THREE ESSAYS ON TIME SERIES ANALYSIS OF CHINESE FINANCIAL MARKETS

A Dissertation

by

LU FANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David A. Bessler
Committee Members,	Ximing Wu
	Henry L. Bryant
	David J. Leatham
Head of Department,	Parr Rosson III

May 2017

Major Subject: Agricultural Economics

ABSTRACT

This dissertation studies three important issues in Chinese financial markets. The interdependence structure and information transmission among Chinese cross-listed stocks in Shanghai, Hong Kong and New York is examined. Results indicate that the home bias hypothesis, which suggests the dominant role of home market in pricing information transmission, is strongly supported in contemporaneous time, modestly supported at the short horizon and not supported at the long horizon. The Shanghai market as the home market is highly exogenous at all horizons. Moreover, the Hong Kong market leads the New York market in contemporaneous time.

Whether interest rates help to forecast stock returns in China is studied using the prequential approach. With respect to calibration (reliability), it is found that including interest rates in the model improves the model's ability to issue realistic probability forecasts of stock returns – a model of stock returns that does not include interest rates as an explanatory variable is not as well calibrated as a model that does include interest rates in the stock returns equation. With regard to sorting (resolution), results suggest that the model that includes interest rates performs better in distinguishing stock returns that actually occur and stock returns that do not occur when compared to a model that does not include interest rates in the stock returns equation. Overall, the interest rates help in forecasting stock returns in China in terms of both calibration and sorting.

Two factor analysis methods are investigated through forecasting Chinese interest rate based on a factor-augmented vector autoregression (FAVAR). Factors are

estimated from 288 Chinese security series to reflect the common forces that drive the movements and dynamics in the Chinese equity market. As a result, the factor estimation method by Lam and Yao outperforms the traditional principal components analysis (PCA) in terms of forecasting accuracy, especially at the short horizons.

DEDICATION

To my father, Luming Fang,

to my mother, Xuehong Jiang,

to my husband, Ke Wang,

and to my daughter

ACKNOWLEDGEMENTS

I would like to express my first and deepest gratitude to my advisor, Dr. David Bessler. Without his support and guidance, I have no way of achieving what I have accomplished. Dr. Bessler has extensive and deep knowledge of research. He mentors and encourages my research in many ways. Dr. Bessler has very rigorous and serious research attitude. His passion and persistence to research inspire me through my Ph.D. journey and will continue encouraging me to be a diligent and active scholar in my professional career. The great personality of Dr. Bessler such as humbleness, patience and kindness has also motivated me and will benefit me for my whole life.

My thanks also go to the members of my committee, Dr. Ximing Wu, Dr. Henry Bryant and Dr. David Leatham and the former member of my committee, Dr. Keli Xu. They have provided me with valuable research comments in improving my presentation and dissertation. Their in-depth advice is also helpful to my future research.

My sincere thanks is also extended to my classmates and colleagues in Texas A&M University, whom have shared this precious and memorable experience with me. Thank them for their friendship, encouragement and support.

Last but not the least, I want to express my greatest gratitude to my families. My father has been working as an academic scholar for the past thirty years. He is the first role model in my life and has also encouraged me to pursue an academic career. My mother offers me the most love and caring in my life. The support from my parents and parents-in-law have made me determined in fulfilling my dream. I also would like to thank my husband and my daughter. They bring me the greatest joy and motivate me to be a more grateful, responsible and reliable person.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor David A. Bessler and Ximing Wu, Henry L. Bryant of the Department of Agricultural Economics and Professor David J. Leatham of the Department of Agricultural Business.

All work for the dissertation was completed by the student, under the advisement of David A. Bessler of the Department of Agricultural Economics.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Rapid economic development and expanding financial openness in China bring forth many issues concerning financial aspects of this emerging market. Three of these issues are explored in this dissertation.

In equity markets, there is an increasing trend of Chinese firms choosing to cross-list their stock in offshore markets, aiming to take advantage of overseas financing channels, lower capital cost and enhancement of corporate governance, as well as other considerations. Chapter II examines the information transmission among Chinese cross-listed stocks in Shanghai, Hong Kong and New York. Much literature studying cross-listed stocks indicate that the home market of the stocks' underlying company will have dominant impacts on information transmission (Frijns et al., 2010, 2015; Grammig et al., 2005; Hauser et al., 1998; Kim et al., 2000;). This is consistent with the home bias hypothesis, suggesting that cross-listed stocks are expected to be heavily traded in the home market, which is geographically proximate to the headquarters of the underlying firms. There are also some literature suggests that market quality may also influence a market's contribution in price discovery of cross-listed stocks, besides the geographical proximity to the headquarters (Chen et al., 2010; Frijns et al., 2015). Therefore, for Chinese cross-listed stocks, whether the home market influence dominates and yields results supporting the home bias hypothesis or whether the inferior market quality of the

Shanghai market impeds its leading role in information transmission, leading to results against the home bias hypothesis is worth exploring.

