (%)- *Panel A*		**Excess Returns Per Unit of Market Risk (%)- *Panel B***	
	CEI-DJGI	**IEI-DJIM**	**CEI-DJGI**	**IEI-DJIM**
		Total Sample: 2001– 2015		
Complete	-269.25	-272.25	-0.64	-0.62
Up	49.78	48.70	0.16	0.15
Down	-358.12	-355.05	-1.08	-1.03
		Pre-Crisis: 2001– 2007		
Complete	-632.31	-590.16	-1.32	-1.21
Up	26.69	24.81	0.13	0.10
Down	-663.77	-612.96	-1.50	-1.38
		Crisis:2008– 2011		
Complete	19.45	20.48	0.032	0.033
Up	81.78	78.55	0.13	0.12
Down	-51.89	-50.00	-0.14	-0.12
		Post-Crisis:2012– 2015		
Complete	-37.38	-36.85	-0.10	-0.09
Up	68.01	67.08	0.17	0.16
Down	-104.30	-104.20	-0.42	-0.40

In order to test the stability of the parameters in different phases, a rolling base analysis is carried out with 30 and 120 days windows to identify deviation of the performance measures. Table 2.1 shows that Sharpe and Treynor ratios based on 30 days and 120 rolling windows are largely consistent across each of the phases further supporting Hypotheses H3, H4, and H5.

In Panel A (based on monthly data), Sharpe Ratio is computed as $SR = \frac{E(R)-rfr}{\sigma}$, in Panel B, Treynor Ratio is computed as $TR = \frac{E(R)-rfr}{\beta}$ where beta is the market risk factor, i.e., $\beta = \frac{Cov(R_i,mkt)}{var(mkt)}$. The IEI represents the Islamic Equity Index while CEI represents the Conventional Equity Index. The mean β has been test to be significant at 5% level. The IEI represents the Islamic Equity Index while CEI represents the Conventional Equity Index. The rolling windows are based on 30 days and 120 days of historical data. The mean β has been tested to be significant at 5% level. The up market is when $ERM_t > 0$ and down market is when $ERM_t < 0$. The ERM_t represents the excess return on the market portfolio. The grey color shows the greater value among the two.

Figure 1.2 plots the rolling returns, standard deviation and beta over 120 days moving window. It shows absence of any instability in returns. Furthermore, the market beta between the IEI and CEI reflect the riskiness of the respective markets. However, the pattern between the two appears to be consistent, lending support to the hypotheses H3, H4, and H5 as in Table 2.1.

6.2.2 Discussion on Liquidity Adjusted Risk Return Analysis

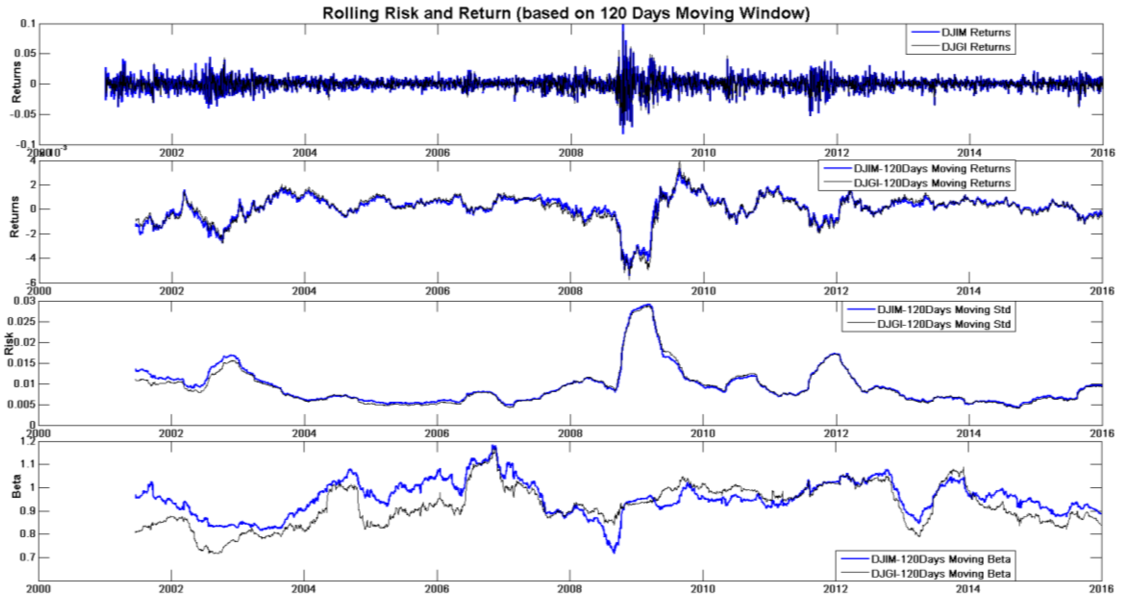
6.2.2.1 Liquidity Risk Factors

We further test for the robustness of the results for the hypotheses H3 to H5 after controlling for different channels of liquidity. For this purpose, we estimate the commonality among the respective variables given in Eq. (13) to Eq. (16) for IEI-DJIM and CEI-DJGI markets with respect to benchmark index of CEI-DJGT. The analysis essentially identifies the spread of risk across different factors. This allows us to identify the relevant risk in a comparative framework. Table 2.2 presents the results

Table 2.1: Rolling Risk-Adjusted Return Analysis

	Panel A-Excess Returns Per Unit of Total Risk (%)		Panel B-Excess Returns Per Unit of Market Risk (%)	
	CEI-DJGI	**IEI-DJIM**	**CEI-DJGI**	**IEI-DJIM**
30 Days Rolling Period				
Total Sample: 2001– 2015				
Complete	-62.57	-61.04	-0.60	-0.58
Up	15.21	14.39	0.13	0.12
Down	-101.18	-97.00	-1.04	-0.99
Pre-Crisis: 2001– 2007				
Complete	-146.78	-137.79	-1.25	-1.15
Up	20.51	16.78	0.15	0.12
Down	-165.28	-153.84	-1.43	-1.31
Crisis: 2008-2011				
Complete	-9.04	-8.39	-0.12	-0.11
Up	14.07	14.04	0.15	0.14
Down	-27.78	-26.30	-0.48	-0.47
Post Crisis:2012– 2015				
Complete	4.36	4.35	0.03	0.03
Up	13.88	12.88	0.09	0.08
Down	-11.12	-9.68	-0.09	-0.08
120 Days Rolling Period				
Total Sample: 2001– 2015				
Complete	-59.97	-58.54	-0.59	-0.57
Up	9.49	0.05	0.08	0.07
Down	-99.95	-95.72	-1.08	-1.02
Pre-Crisis: 2001– 2007				
Complete	-141.29	-132.77	-1.24	-1.16
Up	17.63	14.27	0.14	0.11
Down	-163.75	-152.50	-1.45	-1.35
Crisis: 2008-2011				
Complete	-8.45	-7.78	-0.11	-0.11
Up	11.05	11.01	0.11	0.11
Down	-21.24	-20.12	-0.36	-0.35
Post Crisis:2012– 2015				
Complete	4.03	3.89	0.03	0.03
Up	7.08	6.42	0.05	0.04
Down	-4.80	-3.25	-0.04	-0.03

Figure 1.2: Rolling Risk and Return



The market risk (β^{1i}) and the liquidity risk (β^{2i}) are decoupled. Although significant at 5% confidence level, the average market risk adjusted for trading cost appears similar in both the periods as well as across the sample. This may be due to the fact that the sample period is essentially marked by high volatility, i.e., GFC and subsequent recession and hence the risk appears to shift from market beta to liquidity beta which is the more relevant factor to capture the crisis effect.

In line with H3, in the down market, the liquidity risk (β^{2i}) is higher for the IEI compared to CEI. However, the higher risk translates into the higher returns as shown by Treynor ratio. Similarly, the performance of the two markets remain the same in the absence of the core-crisis period i.e., 2008. Analyzing the up market, the liquidity risk (β^{2i}) continues to be higher for IEI compared to CEI, however, Treynor ratio shows higher liquidity-adjusted excess returns for CEI compared to IEI.

Table 2.2: Index Risk Performance

The $\beta^{1i} \equiv cov(r_t^i, r_t^M)$ represents the commonality in returns of portfolio and the market index, $\beta^{2i} \equiv cov(c_t^i, c_t^M)$ is the commonality in liquidity of portfolio and the market index, $\beta^{3i} \equiv cov(r_t^i, c_t^M)$ is commonality in returns of portfolio and the market index, $\beta^{4i} \equiv cov(c_t^i, r_t^M)$ is the commonality in returns of portfolio and the market index. The $\beta^{5i} = (\beta^{2i} - \beta^{3i} - \beta^{4i})$ shows the aggregate liquidity risk and $\beta^{6i} = (\beta^{1i} + \beta^{2i} - \beta^{3i} - \beta^{4i})$ expresses the aggregate systematic risk. Each of these coefficients are estimated using respective monthly data for two paired relationships: i) where DJIM is the portfolio and DJGT serves as market index ii) where DJGI is the portfolio and DJGT serves as market index. Each beta is tested against the null hypothesis of zero mean at a 5% significant at level. The expected signs are based on the theoretical framework. The Treynor ratio is the excess return per unit of liquidity risk premium (β^{2i}). The up market is when $ERM_t > 0$ and down market is when $ERM_t < 0$. ERM_t represents the excess return on the market portfolio. Treynor Ratio is computed as $TR = \frac{E(R)-rfr}{\beta^{2i}}$ where The $E(R)$ show the average monthly index return and β^{2i} is the liquidity risk factor. The grey color shows the greater value among the two.

		Time Span: 2008-2012					
Coeff.	***Exp.***	***Complete Market***		***Up Market***		***Down Market***	
	Signs	***DJIM***	***DJGI***	***DJIM***	***DJGI***	***DJIM***	***DJGI***

β^{1i}	+	0.0004*	0.0004*	0.0001*	0.0001*	0.0003*	0.0003*
β^{2i}	+	0.7888*	0.7457*	0.2867*	0.2730*	0.5021*	0.4728*
β^{3i}	-	-0.0009	-0.0009	0.0009*	0.0009*	-0.0018*	-0.0019*
β^{4i}	-	-0.0010	-0.0008	0.0011*	0.0010*	-0.0021*	-0.0018*
β^{5i}	+/-	0.7907*	0.7474*	0.2847*	0.2710*	0.5060*	0.4764*
β^{6i}	+/-	0.7910*	0.7478*	0.2847*	0.2711*	0.5063*	0.4767*
Treynor(β^{2i})		-0.47%	-0.51%	1.21%	1.32%	-1.43%	-1.56%
				Time Span: 2009-2012			
Coeff.	***Exp.***	***Complete Market***		***Up Market***		***Down Market***	
	Signs	***DJIM***	***DJGI***	***DJIM***	***DJGI***	***DJIM***	***DJGI***
β^{1i}	+	0.0002*	0.0003*	0.0001*	0.0001*	0.0002	0.0002*
β^{2i}	+	0.8294*	0.7809*	0.3584*	0.3412*	0.4710	0.4397*
β^{3i}	-	-0.0002	-0.0003	0.0012*	0.0012*	-0.0014	-0.0015*
β^{4i}	-	-0.0001	-0.0003	0.0013*	0.0013*	-0.0014	-0.0015*
β^{5i}	+/-	0.8297*	0.7815*	0.3558*	0.3388*	0.4739	0.4427*
β^{6i}	+/-	0.8299*	0.7817*	0.3559*	0.3389*	0.4740	0.4428*
Treynor(β^{2i})		-0.08%	-0.09%	1.01%	1.10%	-0.91%	-1.02%

The result show that during the up market, the IEI investor takes more risk at a lesser return (sub-optimal mean variance solution) compared to the CEI investor. This means that the absence or comparatively lower level of debt in IEI result in loss of returns in the up market and vice versa in the case of CEIs. Overall the results support hypothesis H4. Similarly, in case of a crisis, the IEI outperforms the CEI based on Treynor ratio supporting hypothesis H5.

6.2.2.2 Liquidity Analysis

Since the market and liquidity risk-pattern are the same in the total sample and post-crisis in Table 2.3, we test the presence of different risk premiums based on the whole sample period. Using data over 2008-2012, we run time series regression in order to identify that the market and liquidity risks[119] are "priced" i.e., the respective market wide risk-equilibrium exists. For this purpose, we estimate the time series regression given in Eqs. (3), (12), (18) and (20) based on the dummy variable for up and down markets in line with Eq. (9) for each market in the sample. The coefficients of the risk factors represent the premium earned by an investor on its investment for their exposure to the respective factor. Table 2.3 gives the comparative return per unit of risk using Treynor ratio.

In the down market, the IEI offers lower market (Panel A: IEI -2.366 and CEI -2.870) as well as liquidity risk compared to CEI (in Panel B). The index performance as shown by the Treynor ratio favors IEI. The results show that after controlling for liquidity risk channels, the IEI outperforms CEI during the down markets, thereby supporting hypothesis H3.

In contrast, in the up market, the CEI has the performance edge. The market risk-adjusted return is higher for CEI compared to IEI during up market conditions as confirmed by the results in panel A. Studying the results in panel B, i.e., when controlling for liquidity factors, the IEI and CEI results are found not significant at 5% level. This may be because the liquidity plays a more important role during the crisis period. Looking at the Treynor ratio based on liquidity factor, we find that the CEI outperforms the IEI thus supporting H4. The CEI behavior can be attributed to the higher debt levels which may provide additional gains over and above the IEI offered risk-adjusted returns.

[119] It is important to review λ^2 (the liquidity factor). The lambdas are essentially the price of the beta risk. For instance, the coefficient λ^1 and λ^2 represent the 'price of the market (beta) risk' and 'price of the liquidity (beta) risk,' respectively. Intuitively, these coefficients reflect market wide risk equilibrium over the given period.

6.2.2.3 Size Analysis

In order to test the validity of the size effect, we estimate the cross-sectional regression given in Eqs. (3), (12), (18) and (20) for IEI and CEI respectively with an additional variable, i.e. $lnSize$. The size effect is captured by taking the log-natural of the market capitalization of the benchmark index, i.e., DJGT. We find that the size effect has a negative correlation with the liquidity, in the case of CEI-DJGI and DJGT, it is -0.045, and in the case of IEI-DJIM and DJGT, it is -0.166.

The size effect has been largely found to be significant at 5% confidence level. The result in Table 2.4 shows that the IEI continue to outperform CEI in the presence of liquidity as well as size effect based on market and liquidity risk-adjusted returns lending support to H3. The result essentially means that the firm size does not have an impact on the IEI behavior during down market conditions. This behavior can be attributed again to the low debt level of IEI, effectively reducing its exposure to losses.

Once the firm size is taken into account, the risk parameters in the up market are found not significant at 5% confidence level. Table 2.4 shows that IEI outperforms the CEI during complete market, i.e., under crisis, hence supporting H5. The result shows that the firm size does not affect the IEI return behavior compared to CEI. Since it is known that small firms suffer from certain risks such as low liquidity, they offer high risks as well as high returns. Once the size factor is separately accounted for, given that IEI have a large number of small size firms, the result continues to show that the IEI have superior performance.

Table 2.3: Liquidity Analysis

The table provides the results of time series regression results for the 2008-2012. The lambdas in Panel A to C are given as: λ^1 represents the market risk premium indicated by $\beta^{1i} \equiv cov(r_t^i, r_t^M)$, λ^2, λ^3, and λ^4 represents liquidity channels indicated by $\beta^{2i} \equiv cov(c_t^i, c_t^M)$, $\beta^{3i} \equiv cov(r_t^i, c_t^M)$, $\beta^{4i} \equiv cov(c_t^i, r_t^M)$ while λ^5 and λ^6 represents the aggregate liquidity risk given by $\beta^{5i} = (\beta^{2i} - \beta^{3i} - \beta^{4i})$ and aggregate systematic risk $\beta^{6i} = (\beta^{1i} + \beta^{2i} - \beta^{3i} - \beta^{4i})$. The up market is when $ERM_t > 0$ and down market is when $ERM_t < 0$. ERM_t represents the excess return on the market portfolio. Treynor Ratio is computed as $TR = \frac{E(R)-rfr}{\beta}$ where $E(R)$ shows the average monthly index return and β is the respective risk factor. The US 3-month T-bill is used as the risk free rate. The expected signs are based on the theoretical framework. The * represents the values are significant (at least) at 5% significance level** says significant at 10%.

		Combine Effect			*Separate Effect*		
Coeff.	***Exp. Sign***	***DJIM***	***DJGI***	***DJIM-Up*** (δ)	***DJIM-Down*** $(1-\delta)$	***DJGI-Up*** (δ)	***DJGI-Down*** $(1-\delta)$
Panel A							
$\boldsymbol{\alpha}$		-0.003*	-0.003*	-0.003*		-0.004*	
$\boldsymbol{\lambda^1}$	+/-	-1.986**	-2.344**	9.105*	-2.366*	8.300*	-2.870
Treynor($\boldsymbol{\lambda^1}$)		18.58	15.74	3.83	30.33	4.20	25.00
Panel B							
$\boldsymbol{\alpha}$		-0.015*	-0.011*	-0.015*		-0.011*	
$\boldsymbol{\kappa}$		0.034	-0.018	0.006	0.052	-0.011	-0.070
$\boldsymbol{\lambda^1}$	+	-2.205*	-2.409*	-0.118	-2.126**	1.257	-2.529*
$\boldsymbol{\lambda^2}$	+	0.016*	0.011*	0.017*	0.014*	0.012*	0.009*
$\boldsymbol{\lambda^3}$	-	0.546*	0.166	0.569	0.448	0.252	0.099
$\boldsymbol{\lambda^4}$	-	-0.016	0.530*	-0.050	-0.120	0.498	0.336
Treynor($\boldsymbol{\lambda^1}$)		16.74	15.65	-	33.75	-	29.20
Treynor($\boldsymbol{\lambda^2}$)		-23.06	-34.27	20.50	-51.26	30.12	-82.04
Panel C							
$\boldsymbol{\alpha}$		-0.014*	-0.009*	-0.015*		-0.009*	
$\boldsymbol{\kappa}$		0.032	-0.013	-0.003	0.028	-0.030	-0.069
$\boldsymbol{\lambda^1}$	+	-2.925*	-3.194*	-1.138	-2.181*	1.861	-2.680*
$\boldsymbol{\lambda^5}$	+/-	0.015*	0.009*	0.019*	0.012*	0.014*	0.006**

Treynor(λ^1)		12.62	11.80		32.90		25.65
Panel D							
α		-0.014*	-0.009*	-0.014*		-0.009*	
κ		0.035	-0.008	-0.005	0.038	-0.025	-0.056
λ^6	+/-	0.013*	-0.006	0.018*	0.010*	0.014*	0.004

In the case of the crisis period, the result shows that the IEI outperformed the CEI based on models given in panel A, B, and C respectively, hence supporting hypothesis H5.

Table 2.4: Size Analysis

The table provides the results of time series regression results for the 2008-2012. The $lnSize$ is defined as the log of index market capitalization. The lambdas in Panel A to C are given as: λ^1 represent the market risk premium indicated by $\beta^{1i} \equiv cov(r_t^i, r_t^M)$, λ^2, λ^3, and λ^4 represent liquidity channels indicated by $\beta^{2i} \equiv cov(c_t^i, c_t^M)$, $\beta^{3i} \equiv cov(r_t^i, c_t^M)$, $\beta^{4i} \equiv cov(c_t^i, r_t^M)$ while λ^5 and λ^6 represent the aggregate liquidity risk given by $\beta^{5i} = (\beta^{2i} - \beta^{3i} - \beta^{4i})$ and aggregate systematic risk $\beta^{6i} = (\beta^{1i} + \beta^{2i} - \beta^{3i} - \beta^{4i})$. The up market is when $ERM_t > 0$ and down market is when $ERM_t < 0$. ERM_t represents the excess return on the market portfolio. Treynor Ratio is computed as $TR = \frac{E(R)-rfr}{\beta}$ where $E(R)$ show the average monthly index return and β is the respective risk factor. The US 3-month T-bill is used as the risk free rate. The expected signs are based on the theoretical framework. The * represents that the values are significant (at least) at 5% significance level** says significant at 10%.

		Combine Effect		***Separate Effect***			
Coeff.	***Exp. Sign***	***DJIM***	***DJGI***	***DJIM -Up*** (δ)	***DJIM - Down*** $(1-\delta)$	***DJGI -Up*** (δ)	***DJGI – Down*** $(1-\delta)$
Panel A							
α		0.255*	0.272*	0.189**		0.199**	
$lnSize$	-	-0.016*	-0.017*	-0.012	-0.012**	-0.013**	-0.013**
λ^1	+/-	-3.341*	-3.913*	-2.485	-2.238**	-2.588	-2.721*
Treynor(λ^1)		11.05	9.43	-14.03	27.73		26.37
Panel B							
α		0.295*	0.258*	0.282*		0.209	
κ		0.025	-0.015	-0.006	0.061	0.005	-0.059
$lnSize$	-	-0.020*	-0.017*	-0.018*	-0.019*	-0.013	-0.014**
λ^1	+	-3.877*	-3.696*	-4.088	-3.582*	-2.355	-3.399*
λ^2	+	0.015*	0.010*	0.002	0.016*	0.001*	0.009*
λ^3	-	0.809*	0.425**	0.163	0.779*	0.299	0.309
λ^4	-	-0.196	0.276	0.129	-0.284	-0.005	0.204
Treynor(λ^1)		9.52	10.20	-	20.03	-	21.72
Treynor(λ^2)		-24.60	-37.69	-	-44.85	-	-82.04
Panel C							
α		0.173	0.221*	0.132		0.135	
κ		0.022	-0.015	-0.005	0.022	-0.005	-0.063
$lnSize$	-	-0.012**	-0.015*	-0.008	-0.009	-0.009	-0.009
λ^1	+	-3.845*	-4.361*	-1.635	-2.879*	-1.143	-3.307*
λ^5	+/-	0.014*	0.007*	0.003	0.014*	-0.001	0.007**
Treynor(λ^1)		9.60	8.64	-	20.03	-	22.33
Panel D							
α		0.036	0.075	0.039		0.046	
κ		0.032	-0.009	-0.002	0.034	-0.004	-0.051
$lnSize$	-	-0.003	-0.005	-0.003	-0.004	-0.003	-0.004
λ^6	+/-	0.012*	0.006	0.003	0.012*	-0.001	0.005

7. Conclusion

This paper tests for the externality of debt in the equity markets. It argues that in the up market, equity portfolios and indices with high debt levels should outperform those with low debt, while during the down market, low debt level portfolios and indices would perform better. We conduct two levels of analyses. Firstly, we use firm level data and construct high leverage portfolios (HLP) and low leverage portfolios (LLP), to compare their performance in up and down markets (Hypotheses H1 and H2). Secondly, at an index level, we use Islamic indices as a proxy for low debt indices and test their performance compared to conventional indices in the down and up markets during the non-crisis and crisis periods (Hypotheses H1 to H3).

The paper employs the Fama Macbeth, 1973 two-step regression while controlling for excess market return, size, value, momentum, and leverage risk. For the index level analysis, we use risk-adjusted return analysis based on Sharpe and Treynor ratio to capture performance differences along with a number of robustness tests.

The firm level results indicate that the HLPs outperform LLPs in the up markets while the reverse is true in the down markets. The results support our Hypotheses H1 and H2. The index level results suggest that Islamic equity indices perform better than conventional equity indices during the down markets. This supports our Hypothesis H3. In contrast, the conventional equity indices outperform Islamic equity indices in the up markets. This supports our Hypothesis H4. The results in the complete market conditions during the crisis period suggest that the Islamic equity indices outperform the conventional equity indices. This supports our Hypothesis H5. These results are robust to rolling analysis and hold even after incorporating liquidity and size effects.

Our results support both the conventional theories of capital structure which suggest that high debt stocks should outperform low debt ones, along with the literature on the externality of debt which suggests that low debt stocks would outperform high debt stocks. We show that the advantage of high debt is reflected in the equity index performance when the markets are experiencing an upturn, while during the downturns low debt indices would perform better. We show that this advantage of low debt indices would hold true in all down markets. The gains of low debt, however, are significant during a financial crisis, when the market experiences a significant plunge. Overall, our findings suggest that debt externality is prevalent in the equity markets.

The previous literature had shown that debt externality may exist in the case of households. It explains why households may have over borrowed during the global financial crisis. At the macroeconomic level, the literature had also suggested that debt may have an adverse impact on growth with a lagged effect, which many countries tend to ignore. Our results show that the externality of debt may exist in the equity market, where low debt equity portfolios and indices may perform better than high debt portfolios and indices. We argue that this externality cannot be neutralized by arbitrage as it may be rooted in the investor’s myopia.

References

Acharya, V.V., Pedersen, L.H., 2005. Asset pricing with liquidity risk. Journal of Financial Economics. 77(2), 375–410.
Albaity, M., Mudor, H., 2012. Return performance, cointegration and short run dynamics of Islamic and non-Islamic indices: Evidence from the U.S. and Malaysia during the subprime crisis. Atlantic Review of Economics. 1, 1–22.
Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets. 5(1), 31–56.
Amihud, Y., Mendelson, H., 1986b. Liquidity and stock returns. Financial Analysts Journal. 42(3), 43–48.
Amihud, Y., Mendelson, Haim., Pedersen, Lasse Heje., 2006. Liquidity and Asset Prices. Foundations and Trends in Finance. 1(4). 269-364.
Arouri, M. E., H. Ameur, Ben., Jawadi, N., Jawadi, F. and Louhichi, W., 2013. Are Islamic finance innovations enough for investors to escape from a financial downturn? Further evidence from portfolio simulations. Applied Economics. 45(24).
Ashraf, D., Nazeeruddin, M., 2014. Matching perception with the reality-Performance of Islamic equity investments. Pacific-Basin Finance Journal. 28, 175–189.
Ayub, M., 2009. Understanding Islamic Finance. John Wiley and Sons Ltd.
Banz, R.W., 1981. The relationship between return and market value of common stock, Journal of Financial Economics. 9(1), 3-18.
Benartzi, S., Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. Quarterly Journal of Economics. 110(1), 73-92.
Berk, J., 1995. A critique of size-related anomalies. Review of Financial Studies. 8(2), 275–286.
Brennan, M.J., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics. 49(3), 345–373.
Carhart , M., 1997. On persistence in mutual fund performance, Journal of Finance. 52(1), 57-82.
Charles, A., Darné, O., Pop, A., 2015. Risk and ethical investment: Empirical evidence from Dow Jones Islamic indexes. Research in International Business and Finance. 35, 33-56.
Cochrane, J., 2001. Asset Pricing. Princeton University Press, New Jersey.
Crain, M. A., 2011. A Literature Review of the Size Effect. (Retrieved on May 20, 2016, http://ssrn.com/abstract=1710076)
Datar, V.T., Naik, Y.N., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets. 1(2), 203–219.
Derigs, U., Marzban, S., 2009. New strategies and a new paradigm for Shariah-compliant portfolio optimization. Journal of Banking and Finance. 33(6), 1166-1176.
Ebrahim, M.S., Jaafar, A., Omar, F., Salleh, M. O., 2016. Can Islamic injunctions indemnify the structural flaws of securitized debt? Journal of Corporate Finance. 37, 271-286.
El Alaoui, A.O., Bacha, O.I., Masih, M., Asutay, M., 2016. Shari'ah screening, market risk and contagion: A multi-country analysis, Journal of Economic Behavior and Organization. 132, 93-112.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance. 47(2), 427-465.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics. 33, 3-56.
Fama, E.F., French, K.R., 2002. Testing trade-off theory and pecking order predictions about dividends and debt. Review of Financial Studies. 15, 1-33.

Fama, E.F., Macbeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy. 81(3), 607-636.

Gennaioli, N., Andrei S., Robert V., 2012. Neglected risks, financial innovation, and financial fragility. Journal of Financial Economics. 104(3), 452-468.

Gibbons, M. R., Ross, S., and Shanken J., 1989. Test of the efficiency of a given portfolio. Econometrics. 57(5), 1121-1152.

Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics. 92(2), 153–181.

Hansen, L. P., Hodrick, R. J., 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. Journal of Political Economy. 96, 829-853.

Hasbrouck, J., 2009. Trading costs and returns for U.S. equities: estimating effective costs from daily data. Journal of Finance. 64(3), 1445–1477.

Hayat, R., Kraeussl, R., 2011. Risk and return characteristics of Islamic equity funds. Emerging Markets Review. 12(2), 189-203.

Ho, C.S.F., Abd-Rahman, N.A., Yusuf, N.H.M., Zamzamin, Z., 2014. Performance of global Islamic versus conventional share indices: International evidence. Pacific-Basin Finance Journal. 28, 110-121.

Hoepner, A.G.F., Rammal, H.G., Rezec, M., 2011. Islamic mutual funds' financial performance and international investment style: Evidence from 20 countries. European Journal of Finance. 17(9-10), 829-850.

Jawadi, F., Jawadi, Nabila., Louhichi, Waël., 2014. Conventional and Islamic stock price performance: An empirical investigation, International Economics. 137, 73-87.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica. 47, 2, 263.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica. 53(6), 1315-1335.

Mian, A., and Sufi, A., 2010. The Great Recession: Lessons from Microeconomic Data. The American Economic Review. 100(2), 51–56.

Mian, A., and Sufi, A., 2011. House prices, home equity–based borrowing, and the US household leverage crisis. The American Economic Review. 101(5), 2132-2156.

Mian, A., and Sufi, A., 2015. House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again. University of Chicago Press.

Mian, A., Sufi, A., and Verner, E., 2017. Household debt and business cycles worldwide. Quarterly Journal of Economics.

Myers, S.C., Majluf, N.S., 1984. Corporate financing and investment decisions when firms have information that investors do not have. Journal of Financial Economics. 13(2), 187-221.

Nainggolan, Y., How, J., Verhoeven, P., 2015. Ethical screening and financial performance: The case of Islamic equity funds. Journal of Business Ethics. 137(1), 1-17.

Narayan, P.K., Phan, D.H.B., 2016. Momentum strategies for Islamic stocks, Pacific-Basin Finance Journal. 42, 96-112

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica. 55(3), 703–708.

Pástor, L.U., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy. 111(3), 642-685.

Peterkort, R. F., and Nielsen, J. F., 2005. Is the book-to-market ratio a measure of risk? Journal of Financial Research. 28, 487–502

Pettengill., G. N., Sundaram, S., Mathur, I., 1995. The conditional relation between beta and returns. The Journal of Financial and Quantitative Analysis. 30(1), 101-116.

Reinganum, Marc R., 1981. Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. Journal of Financial Economics. 9(1), 19-46.

Liquidity in Asian Markets: Intensity of Regional and Global Linkages

Suraj Kumar, Department of Management Studies, Indian Institute of Technology Madras, India

Dr. Krishna Prasanna, Department of Management Studies, Indian Institute of Technology Madras, India

ABSTRACT

This study investigates cross market linkages and measures the intensity of liquidity spillovers across 9 Asian markets and 5 developed markets during 2006 to 2016. The direction and intensity of spillovers has been measured using forecast error variance decomposition method as suggested by Diebold and Yilmaz (2012). Among the developed markets US, Germany and UK significantly affect liquidity changes in Asian markets like India, China, Singapore and Japan. The result revels that on average, each Asian market receives 7% spillover from the global markets and 16% from regional markets. During the financial crisis, it increased to 11% and 20% respectively.

Keywords: Liquidity, Intensity, Spillover, Asian markets, Global Financial crisis
JEL Classification: F21, F36, F65, G01, G15, G29

1. Introduction

Globalisation and the resultant financial market integration is rising liquidity commonality across regional markets (Brockman et al., 2009). Liquidity commonality refers to the extent to which each country's illiquidity premium co-varies with that of the global and regional average (Amihud et al., 2015).While the supply side hypothesises proposes funding constraints of financial intermediaries as the key driver for liquidity commonality, the demand side hypothesis suggests that investor sentiments and correlated trading activities are the key contributors (Karolyi et al., 2012).

Extending the liquidity co-movement literature, this study investigates cross market linkages and measures intensity of liquidity spillovers across 9 Asian markets and 5 developed markets. Further the contagion caused by recent global financial crisis and its impact on the market liquidity has also been examined. Liquidity spillovers have potential implications for international investors, economic policy makers, and market regulators.

The financial crisis, which began in the United States in 2008, spread rapidly across the globe, destabilizing the financial market stability through illiquidity. Liquidity of the stock markets across the countries was severely affected during the crisis. This global event instigated finance researchers to investigate global market integration from all possible angles. It is not just the United States or other developed markets that influence liquidity in emerging markets; regional influences are equally significant due to trade and investment inter-linkages. Notably, the liquidity commonality in Asia increased significantly during and after the recent global financial crisis (Jian-Xin Wang, 2010).

Liquidity is a measure of market quality and is a critical factor for growth and development of emerging markets. It impacts cost of capital and influences asset prices (Amihud and Mendelson 1986, Chordia et al., 2008). Liquidity plays a major role in hedging and risk management strategies of portfolio investors (Acharya et al., 2015). It is a critical factor for propagating financial crisis (Borio, 2004); in fact, even the recent financial crisis was the result of illiquidity (Brunnermeier, 2009; Gorton, 2009).

The 2008 global financial crisis clearly demonstrated that financial events and national policies have important cross-border effects (Chang et al., 2015). Liquidity commonality increases during market downturns, peaks at major crisis events and becomes weaker in the subsequent periods (Rosch and Kaserer, 2014). Commonality in the liquidity premium is not affected by domestic market conditions such as market returns and volatility (Amihud et al., 2015).

Liquidity proxies such as market spread and depth are found to be common within and across the 27 developed markets, and the 20 emerging markets as well (Brockman et al., 2009). Manicini et al., (2013) reported that there is commonality in liquidity across currencies, equity and bond markets. Smimou and Khallouli (2016) explored shift contagion and spillovers in the Eurozone and found that shocks get transmitted through the liquidity channel.

It was found that illiquidity variations in the Asian equity markets are increasingly driven by common factors, volatility being one of them. During the Asian crisis, a significant increase in co-movements of stock returns which increased the contagion among south East Asian countries was noted (Chiang et al., 2007, Carporale et al., 2005). Shocks from one nation to another gets transmitted through real and financial linkages. Trade alone cannot explain all forms of propagation and spillover phenomenon (Bekaert et al., 2005).

In the last decade, Asian markets attracted a large number of global investors as emerging economies were dominating the region. But during the 2008 financial crisis, most of the Asian markets suffered due to illiquidity. China is one of the most influential economies among Asian countries, but the recent volatility in the Chinese equity market created stress for investors and lead to a flight-to-quality situation even in other Asian markets. Investors became cautious of possible margin calls due to spillover effects of liquidity and return volatility.

This study has examined liquidity spillovers across 14 markets. While the Asian markets include China, Japan, Singapore, India, South Korea, Philippines, Malaysia, Thailand and Taiwan, the outside developed markets include the United States, the United Kingdom, Germany, France and Australia.

The study period is from January 2006 to December 2016. This period had numerous phases of turfs and turns in the financial markets and thus provided a suitable database to examine liquidity spillovers across various phases in the market.

The inter-linkages between the markets have been examined primarily using the Granger Causality tests. The intensity of spillovers was measured using forecast error variance decomposition method as suggested by Diebold (2012). Structural breaks in the liquidity have been identified using the Bai-Perron (1998) test for analysing the liquidity shifts during, and post the crisis period.

Our empirical results using the Amihud illiquidity ratio revealed that on an average, each Asian market receives 7% spillover from global developed markets, and 16% from within the region. The average regional spillover increased to 20%, and spillover from developed markets increased to 11%, during the 2008 global financial crisis. Thus, the regional spillover was higher than the global spillover in Asia. Emerging markets like China and India received higher spillover, both from the regional and the global markets. Our results are in line with the observations that developed markets exhibit higher sensitivity of domestic liquidity to global liquidity, as compared to emerging markets (Brockman et al., 2009).

Smimou and Khallouli (2015) noticed a pattern of liquidity spillover from small markets to the big markets in Eurozone. The results of this study show that on an average, Asian markets cause 11% changes in each of the 5 developed markets. Thus, while Asian markets receive 7% spillovers, they contribute to 11% changes in the liquidity of developed markets.

On an average, 70% of the liquidity changes that occur in Asian countries are due to own country’s spilllovers driven by domestic factors. This is because, despite going global, emerging markets are still operating in a regulated environment.

Liquidity spillovers provide very helpful insights about the interlinkages between financial markets in a globalized economy, through investment channels. The degree of financial market integration is an important link in explaining the spillover from one country to another due to the financial shock.

Our research contributes and extends liquidity commonality literature. While other studies established liquidity commonality, we have measured the direction and magnitude of spillovers across 14 markets. While some previous studies have focused mainly on return and volatility spillovers to explore the relationship between developed and emerging Asian countries (Bong-Han Kim et al., 2015), othershavetalked about the liquidity contagion from developed to emerging markets. Very few studies have analysed the regional influences in Asia and none has explored liquidity linkages.

Rest of the paper proceeds as follows: Section 2 discusses the methodology. Section 3 discusses the empirical results along with data description and Section 4 ends the paper with summary and conclusion.

2. Methodology

This study examines the direction and magnitude of liquidity spillovers to and from Asian economies. In addition, it has also explored the liquidity contagion in emerging Asian economies during the 2008 global financial crisis.

2.1 Granger Causality Test

The Granger Causality Test is used to investigate cross country liquidity linkages. As an econometric procedure, this test yield gross statistical association that only indicates economic causation, but does not provide the proof of the association. It is used to identify unique time-ordered and signed relationships among pairs of countries. Mathematically this can be expressed as:
If Pr $(Y_{t+n}|Y_{t-k})$ = Pr $((Y_{t+n}|Y_{t-k}, X_{t-k})$
Where Pr ( ) is the conditional probability, $X_{t-k} = (X_t, X_{t-1}, X_{t-2} \ldots\ldots X_{t-n})$ and
$Y_{t-k} = (Y_t, Y_{t-1}, Y_{t-2} \ldots\ldots Y_{t-n})$
The null hypothesis is X, which does not Granger cause y. The null hypothesis is rejected if a_{21} and a_{22} coefficients in the above equation are jointly different from Zero. This means that the result 'X Granger causes Y' does not imply that y is the result or effect of x.

2.2 Measuring Spillover Effects

In comparison to previous liquidity spillover and contagion literature, our methodology is different. We adopted the measure proposed by Diebold and Yilmaz (2012) to capture cross country spillovers, since it has the advantage of being dynamic as well as directional.

Diebold and Yilmaz (2009) developed a spillover index based on vector autoregressive (VAR) models. Construction of the index is based on forecasted error variance decomposition. The proposed spillover index shows the proportion of the movement in a variable over time due to its own shocks as well as shocks in other variables in a VAR framework. It quantifies how much of the total forecasted variance is attributed to each variable. A higher spillover index indicates that a larger portion of shocks in the market as a whole is due to cross-variable shocks, than from own variable shocks.

The Diebold and Yilmaz (2009) variance decomposition model utilises orthogonal innovation, which is achieved by using the Cholesky decomposition. The major drawback of such orthogonalisation is that it leads the variance decomposition which depends on the ordering of the variable.

Diebold and Yilmaz (2012) used the generalised VAR framework of Koop et al., (1996) and Pesaran and Shin (1998), referred as KPSS, which produces variance decomposition as order invariant. The KPSS variance decomposition attempts to correlate the shocks by accounting for them appropriately using historical distribution of errors instead of orthogonalized shocks. As the shocks are not orthogonalised here, the sum of the contributions will not be equal to one. The covariance of stationary N variable VAR (k) is:

$$x_t = \sum_{n=1}^{k} \Phi x_{t-m} + \varepsilon_t$$

Where $\varepsilon_t \sim (0, \sum)$ is a vector of independently and identically distributed disturbances. The moving average representation is written as:

$$x_t = \sum_{k=0}^{n} A_n \varepsilon_{t-m}$$

The moving average coefficients are useful in understanding the dynamics of the system. The variance decomposition allows the error variance forecasts of each variable to be divided into parts that are separable.

2.3 KPSS Variance Decomposition

The own variance share is the fraction of H step ahead error variance in forecasting x_i which is due to shocks of x_i for i= 1,2,3…n. Cross variance share or spillover in forecasting x_i is due to shock of x_j for j = 1,2,3,…n.
The KPSS H step ahead error variance is:

$$\widetilde{\theta}_{ij}^{g} = \frac{\sigma_{ii}^{-1} \sum_{h=0}^{H-1} (e_i' A_h \sum e_j)^2}{\sum_{h=0}^{H-1} (e_i' A_h \sum A_h' e_j)^2}$$

Where, $\sum$ is the variance matrix for the error vector, σ_{ii} is the standard deviation of error terms for the *i*th equation and $e_i$is the selection vector of the *i*th term. As discussed above, the sum of each row in the table represents variance decomposition that is not equal to one ($\sum_{j=1}^{N} \widetilde{\theta}_{ij}^{g} (H) \neq 1$). Hence, it is normalized to use the information in calculating the spillover index. Normalised representation for each entry in the variance decomposition Table is:

$$\widetilde{\theta}_{ij}^{g}(H) = \frac{\theta_{ij}^{g}(H)}{\sum_{j=1}^{N} \theta_{ij}^{g}(H)}$$

The total spillover index is given by:

$$S_i^g(H) = \frac{\sum_j^N = 1 \theta_{ij}^{g(H)}}{N} * 100$$

This spillover index measures the total spillover of liquidity among developed as well as emerging countries.

2.4 Directional Spillovers

The directional spillover helps to find the direction of spillover from various countries (Sowmya et. al., 2016). The directional spillover measures the spillovers received by the country *i* from all other countries *j* as:

$$S_i^g(H) = \frac{\sum_j^N = 1 \theta_{ij}^{g(H)}}{N} * 100$$

Net spillover from the market is obtained from the country *i* to all other countries *j* as:

$$S_i^g(H) = S_j^g(H) - S_i^g(H)$$

2.5 Structural Breaks

The Bai and Perron (1998) test is used to find the structural breaks in the individual time series data. Bai and Perron advocated the $SupF_t(L)$ F-statistics and double maximum tests. The $SupF_t(L)$ F-statistics test the null hypothesis of no structural break against the alternative hypothesis that there are K breaks. The double maximum test considers the null hypothesis of no breaks against 1 break upto K structural breaks. Here, maximum K is kept as 5, i.e maximum upto 5 breaks in one series. Hence, it is a sequential $SupF_t(L)$ procedure that will determine the optimal number of breaks with K and location of brakes.

3. Empirical Results

3.1 Data description

The study considered 14 markets, including 7 emerging and 2 developed markets from Asia and five developed markets from other regions. Asian emerging markets include China, India, South Korea, Philippines, Malaysia, Thailand and Taiwan; while Japan and Singapore represent developed markets of the region. Other developed markets include the United States, the United Kingdom, Germany, France and Australia. The emerging market sample selection is based on MSCI market classification (May 2017). Developed markets sample is based on trade and investment linkages with Asian markets.

Daily data for all individual stocks of key indices across the sample countries were sourced from the Bloomberg database in US dollar terms. The selected indices represent most liquid stocks of their country.

Daily returns of the Index constituent stocks have been computed for the sample period January 2006 to December 2016. Stocks not having historical data during this time period were replaced by the next higher stock in next major index, based on their weightage in the Index.

We calculated liquidity proxies from the daily data for each stock and computed the daily value weighted average representing the index, which was then averaged for the week. The liquidity proxy of each index is therefore the value weighted weekly average of all its constituent stocks. Thus, the data represents a total of 574 weekly observations for the years 2006 – 2016, and for each country. We have computed two widely used liquidity measures; i) Amihud Illiquidity Ratio was calculated as $\frac{|R_{it}|}{\text{Volume}}\times 10^6$, where $R_{it}$is the absolute return of stock *i* on day *t* and volume was the Dollar trading volume of the stock on day *t*. and it was multiplied by 10^6 to scale up the value, ii) Quoted Spread was calculated at the end of each trading day taking the difference of ASK and BID prices divided by its midpoint. While these two variables represent liquidity in the market, they reflect different characteristics of a liquid market. The Amihud Illiquidity Ratio is based on market activity. It represents return elasticity and measures the change in the price in relation to change in the trading volume. Thus, it measures price impact and is a proxy for market depth. Quoted Spread, on the other hand represents transaction cost in the market. We measured Quoted Spread from bid ask quotes at the closing market hours on each trading day. The trading cost and the spread were determined by the type of market, trading mechanism and regulations. Though both the liquidity proxies are different, they never the less revealed the causal relationship and spillovers across the markets. Since the results are similar for both the liquidity proxies, the results of Amihud illiquidity ratio are reported in the paper for brevity.

[Insert Figure 1a, 1b near here] [Insert Figure 2a, 2b near here]

Table 1 report the descriptive statistics of Amihud Illiquidity Ratio. The descriptive indicate that US has higher liquidity with lowest Amihud Illiquidity Ratio. Also, UK experienced highest illiquidity during the study period. Among the Asian markets, Philippines and Thailand had lowest liquidity, while Japan and South Korea exhibited higher liquidity. Figures 1 and 2 presents cross country time varying illiquidity ratios. In addition, liquidity proxy exhibit higher Skewness and excess Kurtosis. J-B statistics confirmed that the variable is non-normal.

[Insert Table 1 near here]

Karolyi et al., (2012) developed supply side and demand side hypothesises for liquidity commonality. They identified funding constraints of financial intermediaries as the key driver on the supply side; while on the demand side, they suggested that institutional investors, foreign investor involvement, investor sentiments and correlated trading activities are the key contributors (Karolyi et al., 2012).

Figure 3 depicts Total Portfolio Investments ($ million) in the Asian markets by the other 13 countries considered in the study during June 2016. China, South Korea and India represent three emerging Asian markets receiving higher cross-border portfolio investments.
Figure 4 represents the trade relationship of Asian countries with other countries considered in the sample. Apart from Japan and Singapore, which are developed markets in the region, China, Korea, India and Taiwan are the emerging markets having higher international trade relationships in the region. While China exhibits higher trade relationship, Korea and India sourced higher portfolio investments compared to their trade values.

[Insert Figure 3 and Figure 4 near here]

3.2 Structural Breaks
In order to identify the effect of the financial crisis, structural breaks in the market liquidity were identified using the Bai and Perron (1998) test. Significant break points are reported in Table2. The table shows that Australia, UK and Japan had the first structural break in the first quarter of 2008, which was the first shock of the global financial crisis to liquidity in the market. Market liquidity had a lag effect of the crisis, as most of the Asian emerging markets along with US, France, Germany and Singapore had the structural break in 2009. Liquidity in Australia and India had shown resistance to all the shocks and did not have any significant break. China reported four brakes in the sample period. Considering the common liquidity break points across the countries, the data was subdivided, taking 1st April 2009 as the basis to assess the impact of the global financial crisis. We have defined the crisis period as the time up to 31st march 2009 and the post-crisis period as the time from 1st April 2009 till the end of our sample period. Separate analysis of both the periods provides the spillover effects in the market liquidity during the crisis.