Ten Chinese cross-listed stocks are examined using vector error correction model (VECM). The recent advance in inductive causal graphs (Pearl, 1995, 2000; Spirtes et al., 2000) are used to explore the contemporaneous causal pattern of price transmission for cross-listed stocks. Particularly, the recently developed linear non-Gaussian acyclic models (LiNGAM) algorithm (Shimizu, Hoyer et al., 2006) is applied. Compared to the commonly followed PC algorithm, LiNGAM algorithm is able to exploit the non-Gaussian nature in data, to infer stronger contemporaneous causal relations. The contemporaneous structure generated by LiNGAM is used in forecast error variance decompositions, which is adopted to further explore the interdependence structure among the cross-listed stocks. The robustness of the estimated model is checked through forecasting exercises. The generated one-step-ahead out-of-sample forecasts are evaluated using both d-separation and root mean squared errors (RMSEs).

Besides the equity market, the interest rate market in China has also experienced fast development. A series of implementations have been carried out to improve the liberalization of interest rates in China. Many studies have investigated the contribution of interest rates in explaining the predictable movements of stock returns and indicate that short-term interest rates help in forecasting stock returns (Campbell, 1987, 1991; Fama and French, 1988b). One characteristic of this existing literature is that most studies conduct examinations based on in-sample tests of model fit. A further characteristic of this existing literature is that it focuses on developed economies such as the U.S. market, rather than developing economies such as China. Motivated by these two characteristics, Chapter III examines whether interest rates improves the forecasts of stock returns in China by emphasizing the role of out-of-sample forecasts instead of in-sample tests of model fit.

The prequential analysis is used to study the issue from both calibration (reliability) and sorting (resolution) perspectives, following Bessler and Ruffley (2004). A VAR model on stock returns and interest rates and a univariate autoregressive (AR) model on each series are examined. If the VAR model outperforms the AR model for stock returns, it is concluded that interest rates help to predict stock returns. Both a bootstrap-like simulation method and a nonparametric kernel-based simulation method are used to generate the probability forecasts, which enables the robustness check of the results. Several data-driven methods combined with the arbitrary selection method are considered for settings of subintervals and events when evaluating probability forecasts. In terms of calibration, calibration plots and calibration tests are considered to examine whether including interest rates in the model helps to issue realistic forecasts of stock returns (be well-calibrated). In terms of sorting, the Brier score and its Yates partition are applied to assess whether incorporating interest rates in the model improves the model's ability in distinguishing stock returns that actually obtain (occur) from stock returns that do not obtain.

Different from Chapter III, which uses the compiled index to represent the main information of Chinese equity market, Chapter IV studies two ways of factor estimation that can extract a few factors for representing the general movements and dynamics in

Chinese equity market. Considering the increasing availability of high-dimensional data sets consisting of thousands of economic and financial time series, the question of how to summarize the main variation contained in the high-dimensional data through just a few factors is of great interest and importance. Factor estimation is capable of achieving dimension reduction. In Chapter IV, we aim to compare the commonly followed factor estimation method of principal components analysis (PCA) with the new method developed by Lam and Yao (2012) through forecast performance evaluation.

Specifically, a few factors are estimated from 288 return time series of Chinese equities. The first group of factors are estimated through PCA based on the number of factors selected by the method of Bai and Ng (2002). The second group of factors are estimated through the method by Lam and Yao (2012). These two groups of factors are fitted with Chinese interest rate in a factor-augmented vector autoregression (FAVAR) model respectively. Two sets of forecasts of interest rate are obtained accordingly and are evaluated through various statistical measures and tests of forecast accuracy.

Chapter V summarizes the main results and findings of the above three chapters.

CHAPTER II

INTERDEPENDENCE STRUCTURE OF CHINESE STOCKS CROSS-LISTED IN SHANGHAI, HONG KONG AND NEW YORK

2.1. Introduction

In recent years, China has experienced dramatic economic growth and openness. Equity markets in China have experienced rapid expansion since the 1990s. In fact, the two stock exchanges in mainland China, Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) were established in 1990 and 1991, respectively. As indicated by China Securities Regulatory Commission (CSRC), up to January 2016, the total market capitalization of SSE and SZSE has been 40393.6 billion Chinese yuan with 2828 listing stocks.

With the globalization of world financial markets, an increasing number of Chinese firms choose to cross-list in multiple stock markets. There are various advantages of cross-listing stocks in offshore markets (Karolyi, 2006; Roosenboom and Van Dijk, 2009; Su and Chong, 2007; Xu and Fung, 2002). First, cross-listing can take advantage of oversea financing channels and lower the capital raising cost by overcoming the investment barriers due to market segmentation. Second, companies can benefit from improved liquidity by cross-listing their stocks in more liquid equity markets. Third, cross-listing can strengthen corporate governance and enhance information quality because of higher level of information disclosure in offshore markets.

Other benefits of cross-listing include increasing global awareness, facilitating foreign mergers and acquisitions and so on.