3.3 Liquidity Changes and Causality across the Countries
We have reported the bi-directional Granger Causality F statistics in Table 3. In order to visualise causality direction during and after the crisis, we have also analysed it for both the period separately for Amihud illiquidity ratio as well as for Quoted spread. However Table 3 reports Granger Causality of full sample period for Amihud illiquidity proxy only to ensure brevity.

[Insert Table 3 near here]

The US, UK and Germany leads the market liquidity of all the countries except Australia, Malaysia and China. UK had an influence on Malaysia and China also in the post-crisis period. The emerging markets also causes liquidity changes in the developed markets. Among the developed markets, Singapore in Asia and UK in the west have been influenced significantly by the emerging markets.

Australia has been found to be independent. It does not cause liquidity changes in other countries. Germany and Singapore caused liquidity changes in Australia. France also has a lead effect on six out of the nine Asian markets. These results indicate that Asian markets are at the receiving end for liquidity spillover from US and other developed European markets.

China, which experienced a series of market crashes after 2009, had a significant influence on the liquidity of other markets, particularly on neighbourhood countries like India, Malaysia and Singapore, only after the financial crisis.

Within Asia, Japan and Korea caused significant liquidity changes in the region. Japan being a developed Asian market, has a significant influence on all other developed markets expect Australia.

In developed markets like US, UK, Germany and France, it is their Amihud Illiquidity Ratio that exhibited significant lead effects. Quoted Spread ratio exhibited higher cross- country linkages across Asian markets (results not reported in the paper).

In summary, Granger Causality shows that there is a bi-directional relationship among developed and emerging Asian markets. This result is consistent with that of Smimou and Khallouli (2015), who reported that even small developing markets causes liquidity changes in the developed markets. The regional influence is found higher than the global influence. In the next section, we measure the direction of spillover between the countries using the spillover index and also its change during and post the crisis period.

3.4 Spillover Effects in Market liquidity

Liquidity Commonality and spillovers arise due to supply or demand side sources (Coughenour and Saad 2004; Forde et al., 2010, Karolyi et al., 2012).

Table 4 reports the spillover index across the countries using the Amihud Illiquidity Ratio for the full sample period. Tables 5 and 6 report spillover index for during and post crisis periods. Spillover index was computed as proposed by Diebold and Yilmaz (2012). Seven weeks ahead forecast was used to decompose the error variance. The forecast for alternative time periods of 1 week, 3 weeks, 5 weeks, 7 and 10 weeks have also been estimated for robustness checks but not reported for brevity. Total spillover index value represents total liquidity spillovers across the countries. The estimated total spillover index of 22.3% refers to the forecasted error variance which had spilled from other countries.

[Insert Table 4 near here]

The diagonal elements of the spillover table represent the own-country liquidity spillovers. Table 4 shows that Australia has highest own country spillover (98%), followed by Philippines (93%) and China (83.3%). This conveys that their market liquidity is more driven by local factors.

Changes in US liquidity caused 38%, and changes in Germany caused 47% of variance in the market liquidity across all the countries. Australia had the lowest contribution to the cross-country liquidity spillover, followed by South Korea and China. As noted earlier in our results, Australia neither influences others, nor is it influenced by others.

However, liquidity spillovers in the regional markets are larger than that of developed markets. Taiwan contributes to 60% spillover, which is the highest across the countries. Apart from Taiwan, Japan and Philippines are individually responsible for 31% of liquidity changes among the Asian markets.

The developed markets, being market driven economies, not only contribute to the liquidity changes to the other markets, but they also receive significant spillovers from emerging countries.UK receives 25% of spillover from Germany, Japan and Taiwan. The US receives 30% and Germany receives 33% from other countries. However, they have a net positive spillover of 8% and 14% respectively, which indicates that their influence on others is more than what they receive from others. The liquidity changes in these two markets get transmitted to other markets.

Among Asian markets, again Taiwan and Japan receive spillover from other markets and their net spillover is also positive. Among all the countries considered in the study, India receives highest spillover of 40% from others. It receives 26% from regional neighbourhood countries like Malaysia, Philippines, Taiwan and Thailand. In addition, India also receives another 10% from US and Germany.

3.5 Cross Country Linkages during the Crisis

Table 5 reports the liquidity spillover index during the crisis. The crisis period refers to the period starting Jan 2006 to March 2009. It shows that the total spillover index increased to 36.3%, which is much higher than the total index of 22.3% reported for the total sample period. The surprising observation was made that during the crisis, Australia, which was noted to be independent in earlier analysis, contributes to 29% and receives 45% of the liquidity spillover. Consistent with the previous results, Germany and US in developed markets, and Japan and Taiwan in Asia, continued to be the highest contributors in liquidity spillover. Even during the crisis, which erupted in developed markets and spiralled to other markets, regional influence was noted to be higher than the global influence on the liquidity of Asian markets.

[Insert Table 5 near here]

The consistent influence of US on the liquidity of the Asian countries during the crisis period shows its dominance in the region. The total spillover from US during the crisis increased from 38% to 54%, out of which Asian markets received 29% of the spillover. This result also suggests that liquidity is the channel of crisis contagion, since the financial crisis erupted in the US with the fall out of Lehman brothers.

Table 6, which reports the spillover for post crisis period, is similar to Table 4, and has the total spillover index of 23.1%. This confirms higher liquidity spillovers during the crisis period. The same is also confirmed from yearly spillover results (not reported in the paper), that the spillover index was maximum in 2008, due to the crisis impact.

[Insert Table 6 near here]

In the post crisis period, China's contribution to the liquidity changes in other countries has increased. The frequent trading halts and downturns in the Chinese stock market have a higher impact on the liquidity of other Asian markets. The existing spillover, both during the crisis as well as in the post crisis period, confirms interdependence in the liquidity of the countries. The results reiterate the liquidity commonality phenomena.

The regional contagion during south East Asian crisis did have a small impact on developed financial market's asset returns (Carporale et al., 2005). The crisis that began in west in 2007 had a significant impact on Asian markets. It is also noted that during this phase, emerging Asian markets like South Korea, Philippines and India also caused liquidity changes in the developed markets.

Our results are in line with empirical observations in the literature that trade and financial sector linkages cause transmission of shocks between the countries (Kaminsky and Reinhart, 2000). These are the potential reasons for liquidity spillovers. Demand for liquidity and supply of liquidity of each country is interrelated as dealers and traders, who are liquidity providers of asset markets, trade in multiple markets (Coughenour and Saad, 2004). These traders evolve as effective transmitters of market liquidity. The shocks and the losses in one market influence their trading behaviour who instils risk aversion and extract the liquidity in other markets.

4 Summary and Conclusion

In the last decade, stock market liberalization, coupled with rising returns and market valuations, has resulted in Asian markets receiving significant capital inflows. Foreign portfolio investments and the information overflow across the markets are the key drivers in the integration of Asian emerging markets with the developed markets.

The purpose of this study was primarily to investigate the liquidity spillover among the emerging Asian economies. The study specifically addressed three research issues: i) It examined liquidity linkages across the markets, ii) It measured the direction and magnitude of liquidity spillover, and iii) It examined the liquidity shifts during the 2008 global financial crisis. This research used the Amihud Illiquidity ratio and Quoted Spread to measure liquidity and generalise the cross market relationships.

Brockman et al., (2009) reported liquidity commonality across the 47 markets. They found that spillover is higher in Asian emerging markets as compared to others. We extended the empirical work in the Asian region to capture the direction and magnitude of the liquidity spillover to and from individual countries. Covariance of liquidity across the countries was investigated by developing a spillover index that precisely measures pair wise relationship between emerging and developed countries.

Our results revealed the bi-directional causality in market liquidity among Asian emerging markets, as well as the developed markets. Among the developed markets, US, Germany and UK significantly affect liquidity changes in the emerging countries like India and China. These countries also receive spillover from India, china and Taiwan. Our results support the demand side hypothesis and suggest that trade and portfolio investments are the drivers of liquidity spillovers.

Regional spillovers were noted to be larger than those from developed markets. Australia is one country in our sample that exhibited relatively lower liquidity spillover, despite strong trading connections with the developed west and the emerging east. This observation could be of interest to global investors looking for portfolio balancing in terms of liquidity. The speed of learning and sharing the information across the markets drives the direction of spillover

(Cespa and Focoult, 2014). The cross-country linkages and co-movements provide insights to the policy makers and regulators as to which are the countries that have the ability to influence the domestic markets' liquidity.

References

Acharya, V. V., Schaefer, S., Zhang, Y., 2015. Liquidity Risk and Correlation Risk: A Clinical Study of the General Motors and Ford Downgrade of May 2005. Q. J. Finance. 5, 550006.doi:10.1142/S2010139215500068

Amihud, Y., Hameed, A., Kang, W., Zhang, H., 2015. The illiquidity premium: International evidence. J. financ.econ. 117, 350–368. doi:10.1016/j.jfineco.2015.04.005

Amihud, Y., Mendelson, H., 1986. Liquidity and Stock Returns. Financ.Anal. J. 42, 43–48. doi:10.2469/faj.v42.n3.43

Arestis, P., Caporale, G.M., Cipollini, A., Spagnolo, N., 2005. Testing for financial contagion between developed and emerging markets during the 1997 East Asian crisis. Int. J. Financ. Econ.44, 0–16.

Bai, B.Y.J., Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66, 47–78. doi:10.2307/2998540

Bekaert, G., Harvey, C.R., Ng, A., 2005. Market Integration and Contagion. J. Bus. 78, 39–69. doi:10.1086/426519

Borio, C., 2004. Market Distress and Vanishing Liquidity: Anatomy and Policy Options. Bank Int. Settlements Work.Pap. 158, 1–35.

Brockman, P., Chung, D.Y., 2002. Commonality in Liquidity: Evidence from an Order-Driven Market Structure. J. Financ. Res. 25, 521–539. doi:http://www.blackwellpublishing.com/journal.asp?ref=0270-2592

Brockman, P., Chung, D.Y., Pérignon, C., 2009. Commonality in Liquidity: A Global Perspective. J. Financ. Quant. Anal. 44, 851.doi:10.1017/S0022109009990123

Brunnermeier, M.K., 2009. Market Liquidity and Funding Liquidity Market Liquidity and Funding Liquidity. Rev. Financ. Stud. 22, 2201–2238. doi:10.1093/rfs/hhn098

Cespa, G., Foucault, T., 2014. Illiquidity contagion and liquidity crashes, in: Review of Financial Studies. pp. 1615–1660. doi:10.1093/rfs/hhu016

Chiang, T.C., Jeon, B.N., Li, H., 2007. Dynamic correlation analysis of financial contagion: Evidence from Asian markets. J. Int. Money Financ. 26, 1206–1228. doi:10.1016/j.jimonfin.2007.06.005

Chiang, T.C., Zheng, D., 2015. Liquidity and stock returns: Evidence from international markets. Glob.Financ. J. 27, 73–97. doi:10.1016/j.gfj.2015.04.005

Chordia, T., Roll, R., Subrahmanyam, A., 2008.Liquidity and market efficiency.J. financ.econ. 87, 249–268. doi:10.1016/j.jfineco.2007.03.005

Comerton-Forde, C., Hendershott, T., Jones, C.M., Moulton, P.C., Seasholes, M.S., 2010. Time variation in liquidity: The role of market-maker inventories and revenues. J. Finance 65, 295–331. doi:10.1111/j.1540-6261.2009.01530.x

Coughenour, J.F., Saad, M.M., 2004. Common market makers and commonality in liquidity.J. financ.econ. 73, 37–69. doi:10.1016/j.jfineco.2003.05.006

Diebold, F.X., Yilmaz, K., 2009. Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets.Econ. J. 119, 158–171. doi:10.1111/j.1468-0297.2008.02208.x

Diebold, F.X., Yilmaz, K., 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. Int. J. Forecast. 28, 57–66. doi:10.1016/j.ijforecast.2011.02.006

Financial, C., Policy, M., Stability, F., 2015. BIS Papers Cross-border Financial Linkages : Challenges for Monetary Policy and Financial Stability.

Gorton, G., 2009. Information, liquidity, and the (Ongoing) panic of 2007, in: American Economic Review. pp. 567–572. doi:10.1257/aer.99.2.567

Kaminsky, G.L., Reinhart, C.M., 2000. On crises, contagion, and confusion, in: Journal of International Economics. pp. 145–168. doi:10.1016/S0022-1996(99)00040-9

Karolyi, G.A., Lee, K.H., Van Dijk, M.A., 2012. Understanding commonality in liquidity around the world.J. financ.econ. 105, 82–112. doi:10.1016/j.jfineco.2011.12.008

Kim, B.-H., Kim, H., Lee, B.-S., 2015. Spillover effects of the U.S. financial crisis on financial markets in emerging Asian countries. Int. Rev. Econ. Financ. 39, 192–210. doi:10.1016/j.iref.2015.04.005

Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. J. Econom. 74, 119–147. doi:10.1016/0304-4076(95)01753-4

Mancini, L., Ranaldo, A., Wrampelmeyer, J., 2013. Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. J. Finance 68, 1805–1841. doi:10.1111/jofi.12053

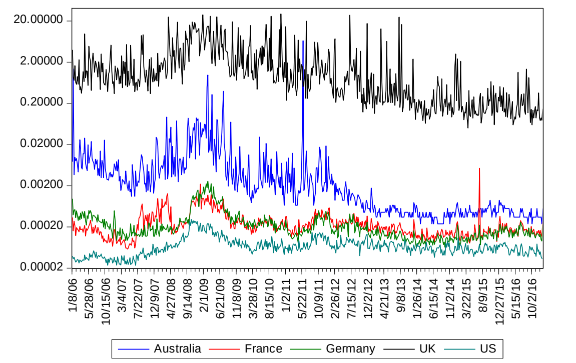
Pesaran, M.H., Shin, Y., 1998. Generalized Impulse Response Analysis in Linear Multivariate Models.Econ.Lett. 58, 17–29. doi:10.1016/S0165-1765(97)00214-0

Rosch, C.G., Kaserer, C., 2014. Reprint of: Market liquidity in the financial crisis: The role of liquidity commonality and flight-to-quality. J. Bank. Financ. 45, 152–170. doi:10.1016/j.jbankfin.2014.06.010

Smimou, K., Khallouli, W., 2016. On the intensity of liquidity spillovers in the Eurozone. Int. Rev. Financ. Anal. 48, 388–405. doi:10.1016/j.irfa.2015.03.009

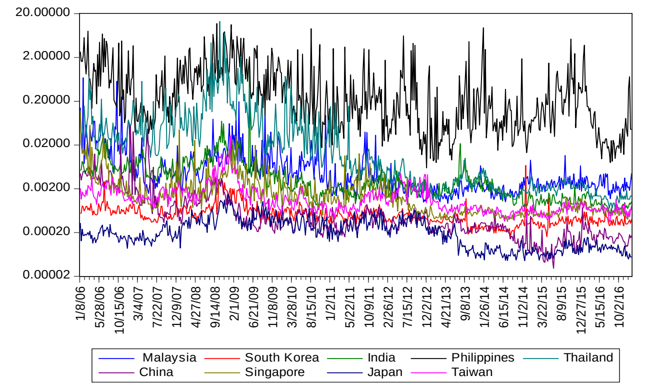
Wang, J.X., 2010. A multi-factor measure for cross-market liquidity commonality. ADB Econ. Work.Pap. Ser. 230, 1–33.

FIGURE 1A: AMIHUD ILLIQUIDITY RATIO OF DEVELOPED MARKETS



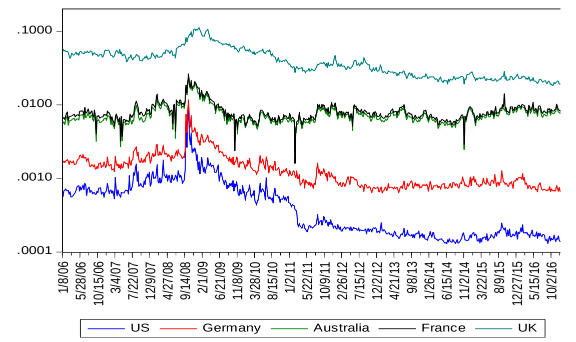
Notes: The graph presents the weekly Amihud illiquidity ratio during Jan 2006 to Dec 2016.

FIGURE 1B: AMIHUD ILLIQUIDITY RATIO OF ASIAN MARKETS



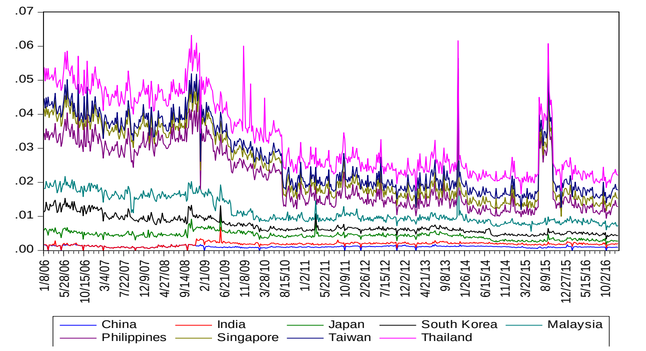
Notes: The graph presents the weekly Amihud illiquidity ratio during Jan 2006 to Dec 2016.

FIGURE 2A: QUOTED SPREAD OF DEVELOPED



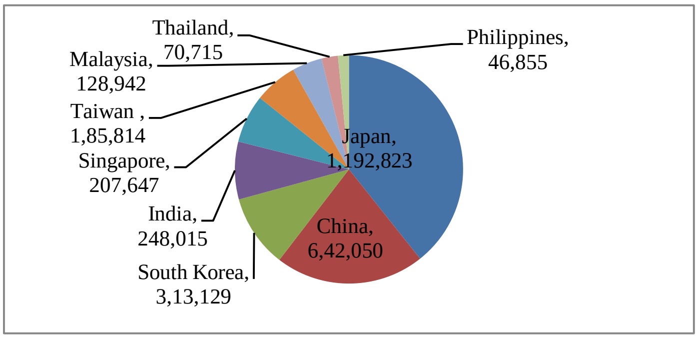
Notes: The graph presents the weekly Quoted Spread during Jan 2006 to Dec 2016.

FIGURE 2B: QUOTED SPREAD OF ASIAN MARKETS



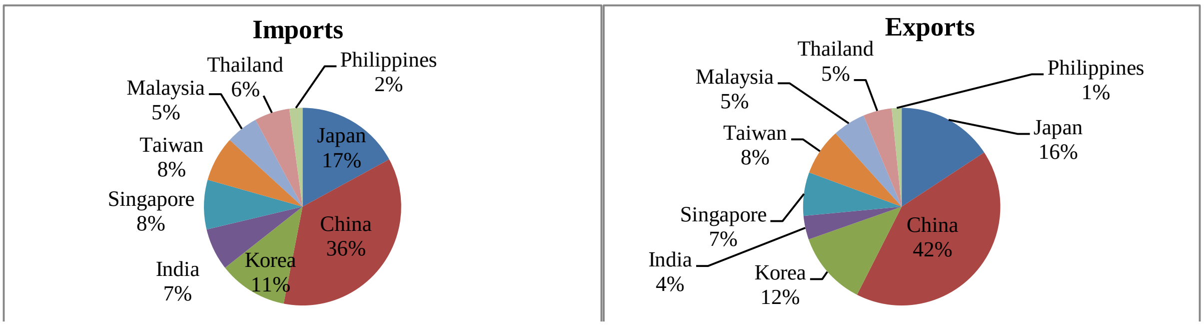
Notes: The graph presents the weekly Quoted Spread during Jan 2006 to Dec 2016.

FIG 3: PORTFOLIO INFLOWS INTO ASIAN MARKETS



Notes: The Figure depicts, Total Portfolio Investment ($ million) into the Asian markets from other countries in sample during June 2016. Source: Coordinated Portfolio Investment (CPIS) from IMF.

FIGURE 4: INTERNATIONALTRADE RELATIONSHIPS OF ASIAN COUNTRIES



Notes: We compiled the trade values in USD for the year 2015 provided by IMF for each Asian market from other countries considered in the sample and the relative share is depicted in the figure.

TABLE 1 : DESCRIPTIVE STATISTICS OF AMIHUD ILLIQUIDITY RATIO

	AIM (UK)	ASX (Australia)	CAC (France)	DAX (Germany)	DJIA (US)	STI (Singapore)	TOPIX (Japan)	KLCI (Malaysia)	KOSPI (Korea)	NIFTY (India)	PSEI (Philippines)	SET (Thailand)	SSE (China)	TWSE (Taiwan)
Mean	1.31962	0.02135	0.00024	0.00021	0.00007	0.00255	0.00025	0.01099	0.00055	0.00756	0.77681	0.11474	0.00115	0.00145
Median	0.40903	0.00064	0.00019	0.00015	0.00007	0.00104	0.00019	0.00327	0.00046	0.00242	0.19433	0.00816	0.00033	0.00094
Max.	15.92964	6.64343	0.00530	0.00223	0.00020	0.12185	0.00148	0.68546	0.00683	1.77282	11.50305	13.28881	0.05421	0.03269
Min.	0.05051	0.00024	0.00008	0.00006	0.00003	0.00036	0.00005	0.00093	0.00016	0.00053	0.00598	0.00072	0.00003	0.00003
SD	2.43666	0.33097	0.00029	0.00021	0.00003	0.00678	0.00020	0.04138	0.00046	0.07406	1.50513	0.67203	0.00390	0.00202
Skew.	3.54	19.86	14.13	5.07	1.58	5.69	1.19	8.75	14.00	20.04	3.56	15.39	9.69	8.57
Kurt.	16.96	397.66	247.32	38.85	6.51	48.74	4.21	100.78	246.64	402.66	18.67	272.31	112.54	111.86
JQ	4014	2654952	1020818	23424	376	37493	120	166498	1014966	2722447	11696	42130	299	1467

Notes: This table reports descriptive statistics of Amihud Illiquidity Proxy for stock market represented by their major indexes, which includes FTSE AIM UK 50 for London Stock Exchange, ASX 50 for Australia, CAC for France, DAX for Germany, DJIA for United states, STI for Singapore, Topix Core 30 for Japan, KLCI for Malaysia, KOSPI 50 for South Korea, Nifty for India, PSEI (PCOMP) for Philippines, SET 50 for Thailand, SSE 50 for China, TWSE 50 for Taiwan

TABLE 2: STRUCTURAL BREAKS IN AMIHUD ILLIQUIDITY RATIO DURING 2006 TO 2016

year	UK	Australia	France	Germany	US	Malaysia	Korea	India	Philippines	Thailand	China	Singapore	Japan	Taiwan
2006		-						-						
2007		-	12/30/07			09/02/07		-			9/16/07			
2008	4/13/08	-			3/16/08			-					1/20/08	
2009		-	9/20/09	12/06/09	11/08/09		08/02/09	-	7/26/09	4/26/09	5/17/09	7/19/09		04/12/09
2010	3/14/10	-				12/05/10		-						
2011	1/09/11	-						-		2/20/11				07/10/11
2012		-		12/02/12			08/12/12	-			11/18/12	3/18/12		
2013	5/19/13	-			10/27/13			-					3/17/13	03/03/13
2014		-					11/23/14	-			07/06/14			
2015		-		5/17/2015				-						

Note: This table reports the structural break points in the Amihud illiquidity ratio during the sample period based on the Bai and Perron (1998) test during 2006 to 2016

This table reports bi-directional Granger causality test F-statistics between countries. Every row of the table represents null hypothesis as illiquidity of first column country does not granger causes the illiquidity of other country in the same row. A significant Value of F- statistics rejects the null and it is marked as *, **and *** which represents p-values <0.10, <0.05 and <0.01 respectively. Lag length of Granger causality is 1 based on AIC and SBC criteria.