Cross-listed stocks refer to stocks of the same underlying company that are listed on multiple stock markets. Much literature indicates that although both home market factors and offshore market factors can influence the returns of cross-listed stocks, the home market of the stocks' underlying company acts as the dominant market and plays a leading role in information transmission, since the underlying company is headquartered in the home market and most information concerning the company such as business activities and revenues is generated in the home market (Frijns et al., 2010, 2015; Grammig et al., 2005; Hauser et al., 1998; Kim et al., 2000; Lau and Diltz, 1994; Pascual et al., 2006). This is consistent with the home bias hypothesis. The hypothesis implies that home market will make the dominant contribution to price discovery and information transmission of cross-listed stocks, for cross-listed stocks are expected to be heavily traded in the home market, which is geographically proximate to the headquarters of the underlying companies, rather than the offshore market, which is geographically far away from the headquarters of the underlying companies (Chen et al., 2010; Xu and Fung, 2002).[1]

However, as suggested by Chen et al. (2010) and Frijns et al. (2015), besides the geographical proximity to the headquarters, aspects related to market quality may also

[1] Home bias hypothesis highlights investors' preference of investing close to home (French and Poterba, 1991; Tesar and Werner, 1995). Focusing on domestic equity portfolios, Coval and Moskowitz (1999) indicate that within a country, asymmetric information may motivate investors' preference of investing in companies with geographically proximate headquarters and lead them to strongly place weight on equities with locally headquartered firms rather than those with overseas headquartered firms.

affect a market's role in price discovery of cross-listed stocks, implying it is possible for an offshore market with superior market quality to have overwhelming influence, instead of the home market. In fact, a significant part of literature that suggests the leading role of home market focuses on studying cross-listed stocks originating from developed economies with premium market quality. Will an emerging home market with lower level of market quality such as China tell the same story? To answer the question, this chapter examines the interdependence structure of Chinese stocks cross-listed in SSE (the home market), Hong Kong Stock Exchange (HKEx) and New York Stock Exchange (NYSE). HKEx and NYSE are the two most popular offshore markets for Chinese firms to cross-list in besides the mainland China.

The stock market in mainland China has a series of possible insufficiencies and restrictions when compared to more developed stock markets like HKEx and NYSE. For instance, HKEx and NYSE have higher level of information disclosure and more advanced market mechanisms than SSE. The A-share market at SSE is dominated by individual investors lacking of investment expertise, while the other two markets contain a larger proportion of institutional investors. The A-shares issued at SSE are predominately traded by domestic investor. Nevertheless, HKEx and NYSE are open to both domestic and international investors (Tan et al., 2008; Wang and Jiang, 2004). Thus, the market quality of SSE is not as exceptional as that of HKEx and NYSE (Chen et al., 2010). Frijns et al. (2015) suggest that better market quality may boost the incorporation of new information, improve the efficiency of information transmission and promote a market's contribution in price discovery of cross-listed stocks. Therefore, whether the inferior market quality of SSE will impede its leading role in information transmission and lead to results against the home bias hypothesis is worth exploring. Such study should allow comment on the efficiency of asset allocation (Xu and Fung, 2002).

This paper contributes to the literature in several ways. First, most existing studies on Chinese cross-listed stocks focus on just two markets such as Shanghai and Hong Kong (Cai et al., 2011; Li et al., 2006; Wang and Jiang, 2004) or Hong Kong and New York (Su and Chong, 2007; Xu and Fung, 2002). We aim to fill the gap by expanding to a multi-market estimation and including all three markets.

Second, this is the first paper to explore the contemporaneous causal pattern of price transmission for cross-listed stocks. Price dynamics across markets are often addressed through significance tests of coefficients in the lead-lag models of stock returns (Lau and Diltz, 1994; Wang and Jiang, 2004; Xu and Fung, 2002). In this paper, the recent advance in inductive causal graphs (Pearl, 1995, 2000; Spirtes et al., 2000) are used to examine the contemporaneous causal structure across markets for cross-listed stocks.

Third, two machine learning algorithms are considered to generate the graphical representation of contemporaneous causality. In considerable literature, PC algorithm has been applied (Awokuse, 2006; Awokuse et al., 2009; Bessler and Yang, 2003; Bessler et al., 2011). Oftentimes, however, PC algorithm is not able to distinguish observationally equivalent structures due to its assumption of Gaussian distributed data, leading to undirected causal relations. In fact, non-Gaussian data is quite common in most empirical studies, especially those involve asset prices (Moneta et al., 2013). In this paper, the recently developed linear non-Gaussian acyclic models (LiNGAM) algorithm (Shimizu, Hoyer et al., 2006) is applied, which is able to exploit the non-Gaussion nature in data to infer stronger contemporaneous causal relations.

Fourth, a significant part of the cross-listed stock literature aims to explore each market's contribution to the price discovery of cross-listed stocks. Two extensively used models are the information shares model of Hasbrouck (1995) and the permanent-transitory model of Gonzalo and Granger (1995) (Frijns et al., 2010; Pascual et al., 2006; Su and Chong, 2007). Hasbrouck's (1995) method considers correlated innovations and incorporates contemporaneous correlations in his metric. However, this method has several shortcomings in dealing with the contemporaneous correlations among variables. In this paper, we apply the forecast error variance decompositions based on vector error correction model (VECM) to explore each market's role in price discovery. Several techniques are combined to address the shortcomings of Hasbrouck's (1995) method. First, Hasbrouck (1995) adopts a commonly followed Choleski factorization, which is a lower triangular causal ordering, to contemporaneous correlations. This may cause problems such as unrealistic causal assumptions and misleading innovation accounting results (Bernanke, 1986; Sims, 1986; Swanson and Granger, 1997; Yang and Bessler, 2004). Hence, we use a structural factorization, referred to as the "Bernanke ordering", to allow for more general causal ordering (Bernanke, 1986; Sims, 1986). Second, the Choleski factorization depends on subjective analysis or prior economic theories to issue causal ordering. To minimize the subjective assumptions or prior knowledge, this paper applies the inductive causal graphs as a data-determined approach to model the contemporaneous causal structure. This is consistent with the main idea of vector autoregression (VAR)-type models to rely on data and diminish *a priori* restrictions (Motena et al., 2013; Sims, 1980a). Third, instead of offering a unique measure of each market's contribution in price discovery, Hasbrouck (1995) gives a range with upper and lower bounds, since the information share for a specific market will vary by its order in the Choleski factorization (Grammig and Peter, 2013). Lanne and Lütkepohl (2012) argue that a unique measure can be obtained through non-Gaussian innovations. Because of the LiNGAM algorithm, we are able to discover a unique contemporaneous causal ordering and obtain a unique set of forecast error variance decompositions by utilizing the non-Gaussian nature of the data.

The fifth contribution is that the long-run relationship of cross-listed stocks is explored using the VECM and cointegration analysis. As suggested by Jorion and Schwartz (1986), the same long-run fundamental value will drive cross-listed stocks to generate the same risk-adjusted expected returns under perfect market integration. For a single stock traded in multiple markets, Hasbrouck (1995) also indicates that the stocks have a common implicit efficient price. Stock prices in different markets should be mutually linked and adjusted because of arbitrage. Therefore, prices of the cross-listed stocks ought to be cointegrated. The issue of cointegration has not been widely addressed in the existing literature of cross-listed stocks.

The remainder of the paper is organized as follows. Section 2 is the literature review. Section 3 discusses the methodology of VECM and inductive causal graphs.

Section 4 describes the data. The empirical results are presented in Section 5. Section 6 concludes the chapter.

2.2. Literature Review

One of the earliest papers on cross-listed stocks is that of Garbade and Silber (1979). They examine information transmission among multiple markets using dually-traded assets and characterize the relations of these markets as dominant-satellite relations. They suggest that some markets have dominant influences on determining the returns of dually-listed stocks, while the other markets perform as satellite markets and exhibit very limited impacts on price adjustment.

Focusing on cross listed stocks originating from developed markets, many studies highlight the importance of the home market in information transmission. Grammig et al. (2005) study three German stocks cross-listed in Frankfurt and New York. Focusing on only the overlapping trading hours, it is found that the home market accounts for most of the price discovery of cross-listed stocks. Kim et al. (2000) analyze the foreign securities traded in the US in the form of American Depository Receipts (ADRs). Although the prices of ADRs are influenced by both home market and US market factors, the home market factor explains more variations in the corresponding ADRs. Lieberman et al. (1999) examine Israeli stocks dually-listed in Israel and the U.S. It is suggested that the home market plays as a dominant market and has substantial influence on the pricing information transmission.

Some studies examine the Chinese cross-listed stocks. Wang and Jiang (2004) investigate the interrelations of Chinese stocks dually-listed as A-shares in mainland China and as H-shares in Hong Kong. As the home market of the selected Chinese stocks, the A-share market is highly exogenous and exclusively affected by home market factors. For H-share market, both home market factors and Hong Kong factors help explain the returns of associated stocks. Chelley-Steeley and Steeley (2012) explore Chinese cross-listed stocks in multiple exchanges. They also shed some light on the importance of A-share market in information transmission. In addition, they argue that the A-share market is relatively isolated without considerable exposure to other international markets. Chen et al. (2010) find that the home market plays a dominant role in both price discovery and volatility spillover for Chinese stocks cross-listed in mainland China and New York.