TABLE 3: GRANGER CAUSALITY OF AMIHUD ILLIQUIDITY PROXY FOR FULL SAMPLE PERIOD

	U K	Australia	China	France	Germany	India	Japan	Korea	Malaysia	Philippines	Singapore	Taiwan	Thailand	U S
U K	-	1.47	0.10	27.21***	11.95***	7.24***	7.96***	32.86***	2.18	13.06***	13.74***	9.34***	51.08***	4.38**
Australia	0.77	-	0	0.24	0.16	0	0.01	0.19	0.02	0	0.37	0.04	0.04	1.29
China	0.25	0.01	-	1.09	0.59	0.01	0.09	1.24	4.51**	1.97	2.30	0.41	0.60	1.09
France	9.70***	0.835	0.68	-	0.44	3.94**	11.42***	14.63***	0.22	9.30***	8.41***	0.84	8.65***	0.11
Germany	57.38***	4.92**	0.31	77.92***	-	8.33	19.49***	65.24***	11.01***	51.43***	36.89***	14.37***	24.67***	11.15***
India	51.99***	0.06	0	0.14	0.21	-	1.35	1.05	0.03	0.05	0.02	0.01	0.56	1.46
Japan	43.22***	1.11	0.12	60.32***	23.88***	1.86	-	47.53***	2.85*	24.19***	18.65***	83.64***	23.34***	22.67***
Korea	21.16***	0.87	1.16	28.39***	6.74***	1.44	13.83***	-	5.02**	30.96**	29.56***	19.91***	47.27***	12.96***
Malaysia	0.22	0.36	8.54***	0.05	1.45	0.02	0.14	2.81*	-	15.94***	7.22***	2.01	0.17	0.02
Philippines	15.24***	0.56	0.18	12.49***	6.22**	0.88	1.19	17.02**	6.51***	-	16.35***	0.91	14.47***	25.48***
Singapore	0.36	14.77***	0.73	5.12**	1.03	0.27	2.01	12.73***	5.80**	18.13***	-	1.88	1.07	0.01
Taiwan	44.32***	0.17	0.64	25.73***	10.01***	0.69	26.08***	57.69***	2.41	34.35***	20.40***	-	48.78***	2.64
Thailand	11.13***	0.19	0.13	12.08***	17.64***	0.26	4.04**	24.79***	0.25	4.82**	4.00**	26.77***	-	0.30
U S	62.09***	0.74	4.0 5**	80.37***	33.84***	3.99**	55.26***	73.03***	0.01	25.98***	4.67**	33.98***	72.02***	-

TABLE 4: LIQUIDITY SPILLOVER USING AMIHUD RATIO: FULL SAMPLE PERIOD

	UK	AUS	FRA	GER	US	MAY	SKR	IDN	PHP	THI	CHN	SGP	JPY	TAI	From Others
UK	75.2	0.1	0.7	9.9	2	0.6	0.3	1.8	0.3	0.3	0	0.3	4.7	3.8	25
AUS	0.1	98	0	0.4	0.1	0.1	0.1	0	0.1	0	0	0.9	0.2	0.1	2
FRA	1	0.1	82.2	5.8	4.5	1.3	0.2	0.1	0.5	0.2	0.5	0.1	1.6	2	18
GER	5.9	0.4	2.1	66.6	7.7	0.9	0.8	1.1	0.1	0.9	0.1	1.5	2.6	9.3	33
US	1.5	0.6	1.8	7.2	69.5	1.4	0.2	0.1	0.4	0.6	2.8	0.2	4.6	9.1	30
MAY	0.1	0	0	3.2	1.7	71.2	0.1	0.9	20.2	0.1	0.4	1.4	0.2	0.4	29
SKR	0.6	0.1	0.2	3.8	0.1	1.2	78	5.2	0.6	1.6	0.3	0.6	1.2	6.2	22
IDN	0.6	0	0.1	5.8	4.2	9.5	0.2	60.2	6.3	4.1	1	1.3	0.5	6.2	40
PHP	0.1	0	0	0.8	0.5	1.6	0.1	1.4	93	0.5	0.1	0.2	0.2	1.4	7
THI	1.4	0	0.1	1.2	1.8	0.3	0.2	3.1	0.4	83	0.2	0.1	1.2	7.6	18
CHN	0.1	0.1	0.3	0.9	6.8	4.2	0.2	0.9	1.2	0.7	84	0.2	0	0.7	16
SGP	0.5	0	0.1	5.6	3.2	2.6	0.4	2.3	0.3	0.6	0.3	78.3	1.5	4.1	22
JPY	1.3	0.1	1.3	1.3	5.1	1.1	0.3	0.2	0.3	0.4	0.1	0.2	78.9	9.4	21
TAI	0.5	0.1	0.4	1.1	0.5	0.1	0.9	2.9	0.2	9.9	0.3	0.7	12.7	69.6	30

Contribution to others	14	2	7	47	38	25	4	20	31	20	6	8	31	60	313
Contribution including own	89	100	90	114	108	96	82	80	124	102	90	86	110	130	Spillover index: 22.3%
Net Spillover	-11	0	-11	14	8	-4	-18	-20	24	2	-10	-14	10	30	

Note: This table Reports the spillover of the Illiquidity among the countries. The diagonal value represents the spillover of own country, off diagonal element for each column represents spillover to other countries and every row represents Spillover received from other countries. Spillover index is ratio of total contribution to others divided by total contribution including own. Table represents abbreviated form of counties name where, UK stands for United Kingdom, AUS for Australia, FRA for France, GER for Germany, US for united states, May for Malaysia, SKR for South Korea, IDN for India, PHP for Philippines, THI for Thailand, CHN for China, SGP for Singapore, JPY for Japan, TAI for Taiwan.

TABLE 5: LIQUIDITY SPILLOVER USING AMIHUD RATIO: DURING CRISIS															
	UK	AUS	FRA	GER	US	MAY	SKR	IDN	PHP	THI	CHN	SGP	JPY	TAI	From Others
UK	60.5	4.7	1.4	6	11.1	1	0.3	1.1	0.9	2.6	0.3	1.7	4.6	3.8	39
AUS	1.6	54.9	1.2	13.5	5.4	1.1	0.9	0.5	0.8	1.9	0.2	14.5	2	1.4	45
FRA	1.1	0.3	70.2	5.5	2.2	2.8	1.5	0	1.1	0.9	0.8	1.4	12	0.3	30
GER	0.2	10	8.8	48.4	6.8	2.8	3.2	0.3	0.7	0.2	0.4	1.2	8.2	8.8	52
US	0.5	4.4	4.4	4.4	45.5	0.7	0.5	0.7	1	0.6	1	0.4	23.4	12.5	54
MAY	1.4	0.9	0.5	7.4	1	81.9	0.9	0.2	2.3	0.1	0.2	0.9	1.8	0.5	18
SKR	0.2	0.9	2.5	7.5	2.3	1.1	56.4	9.6	5.3	1.2	0.8	1	2.6	8.5	44
IDN	0.8	1.4	0.7	2.8	3.7	5.8	0.3	56.7	22.2	0.6	1.2	0.5	1.6	1.8	43
PHP	0.2	0.8	0.5	0.9	1.7	4.9	1.5	6.7	79.7	1.1	0	0.9	0.2	1	20
THI	2.5	2.4	5	2	11.9	1.9	0.6	0.5	1.6	65.6	0.2	0.5	0.9	4.5	34
CHN	0.5	0.1	1.2	1.5	1.4	1.4	1.4	1.3	0.4	0.8	88.2	0.9	0.2	0.7	12
SGP	2.6	0	1.8	5	3.7	2.1	2.1	0.7	0.5	1.2	0.9	69.9	5.2	4.3	30
JPY	0.9	1	6.4	5.3	1.4	2	0.4	0.5	5.1	0.2	0.5	0.1	71.7	4.7	28
TAI	0.2	1.9	1.2	2.8	1.6	2.1	2.3	1.7	3.3	3.7	0.1	0.2	36.3	42.6	57
Contribution to others	13	29	36	65	54	30	16	24	45	15	6	24	99	53	508
Contribution including own	73	84	106	113	99	112	72	81	125	81	95	94	171	95	Spillover index: 36.3%
Net Spillover	-26	-16	6	13	0	12	-28	-19	25	-19	-6	-6	71	-4	

Note: This table Reports liquidity spillover during the crisis. The period between Jan 2006 to April 2009 has been identified as crisis period using structural breaks. The diagonal value represents the spillover of own country, off diagonal element for each column represents spillover to other countries and every row represents Spillover received from other countries. Spillover index is ratio of total contribution to others divided by total contribution including own. Table represents abbreviated form of counties name where, UK stands for United Kingdom, AUS for Australia, FRA for France, GER for Germany, US for united states, May for Malaysia, SKR for South Korea, IDN for India, PHP for Philippines, THI for Thailand, CHN for China, SGP for Singapore, JPY for Japan, TAI for Taiwan.

TABLE 6: LIQUIDITY SPILLOVER USING AMIHUD RATIO: POST CRISIS

	UK	AUS	FRA	GER	US	MAY	SKR	IDN	PHP	THI	CHN	SGP	JPY	TAI	From Others
UK	74	0.1	0.5	11.8	1.7	0.3	0.3	1.9	0.6	0.1	0.9	1.2	5.9	0.7	26
AUS	0.1	97.9	0	0.1	0.1	0	0.1	0	0.1	0	0.5	0.2	0.8	0.1	2
FRA	1.3	0	86.7	6.5	2.9	0.1	0	0.3	0.1	0.3	0.2	0.4	0.5	0.6	13
GER	12.6	0	1.4	64.9	9.1	0.4	0.2	2.7	0.1	0.1	1	3.3	2.1	2	35
US	2.1	0.2	1.2	14.1	60.4	0.2	0.2	3.4	0.4	0.4	3.9	1.3	2.6	9.6	40
MAY	0.6	0.1	0.1	1.5	1	82.3	0.4	0.4	0.3	0.3	1.1	7.7	3.2	1	18
SKR	0.6	0.1	0.1	1.5	0.4	0.3	93.5	0.1	0	0.6	0.6	0.3	0.9	1	7
IDN	2.6	0.1	0.1	7.7	2.9	0.1	0.1	77.6	0.1	0.5	3.2	2.9	1.2	1	22
PHP	1	0.1	0.1	2.3	0.5	0.8	0.1	0.6	91.8	0.1	0.6	0.4	0.9	0.7	8
THI	3.6	0	0.1	2.9	1.1	0.4	0.6	3.9	0.1	81.3	1.6	2.1	1.7	0.7	19
CHN	0.7	0.7	0.2	1.3	5.5	0.7	2.8	1.6	0.3	0.6	68.9	0.2	7.8	8.8	31
SGP	5.5	0.1	0.3	7.3	2.5	6.6	0.2	1.7	0.2	1	0.7	70.5	2.8	0.5	30
JPY	3.5	0.4	0.4	2.6	8.2	3.8	1.1	2.9	0.2	0.9	8.9	1.9	53.1	12.1	47
TAI	0.2	0	0.1	2	4.5	0.7	0.4	0.6	0.7	1.1	9.6	0.1	5.6	74.2	26
Contribution to others	34	2	4	62	40	14	7	20	3	6	33	22	36	39	323
Contribution including own	108	100	91	127	101	97	100	98	95	87	102	92	89	113	Spillover index: 23.1%
Net Spillover	8	0	-9	27	0	-4	0	-2	-5	-13	2	-8	-11	13	

Note: This table Reports liquidity spillover in Post crisis Period. The period between April 2009 and December 2016 has been considered as Post-crisis period using structural breaks. The diagonal value represents the spillover of own country, off diagonal element for each column represents spillover to other countries and every row represents Spillover received from other countries. Spillover index is ratio of total contribution to others divided by total contribution including own. Table represents abbreviated form of counties name where, UK stands for United Kingdom, AUS for Australia, FRA for France, GER for Germany, US for united states, May for Malaysia, SKR for South Korea, IDN for India, PHP for Philippines, THI for Thailand, CHN for China, SGP for Singapore, JPY for Japan, TAI for Taiwan.

Liquidity Pull-Back and Predictability of Bond Yield Volatility: Evidence from India

Radeef Chundakkadan, Department of Humanities and Social Sciences, Indian Institute of Technology Madras, Chennai, 600036, Tamil Nadu, India

Subash Sasidharan, Department of Humanities and Social Sciences, Indian Institute of Technology Madras, Chennai, 600036, Tamil Nadu, India

ABSTRACT

This paper investigates the forecasting power of liquidity position on volatility of government securities' yields using daily data. We introduce a novel measure of liquidity called term repo spread, and call money rate in the empirical analysis. The result indicates that both the liquidity measures have significant predictive power on volatility of yields and term repo spread outperforms call money rate in most of the cases.

Keywords: Liquidity; Volatility; Government Securities; Forecasting.

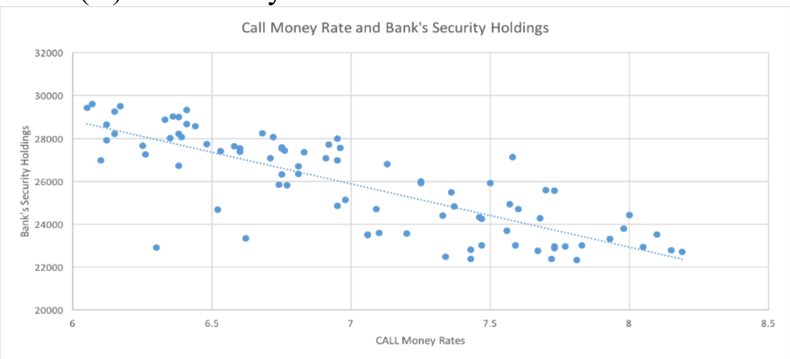
1. Introduction

Recent literature in the financial markets demonstrate the impact of liquidity swing of one interbank market on other financial markets (including other interbank markets). Eross et al. (2016) find evidence for inter-linkages between interbank markets such as LIBOR-OIS spread, Euro Fixed-Float OIS swap rate and three-month US-German Bond spread. Nyborg and Östberg (2014) show the relationship between daily interbank and stock markets. Jin (2015) reports volatility spillover between interbank and exchange T-bond markets. However, the relationship between interbank and bond market volatility is hardly explored in the financial literature. To fill this gap, our study intends to explore whether the volatility in Indian bond market can be explained by movements in interbank market rate. Further, Garcia de Andoain et al. (2016) shows that central bank as a lender-of-last-resort or liquidity provider has a significant influence on interbank markets. Therefore, our study extends to explore the impact of central bank's liquidity provision on bond yields.

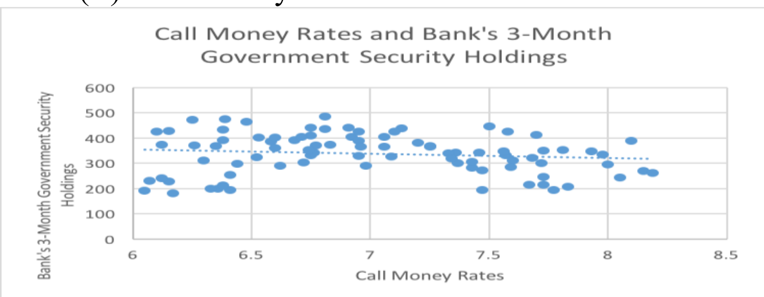
The present study is motivated by the work of Nyborg and Östberg (2014) on liquidity pull-back hypothesis. They argue that banks sell financial assets to meet day-to-day financial requirements during liquidity crunch in the interbank market. Since Indian banks have restrictions to invest in the equity market and they are dominant government security holder, we expect that banks sell off their securities to meet liquidity shortages. For this purpose, we depict bank’s government security holdings (in Rupees) against liquidity measures in Figure 1 and Figure 2 for a period ranging from October 2013 to October 2016. Figure 1 illustrates the relationship of call money rate (hereafter CALL) with bank’s investments in central and state government securities (Panel A), 3-month government securities (Panel B) and 1-year government securities (Panel C) along with trendline.

Figure 1. Call Money Rate and Bank’s Government Securing Holdings

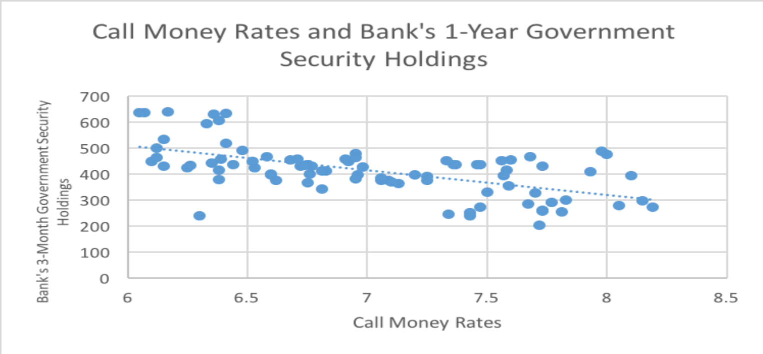
Panel (A) Call Money Rate and Bank’s Central and State Security Holdings



Panel (B): Call Money Rates and Bank’s 3-Month Government Security Holdings



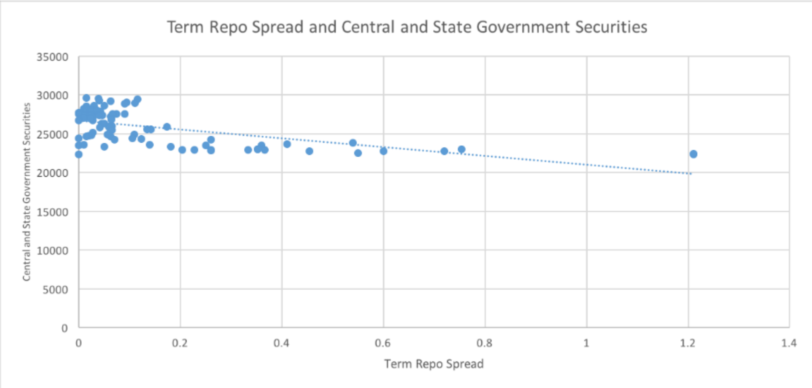
Panel (C): Call Money Rates and Bank’s 1-Year Government Security Holdings



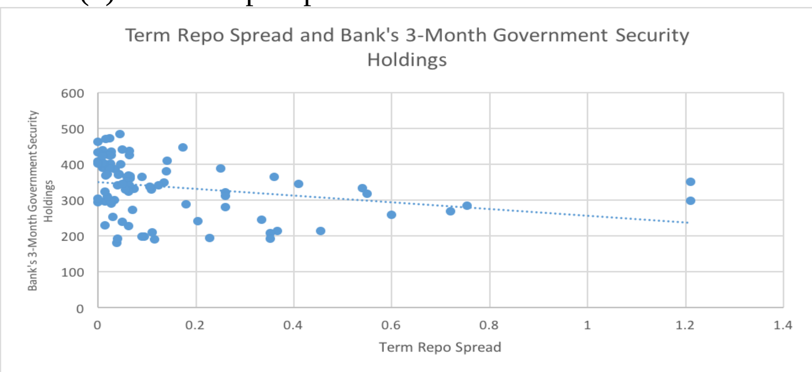
Note: Panel A depicts call money rate (in %) against bank's government securities holdings (in billion Rupees). Panel B and C depicts call money rate (in %) against bank's investments in 3-month and 1-year government securities (in billion Rupees) respectively.

Figure 2. Term Repo Spread and Bank's Government Securing Holdings

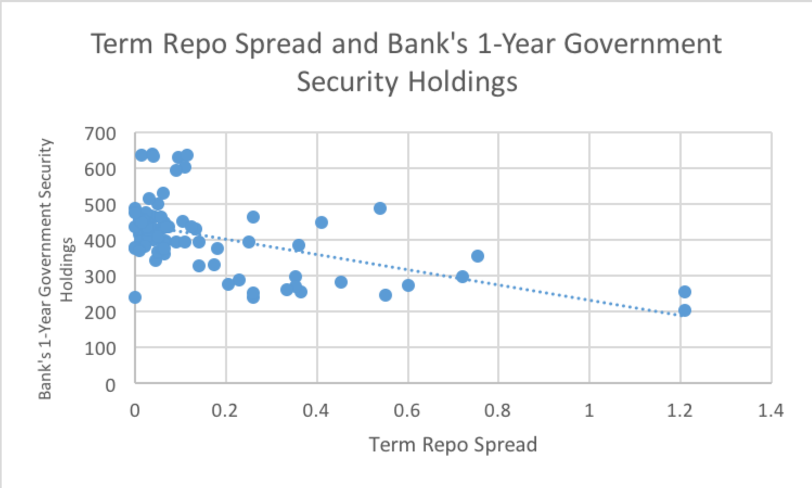
Panel (a) Term Repo Spread and Bank's Central and State Security Holdings



Panel (b): Term Repo Spread and Bank's 3-Month Government Security Holdings



Panel (c): Term Repo Spread and Bank's 1-Year Government Security Holdings



Note: Panel A depicts term repo spread (in %) against bank's government securities holdings (in billion Rupees). Panel B and C depicts term repo spread (in %) against bank's investments in 3-month and 1-year government securities (in billion Rupees) respectively.

All the trendlines in figure 1 are downward sloping indicating that when there is liquidity crunch in the interbank market or increase in the cost of borrowing from interbank market, government securities holding of banks is falling. Similarly, Figure 2 exhibits the relationship of term repo spread (hereafter REPO, which will explain in the latter part of this paper) with bank's investments in central and state government securities (Panel A), 3-month government securities (Panel B) and 1-year government securities (Panel C) along with trendline. The figure shows a downward sloping trendline as we see in Figure 1. Given these facts, we econometrically explore how well the liquidity measures (REPO and CALL) predict the daily government securities yields. If Nyborg and Östberg (2014)'s liquidity pull back hypothesis is true for India, we expect prices of the securities goes down or yield goes up. Moreover, liquidity pull-back have impact on all the government securities irrespective of their maturity period. In other words, liquidity in the market affects both short and long terms bonds. Therefore, we use 3-month, 1-year, 5-year and 10-year government securities in the empirical analysis.

As mentioned above, to investigate the predictability of government securities, we employ two set of liquidity proxies namely CALL and REPO rates. CALL is widely used overnight interbank rate the literature (Aleen, 2010; Kanjilal, 2011; Nath, 2015). However, REPO is newly used measure in this study and requires much attention. This measure is constructed from Reserve Bank of India (RBI) newly introduced liquidity operation called Term Repo Operation (hereafter TRO).[120] TRO allows banks to borrow from RBI at market rates, term repo rate, determined through auction against government securities, as a collateral, with a promise to repurchase the same on a predetermined date. Unlike interbank operations in many countries and Federal Reserve discount window in the US, term repo auctions are conducted for an amount notified by the RBI prior to the auction day and only interest rates (term repo

[120] Unlike open market operations (OMOs), both TRO have two legs- purchase and sale of securities- whereas OMOs are outright sale and purchase of securities. RBI conducts TRO through auction where the notified amount is announced prior. The minimum rate the banks can bid is called fixed repo rate which is fixed by RBI during the monetary policy committee meeting. The maturity period of these operations varies from overnight to 28 days. The auction generally takes place between 11:00 AM and 11:30 AM IST[120] or RBI announces it in advance. To know more about the operational issues, follow
https://rbi.org.in/scripts/NotificationUser.aspx?Id=8501&Mode=0.

rates) are determined on the days of auction. The benchmark rate for the auctions is called fixed repo rate which is fixed during monetary policy committee meeting and subject to change according to the macroeconomic conditions. Therefore, we use the weighted average of the spread between term repo rate and fixed repo rate of all outstanding TROs as REPO in our analysis instead of term repo rate.

In the empirical analysis, we use GARCH family models to predict volatility of government securities yields. It is well known that GARCH models have mean and variance equations which are to be jointly estimated. To investigate the main objective of the study, we augment REPO and CALL rates into the variance equations. Our analysis begins by implementing standard GARCH (1,1) model (Bollerslev, 1986) along with GARCH-REPO (1,1) and GARCH-CALL (1,1) models to grasp the significance of liquidity measures in explaining yield volatility. Then we extend our model to EGARCH (Nelson, 1991) and GJRGARCH models (Glosten et al., 1993) to accommodate asymmetric effects of positive and negative shocks on yields. Besides, we compare the persistence of volatility in standard GARCH family models with liquidity augmented GARCH family models. This activity helps to explore the contribution of liquidity measures in explaining the volatility of government securities' yields. Further, we use three variants of each GARCH family models to forecast the daily volatility in the government securities' yields. Finally, the forecasting performance of each model is evaluated by Sum of Squared Forecasting Errors, Theil's Inequality Coefficients and Out of Sample R-Squared.

A few notable recent works on Indian government securities are as follows. A set of studies by Kanjilal (2011; 2013) shows macroeconomic factors such as growth, inflation and monetary policy indicator (call money rate) has a significant effect on the shape of yield curve, especially after the introduction of Liquidity Adjustment Facility (LAF). Using a narrative-based measure of monetary policy, Sahoo and Bhattacharyya (2012) find yield curve is subject to change to countries monetary policy. More recently, Sensarma and Bhattacharyya (2016) derive a composite measure of monetary policy using principal component analysis and provide supporting evidence of the previous studies. Prasanna and Soumya (2017) demonstrate the impact of US bond yields on Indian securities.

This study makes value addition to the existing literature on the following grounds. *First*, to the best of our knowledge, this study is perhaps the first attempt to examine the relationship between daily bond yield volatility and interbank liquidity (using call money rate as the proxy). Such an empirical analysis helps to understand the liquidity pull-back hypothesis that banks sells financial assets during the time of liquidity shortage. If such a relationship exists, the magnitude persistence of volatility reduces after incorporating the call money rate into the variance equation. In that way, our study extends the literature which establishes relationship between interbank market and other financial markets (Nyborg and Östberg, 2014; Eross et al., 2016). Further, using of daily frequency data has its own advantages demonstrated in the financial literature. It includes capturing more information (Bollerslev and Wright, 2001), higher predictability (Narayan et al, 2013) and higher utility gain (Narayan and Sharma, 2015). *Second*, we extend our study to examine the relationship between central bank's liquidity provision on daily bond yield volatility. Since central bank is the liquidity provider at the end, their operations provide more precise information about liquidity position of the economy. Therefore, we use RBI's recent liquidity activity called TROs and create a novel liquidity measure called term repo spread. In that way, our study is a value addition to the literature of central banks' liquidity operations (McAndrews et al., 2017; Garcia de Andoain et al., 2016). *Finally*, we contribute to the literature on predictability of financial assets volatility (Vlastakis and Markellos, 2012; Chronopoulos et al., 2016; Luo and Zhang, 2017;). We forecast the daily volatility of government securities' yields based on interbank market liquidity (call money rate) and central bank's liquidity provision (term repo spread. For that, our study adopt out-of-sample forecasting technique similar to Narayan and Sharma (2015) in the empirical analysis.

The findings of the empirical analysis shows that the liquidity measures employed in the study have significant role in explaining the daily volatility of Indian government securities' yields. The out-of-sample forecasting performance indicates using CALL and REPO improves the predictability of yield volatility. Among the two liquidity measures, REPO outperforms CALL in the majority of cases.

The rest of this paper is organized as follows. Section 2 provides the data descriptions, summary statistics and results of unit root test. Section 3 deals with the methodology used in the empirical analysis. Section 4 presents empirical results. Finally, Section 5 concludes the study.

2. Data

Our study employs daily frequency data ranging from October 2013 to October 2016. The sample period begins when TRO got operationalized. This study analyzes the volatility of 3-month (G3M), 1-year (G1Y), 5-year (G5Y) and 10-year (G10Y) government security yields obtained from Bloomberg. Yields of the securites are used in their log-difference forms to obtain stationary process. There are missing observations for some securites and therefore number of observations are not same for all securities. Call money rates (CALL) are obtained from RBI database. Since there are more than one TROs and outstanding operations in a single day, term repo spreads (REPO) are constructed as the spread between term and fixed repo rates of all outstanding TROs and calculate their term-weighted averages. Data on TROs are obtained from RBI's daily press release on money market operations.

Table 1 reports the descriptive statistics and results of unit root test of the variables under study. Overall, all the four government yields register negative mean daily changes, with higher changes for short-term bonds. The mean value of REPO and CALL are 0.16% and 7.28% and the standard deviation is 0.22 and 0.87 respectively. Unit root test using Augmented Dickey-Fuller test rejected the null hypothesis that variables are nonstationary at 1% level for all the variables.

Table 1. Descriptive Statistics and Unit Root Test

VARIABLE	MEAN	STD-DEV	MIN	MAX	ADF	N
G3M	-0.00049	0.00554	-0.04654	0.02951	-9.23***	718
G1Y	-0.00045	0.01189	-0.14335	0.14046	-10.98***	686
G5Y	-0.00032	0.00522	-0.03282	0.02159	-10.75***	720
G10Y	-0.00032	0.00492	-0.03898	0.01959	-10.02***	722
REPO	0.15573	0.22058	0.01	1.3	-6.76***	722
CALL	7.27506	0.86927	4.92	11.13	-8.24***	722

Note: This table describes the summary statistics and results of Augmented Dickey-Fuller (ADF) test of the variables employed in the study. G3M, G1Y, G5Y and G10Y are log difference of 3-month, 1-year, 5-year and 10-year government security yields respectively. '***' indicates significance at 1% level.

3. Methodological Framework

3.1. Volatility Forecasting Models

To analyze the predictability of government security yields, we turn to the GARCH family models (Chronopoulos et al., 2017). These models jointly estimate both conditional mean and conditional variance equations. We begin our empirical analysis using following GARCH (1,1) model. The conditional mean equation is given as:

$$y_t = \mu + \theta y_{t-1} + \varepsilon_t \qquad (1)$$

where y_t is the log difference of bond yields at time t and $\varepsilon_t \sim N(0, \sigma_t^2)$ is the error term. Conditional variance equation is given as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \epsilon_t \qquad (2)$$

where $\alpha, \beta \geq 0$; $\omega > 0$ to ensure the positiveness of the conditional variance. $\alpha + \beta < 1$ to ensure unconditional variance is to exist. To examine the main objective of our study whether REPO and CALL can improve yield forecast, we extend our standard GARCH (1,1) model to GARCH-REPO (1,1) and GARCH-CALL (1,1) models. The conditional variance of the former and the latter is given as equation 3 and 4 respectively.

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta REPO_{t-1} + \epsilon_t \qquad (3)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta CALL_{t-1} + \epsilon_t \qquad (4)$$

where REPO and CALL are term repo spread and call money rate respectively. To accommodate the asymmetric response of volatility for positive and negative shocks, we employ GJRGARCH model proposed by Glosten et al. (1993) and EGARCH model proposed by Nelson (1991) in our study. The conditional variance equations of different variants of GJRGARCH (1,1) and EGARCH (1,1) models are given in equation (5-7) and (8-10) respectively.

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I_{t-1} + \epsilon_t \qquad (5)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta REPO_{t-1} + \gamma\varepsilon_{t-1}^2 I_{t-1} + \epsilon_t \qquad (6)$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \delta CALL_{t-1} + \gamma\varepsilon_{t-1}^2 I_{t-1} + \epsilon_t \qquad (7)$$

where I_{t-1}=1, if $\varepsilon_{t-1}^2 < 0$ and I_{t-1}=0, if $\varepsilon_{t-1}^2 \geq 0$

$$\ln(\sigma_t^2) = \omega + \gamma\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha\left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] + \epsilon_t \qquad (8)$$

$$\ln(\sigma_t^2) = \omega + \gamma\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha\left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] + \delta REPO_{t-1} + \epsilon_t \qquad (9)$$

$$\ln(\sigma_t^2) = \omega + \gamma\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha\left[\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}}\right] + \delta CALL_{t-1} + \epsilon_t \qquad (10)$$

3.2. Volatility Forecast performance

To analyze the forecasting performance of each model, we follow Chronopoulos et al. (2016) and obtain one-period ahead forecasts of conditional volatility. This procedure involves estimating the parameters for first $\ddot{T}$-periods of the sample and forecasting one-period ahead conditional volatility. Then, we re-estimate the parameters with an additional period added to the sample and forecasts one-period ahead conditional volatility. This process repeats itself till the sample period gets completely exhausted. Finally, we rely on Sum of Squared Forecasting Errors (SSFE), Theil's Inequality Coefficient (TIC), and Out of Sample R-squared (OSR) to evaluate forecasting performance.

$$SSFE = \sum_{t=\ddot{T}+1}^{\ddot{T}+h} (P_t - A_t)^2$$

where, P_t and A_t are the predicted and actual value of yield volatility respectively. Since A_t is not directly observable, our study uses actual variance of yields $[var(y_t)]$. We calculate

SSFE for three variants of each GARCH family models and choose lowest SSFE as better forecasting model.

$$TIC = \frac{\sqrt{\sum_{t=\ddot{T}+1}^{\ddot{T}+h} \frac{(P_t - A_t)^2}{h}}}{\sqrt{\sum_{t=\ddot{T}+1}^{\ddot{T}+h} \frac{{P_t}^2}{h}} + \sqrt{\sum_{t=\ddot{T}+1}^{\ddot{T}+h} \frac{{A_t}^2}{h}}}$$

The model which shows TIC less than one is considered as better forecasting model. In addition, the forecasting performance of the liquidity augmented models is also compared with the standard model. Therefore, Out of Sample R-squared (OSR) is used to evaluate the relative performance of liquidity augmented and the standard model.

$$OSR = 1 - \frac{\sqrt{\sum_{t=\ddot{T}+1}^{\ddot{T}+h} \frac{(\hat{P}_t - A_t)^2}{h}}}{\sqrt{\sum_{t=\ddot{T}+1}^{\ddot{T}+h} \frac{(\ddot{P}_t - A_t)^2}{h}}}$$

where, $\hat{P}_t$ and $\ddot{P}_t$ are the predicted values of liquidity augmented model and the standard model. A positive OSR indicates higher predictability of liquidity augmented model compared with the standard model.

4. Empirical Result

4.1. Volatility Forecasting Results

We begin by reporting the results of different GARCH model variants. Table 2 reports the GARCH (1,1), GARCH-REPO (1,1) and GARCH-CALL (1,1) models in Panel A, B and C respectively for four types of government securites. Panel A indicates that both the ARCH (α) and GARCH coefficients (β) are statistically significant at 1% level in GARCH (1,1) model. To examine whether liquidity measures have a significant effect on the volatility of bond yields, we examine the coefficients of liquidity measures (δ) in Panel B and C.

Table 2. Garch Results

Panel A: GARCH						
	μ	**θ**	**ω**	**α**	**β**	
G3M	-0.00030*	-0.17820***	0.00001***	0.23076***	0.47225***	
G1Y	-0.00053***	-0.23935***	0.00000	0.04348***	0.95551***	
G5Y	-0.00036***	0.00902	0.000005***	0.26951***	0.56319***	
G10Y	-0.00039**	-0.03212	0.000001**	0.05154***	0.90649***	
Panel B: GARCH-REPO						
	μ	**θ**	**ω**	**α**	**β**	**δ**
G3M	-0.00028*	-0.17006***	0.00001***	0.20294***	0.41555***	0.00002***
G1Y	-0.00008	-0.24101***	0.0000007	0.99743***	0.561697***	0.000007**
G5Y	-0.00040**	0.02353	0.000005***	0.31794***	0.41657***	0.00002***
G10Y	-0.00036**	-0.00484	0.00001***	0.09214**	3.92E-08	0.00007***
Panel C: GARCH-CALL						
	μ	**θ**	**ω**	**α**	**β**	**δ**
G3M	-0.00012	-0.16961**	2.24E-16	0.59581***	0.00049	0.000002***
G1Y	-0.00018***	-0.20364***	3.49E-09	0.99251***	0.57065***	0.0000001***
G5Y	-0.00031*	-0.00118	1.68E-10	0.32193***	0.45326***	0.000001***
G10Y	-0.00042***	-0.03675	5.65E-08	0.03221	0.93217***	0.0000001

Note: This table reports the results different variants of GARCH model. G3M, G1Y, G5Y and G10Y are 3-month, 1-year, 5-year and 10-year government security yields respectively.The mean equation in each panel is estimated using equation 1; variance equation is estimated using equation 2, 3 and 4 in Panel A, B and C respectively. '***', '**' and '*' indicate 1%, 5% and 10% significance level.

GARCH-REPO (1,1) suggest that REPO have significant effect on four types of government bond yields (prices). GARCH-CALL (1,1) reports a significant effect of CALL on all government bonds yields, except for G10Y. These findings imply that both liquidity measures used in the present study possess significant in-sample predictive power for government security yield volatility. To accommodate asymmetry, we report different variants of EGARCH and GJRGARCH in Table 3 and Table 4. All the coefficients of asymmetry (γ) in Table 3 are statistically significant at 1% level whereas Table 4 reports 8, out of 12, significant coefficients (γ). These results indicate the existence of asymmetry in the conditional yield distributions of government bonds. Turning towards the effect of REPO and CALL on yield volatility, all the coefficients of liquidity measures (δ) are positive. 7 out of 8 δs are statistically significant at conventional levels in Table 3 whereas 6, out of 8, δs are significant in Table 4.

Table 3. Egarch Results

Panel B: EGARCH							
	μ	**θ**	**ω**	**α**	**β**	**γ**	
G3M	-0.00028**	-0.19241***	-2.59845***	-0.21168***	0.74980***	0.21475***	
G1Y	-0.00029***	-0.37273***	-0.37171***	-0.07248**	0.94578***	0.84347***	
G5Y	-0.00035***	0.00880	-1.9983***	0.06483*	0.81012***	0.42662***	
G10Y	-0.00032***	-0.05850*	-0.37994***	0.02354*	0.96345***	0.11368***	
Panel C: EGARCH-REPO							
	μ	**θ**	**ω**	**α**	**β**	**γ**	**δ**
G3M	-0.00041***	-0.16219***	-3.15308***	-0.27887***	0.70473***	0.14063***	0.40949***
G1Y	-0.00030	-0.36782***	-0.45001	-0.08111	0.93801***	0.87533***	0.06311
G5Y	-0.00035***	0.01521	-2.67435***	0.05995	0.75280***	0.44718***	0.38427**
G10Y	-0.00040***	-0.03663	-1.31862***	-0.01092	0.88052***	0.08768***	0.27503***
Panel A: EGARCH-CALL							
	μ	**θ**	**ω**	**α**	**β**	**γ**	**δ**
G3M	-0.00013*	-0.21415***	-8.32267***	-0.38443***	0.49046***	0.55886***	0.40250***
G1Y	-0.00034***	-0.32398***	-4.82804***	-0.11355***	0.81163***	0.65845***	0.41810***
G5Y	-0.00032***	-0.00542	-3.8353***	0.04762	0.73454***	0.46507***	0.14395***
G10Y	-0.00037**	-0.02886	-1.07882***	0.00617	0.92445***	0.08799***	0.03823***

Note: This table reports the results different variants of EGARCH model. G3M, G1Y, G5Y and G10Y are 3-month, 1-year, 5-year and 10-year government security yields respectively. The mean equation in each panel is estimated using equation 1; variance equation is estimated using equation 5, 6 and 7 in Panel A, B and C respectively. '***', '**' and '*' indicate 1%, 5% and 10% significance level.

Table 4. Gjrgarch Results

Panel A: GJRGARCH							
	μ	θ	ω	α	β	γ	
G3M	-0.00040**	-0.13539**	0.00001***	0.026291*	0.50235***	0.38113**	
G1Y	-0.00001	-	0.000003*	0.32097**	0.65611***	0.04382	
G5Y	-0.00029**	0.02418	0.000006***	0.36320**	0.52424***	-0.19992**	
G10	-	-0.04075	0.00000003*	0.01073**	0.99719***	-	
Panel A: GJRGARCH-REPO							
	μ	θ	ω	α	β	γ	δ
G3M	-0.00038**	-0.14416**	0.000009***	0.0000000	0.43165***	0.44600**	0.00002**
G1Y	-0.00009	-	0.0000006	0.98674**	0.56111***	0.02189	0.000007
G5Y	-0.00032*	0.05674	0.000007***	0.49476**	0.31395***	-0.30998**	0.00002***
G10	-0.00037**	-0.01135	0.00001***	0.035082	4.38E-09	0.09546	0.00007***
Panel C: GJRGARCH-CALL							
	μ	θ	ω	α	β	γ	δ
G3M	-	-	1.12E-11	0.29403**	0.04148	0.99818**	0.000002***
G1Y	-0.00017	-	6.49E-13	0.98927**	0.57195***	-0.00439	0.0000001**
G5Y	-0.00025	0.00077	2.22E-16	0.34684**	0.519830**	-0.11813*	0.0000008**
G10	-0.00034**	-0.03089	4.08E-08	0.03240**	0.97062***	-	4.12E-08

Note: This table reports the results different variants of GJRGARCH model. G3M, G1Y, G5Y and G10Y are 3-month, 1-year, 5-year and 10-year government security yields respectively.The mean equation in each panel is estimated using equation 1; variance equation is estimated using equation 8, 9 and 10 in Panel A, B and C respectively. '***', '**' and '*' indicate 1%, 5% and 10% significance level.

4.2. Reduction in Volatility Persistence

Table 5 reports persistence of volatility ($\alpha + \beta$) in each model employed above and reduction in the persistence of volatility after the inclusion of REPO and CALL in the models. In general, the results indicate that the inclusion of liquidity variables reduced persistence significantly in most of the cases. 20 out of 24 cases liquidity measures reduce the persistence of volatility in government bond yields. Only G1Y in GARCH and GJRGARCH models have positive change in persistence of volatility after the addition of REPO and CALL in the models. Keeping G1Y aside, REPO and CALL reduces an average of 28% and 17.75% of volatility persistence respectively. GJRGARCH model account highest (27.18%) in reducing persistence compared to GARCH (22.6%) and EGARCH (19.8%) models.

Table 5. Persistence of Volatility

Panel A: GARCH					
	$\alpha + \beta$			**Reduction in Persistence (%)**	
	GARCH	**GARCH-REPO**	**GARCH-CALL**	**GARCH-REPO**	**GARCH-CALL**
G3	0.703	0.618	0.596	-12.02%	-15.18%
G1Y	0.999	1.559	1.563	56.07%	56.47%
G5Y	0.833	0.735	0.775	-11.79%	-6.91%
G10	0.958	0.092	0.964	-90.38%	0.66%
	EGARCH	**EGARCH-REPO**	**EGARCH-CALL**	**EGARCH-REPO**	**EGARCH-CALL**
G3	0.538	0.426	0.106	-20.86%	-80.30%
G1Y	0.873	0.857	0.698	-1.88%	-20.06%
G5Y	0.875	0.813	0.782	-7.11%	-10.61%
G10	0.987	0.870	0.931	-11.89%	-5.71%
Panel C: GJRGARCH					
	GJRGARC	**GJRGARCH-**	**GJRGARCH-**	**GJRGARCH-**	**GJRGARCH-**
G3	0.529	0.432	0.336	-18.35%	-36.53%
G1Y	0.977	1.548	1.561	58.42%	59.78%
G5Y	0.887	0.809	0.867	-8.87%	-2.34%
G10	1.008	0.035	1.003	-96.52%	-0.49%

Note: This table shows the persistence of volatility, calculated by the sum of ARCH and GARCH coefficients, for three variants of each GARCH family models, and percentage change in persistence after using liquidity augmented models. G3M, G1Y, G5Y and G10Y are 3-month, 1-year, 5-year and 10-year government security yields respectively.The variance equation is estimated using equation 2-4, 5-7 and 8-10 in Panel A, B and C respectively. '***', '**' and '*' indicate 1%, 5% and 10% significance level.

4.3 Forecasting Performance

Panel A of Table.6 presents the sum of SSFE of the volatility of different government bond yields using different variants of volatility models. Boldface indicates lowest forecasting errors and superior forecasting performance. The result suggests that different types of bond yields are better forecasted by different variants of GARCH models. However, REPO has lowest forecasting error in most of the cases (G3M, G5Y and G10Y). Interestingly, CALL outperforms none of the yield volatility forecast errors in our analysis. Therefore REPO is superior to CALL in the prediction of yield volatility in term of SSFE.

Panel B reports the TIC for predictions of yield volatility using different variants of GARCH. The results show that all the TICs are less than one, suggesting the predictability of yield volatility using GARCH family models. Boldface indicates the lower TIC for each government bonds. As reported above, different types of bond yields are better forecasted by different variants of GARCH models. The lowest TIC for G3M, G5Y and G10Y are accounted by REPO augmented models whereas lowest TIC for G1Y is by CALL augmented model. Panel C presents our final forecasting performance measure, OSR. In our analysis, OSR compares augmented GARCH and GARCH models.

Table 6. Forecasting Performance

	Panel A: Sum of Squared Forecast Errors								
	GARCH	**GARCH-REPO**	**GARCH-CALL**	**EGARCH**	**EGARCH-REPO**	**EGARCH-CALL**	**GJRGARCH**	**GJRGARCH-REPO**	**GJRGARCH-CALL**
G3	0.00	**0.00034**	0.00285	0.002	0.00555	1.81	0.0033	0.0078	0.0254
G1	0.17	0.138	0.167	0.154	0.153	0.14	0.165	**0.135**	0.177
G5	0.00	0.00582	0.00389	0.002	0.00267	0.00394	**0.0018**	0.0066	0.0038
G1	0.00	0.0139	0.00128	0.001	**0.00073**	0.00168	0.0022	0.00538	0.00202

Panel B: Theil's Inequality Coefficients

	GARC	**GARCH-**	**GARCH-**	**EGARCH**	**EGARCH-REPO**	**EGARCH- CALL**	**GJRGARCH**	**GJRGARCH-REPO**	**GJRGARCH- CALL**
G3	0.13	**0.093**	0.255	0.235	0.299	0.907	0.277	0.318	0.500
G1	0.75	0.564	0.639	0.612	0.610	**0.517**	0.705	0.551	0.636
G5	0.30	0.309	0.313	0.305	**0.270**	0.317	0.272	0.320	0.297
G1	0.27	0.483	0.254	0.289	**0.182**	0.333	0.413	0.462	0.395

Panel C: Out-of-Sample R-Squared

	GARCH-REPO	**EGARCH-REPO**	**GJRGARCH-REPO**	**GARCH-CALL**	**EGARCH-CALL**	**GJRGARCH-CALL**
G3	0.287	-0.570	-0.523	**0.470**	-4.911	0.062
G1	**0.098**	0.004	0.098	0.027	0.071	0.027
G5	-0.487	-0.084	-0.872	0.046	0.010	**0.071**
G10	-1.882	**0.288**	-0.561	0.069	-0.037	-0.013

Note: This table shows forecast evaluation measures. G3M, G1Y, G5Y and G10Y are 3-month, 1-year, 5-year and 10-year government security yields respectively. Boldface indicates superior performance for each security. Panel A, B and C report Sum of Squared Forecast Errors (which is reported as 105 of the actual value), Theil's Inequality Coefficients, and Out-of-Sample R-squared respectively.

The result indicates that CALL augmented model outperform standard GARCH model for G3M and G5Y while REPO augmented model outperform standard GARCH model for G1Y and G10Y. Moreover, given the three forecasting evaluation measures, REPO outperformed CALL in 8 out of 12 cases whereas CALL outperformed REPO only 3 cases. Therefore, to conclude, even though both the liquidity variables have predictive power over daily yield volatility, REPO is superior to CALL.