2.3. Empirical Methodology

2.3.1. Vector Error Correction Model (VECM)

Existing literature has well discussed cointegration analysis and the VECM (Johansen, 1991; Johansen and Juselius, 1990, 1994). Here we offer a short description of a VECM on stock prices that are traded in three markets, Shanghai, Hong Kong and New York. A VECM with *k*-1 lags is presented as:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu + e_t, \qquad (t = 1, \dots, T) \tag{1}$$

where X_t is a (3×1) vector of price series at time *t*, ΔX_t is the first difference between X_t and X_{t-1}, Π and Γ_i are (3×3) coefficient matrices, μ is a (3×1) vector of constants, and e_t is an iid (3×1) vector of innovations. Eq. (1) is derived from a levels VAR with *k* lags.

The coefficient matrix Π can be factorized as $\Pi = \alpha\beta'$, where β $(3\times r)$ is the cointegrating vectors capturing the long-run dynamics and α $(3\times r)$ is the short-run response to the long-run relations. The rank of Π is *r*, the number of cointegrating vectors ($r \leq 3$). Testing hypotheses on β can be used to identify the long-run structure of market interdependence, and the short-run structure can be identified by hypothesis tests on α and Γ_i (Johansen, 1995; Johansen and Juselius, 1994). The contemporaneous structure can be identified through inductive causal graphs applied to observed innovations, $\hat{e}_t$ (Pearl, 1995, 2000; Spirtes et al., 2000).

Given the number of cointegrating vectors, further restriction tests are conducted on both β and α. For β, the hypothesis test of exclusion is formulated as follows:

$$\mathcal{H}_1\colon\ \beta = H\varphi \tag{2}$$

The null hypothesis is that a particular series is excluded from the long-run equilibrium or the *i*th row of β is all zero. The hypothesis test of weak exogeneity is performed on α to determine whether a particular series adjusts to the deviation from long-run equilibrium. The test hypothesis is as follows:

$$\mathcal{H}_2\colon \alpha = H\psi \tag{3}$$

The null hypothesis is that a particular series does not respond to the perturbation in the cointegrating space spanned by β or the *i*th row of α has all zero elements.

Further analysis of short-run structure involves the examination of Γ_i. However, the individual coefficients of the VECM are difficult to interpret, particularly the short-run coefficients (Sims et al., 1990). Innovation accounting procedures are suggested as helpful to summarize the short-run dynamic structure (Lütkepohl and Reimers, 1992; Sims, 1980a; Swanson and Granger, 1997). Specifically, the VECM in Eq. (1) with *k*-1 lags is converted to its corresponding levels VAR with *k* lags:

$$X_t = (1 + \Gamma_1 + \Pi)X_{t-1} - \sum_{i=1}^{k-2}(\Gamma_i - \Gamma_{i+1})X_{t-i-1} - \Gamma_{k-1}X_{t-k} + \mu + e_t \; (t = 1, \dots, T) \quad (4)$$

Based on the equivalent levels VAR representation (Eq. (4)) with restrictions on α and β (summarized in Table 2.3 and Table 2.4), the forecast error variance decompositions is applied to study the share of each market in the variation of other markets.

Here one critical problem involved is the way of treating contemporaneous correlations of innovations. The observed innovations e_t in Eq. (4) is generally not orthogonal for there may be contemporaneous innovation correlations among variables. Fail to treat the non-orthogonal innovations may cause problems to accurately characterize the data dynamics (Sims, 1980). Thus, the non-orthogonal innovations e_t are transformed into orthogonal innovations $v_t(3\times1)$, which are the driving sources of variation in the data:

$$Ae_t = v_t \quad (5)$$

where A is a (3×3) matrix of structural parameters. To set up the contemporaneous correlations of the variables, a common way is to use the Choleski factorization, which is a lower triangular causal ordering. However, Choleski factorization may cause problems such as unrealistic causal assumptions and misleading innovation accounting results (Bernanke, 1986; Sims, 1986; Swanson and Granger, 1997; Yang and Bessler, 2004). In this chapter, a structural factorization is applied to permit a more general causal ordering (Bernanke, 1986; Sims, 1986). Accordingly, matrix A is a matrix with diagonal of ones rather than a matrix with lower triangular structure:

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \tag{6}$$

By pre-multiplying Eq. (4) by matrix A, Eq. (7) is obtained for the estimation of innovation accounting procedures:

$$AX_t = A(1 + \Gamma_1 + \Pi)X_{t-1} - A\sum_{i=1}^{k-2}(\Gamma_i - \Gamma_{i+1})X_{t-i-1} - \mathrm{A}\Gamma_{k-1}X_{t-k}$$
$$+A\mu + Ae_t \ (t = 1, \dots, T) \tag{7}$$

Instead of relying on subjective assumptions and prior economic theories, the contemporaneous structure is modeled by inductive causal graphs as a data-determined approach. LiNGAM algorithm is used to identify contemporaneous causal ordering and impose zero restrictions on matrix A. According to Doan (1992, pp. 8-10), if there is no combination of a_{ij} and a_{ji} $(i, j = 1, \dots, 3)$ that are both nonzero, the matrix A can be identified. Since the data in this chapter are non-Gaussian distributed, the contemporaneous structure is also uniquely identified under LiNGAM algorithm. As a result, a unique measure of each market's contribution in price discovery can be summarized using forecast error variance decompositions.

2.3.2. Inductive Causal Graphs

A directed graph is a graph summarizing the causal flows among a set of variables (Pearl, 2000). A more vivid description is a pipeline system transferring water, representing dependence and independence in information flows (Spirtes et al., 2000).

The directed graph consists of vertices (variables), marks (symbols attached to the end of undirected edges), and ordered pairs (directed edges or arrows). Arrows exhibit the direction of information flow in directed graphs. If there are no edges connecting variable X and variable Y, the two variables are conditionally uncorrelated. An undirected edge as $X - Y$, indicates variable X and variable Y are conditionally correlated. However, whether X causes Y or vice versa could not be determined. If there is a directed edge connecting variable X and variable Y as $X \rightarrow Y$, not only correlation but also causation could be inferred (variable X causes variable Y). $X \leftrightarrow Y$ describes a bi-directed edge, indicating that there is an omitted variable which causes both X and Y. In this chapter, only directed acyclic graphs (DAGs) are considered to describe the contemporaneous causal relations among variables

2.3.2.1. PC Algorithm

Several machine learning algorithms of inductive causation have been developed. As the most widely used algorithm, PC algorithm (Spirtes et al., 2000) incorporates the notion of d-separation to direct causal flows. Starting with a complete undirected graph, PC algorithm searches causal flows based on conditional independence. The edges of the graph are then deleted sequentially if the correlation or conditional correlation of variable pairs is zero. The significance of the estimated correlation or conditional correlation is tested using Fisher's z statistic. If there are *M* variables, the procedure continues until the *M*-2 order conditional correlation is checked. For the remaining edges, the notion of d-separation (Pearl, 1995), also known as sepset, is used to determine causal directions. If the edges between two variables are removed conditioning on certain variables, these conditioning variables are called the sepset of the variables whose edge has been removed. Specifically, for the triples $\mathrm{A} - \mathrm{B} - \mathrm{C}$, it is directed as $\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$ is B is not in the sepset of A and C. If there is an directed edge as $\mathrm{A} \rightarrow \mathrm{B}$, then the causal flow between B and C is $\mathrm{B} \rightarrow \mathrm{C}$. If there is an directed edge as $\mathrm{B} \leftarrow \mathrm{C}$, then the causal relation between A and B is directed as $\mathrm{A} \leftarrow \mathrm{B}$.

One type of causal relation is the causal chain ($\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C}$). Another causal relation is the causal fork ($\mathrm{A} \leftarrow \mathrm{B} \rightarrow \mathrm{C}$). In both of these cases, A and C are unconditionally correlated and become uncorrelated or d-seperated after conditioning on B. For the inverted fork ($\mathrm{A} \rightarrow \mathrm{B} \leftarrow \mathrm{C}$), conditioning on B opens the causal flow between A and C.

PC algorithm is based on the assumption that the variables are Gaussian distributed. Under this assumption, no higher-order moment structures are needed since second-order moments can provide required information of probability distribution and conditional correlation (Shimizu, Hoyer et al., 2006; Shimizu et al., 2012). However, it is possible for more than one graph to lead to the same joint probability distribution (e.g. $A \leftarrow B \rightarrow C$ and $A \rightarrow B \rightarrow C$). These observationally equivalent graphs cannot be distinguished by PC algorithm (Kwon and Bessler, 2011; Moneta et al., 2013; Pearl, 2000). In the graph, the indistinguishable structure is illustrated as undirected edge. PC algorithm also assumes "causal sufficiency", that there is no omitted variable that causes two (or more) variables included in the analysis. Clearly, causal sufficiency can be violated. Results given below need to be conditioned on such violations.[2]

2.3.2.2. Linear Non-Gaussian Acyclic Models (LiNGAM)

Different from PC algorithm, the recently developed LiNGAM algorithm (Shimizu, Hoyer et al., 2006) takes advantage of the non-Gaussian nature of data, upon which higher-order statistics can be used to yield stronger causal identifications (Dodge and Rousson, 2001; Shimizu and Kano, 2008).

As it is our purpose to apply LiNGAM and not to develop further algorithmic extensions, accordingly we offer a brief summary. Papers by Shimizu, Hoyer et al. (2006) and Moneta et al. (2013) can be consulted for details. LiNGAM algorithm is conducted by applying independent component analysis (ICA) (Hyvärinen et al., 2004; Jutten and Herault, 1991; Moneta et al., 2013) to estimate a matrix of mixing coefficients. In the ICA model, the independent signals $s_j (j = 1, \dots, m)$ are summed with coefficients a_{ij} to generate observed variables $x_i (i = 1, \dots, m)$:

$$x = As = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix} \tag{8}$$

[2] PC algorithm also assumes the "causal Markov condition": the joint probability of the data can be factored into the product of conditional variables where the condition is taken with respect to each variable's "causal parents". For exogenous variables, the marginal (unconditional) distribution applies. Finally, PC assumes "faithfulness": if we observe a zero correlation between two variables, this zero arises because there is no causation between them and not because of cancellation of deeper parameters (see Spirtes et al. 2000 on these assumptions).

where x is the collection of observed variables, s is the collection of mutually independent components, and A is a matrix of mixing coefficients $a_{ij}(i, j = 1, \dots, m)$.