5. Conclusion

The present paper examines the impact of liquidity in Indian interbank market on the volatility of government securities' yields and their predictability. We use two liquidity measures in the study namely term repo spread and call money rate to access daily yields of 3-month, 1-year, 5-year and 10-year bonds. A family of GARCH models such as standard GARCH, EGARCH and GJRGARCH models is employed in the empirical analysis. To access the forecasting performance, we use measures such as Sum of Squared Forecasting Errors, Theil's Inequality Coefficients and Out of Sample R-Squared.

The results of our study indicate that both the liquidity variables employed in the study have significant influence of daily bond yields. The inclusion of these variables reduces the persistence of volatility in government securities' yields. Our forecasting performance suggests that the liquidity augmented models outperform the standard GARCH family models. In nutshell, liquidity in interbank market have significant predictable power on the volatility of daily bond yields and term repo spread outperform call money rate in most of the cases.

References

Aleem, A. (2010). Transmission mechanism of monetary policy in India. Journal of Asian Economics, 21(2), 186-197.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Bollerslev, T., & Wright, J. H. (2001). High-Frequency Data, Frequency Domain Inference, and Volatility Forecasting. Review of Economics and Statistics, 83(4), 596-602.

Chronopoulos, D. K., Papadimitriou, F. I., & Vlastakis, N. (2017). Information demand and stock return predictability. Journal of International Money and Finance.

Eross, A., Urquhart, A., & Wolfe, S. (2016). Liquidity risk contagion in the interbank market. Journal of International Financial Markets, Institutions and Money, 45, 142-155.

Garcia-de-Andoain, C., Heider, F., Hoerova, M., & Manganelli, S. (2016). Lending-of-last-resort is as lending-of-last-resort does: Central bank liquidity provision and interbank market functioning in the euro area. Journal of Financial Intermediation, 28, 32-47.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. The journal of finance, 48(5), 1779-1801.

Jin, X. (2015). Asymmetry in return and volatility spillover between China's interbank and exchange T-bond markets. International Review of Economics & Finance, 37, 340-353.

Kanjilal, K. (2011). Macroeconomic factors and yield curve for the emerging Indian economy. Macroeconomics and Finance in Emerging Market Economies, 4(1), 57-83.

Kanjilal, K. (2013). Factors causing movements of yield curve in India. Economic Modelling, 31, 739-751.

Luo, X., & Zhang, J. E. (2017). Expected stock returns and forward variance. Journal of Financial Markets, 34, 95-117.

Narayan, P. K., Narayan, S., & Sharma, S. S. (2013). An Analysis of commodity Markets: What Gain for Investors?. Journal of Banking & Finance, 37(10), 3878-3889.

Narayan, P. K., & Sharma, S. S. (2015). Does Data Frequency Matter for the Impact of Forward Premium on Spot Exchange Rate?. International Review of Financial Analysis, 39, 45-53.

Nath, G. C. (2015). Repo market–A tool to manage liquidity in financial institutions. Macroeconomics and Finance in Emerging Market Economies, 8(3), 286-305.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 347-370.

Nyborg, K. G., & Östberg, P. (2014). Money and Liquidity in Financial Markets. Journal of Financial Economics, 112(1), 30-52.

Prasanna, K., & Sowmya, S. (2017). Yield curve in India and its interactions with the US bond market. International Economics and Economic Policy, 14(2), 353-375.

Sahoo, S., & Bhattacharyya, I. (2012). Yield Curve Dynamics of the Indian G-Sec Market: A Macro-Finance Approach. Indian Economic Review, 157-182.

Sensarma, R., & Bhattacharyya, I. (2016). Measuring monetary policy and its impact on the bond market of an emerging economy. Macroeconomics and Finance in Emerging Market Economies, 9(2), 109-130.

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. Journal of Banking & Finance, 36(6), 1808-1821.

Marshall Lerner Condition and the Balance of Payments Constrained Growth: The Spanish Case

Luis Sastre, Departamento de Teoría Económica y Economía Matemática

ABSTRACT

The analytical reformulation of the Marshall-Lerner condition developed in this paper has a series of implications for empirical studies about the Foreign Sector: export and import functions must be estimated considering the possibility of high cross correlations between them in the modelled countries. The analytical development lead us to reformulate also the Thirwal´s model in open economies, in the following terms: in the long run, if the Marshall-Lerner condition is maintained, the balance of payments constrained growth income not only depends on export and import income elasticities but also on the cross elasticities values between exports and imports.

JEL Classification: F41
Keywords: Marshall-Lerner condition, export and import flow simultaneity, price-elasticities, cross elasticities

1. Introduction.

The economic globalisation process that has been characteristic of the evolution of the world's economic system in the last few decades has been analysed in depth from a financial perspective, but studies of its implications for the real economy, and particularly the foreign sector of national economies, have not been as plentiful. Traditionally, economic theory has analysed the foreign sector in relation to compliance with the Marshall-Lerner condition (Lerner, 1934, 1952), according to which "for a currency devaluation to have a positive impact on trade balance, the sum of price elasticity of exports and imports (in absolute value) must be greater than 1". This condition implicitly assumes that the GDP is independent from the exchange rate. In a globalised economy in which trade has heavily increased, there are many countries in which the ratio between the trade balance and the GDP is very high, so this assumption is not sustained. The theory of flexible exchange rates was developed, and it was shown that if the real exchange rate is flexible and the so-called Marshall- Lerner condition is satisfied, the balance of payments will equilibrate, without income adjustment. This may not be the case in the short run, or because of the nature of goods exported and imported by a particular country.

The economic literature includes numerous theoretical and empirical studies of the impact of exchange rate variations on the balance of trade; despite their number, they fail to agree on the effect of currency devaluation on trade balance, so it is an open question.
Thirwall (1979) developed a model started from proposition that no country can grow faster than that rate consistent with balance of payments equilibrium on current account, unless it can finance ever-growing deficits, which in general cannot. It is also the basis of Krugman´s rule (1989) that one country´s growth rate relative to another´s will be equiproportional to the ratio of its income elasticities of demands for exports and imports if the real exchange rate is constant.

Oskooee-Bahmani (1998) employs a long-run method, cointegration technique, to estimate trade elasticities in less developed countries. In most cases the results reveal that indeed trade elasticities are large enough to support devaluation as a successful policy for improving the balance of trade.

Wilson (2001) analyses the impact of currency devaluation on the trade balances of Malaysia, Korea and Singapore, concluding that there is not *J-curve* for these countries, where the Marshall-Lerner condition is not met. Mahmud, Ullah and Yucel (2004), using non-parametric techniques to estimate the price elasticity of the exports and imports of six developed countries, find that the Marshall-Lerner condition is only partly met in some sub-sample periods. Mancies (2005), based on recent estimations of the Australian balance of trade, finds that the Marshall-Lerner condition is met in the 1999-2001 period. Pierdzioch (2005), using a general equilibrium model, reaches the conclusion that international capital mobility only increases the short-term effects on output if the Marshall-Lerner condition is met.

Sastre (2005), based on his estimations of the Spanish trade balance, concludes that an analysis of the effect of a currency devaluation upon trade balance should not only consider export and import price elasticities but also their cross elasticities. Matesanz and Fumorolas (2009), using multivariate cointegration tests and error-correcting models to obtain the determinants of the Argentinean balance of payments, find no empirical support for maintaining the Marshall-Lerner condition or existence of the *J-curve* in the short term.

Hsing (2010) tests for the Marshall-Lerner condition in eight selected Asian countries and policy implications. Applying a general functional form, the Marshall-Lerner condition of the bilateral trade between the US and Hong Kong, India, Japan, Korea, Malaysia, Pakistan, Singapore, or Thailand is examined. In deriving the real exchange rate, both the relative consumer price index (CPI) and the producer price index (PPI) are considered. The results show that the widely used log-log form can be rejected for Singapore and Malaysia using either the relative CPI or PPI, and is also inappropriate for India and Pakistan using the relative PPI. The Marshall-Lerner condition holds for India, Korea, Japan and Pakistan, is confirmed for Hong Kong, Singapore and Thailand using the relative CPI, and cannot be confirmed for Malaysia.

Welfens (2009) considers the impact of FDI inflows and FDI outflows and shows that the presence of (cumulated) FDI requires higher import elasticities in absolute terms than stated in the standard Marshall Lerner condition. One may derive a range for the elasticity of the ratio of exports to imports with respect to the real exchange rate, namely that the sum of absolute import elasticities at home and abroad must exceed unity plus an additional parameter.

2. The Theoretical Model.

In modern macroeconomic literature, the determinants of export and import flows in small countries with open economies are derived from models contemplating trade between two countries with a representative agent (see Ostry 1988, Obstfeld and Rogoff 1995, Reinhart 1995, Lombardo, 2011). The export and import demand functions are obtained by a dynamic optimisation process, in which the agent maximises his intertemporal utility for the consumption of two types of goods: one produced on site (not marketable) and another that is imported (marketable), subject to an intertemporal budgetary constraint[121]. The export and import demand for small, open economies would be, respectively:

$$x = \varphi\,(G^f, m, tcr) \qquad (1)$$

where $\dfrac{\partial G^f}{\partial tcr} = 0$, $\dfrac{\partial m}{\partial tcr} \neq 0$ and $\dfrac{\partial x}{\partial m} \neq 0$

$$m = \varphi\,(G, x, tcr) \qquad (2)$$

where $\dfrac{\partial G}{\partial tcr} = 0$, $\dfrac{\partial x}{\partial tcr} \neq 0$ and $\dfrac{\partial m}{\partial x} \neq 0$

G is the quantity of goods produced in the country (non-marketable); G^f is the quantity of non-marketable goods produced abroad and *tcr* is the real effective exchange rate or the ratio between foreign and domestic prices.

These equations, which express the simultaneity found in open economies between export and import flows, imply non-independence between the GDP and the exchange rate, and can thus be used for reformulation of the Marshall-Lerner condition, according to the classification of countries in relation to cross elasticities between exports and imports. We use the above export and import equations to analyse the Marshall-Lerner condition, starting with the general case of export and import simultaneity and proceeding to specific cases, including total independence between export and import flows, in which the Marshall-Lerner condition is maintained.The balance of trade (BC) would be:

[121] Krugman (1995), in order to focus on the effects of Newly Industrializing Economies (NIEs), assumes a model consisting of only two economies: one that is intended to represent the OECD, the other to represent the aggregate of NIEs and assuming that the OECD faces a rest-of-world offer curve m=f(x).

$$BC = x - m = \varphi(Y^f, tcr, m) - tcr\,\varphi(Y, tcr, x) \qquad (3)$$

Calculating the total impact of an exchange rate variation on the trade balance, we would have

$$\frac{dBC}{dtcr} = \frac{dx}{dtcr} - \frac{dm}{dtcr}$$

where $\frac{dx}{dtcr} - \frac{dm}{dtcr} = \left(\frac{\partial x}{\partial m}\right)\left(\frac{\partial m}{\partial tcr}\right) + \frac{\partial x}{\partial tcr} - \left(m - tcr\left(\left(\frac{\partial m}{\partial x}\right)\left(\frac{\partial x}{\partial tcr}\right) + \left(\frac{\partial m}{\partial tcr}\right)\right)\right)$

Considering the price elasticities of exports and imports and their cross elasticities:

$$e_{x,tcr} = \left(\frac{\partial x}{\partial tcr}\right)\left(\frac{tcr}{x}\right) \qquad \frac{\partial x}{\partial tcr} = e_{x,tcr}\left(\frac{x}{tcr}\right)$$

$$e_{m,tcr} = \left(\frac{\partial m}{\partial tcr}\right)\left(\frac{tcr}{m}\right) \qquad \frac{\partial m}{\partial tcr} = e_{m,tcr}\left(\frac{m}{tcr}\right)$$

$$e_{m,x} = \left(\frac{\partial m}{\partial x}\right)\left(\frac{x}{m}\right) \qquad \frac{\partial m}{\partial x} = e_{m,x}\left(\frac{m}{x}\right)$$

$$e_{x,m} = \left(\frac{\partial x}{\partial m}\right)\left(\frac{m}{x}\right) \qquad \frac{\partial x}{\partial m} = e_{x,m}\left(\frac{x}{m}\right)$$

And that at equilibrium, BC=0, so m = x/tcr. Replacing these expressions in (3), we obtain

$$\frac{dBC}{dtcr} = e_{x,m}\left(\frac{x}{m}\right)e_{m,tcr}\left(\frac{m}{tcr}\right) + e_{x,tcr}\left(\frac{x}{tcr}\right) - m + tcr\left[e_{m,x}\left(\frac{m}{x}\right)e_{x,tcr}\left(\frac{x}{tcr}\right) + e_{m,tcr}\left(\frac{m}{tcr}\right)\right]$$

$$m\,e_{x,m}e_{m,tcr} + m\,e_{x,tcr} - m + m\,e_{m,x}e_{x,tcr} + m\,e_{m,tcr} = 0$$

$$m\left[e_{x,tcr}(1+e_{m,x}) + e_{m,tcr}(1+e_{x,m}) - 1\right] = 0$$

And then:

$$\frac{dBC}{dtcr} = m\left(e_{x,tcr}\left(1+e_{x,m}\right) + e_{m,tcr}\left(1+e_{m,x}\right) - 1\right) = 0$$

According to the above expressions, the balance of trade would be improved by a currency devaluation when dBC/dtcr > 0, and therefore:

$$\left(e_{x,tcr}\left(1+e_{x,m}\right) + e_{m,tcr}\left(1+e_{m,x}\right)\right) > 1 \qquad (4)$$

3. Cross Elasticities Export Import and Balance of Payments Constrained Growth.

The model is based on Krugman´s rule that one country´s growth rate relative to another´s will be equiproportional to the ratio of its income elasticities of demand for exports and imports if the real exchange rate is constant (see Krugman, 1989 and Thirwall, 1991).

The simplest condition for a balance of payments in equilibrium is through the export and import demand functions. It follows that the rule of a balanced current account in the long term is

P_d*X = P_f*M

Where P_d and P_f are the prices of exports and imports in domestic currency

The export function depends on the relative prices of exports, the level of foreign income and imports

$$X = k * Y_f^{\in_f} * (\frac{P_d}{P_f} * tc)^{\in_{x,tc}} * M^{\in_{x,m}}$$

Where $\in_f$ is the income world elasticity of the exports; $\in_{x,tc}$ is the elasticity price of the exports and $\in_{x,m}$ is the cross elasticity export-import.

$$M = k1 * Y_d^{\in_d} * (\frac{P_f}{P_d} * tc)^{\in_{m,tc}} * X^{\in_{m,x}}$$

Where $\in_d$ is the income elasticity of the imports; $\in_{m,tc}$ is the elasticity price of the imports and $\in_{x,m}$ is the cross elasticity import-export. Transforming export and import into growth rates, we have the following system:

$x + p_d = m + p_f$ (5)

$x = \in_{x,tc} * (P_d - P_f - tc) + \in_f * y_f + \in_{x,m} * m$ (6)

$m = \in_{m,tc} * (P_f - P_d + tc) + \in_d * y_d + \in_{m,x} * x$ (7)

Where lower-case letters stand for the growth rates variables. The equilibrium of the current account would be:

Substituting (6) and (7) in (5)

$p_d + \in_{x,tc} * (P_d - P_f - tc) + \in_f * Y_f + m * \in_{x,m} = p_f + \in_{m,tc} * (P_f - P_d + tc) + \in_d * Y_f + x * \in_{m,x}$

The balance of payments equilibrium growth rate (y_d^*) would be:

$$y_d^* = \frac{(1 + \in_{x,tc} + \in_{m,tc}) * (p_d - p_f - tc) + \in_f * y_f + \lfloor m * \in_{x,m} - x * \in_{m,x} \rfloor}{\in_d} \quad (8)$$

This result modify the Thirwal´s model (see Thirwall 2011). Analyzing this formula, we can outline several conclusions about the dynamics of the equilibrium rate.

First is the effect of the different inflation between the local economy and abroad. The effect on the equilibrium growth rate depends on the sum of the price elasticities of exports and imports. If this amount is greater than 1, an increase in the domestic inflation in relation with the abroad inflation will decrease the equilibrium growth rate if the exchange rate is constant.

Second, the rate of growth will depend on the difference between exports and imports elasticities weighted by the level of exports and imports that make up the external sector of each economic system

We can consider the following four propositions for the cross elasticities between exports and imports.

Proposition 1

If $\varepsilon_{m,x}=0$ and $\varepsilon_{x,m}=0$, it characterises an economy that depends little on other countries, with zero correlation between exports and imports. Then we would have dBC/dtcr > 0 when

$$\left(e_{x,tcr} + e_{m,tcr}\right) > 1 \quad (9)$$

In this case, the Marshall-Lerner condition is maintained and the balance of payments equilibrium growth rate (y_d^*) would be:

$$y_d^* = \frac{(1 + \in_{x,tc} + \in_{m,tc}) * (p_d - p_f - tc) + \in_f * y_f}{\in_d} \quad (10)$$

If relative prices in international trade, or real exchange rates, are constant, equation (n) reduces to

$y_d^* = \frac{\epsilon_f * y_f}{\epsilon_d}$ (11)

The strong version of thirwall´s law

Proposition 2

If $\varepsilon_{m,x} \neq 0$ and $\varepsilon_{x,m}=0$, these conditions characterise an economy in which the demand for imports depends on exports, but exports do not depend on imports. In this case dBC/dtcr > 0 when

$$\left(e_{x,tcr} + e_{m,tcr}\left(1+e_{m,x}\right)\right) > 1 \quad (12)$$

This condition would correspond to economies in which many industries import raw materials or intermediate products and then export the final products. Krugman (1995) defines it as "slicing up the production process" and suggests that it is one of the leading causes of growth in world trade. For some countries with very open economies, he proposes import equations like $m=\phi(x,z)$, where *x* represents exports and *z* represents other determinants.

The balance of payments equilibrium growth rate (y_d^*) would be:

$$y_d^* = \frac{(1+\epsilon_{x,tc}+\epsilon_{m,tc}) * (p_d - p_f - tc) + \epsilon_f * y_f - x * \epsilon_{m,x}}{\epsilon_d} \quad (13)$$

If relative prices in international trade, or real exchange rates, are constant, equation (n) reduces to

$$y_d^* = \frac{\epsilon_f * y_f - x * \epsilon_{m,x}}{\epsilon_d} \quad (14)$$

Lower than Thirwall´s law

Proposition 3

If $\varepsilon_{m,x}=0$ and $\varepsilon_{x,m} \neq 0$, this would represent an economy in which exports depend on imports, but imports would not depend on exports. Then, dBC/dtcr > 0 When

$$\left(e_{x,tcr}\left(1+e_{x,m}\right) + e_{m,tcr}\right) > 1 \quad (15)$$

This would correspond to the economies of countries used by multinational corporations as logistic bases for their products. The theory also depends on "slicing up the production process". Multinational corporations do not react to unexpected changes in the demand for their products in the countries in which they operate by varying their production, which would lead to a significant increase in production costs, but by re-allocating their international stocks.

This process could be contemplated by the national accounts as imports and exports in the same period. Castillo and Picazo (1995) propose an indicator to measure "coincident trade", defined as when a company exports and imports the same type of product at the same time, concluding that this type of trade represented nearly 12 per cent of all foreign trade in Spain in 1988. The balance of payments equilibrium growth rate (y_d^*) would be:

$$y_d^* = \frac{(1+\epsilon_{x,tc}+\epsilon_{m,tc}) * (p_d - p_f - tc) + \epsilon_f * y_f + m * \epsilon_{x,m}}{\epsilon_d} \quad (16)$$

If relative prices in international trade, or real exchange rates, are constant, equation (n) reduces to

$$y_d^* = \frac{\epsilon_f * y_f + m * \epsilon_{x,m}}{\epsilon_d} \quad (17)$$

Higher than Thirwall´s law

Proposition 4

If $\varepsilon_{m,x} \neq 0$ and $\varepsilon_{x,m} \neq 0$, these would apply to an economy in which import demand depends on export demand and vice versa. In this case, dBC/dtcr > 0 and equation (4) would be:

$$\left(e_{x,tcr}\left(1+e_{x,m}\right)+e_{m,tcr}\left(1+e_{m,x}\right)\right)>1 \qquad (18)$$

In these economies, the empirical problem of estimating export and import flow determinants should be considered from the perspective of their simultaneity (see Mauleón and Sastre, 1992, 1996).Sastre (2005) estimates a cointegrated simultaneous two-equation model for the balance of trade in Spain, with high explanatory capacity for both export and import flows, as well as the balance of trade and its evolution in 1967-2002.

The balance of payments equilibrium growth rate (y_d^*) would be:

$$y_d^* = \frac{(1+\epsilon_{x,tc}+\epsilon_{m,tc})*(p_d-p_f-tc)+\epsilon_f*y_f+\lfloor m*\epsilon_{x,m}-x*\epsilon_{m,x}\rfloor}{\epsilon_d} \qquad (19)$$

If relative prices in international trade, or real exchange rates, are constant, equation (n) reduces to

$$y_d^* = \frac{\epsilon_f*y_f+\lfloor m*\epsilon_{x,m}-x*\epsilon_{m,x}\rfloor}{\epsilon_d} \qquad (20)$$

Higher o lower than Thirwall´s law, it will depend on the cross elasticities between exports and imports.

If $\frac{m}{x} = \frac{\epsilon_{m,x}}{\epsilon_{x,m}}$ the equation (20) reduces to

$$y_d^* = \frac{\epsilon_f*y_f}{\epsilon_d}$$

The strong version of thirwall´s law

4. Simultaneity between Export and Import flows: the Spanish case.

In relation to the Spanish economy, to study the long-run equilibrium relation between volume of imports and its determinants in one relation and the volume of exports and its determinants in another relation, we assume that the import and export demand equations take the following forms.

l_x = f(lr, lit, lm) (5)

l_m = f(lir, lpr, lx) (6)

Where *m* is the volume of imports of goods and services; *ir* is the national investment; *r* is the GDP of the OECD countries; and finally, *it* and *pr*, are the export and import price competitiveness indicators, respectively. The l stands for logarithm.

To establish whether a long-run equilibrium relationship exists between the variables in equations (5) and (6) for Spain, we use the Maximum Likelihood cointegration procedure proposed by Johansen (1988).

We apply the Johansen and Juselius (1990) method to determine the number of cointegrating vectors. The results of the λ-*max* and the trace showed the null hypothesis of no cointegration (r=0) among all variables that enter into the import and export demand equations can be rejected at the 5% level of significance by setting its estimated coefficient.

In order to interpret the estimated cointegrating vectors, we normalize them on one of the variables by setting its estimated coefficient equal to -1, so we obtain long-run trade elasticities.

This practice enables us to read the elasticities directly from cointegrating vectors. Applying Johanssen (1988) methodology to the relation (5) and (6) for the 1968-2003 period, and assuming that the vector has a VAR (2) structure, the cointegration vector obtained was:

$$lm = \underset{(19.5)}{0.84}lr - \underset{(-18.5)}{0.35}lpr + \underset{(19.5)}{0.51}lx$$

		Osterwald-Lenum 95%
Test λ-max	22.8	20.9
Test trace	30.5	29.7

$$lx = \underset{(15.3)}{1.20}lir - \underset{(-5.05)}{1.80}lit + \underset{(18.0)}{0.64}lm$$

		Osterwald-Lenum 95%
Test λ-max	22.8	21.4
Test trace	30.5	30.3

In the bracket next to each coefficient is the likelihood ratio test for each variable's exclusion from the cointegrating space. Our long-run approach supports the notion that devaluation could improve the Spanish balance of trade.

As the Spanish economy, since the country joined the euro area, has constantly been reducing its competitiveness, linked to a high trade deficit, and considering that it can no longer alter its exchange rate as an economic policy tool, foreign trade balance adjustments necessarily involve a policy based on internal price and salary adjustments.