In LiNGAM, each observed variable x_i $(i = 1, \dots, m)$ can be described as a linear function of the earlier variables and the disturbance e_i:

$$x_i = \sum_{k(j)<k(i)} b_{ij} x_j + e_i \tag{9}$$

where e_i is a mutually independent and non-Gaussian disturbance, b_{ij} represents the connection strength from x_j to x_i, and $k(i)$ is the causal ordering of x_i that can be illustrated by directed acyclic graphs. The matrix form of LiNGAM is generated as:

$$x = Bx + e \tag{10}$$

where B stands for the matrix of connection strength. Solving for x will give Eq. (9):

$$x = (I - B)^{-1} e = Ae \tag{11}$$

Generally, ICA-LiNGAM algorithm is used to estimate a separating matrix $W = A^{-1} = I - B$. After a series of permutation and normalization, a strictly lower triangular estimated matrix $\tilde{B}$ is obtained to generate the causal ordering. The correct mixing matrix can only be identified when data is non-Gaussian, since multiple different mixing matrices may yield the same covariance structure if the disturbances are Gaussian (Hyvärinen et al., 2004). Non-Gaussian data enables the use of higher-order moments to construct tests of model fit and determine causal directions (Moneta et al., 2013; Shimizu, Hoyer et al., 2006; Shimizu, Hyvärinen et al., 2006; Shimizu et al., 2012; Shimizu and Kano, 2008). The significance of the remaining edges is tested using Wald test. A remaining edge will be pruned out if the corresponding estimated coefficient is insignificant under Wald test.

LiNGAM algorithm is appropriate when at most one of the series is Gaussian distributed. The more non-Gaussian the data is, the more accurate the identified causal structure is (Shimizu and Kano, 2008). In this chapter, the innovations of all three series for each stock are non-Gaussian (evidence cited below). Therefore, LiNGAM algorithm is applied, aiming to identify stronger contemporaneous causal relations. LiNGAM, as PC, assumes causal sufficiency. Markov condition and faithfulness are not assumed.

2.4. Data

Up to January 2016, there are ten Chinese stocks cross-listed in Shanghai (A-shares), Hong Kong and New York (Table 2.1). All these ten stocks are included in the estimation. Each stock is associated with three price series representing its prices in all three markets, yielding thirty price series in total. Daily closing prices are collected from Datastream database. As suggested by Bessler et al. (2011), converting prices in different local currencies into prices measured by the same base currency may increase exchange rate risk due to currency fluctuations and lead to confounding market interrelation results in contagion-type studies. Here all stock series are measured in their local currencies and are taken in natural logarithms to account for high volatility in data.

Table 2.1 Ten Chinses stocks cross-listed in Shanghai, Hong Kong and New York

Stocks	In-Sample Estimation	Out-of-Sample Forecast
Aluminum Corporation of China Ltd (ACC)	4/30/2007-9/19/2014	9/22/2014-2/6/2015
China Eastern Airlines Co Ltd (CEA)	11/5/1997-9/19/2014	9/22/2014-2/6/2015
China Life Insurance Co Ltd (CLI)	1/9/2007-9/19/2014	9/22/2014-2/6/2015
China Petroleum & Chemical Co (CPC)	8/8/2001-9/19/2014	9/22/2014-2/6/2015
China Southern Airlines Co Ltd (CSA)	7/25/2003-9/19/2014	9/22/2014-2/6/2015
Guangshen Railway Co Ltd (GSR)	12/22/2006-9/19/2014	9/22/2014-2/6/2015
Huaneng Power International Inc (HNP)	12/13/2001-9/19/2014	9/22/2014-2/6/2015
PetroChina Co Ltd (PTC)	11/5/2007-9/19/2014	9/22/2014-2/6/2015
Sinopec Shanghai Petrochemical Co (SSP)	11/8/1993-9/19/2014	9/22/2014-2/6/2015
Yanzhou Coal Mining Co Ltd (YZC)	7/1/1998-9/19/2014	9/22/2014-2/6/2015

2.5. Empirical Results

2.5.1. Stationarity and Structural Breaks

One important questions involved in the estimation of time series data is the stationarity of data. If the data are nonstationary, the usual ordinary least squares estimation of the autoregressive model may lead to spurious regression results (Granger and Newbold, 1974). The stationarity of each series is tested through the augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) and the Phillips–Perron test (Phillips and Perron, 1988). Except for the Shanghai series of PTC, which is stationary at levels, all other series are nonstationary at levels and stationary after first differencing (integrated of order one).