5. Conclusions

The analytical reformulation of the Marshall-Lerner condition developed in this paper has a series of implications for empirical studies aimed at modelling the export or import flows of a given country, or testing compliance with the Marshall-Lerner condition: export and import functions must be estimated considering the possibility of high cross correlations between exports and imports in the modelled countries. To test for the existence of a positive impact in the long term of a currency devaluation on the balance of trade should be verified for each country.

The paper's analytical development leads us to reformulate the Thirwal´s model in open economies, in the following terms: In the long run, the balance of payments constrained growth income of countries with open economies not only depends on export and import income elasticities but also on the cross elasticities values between exports and imports.

In the case of Spain, the long-term estimations of the price elasticities of exports and imports, and the respective cross elasticities, lead us to conclude that currency devaluation would, in the long term, improve the balance of trade and increase the constrained growth income in relation with the strong version of thirwall´s law.

References

Castillo J., Picazo A., 1995. "El comercio coincidente en la empresa manufacturera española". Información Comercial Española. 746, 79- 88.

Hsing, Y., 2010. "Test of Marshall-Lerner Condition for Eight Selected Asian Countries and Policy Implications". Global Economic Review. 39, 91-98.

Johansen, S., 1988. "Statistical Analysis of Cointegration Vectors". Journal of Economic Dynamics and Control. 12(2-3), 231-254.

Johansen, S., Juselius, K., 1990. "Maximum Likelihood Estimation and Inference on Cointegration – With Application to the Demand for Money". Oxford Bulletin of Economics and Statistics, 52(2), 160-210.

Krugman, P., 1995. "Growing world trade: causes and consequences". Brookings Papers on Economic Activity, 1, 327-62.

Krugman P. (1989): "Differences in the income elasticities and trends in the real exchange rates". European Economic Review, May.

Lerner, A. P., 1934. "The Diagrammatical Representation of Cost Condition in Internacional Trade". Economica. 1, 319-334.

Lerner, A. P., 1952. "Factor Prices and International Trade". Economica. 19, 11-40.

Lombardo G., 2011. "On the trade balance response to monetary shocks: The Marshall-Lerner conditions reconsidered". Discussion Papers 98/5. University of York.

Mahmud, S.F., Ullah, A., Yucel, E.M., 2004. "Testing Marshall-Lerner condition: a non-parametric approach". Applied Economics Letters. 11 (4), 231-236.

Matesanz, D., Fugarolas, G., 2009. "Exchange rate policy and trade balance: a cointegration analysis of the Argentine experience since 1962". Applied Economics. 41(20), 2571-2591.

Mauleón, I., Sastre, L., 1994. "El Saldo Comercial en 1993: Un análisis econométrico". Información Comercial Española. 735, 167-172.

Mauleón, I., Sastre L., 1996 "An empirical model for the Spanish Foreign Trade". Economic and Financial Modelling. 3(3), 101-144.

Menzies, G. D., 2005. "Who is afraid of the Marshall-Lerner condition?". Economic Papers - Economic Society of Australia. 24(4), 309-318.

Obstfeld, M., Rogoff, K., 1995. "The Intertemporal Approach to the Current Account". NBER, working paper, 4893.

Oskooee-Bahmani, M., 1998. "Cointegration approach to estimate the long-run trade elasticities in LCDs". International Economic Journal. 12(3), 89-96.

Ostry, J., 1988. "The balance of Trade, Terms of Trade and Real Exchange Rates, International Monetary Fund. 35, 541-73.

Pierdzioch, C., 2005. "Capital Mobility, Consumption Substitutability and the Effects of Monetary Policy in Open Economies". German Economic Review. 6(1), 79- 87.

Reinhart, C., 1995. "Revaluation, Relative prices and International Trade", International Monetary Fund. 42, 290-312.

Sastre, L., 2005. "Simultaneidad exportaciones e importaciones, curva J y condición de Marshall-Lerner, en España". Información Comercial Española. 824, 209-232.

Thirwal A.P. and Hussein M.N. (1982): "The balance of payments constrain: capital flows and growth rate differences between developing countries". Oxford Economics Paper, November.

Wilson, P., 2001. "Exchange Rates and the trade balance for Dynamic Asian Economies-Does the J-Curve exist for Singapore, Malaysia and Korea?". Open Economic Review 12, 389-413.

Welfens P.J.J., 2011. "Marshall-Lerner Condition and Economic Globalization". International Economics and Economic Policy. DOI: 10.1007/s10368-010-0177-5.

What Determine Household Indebtedness? The Role of Housing Wealth Gain, Future Economic Outlook and Net Worth

Huson Joher Ali Ahmed, Department of Finance, School of Business, Deakin University

IKM Mokhtarul Wadud, Department of Economics, School of Business, Deakin University

Xueli Tang, Department of Economics, Deakin Business School, Deakin University

ABSTRACT

This paper studies the role of future economic outlook and capital gain on household debt in the selected OECD economies. The fascinating aspects for this topic is threefold. First, appreciation in property prices improves monetary value of collateral, which is highly valued by the lending institutions. Second, under the rosy future economic outlook, household financial condition largely improves through higher investment return from financial wealth, increase wages and higher income growth thereby reducing vulnerability of households due to indebtedness. Third, despite the deleveraging measures adopted across the economies since recent global financial crisis, the household leverage remains markedly higher which has a greater level of policy implications for both households and financial system as a whole. We report that the household sector tends to relax borrowing restriction as perceived future economic outlook improves and rise in household equity through rises in property price beyond and above the impact captured by traditional factors for household borrowing. As apparent, improved future economic outlook raises expected monetary values of the housing wealth, and hence higher expected property price can motivate household to improve living standard, hence promote a greater level of borrowing. We also show that there has been an overwhelming positive impact of housing value on household debt irrespective of whether an economy falls in the high or low debt category.

Keywords: household, indebtedness, GMM, OECD.

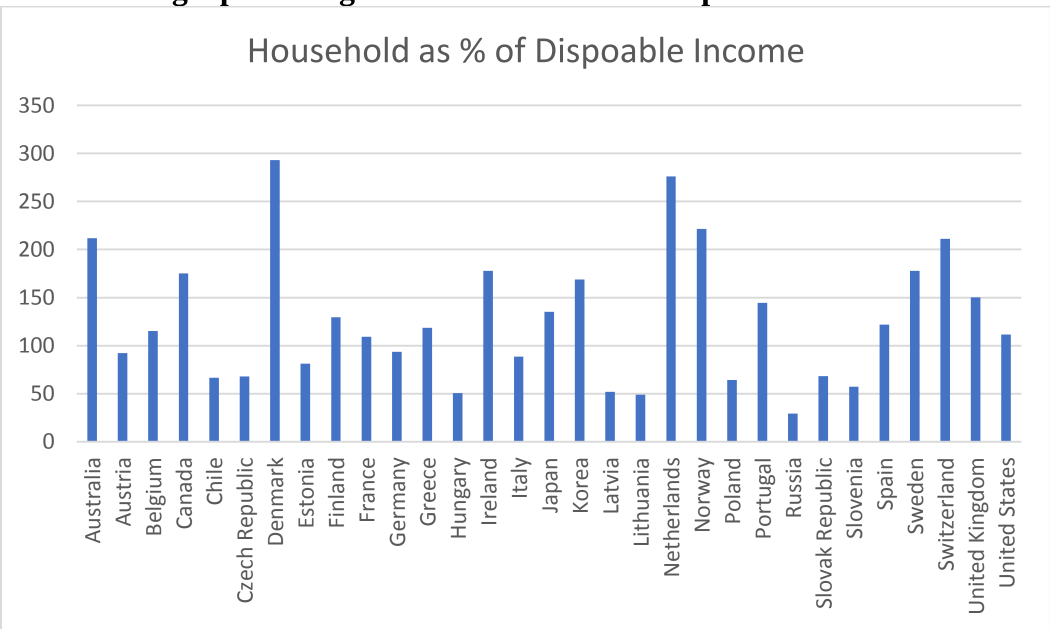
1. Introduction

There has been a plethora of studies on firms' leverage structure, household income growth and consumption decisions. However, little has been addressed on the empirical importance of determinant of household indebtedness. Household debt is defined as all liabilities, built up through borrowing for maintenance of improvement of living standards that require payment or payments of interest or principal by household to the creditor at a date or dates in the future. High level of household indebtedness delay recovery process from recession (Lamont, 1995). Therefore household debt might have major policy implications both at the micro and macro levels. According to life cycle theories, people should borrow in order to maximize overall their life cycle utility and has to borrow up to the limit, comfortable in repayment.

In the developed economies in particular, mortgage borrowing forms a major factor inducing household debt. After global financial crisis, although the deleveraging measures have been in place in many economies, household debt continued to rise markedly and stood at a historically high level in majority of OECD countries. The level of gross household debt ranges from less than 60% of net disposable income in some Central and Eastern European Countries (CEEC) to about 290% in the Netherlands and more than 295% in Denmark (Figure 1).

Thus it may be of interest to delve into the factors that explains household indebtedness beyond the ones that have been generally considered in literature such as the user cost of capital, household disposable income and growth in real wages. More specifically, one could ask if the household wealth gain promotes indebtedness among household and whether future rosy economic outlook promotes more debt among households. The answers of these queries could implicate both private and public policies with lasting effects on any economy.

Fig.1 percentage of household debt to disposable income



To the best of our knowledge there is no specific study conducted so far that substantiate the answers to the questions posed above. First, appreciation in property prices improves monetary value of collateral which is highly valued by the lending institutions. Second, under the rosy future economic outlook, household financial conditions largely improve through higher investment return, increase level of wages and higher income growth and hence reduced vulnerability of the households due to indebtedness. Third, debt can serve as a two-edge sword, if kept unchecked, over-borrowing induces household vulnerability thereby raising the possibility of financial ruin. However, if used wisely based on the household's ability to service the loan, household debt improves household welfare. Thus understanding household debt and what determine household indebtedness is of paramount importance with major implications for both household and policy makers.

Despite the fact that there is no direct relevant literature on the impact of housing wealth and economic outlook on household indebtedness, this study is partly motivated by the study of Main and Sufi (2017); Cambell and Cocco (2007). Cambell and Cocco (2007) suggest that housing is a major component of wealth, therefore it is important to understand how the fluctuations in house price affect households' consumption and apparent borrowing decisions. From households' perspective, home value appreciation experiences a matching increase in housing wealth. Holding everything else constant, any increases in housing value boost consumptions through higher level of borrowing. However the issue remains contested if any increase in nominal housing wealth equally matches the household's cost of living, rendering no expected changes in consumption, saving and borrowing. Main and Sufi (2009), suggest that one of the reasons for the rapid expansion in household leverage during 2002-2007 is that mortgage credit became more easily available to new home buyers. In a more recent article, Main and Sufi (2014) suggest that any increase in home value improves cash-in-hand if the financial institutions are willing to lend against the monetary value of collateral; and thus house value could be an important decision criterion for household spending and the level of debt. In this paper, we promote the idea that the rise in house value combined with rosy economic outlook may boost household to initiate consumptions through more borrowing. Rising house prices may stimulate consumption by increasing households' perceived wealth, or by relaxing borrowing constraints. Also, the higher expected property price could motivate banks to lower lending requirement, and hence could promote a greater level of lending and borrowing activities. Main and Sufi (2017) suggest that housing gains also increase the homeowner's access to "cash on hand" if credit markets are willing to lend against higher collateral value. In this paper we ameliorate these ideas by suggesting that household wealth gain through home value appreciation and the rosy future economic outlook raises the monetary value for collateral. Thus, households relax borrowing restriction based on household-specific ex-ante expectations of future income from rosy economic outlook.

There are two strands of literature which provide a better understanding on household consumption and borrowing decisions. First stream of literature often focuses on income expectation and household consumption and borrowing decisions. Starting point to this literature is the reference of permanent income hypothesis (PIH), pioneered by Milton Friedman (Friedman, 1957), which suggests that people will spend money at a level consistent with long-term average income. However, consumption should respond to an unexpected change in income which deviates from expectation. As suggested by Zeldes (1989), the consumption response could be significant when consumers face borrowing constraints or when precautionary saving motives are strong as the unanticipated rise in income reduces the income uncertainty and encourages immediate spending (Zeldes 1989; Carroll 1992, Carroll and Samwick, 1997). More recently, Jappelli and Pistaferri (2010) document anticipated and unanticipated income shocks bear different implications for the consumption response. Macroeconomic studies have found evidence linking scaled consumption to future income or asset returns, which could be interpreted as households consuming in anticipation of future income growth (Campbell, 1987, Lettau & Ludvigson, 2001). There are also household level studies that test a direct link between current consumption and future income, but show no convincing evidence of this fundamental PIH relationship (Carroll, 1994, Deaton and Paxon (1994) and Alessie & Lusardi, 1997). Nalewaik (2006), who used a synthetic cohort approach to find microdata support for a forward-looking relationship between consumption growth and income growth. Barba and Pivetti (2008) examines the rise in household indebtedness from the point of view of its causes and long-run macroeconomic implications. They suggest that the rising household debt is viewed as the outcome of persistent changes in income distribution and growing income inequalities. More recently, Agarwal and Qian (2014) examine the consumption and debt response to unanticipated income shocks using a natural experiment. They find that consumption rose significantly after the fiscal policy announcement: during the ten subsequent months, with marginal propensity of consumption being 0.8, meaning that for each$1 received, consumers on average spent $0.80. They also find a strong announcement effect—19 percent of the response occurs during the first two-month announcement period via credit cards.

Second stream of literature focuses on rising home value and consumption and the subsequent borrowing decisions. Our study belongs to this second stream of literature, led by the seminal studies of Mian and Sufi (2009; 2017); and Cambell and Cocco (2007). Mian and Sufi (2009) suggest that a single reason for the rapid expansion in household leverage during 2002-2007 is that mortgage credit became more easily available to new home buyers (Mian and Sufi 2009). They further maintain that strong house price appreciation from 2002 to 2006, which may have been fuelled by the availability of mortgage credit to a riskier set of new home buyers, could also have had an important feedback effect on household leverage through existing homeowners. As apparent, aggregate trend in household debt and house price is suggestive of the unobserved macroeconomic factors that affect growth in income expectation (Attanasio and Weber, 1994; Muellbauer and Murphy, 1997), which may jointly explain both household debt and home value. Mian and Sufi (2011) further suggest that home-equity based borrowing is not uniform across households. However, households with high credit card utilization rates and initial credit score has the strongest tendency to borrow against an increase in home equity. Cambell and Cocco (2007) content that housing gains increase the homeowner's access to "cash on hand" if credit markets are willing to lend against higher collateral value. The cash-on-hand effect can be an important driver of household spending, especially for constrained households with low levels of wealth. Cambell and Cocco (2007) further suggest that housing is an asset that can be used as collateral in a loan. Within Euro area, Anderson et al. (2012) analysed an exploratory study using Danish household data. They find that in 2010 about 20 percent of family with highest incomes after tax accounted for 53 per cent of the total family gross debt and the half with the lowest income accounted for 14 per cent in total of the gross debt. The report also suggests that the families with high debt also have the income required to service the debt.

Girouard, Kennedy, and André (2006) review a number of OECD economics, macroeconomics developments in household balanced sheet. The main findings suggest that the rise in household debt to historical level is driven by a combination of favourable financial conditions and buoyant housing markets. Besides, the reviews also report that the households' net worth has risen and provided the households with financial cushion against negative shock. The paper further analyses micro-level information which suggests that most of the debt among OECD economies is held by household with better disposable income.

The missing links in the existing literature on household indebtedness is the absence of accountability of households' wealth gain and the households' financial conditions with expectation of future income flow reflected by the economic outlook, in explaining household indebtedness in a panel framework. Given the limited understanding what determine household indebtedness, this study contributes in several ways over the previous relevant literature. First, we consider both expectation factor and wealth gain factor as two important factors of household borrowing decision criteria, which were largely ignored in the previous literature. It is plausible that prospective future economic outlook raises expected monetary values of the collaterals, housing ownership and more disposable income, and thus motivate households with or without borrowing constraints to relax borrowing restrictions with increased banking activities. The existing studies on household debt offer no pathway to understand how expectation factors are affecting household' borrowing beyond factors captured by borrowing cost and household wages. Second, given that household wealth comprises of both physical and financial assets, any volatility in stock market is likely to have effect on household wealth so thus on household indebtedness. This factor is largely condoned in the existing literature. Third, most of the studies on household debt is focused on a specific country or area, whereas in this study we extend our analyses encompassing 17 OECD economies in a panel framework. Such an approach has the potential to provide a better understanding of how an individual country within OECD makes borrowing decision based on future economic conditions and capital gain, thus confirming or contrasting initial assertions that household give more value to future economic outlooks and its impact beyond and above traditional factors such as user cost capital, wages and current economic outlook.

Our findings reveal some interesting features. Firstly, housing wealth gain consistently motivate household indebtedness after controlling all relevant factors. Secondly, as hypothesized household expectation of future economic outlook, reflecting family financial conditions, decision on buying big ticket items which is proxied by consumer confident index remains largely positively impacting household indebtedness and mostly significant across different scenario. Thirdly, increased financial market volatility (reduced financial stability) create more uncertainty about household wealth which is one of the core factors affecting the household borrowing decision, reduces household indebtedness. As increases in financial wealth enhances household borrowing capacity, household indebtedness increases. Lastly, we also find that the as the real wages increase, household indebtedness falls. These findings are robust to different measures of incomes and different market conditions. The rest of this paper is structured as follows. In section 2 we provide a brief overview of bank lending model and macroeconomic factors driving bank lending behaviour. Section 3 outlines the methods used in this study. Section 4 reports the results of our analysis and section 5 provides some concluding remarks.

2. The Model

2.1 The impact household wealth Gain and Future Economic Outlook

The growth of mortgage could be due to either the increase of supply or demand or both of the credit. By examining the origin of the subprime crisis in the US during 2005-2007, Mian and Sufi (2009) find that the increase of mortgage credit from the financial institutions could be an important reason. They also find that house price expectations-based explanation is less supportive to explain the expansion in supply of credit. In this paper, therefore, we focus on the demand side and examine whether the expectations-based explanation could help explain household borrowing behaviour in response to the future economic outlooks.

As argued in Aguiar and Gopinath (2007) and Mian et al (2017) that household would borrow more if they expect higher income in the future due to technology shocks, natural resource discovery or terms of trade shocks. With the higher credit demand, higher growth is also followed because of these positive shocks. In the theory section, we would like to explore the impact of future economic outlook on household indebtedness through the channel of credit demand and equity prices. It thus could offer insights regarding the household debt with respect to the changes of the expectations on the future economic outlook.

Following Mian et al (2017), we assume that output, y_t, follows a stochastic process due to the technology shocks, natural resource discovery or terms of trade shocks. Households maximize their lifetime utility $E_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$. They can borrow debt, d_t, and pay interest at rate r, in a small open economy, such that their budget constraint is as below,

$$c_t + (1+r)d_{t-1} = y_t + d_t \qquad \text{(a)}$$

Thus, by the standard optimization exercise, it is straightforward to get that household consumption at each period t equals their expected permanent income, $E_t y_t^p$, minus interest payments on outstanding debt, rd_{t-1}. Thus, by Mian et al (2017), the change of the debt at each time period t, Δd_t, will be a positive function of the expected future income. It implies that for given interest rate, any positive shocks will increase household debt due to the expectation of future income.

The more household can borrow from financial institutions, the higher house price increases, as Dynan et al (2012) found. By using the US data over the period of 2006-9 housing collapse, Mian et al (2013) find that the elasticity of consumption with respect to housing net worth could be of 0.6 to 0.8. Mian and Sufi (2011) also find a strong link between asset prices and household borrowing.[122] It means that as the collateral values for borrowing increase, as argued in Sinai and Souleles (2005) and Campbell and Cocco (2007), household, especially low-income families, would tend to borrow more for their consumption because of the concavity of the consumption function as examined in Deaton (1991) and Carroll and Kimball (1996).[123] It thus implies that due to the increase of the expectation on future economic outlook, households would not only borrow more to pay for the increase of the house price due to higher demand, but also borrow more for their consumption because of the higher collateral values of their houses. Thus, the household indebtedness increases significantly.

The house price and mortgage are highly related. Taking the household's leverage function from Dynan et al (2012), we have

$$\frac{d}{a} = 1 - \frac{1}{\frac{a}{nw}} \qquad \text{(b)}$$

where $\boldsymbol{a}$ denotes the house price and $\boldsymbol{nw}$ represents net worth. This function tells that for the whole economy, how the asset price, debt and net worth are correlated. For given net worth, higher debt due to the higher expectations on the future economic outlook leads to higher house price and thus leaving household more leverage than before. In contrast, the net worth

[122] Mian and Sufi (2010) found that household leverage could help explain the large fluctuations of house price and household consumption during the recession.

[123] Mian and Sufi (2014) examined the impact of house price on borrowing among high and low income households and found low income household have higher marginal propensity to consume and thus borrow more than high income household. In this paper, for simplicity, we assume households are homogeneous.

goes down while household borrow and spend more when they are optimistic about future economy due to positive shocks. It can thus drive the leverage ratio even higher. Household may bear such high leverage when they have better expectation on future economic outlook. Carroll, et al (2012), in contrast, claim that household may feel discomfort with high levels of leverage when they are afraid of job loss and feel pessimistic on the future. In addition, due to the recessions or crisis, net worth and house price could drop simultaneously with no change in debt. As argued in Dynan et al (2012), this leverage function also helps explain why the leverage ratio was very high during the subprime crisis.

2.2 Data Descriptions

The data comprises of aggregate annual data series on household debt to income ratio, real wages, short term interest forecast, real house price, consumer confident index, equity return volatility over a period from 1995 through 2015.
Houshold debt (HHD) is defined as all liabilities that require payment or payments of interest or principal by household to the creditor at a date or dates in the future. This indicator is measured as a percentage of Net disposable income (NDI). Data are under 2008 System of National Accounts (SNA 2008) for all countries except few.

Consumer confidence index (CCI) measures future economic outlook to reflect household financial situation over the past year and the coming year, anticipated economic conditions over the coming year and the next five years, and buying conditions for major household items.

Real House price (HV): The Housing indicator shows indices of residential property prices over time. The real house price is given by the ratio of nominal price to the consumers' expenditure deflator in each country, seasonally adjusted from the OECD national accounts database.

Short-term interest rates forecast refers (STI) to projected values of three-month money market rates. It is measured as a percentage. Forecast data are calculated by making an overall assessment of the economic climate in individual countries and the world economy as a whole, using a combination of model-based analyses and statistical indicator models. Share price indices are calculated from the prices of common shares of companies traded on national or foreign stock exchanges of respective economies considered in this analysis.

Household financial assets include (HHFA): currency and deposits securities, shares and other equity, net equity of households in life insurance reserves, net equity of households in pension funds, prepayments of premiums and reserves against outstanding claims, and other accounts receivable. Financial assets held by households form an important part of overall wealth and are an important source of revenue, either through the sale of those assets or refinancing, or as a source of property income (such as interest and dividends). This indicator, which is based on data in USD per capita at current PPPs, shows data by financial asset. Data are under 2008 System of National Accounts (SNA 2008) for all OECD member countries.

Household total net worth (HHNW) is the value of total assets (the total amount of financial assets plus the total amount of non-financial assets; note that this indicator only takes into account the value of dwellings from non-financial assets) minus the total value of outstanding liabilities. Household financial net worth is the balancing item of their financial balance sheet, i.e. total financial assets minus total liabilities, recorded at current market values.

2.3 Hypothesis development

H1: Do household borrowing behaviour change as house price appreciate (wealth gain)?
Based on both life cycle and on cash-on-*hand* models suggest that rising housing wealth is important for spending if it increases access to "cash on hand." These models also predict that the effect of cash-on-hand shocks on spending is strongest for households with low levels of existing cash on hand. Mian and Sufi (2015) find that household with low income zip code are likely to borrow more from increases in home value as compared with their counterpart. Further Rising property price or wealth gain from good future economic outlook improve higher monetary value for collateral, thus more loan is possible.

H2: household expectation of future income with rosy economic outlook may induce more consumption through higher level of borrowing. It is often argued that rational households are forward looking, i.e., they tend to relax borrowing restriction on household if economic outlook is rosy with a better income expectation and greater level of return from financial assets held by household. Likewise, Household may restrict borrowing if future economic activities pose gloomy outlook.

H3: the household net worth which comprises net of both real assets and financial asset may be considered by the lending institutions to be an important decision criteria for lending, thus household may be more embolden to borrow with the accumulated net worth.

3. Econometric Method

GMM based panel Model.

Dealing panel data often encounters the problem of unobserved heterogeneity corresponding to each sample space and problem due to non-zero correlation between the individual-fixed effect and lagged dependent variable. To remedy this, we use panel dynamic GMM to examine the impact of wealth gain and economic outlook and capital gain on household borrowing. GMM model produces consistent parameter estimates for a number of time periods, T, and a large cross-sectional dimension, N (see e.g. Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998).

In this study we adopt the first-difference GMM estimator by Arellano and Bond (1991) involves first differencing to wipe out the individual-specific effect and employing lagged-level variables as instruments to address the correlation between the error terms and the explanatory variables. The Arellano–Bond estimator sets up a generalized method of moments (GMM) problem in which the model is specified as a system of equations.
The base-line dynamic panel mode can be presented as:

$$y_{it} = \gamma y_{it-1} + x'_{it}\beta + \alpha_i + \lambda_t + u_{it}$$
$$\Delta x = \epsilon_{it}$$

Where, i =1,..,……,N and t =1,……..,T.
α_i and λ_t are unobserved individual and times specific effects.
y_{it} and y_{it-1} present ratio of household debt to disposable income at period t and t-1,
u_{it} is the idiosyncratic term with E(u_{it})=0, and (E($u_{it}u_{js}$) $= 0$ and t=s

X presents regressors, a set variables including real house price (HV), consumer confident index (CCI) to proxy for expectation of future economic outlook, stock price volatility (STOCKVOL) proxy for financial stability, Short term financial forecast (STINTF), household financial wealth (HHFW), and household net worth.

First difference GMM of (AB)

$$\Delta y_{it} = \Delta\gamma y_{it-1} + \Delta x'_{it}\beta + \Delta u_{it}$$

$$HHD_{it} = \alpha_i + \beta_1 CCI + \beta_2 \text{STIRF} + \beta_3 Wages_{it} + u_{it} \quad (1)$$
$$HHD_{it} = \alpha_i + \beta_1 HV + \beta_2 \text{STIRF} + \beta_3 Wages_{it} + u_{it} \quad (2)$$

Cross-derivative

$$HHD_{it} = \alpha_i + \beta_1 CCI * HV_{it} + \beta_2 \text{STIRF} + \beta_3 Wages_{it} + u_{it} \quad (3)$$
$$HHD_{it} = \alpha_i + \beta_1 CCI + \beta_2 HV_{it} + \beta_3 \text{STIRF} + \beta_{43} Wages_{it} + \beta_5 \text{STOCKVOL} + \beta_6 GOVD_{it} + u_{it} \quad (4)$$
$$HHD_{it} = \alpha_i + \beta_1 \text{STIRF} + \beta_2 Wages_{it} + \beta_3 \text{HV} + \beta_4 HHFW + u_{it} \quad (5)$$
$$HHD_{it} = \alpha_i + \beta_1 CCI + \beta_2 HV_{it} + \beta_3 \text{STIRF} + \beta_4 Wages_{it} + \beta_5 \text{HHNW} + u_{it} \quad (6)$$

Table1. Independent Variables and The Expected Signs Of The Parameters

Variable	Expected sign
HV	+
CCI	+
WAGES	-
STIRF	-
HV*CCI	+
STOCKVOL	-
HHFW	+
GOVD	+
HHNW	+

Equation (1&2) present base-line models which test whether improve in home value and rosy economic outlook with a greater consumer confident provide a positive feedback on household borrowing after controlling for real wages and interest rates. As discussed earlier, any positive changes in home value combined with a brighter income expectation contribute toward a positive gain in overall wealth household. As Campbell and Cocco (2007) suggest that housing gains increase the homeowner's access to "cash on hand" if credit markets are willing to lend against higher collateral value and that the cash-on-hand effect can be an important driver of household spending, especially for constrained households with low levels of wealth. Campbell and Cocco (2007) further suggest that housing value is an asset that can be used as collateral in a loan. Mian and Sufi (2014) examine the effect of rising U.S. house prices on borrowing and spending from 2002 to 2006. They find that there is strong heterogeneity in the marginal propensity to borrow and spending. Households in low income zip codes aggressively liquefy home equity when house prices rise, and they increase spending substantially. In contrast, for the same rise in house prices, households living in high income zip codes are unresponsive, both in their borrowing and spending behaviour. Households that borrow and spend out of housing gains between 2002 and 2006 experience significantly lower income and spending growth after 2006.

Equation (3) is the derivative of both equation 1 and 2, representing housing value combined with level of consumer confident. We expect a positive responses on household debt to changes in home value under different level of consumer confident level (future economic outlook). To test whether financial stability affect household debt, we include the volatility of share index each year in equation (4), estimated from quarterly index for respective economies. An increase in volatility level suggests a sense of uncertainty, thus may have a negative impact on financial wealth of the household, so thus overall wealth of household. Any increase in the level volatility, reduced level of financial stability, thus, demote level of household debt. In the same equation we also consider additional control factors including government debt. Increases in government debt means, less money available for investment, hence less benefit for household welfare its citizen. Thus household may require more borrowing as government borrowing increases.

In equation 5 we include financial wealth proxied by household financial asset after controlling for real wages and short term interest forecast. Financial assets held by households form an important part of overall wealth and are an important source of revenue, either through the sale of those assets or refinancing, or as a source of income (such as interest and dividends).
In equation 6 we include net worth instead of financial wealth to examine the impact of household net worth on propensity to household borrowing. This may provide an alternative explanation for household indebtedness captured by household financial wealth. We expect a positive and significant impact of household net worth on household borrowing as the accumulated household net worth embolden household sector to borrow more to improve welfare and utility.

As a robustness check, we also use disposable income as a measure of household ability to service the debt in place for real wages to test whether initial estimation stands if we use different measure of income variable. The J−statistic acts as an omnibus test statistic for model misspecification. A large J−statistic indicates a mis-specified model. Models to be well specified if the p-value is more than (p>=0.10).

4. Results and Discussions

In this section we report estimates from GMM model for entire sample and sub-sample based on high and low indebtedness among OECD economies. The estimates of the five models measuring the household indebtedness based on the panel of 17 OECD economies are presented in Table 2. It is evident from the estimates of model 1 and model 2 that upon controlling for real wages and interest rates, both housing gains (HV) and future economic outlook (CCI) exert positive and significant effect on the household debt. Note that while the former is consistent with our expectations, the effect of CCI is not. The increase housing value is supposed to raise creditworthiness of the households with increased housing value used as collateral thereby qualifying them for higher loans (Campbell and Cocco, 2007). However, as per our estimate of Model 1, it seems that rosier economic future as well as such perceptions raises households' confidence of borrowing and investment, raising their indebtedness. Despite the apparent anomaly, this could be plausible for the OECD economies since households in these economies are better endowed with education and skills along with institutional and public support compared to the households in the developing world.

Table 2: Determinants of Household Indebtedness in the OECD Economies

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
HHD(-1)	.910*** (0.000)	0.877*** (0.000)		0.923*** (0.000)	0.910*** (0.000)

CCI	0.004** (0.018)			0.558** (0.002)	
HV		0.231*** (0.000)		0.232* (0.057)	.286*** (0.000)
STIRF	0.003 (0.262)	-0.002 (0.179)	-0.003* (0.062)	-0.112 0.679	-.366* (0.057)
Wages	-0.268** (0.039)	-0.365*** (0.005)	-0.352*** (0.000)	-0.001 0.402	-0.002*** (0.000)
HV*CCI			0.050*** (0.000)		
STOCKVOL				-0.058 (0.353)	
GOVDEB				0.156** (0.025)	
HHFW					0.001** (0.031)**
J-statistics	13.97 (0.378)	14.169 (0.362)	14.249 (0.356)	8.717 (0.559)	11.43 (0.492)

Source: Authors' estimates. Note: ***, ** and * represent significance at 1%, 5% and 10% levels, respectively. We report coefficient and corresponding p-value (in parenthesis) for each variable from GMM estimates. We also report J-statistic for respective model.

Estimates of Model 3 show that after controlling for wages and interest rate, the interaction variable between housing and future economic outlook has positive and highly significant effect on household indebtedness in the OECD economies (Table 2). Ideally this result indicates that better economic outlook based on higher housing value raises household debt. Model 4 incorporates the estimate of financial stability along with other variables. Financial stability seems to have adverse impact on household debt indicating that enhanced financial stability lowers household indebtedness and vice versa. It is perceivable that stable financial conditions induce improved economic conditions of households resulting in lower indebtedness of the households. Estimate of Model 5 indicates that household financial wealth (HHFW) poses significant, *albeit* small positive effect on household debt implying higher debt as HHFW improves (Table 2).

In Table 3, we report the estimates of the models replacing wages by household disposable income and with the rest of the set of the independent variables. Estimates of Model 1 and Model 2 reveal highly significant and positive effect of CCI and housing gain, respectively on the household debt, upon controlling for the effect of other variables. Estimate of Model 3 reveals that better economic outlook with higher housing value has significant and positive effect on household debt. Further, based on estimates of Model 4 and Model 5, we infer that stock market volatility and household financial wealth have insignificant effect on household debt in the OECD economies (Table 3). The effect of household disposable income is significantly negative in all cases, which clearly indicates higher household income reduces household debt, which is an obvious outcome.

While Table 2 and Table 3 provide compelling evidence of the factors affecting the household debt in the 17 OECD economies, there seems to be a need to further examine the nature of such dynamics taking account of the divergence of the household indebtedness within these economies. In view of this we split the sample of 17 OECD economies into two sub samples, viz., high debt economies and low debt economies. We define the high debt economies as the ones with household debt exceeding 130% of their household income, with nine and eight of OECD economies falling in the high debt and low debt categories, respectively.

Table 3: Household Disposable Income and Other Determinants of Household Indebtedness in the OECD Economies

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
HHD(-1)	0.946*** (0.000)	.837*** (0.000)		0.727*** (0.000)	0.847*** (0.000)
CCI	0.844*** (0.000)			1.108 (0.112)	
HV		0.209*** (0.000)		0.573* (0.080)	0.204*** (0.000)
STIRF	0.346* (0.085)	0.402*** (0.000)	0.506*** (0.000)	0.327 (0.514)	0.297*** 0.000
HHDI	-.524*** (0.000)	-.801*** (0.000)	-.903*** (0.000)	-1.249*** 0.007	-0.845*** 0.000
HV*CCI			3.860*** (0.000)		
STOCKVOL				0.287 (0.316)	
GOVDEB				0.245 (0.225)	
HHFW					-.0001 (0.750)
J-statistic	12.97 (0.450)	14.56 (0.416)	13.95 (0.376)	6.580 (0.764)	14.218 (0.287)

Source: Authors' estimates. Note: ***, ** and * represent significance at 1%, 5% and 10% levels, respectively. We report coefficient and corresponding p-value (in parenthesis) for each variable from GMM estimates. We also report J-statistic for respective model.

Estimates of Model 1 and Model 2 presented in Table 4 reveal highly significant and positive impact of consumer confidence and housing value on the household debt in the high debt economies, much in line with the results presented in Table 2 and Table 3. Highly significant positive effect of the housing value based on economic outlook as portrayed by the interaction variable is also evident for these economies, upon controlling for the effect of the other variables. However, insignificant effects of stock volatility, government debt and household financial wealth are obtained, as revealed by the estimates of Model 4 and Model 5 (Table 4).

Table 5 further reports the estimates for the low debt economies. A quick look at the estimates reveal that while housing value remains as a major variable significantly explaining the housing debt, future economic outlook is not, upon controlling for the effect of interest rate and wages. Estimates of Model 3 indicates significant positive impact of economic outlook associated with housing gain. Government debt and HHFW appear to exert significant negative impact on household debt, as indicated by the estimates of Model 4 and Model 5, respectively. Increased government debt may reduce household debt in these economies, perhaps due to lower credit available at personal or household level. In addition, contrary to the expectations, better household financial wealth seem to be reducing debt for the households in these economies. As all the economies in this sub sample is from Europe except for US, it is perceivable that in view of Europe's recent economic growth concerns, household takes recourse to appropriate prudence in accessing loans despite their creditworthiness and loan availability as their financial wealth increases.

Table 4: Determinants Household Indebtedness in the High Debt Economies

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
HHD(-1)	0.902*** (0.000)	0.905*** (0.000)	0.923*** (0.000)	0.914*** (0.0000	0.899*** (0.000)
CCI	0.614*** (0.000)			0.524 (0.124)	
HV		0.241*** (0.000)		0.256*** (0.000)	0.257 (0.000)
STIRF	0.880 0.180	0.099	0.324 0.300	0.240 0.534	0.166 0.581
Wages	-0.001** (0.000)	-0.001** (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.002 (0.017)
HV*CCI			0.164*** (0.000)		
STOCKVOL				0.144 (0.615)	
GOVDEB				0.016 (0.500)	
HHFW					0.001 (0.339)
R^2	.992	.994	.994	.994	0.995

Source: Authors' estimates. Note: ***, ** and * represent significance at 1%, 5% and 10% levels, respectively. We report coefficient and corresponding p-value (in parenthesis) for each variable from panel LS estimates with fixed effect. It is to note that due to lack of observation years under sub-category, we could not use GMM model to estimate the variable of interests. Note: There are eight economies in this sub sample. These economies are classified as the ones with household debt more than 130% of the household income

Table 5: Determinants of Household Indebtedness in the Low Debt Economies

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
HHD(-1)	0.955*** (0.000)	0.931*** (0.000)	0.946*** (0.000)	0.954*** (0.000)	0.944*** (0.000)
CCI	-0.081 (0.717)			-.120 (0.593)	
HV		0.204*** (0.000)		0.146*** (0.000)	0.203*** (0.000)
STIRF	-0.032 (0.840)	-.139 (0.387)	-0.100 (0.529)	-0.400** (0.034)	-0.244 (0.135)
Wages	0.001 (0.105)	-.001*** (0.000)	-0.001 (0.637)	-0.001** (0.006)	-0.001 (0.122)
HV*CCI			0.058*** (0.000)		
STOCKVOL				-0.135 (0.320)	
GOVDEB				-0.078*** (0.001)	

HHFW					-4.743*** (0.000)
R^2	0.983	0.982	0.975	0.985	0.984

Source: Authors' estimates. Note: ***, ** and * represent significance at 1%, 5% and 10% levels, respectively. We report coefficient and corresponding p-value (in parenthesis) for each variable from panel LS estimates with fixed effect. It is to note that due to lack of observation years under sub-category, we could not GMM model to estimate the variable of interests.
Note: There are nine economies in this sub sample. These economies are classified as the ones with household debt less than 130% of the household income

Table 6 report the impact of household net worth on household indebtedness. It is pre-conceived view that family with higher level of net worth more solvent thus, financial institutions will be less hesitant sanctioning more loans. As an effort to understand such phenomena, we separately model household indebtedness and examine the impact of household net worth after controlling for all the relevant factors. GMM estimates tend to indicate the view that household with higher net worth have a greater level propensity to have more debt as net worth improves.

Table 6: Household Net worth and Household Indebtedness

Variables	Coefficient	Std.Error	p-value
HHD(-1)	0.839	0.061	0.000
CCI	-0.222	0.244	0.363
HV	0.291	0.090	0.001
STIRF	-0.002	0.002	0.249
Wages	-0.513	0.165	0.002
HHNW	0.140	0.070	0.046
J-statistics	11.562		0.315

Source: Authors' estimates. Note: ***, ** and * represent significance at 1%, 5% and 10% levels, respectively. We report coefficient and corresponding p-value (in parenthesis) for each variable from GMM estimates. We also report J-statistic for respective model.

5. Conclusions

In this study we make an attempt to examine the household indebtedness in a selected group of OECD economies. In view of the recent concerns of economic and financial stability in many of the OECD economies and the changing phase of sectoral growth, specialisation, globalisation and national income, there seems to be inevitable implications of assessing microeconomic agents such as the households. In this study, we specifically examine the factors that affect household debt in the OECD economies using dynamic GMM estimations for a sample of 17 OECD economies. Our findings reveal that economic outlook measured in terms of consumer confidence index, housing value or wealth gain along with interaction of these two variables exert significant positive effect on the household debt in the OECD economies. These results seems to pose overarching evidence as these remain invariant of the alternative measures of income and the magnitude of household debt in relation to income, estimated in separate sub samples of high debt and low debt economies. Additionally, we find very limited evidence of significant positive effect of government debt and household financial wealth on household debt and insignificant impact of financial stability on the debt. As an expected phenomenon, we also find that both increased household income and wages seem to significantly reduce household debt. Our findings imply that general economic growth as well as property market boom or to be precise, house price growth work as overarching factors contributing to household debt in the OECD economies. Hence a more regulated and well monitored property market could potentially stabilise the household wealth as well as financial positions in the OECD economies thereby benefitting the broader spectrum of these economies ensuring more stable and sustainable economic growth. For both high and low debt economies within OECD, one common striking phenomena is that the housing value consistently explains propensity to have more debt for household.

References

Agarwal, S., and Qian, W., (2014) "Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore", American Economic Review, 104(12), 4205-4230

Aguiar, M., and Gopinath, G., (2007) "Emerging Market Business Cycles: The Cycle Is the Trend," Journal of Political Economy, 115, 69–102.

Alessie, R and Lusardi, A, (1997) "Saving and income smoothing: Evidence from panel data"**,** European Economic Review, Vol.41(7), pp: 1251-1279

Anderson, R, Duru, A and Reeb, D (2012) "Investment policy in family controlled firms", Journal of Banking and Finance, Vol.36(6),pp: 1744-1758

Arellano, M and Bover, O (1995), "Another look at the instrumental variable estimation of error-components models", Journal of Econometrics, Vol.(68(1), pp: 29-51

Arellano, M and Bond, S (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", The Review of Economic Studies, Vol.58(2), pp: 277–297.

Attanasio, OP and Weber, G (1994), "The UK Consumption Boom of the Late 1980s: Aggregate Implications of Microeconomic Evidence", The Economic Journal, Vol. 104 (427), pp. 1269-1302.

Barba, a and Pivetti, M (2009), "Rising household debt: Its causes and macroeconomic implications—a long-period analysis", Cambridge Journal of Economics, Volume 33, Issue 1, Pages 113–137.

Blundell, R and Bond, S, (1998), **"**Initial conditions and moment restrictions in dynamic panel data models"**,** Journal of Econometrics**,** Volume 87, Issue 1, November 1998, Pages 115-143

Campbell, J. and Cocco, J., (2007). "How do house prices affect consumption? Evidence from micro data", Journal of Monetary Economics, 54, 591–621.

Carroll, Christopher and Miles Kimball, (1996). "On the Concavity of the Consumption Function," Econometrica 64: 981-992.

Carroll, Christopher, Jiri Slacalek, and Martin Sommer. (2012). "Dissecting Saving Dynamics: Measuring Credit, Wealth and Precautionary Effects." Johns Hopkins University (January). www.econ2.jhu.edu/people/ccarroll/papers/cssUS Saving.pdf.

Campbell, John Y. (1987) "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis." Econometrica Vol.55:1249-73.

Carroll C.D., (1992). "The buffer stock theory of saving: Some macroeconomic evidence", Brookings Papers on Economic Activity 2, 61 135.

Carroll, C, and Samwick, A (1997), **"**The nature of precautionary wealth", Journal of Monetary Economics, Vol.40 (1), pp:41-71

Deaton, Angus, (1991), "Saving and Liquidity Constraints," Econometrica 59: 1221-1248.

Dynan, Karen, Atif Mian and Karen Pence (2012). Is a Household Debt Overhang Holding Back Consumption? Brookings Papers on Economic Activity, Brookings Papers on Economic Activity (SPRING), pp. 299-362

Friedman, M. (1957), "A Theory of the Consumption Function". Princeton University Press.

Girouard, N. et al. (2006), "Recent House Price Developments: The Role of Fundamentals", OECD Economics Department Working Papers, No. 475, OECD Publishing, Paris. http://dx.doi.org/10.1787/864035447847

Jappelli, T and Pistaferri, L (2010), "The Consumption Response to Income Changes", Annual Review of Economics, Vol 2: pp.479-506

Cambell and Cocco, O (1995) "Corporate-debt overhang and macroeconomic expectations", The American Economic Review, Vol. 85(1995), pp. 1106-1117

Martin Lettau, Sydney Ludvigson (2001), **"**Consumption, Aggregate Wealth, and Expected Stock Returns", Journal of Finance, Vol.56(3),pp. 815–849

Mian, A and Sufi, A (2017), "Household Debt and Business Cycles Worldwide", The Quarterly Journal of Economics, Volume 132 (4), p: 1755–1817,

Mian, A and Sufi, A (2015) "House price gains and u.s. household spending from 2002 to 2006", NBER working paper series, Working Paper 20152 http://www.nber.org/papers/w20152

Mian, A and A Sufi (2014), "House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent it from Happening Again", Chicago, IL: University of Chicago Press.

Mian, A, Sufi, A (2011) "House Prices, Home Equity Based Borrowing, and the U.S. Household Leverage Crisis,'' American Economic Review, 101, pp:2132–2156

Mian, A and Sufi, A (2009) "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis", The Quarterly Journal of Economics, Volume 124 (4) pp: 1449–1496,

Mian, Atif, and Amir Sufi, (2009), "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," Quarterly Journal of Economics, 124, 1449–1496.

Mian, Atif, and Amir Sufi, (2010). "Household Leverage and the Recession of 2007–09". IMF Economic Review, Vol. 58, No. 1, pp. 74-117.

Mian, Atif, and Amir Sufi, (2011). ''House Prices, Home Equity Based Borrowing, and the U.S. Household Leverage Crisis,'' American Economic Review, 101 (2011), 2132–2156.

Mian, Atif, Kamalesh Rao, and Amir Sufi, (2013) "Household Balance Sheets, Consumption, and the Economic Slump," Quarterly Journal of Economics, 128, 1687–1726.

Mian Atif and Amir Sufi (2014). "House Price Gains and U.S. Household Spending from 2002 to 2006". NBER Working Paper No. 20152.

Muellbauer, J and Murphy,A (1997) **"**Booms and Busts in the UK Housing Market", The Economic Journal**,** Vol.107 (445), pp: 1701–1727

Nalewaik, J (2006), **"**Current consumption and future income growth: Synthetic panel evidence"**,** Journal of Monetary Economics, Vol. 53 (8),pp: 2239-2266

Sinai, Todd and Nicholas Souleles, (2005), "Owner-Occupied Housing as a Hedge Against Rent Risk," Quarterly Journal of Economics May: 763-789.

Zeldes,SP (1989), "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence", The Quarterly Journal of Economics, Volume 104, Issue 2, 1 May 1989, Pages 275–298.

Zeldes,SP (2009), "Consumption and Liquidity Constraints: An Empirical Investigation", Journal of Political Economy, Vol.97 (2)