Failure to account for the time-varying instability in stock market may yield unreliable results. Since the data in this chapter cover periods when markets did experience high volatility, Bai-Perron test (Bai and Perron, 2003) is performed to test the potential structural breaks. In general, no breaks exist in most of the thirty series.[3] In this chapter, each set of the three stock series fitted in a VECM is for the same company cross-listed in three different markets, suggesting that they tend to move together and to not be affected by structural changes. Therefore, it is indeed satisfying to find no breaks in general[4].

2.5.2. Cointegration and Order of Lags

Further cointegration tests are performed to detect the potential cointegrating relations. Simply using stationary data achieved by first differencing may fail to capture the long-run information (Engle and Granger, 1987). Following the studies of Phillips (1996) and Wang and Bessler (2005), information criterion (Schwarz-loss and Hannan and Quinn-loss metrics) can be used to test the number of cointegrating vectors, the lag length in VAR representation and the inclusion of a constant in cointegrating space simultaneously. The model used for the simultaneous test is the equivalent levels VAR at various lag orders (1-10) and cointegration ranks (1-3). Since additional regressors may be over-penalized using the Schwarz-loss metric (Geweke and Meese, 1981), only results using Hannan and Quinn-loss metric are considered. Table 2.2 presents simultaneous test results for CEA. To save space, results of other firms are not shown and are available upon request. Accordingly, all ten stocks agree with one cointegration rank with the exception that two cointegration ranks are found for PTC. A constant

[3] No breaks are found for seven out of the ten stocks. Breaks are found for two series of ACC, two series of PTC and one series of YZC respectively based on BIC information criteria. However, the breaks cannot be found when using LWZ information criteria. The volatile results provide stronger support of no breaks.
4 To save space, the detailed results of unit root test, structural break test and the simultaneous test (in Section 2.5.2.) are not reported here in order to save space. But they are available on request.

should be included in the cointegrating space for all ten stocks. Different lag lengths are suggested for different stocks: 2 for ACC, 3 for CEA, 2 for CLI, 2 for CPC, 3 for CSA, 4 for GSR, 3 for HNP, 2 for PTC, 4 for SSP and 4 for YZC.

Table 2.2 Hannan and Quinn-loss metric on VECM for China Eastern Airlines Co Ltd (CEA)[a]

	With a constant[b]			Without a constant		
	r<=1	r<=2	r<=3	r<=1	r<=2	r<=3
lag=1[c]	-21.800	-21.797	-21.796	-21.798	-21.796	-21.796
lag=2	-21.839	-21.837	-21.836	-21.838	-21.836	-21.836
lag=3	-21.846[d]	-21.843	-21.842	-21.844	-21.842	-21.842
lag=4	-21.845	-21.842	-21.841	-21.843	-21.841	-21.841
lag=5	-21.842	-21.839	-21.838	-21.84	-21.838	-21.838
lag=6	-21.839	-21.836	-21.835	-21.837	-21.835	-21.835
lag=7	-21.835	-21.832	-21.831	-21.833	-21.831	-21.831
lag=8	-21.831	-21.829	-21.827	-21.83	-21.828	-21.827
lag=9	-21.827	-21.825	-21.823	-21.825	-21.824	-21.823
lag=10	-21.826	-21.823	-21.822	-21.824	-21.822	-21.822

[a] Hannan and Quinn-loss metric is measured as $\text{HQ} = \log(\det(\Sigma)) + (2.01)(\text{m}\times\text{k})\frac{\log(logT)}{T}$, where Σ is the estimated error variance-covariance matrix, *m* is the number of equations, *k* is the number of regressors in each equation, T is the total number of observations on each series, $det(\Sigma)$ is the determinant of estimated error variance-covariance matrix.
[b] With(without) a constant means does(not) include a constant within the cointegrating vectors.
[c] The lag indicates lag order for the levels vector autoregression (VAR) model. If there are *m* lags in the levels VAR, there will be *(m-1)* lags in the equivalent VECM.
[d] The minimum loss metric.

2.5.3. Identification of Long-Run Structure

A series exploratory tests are performed to identify the long-run structure. First, tests of stationarity are performed to examine whether the cointegrating vector is caused since one of the three series for one stock is itself stationary or since there is a linear combination of two or three series (Consistent with the tests for non-stationarity discussed above. Here the null hypothesis is stationarity). As a result, except for the Shanghai series of PTC, none of the remaining series is stationary by itself.

2.5.3.1. Test of Exclusion

Table 2.3 presents the results of exclusion tests. For CEA, CLI, CSA, HNP and PTC, the null hypothesis that one particular series is excluded from the cointegrating space is rejected for all three series. Therefore, for these stocks, the cointegrating vector arises because of a linear combination of prices in all three markets. For the other five stocks (ACC, CPC, GSR, SSP and YZC), we fail to reject the null hypothesis of exclusion in the Shanghai market, indicating that Shanghai is excluded from the long-run equilibrium. Perhaps, this suggests the Shanghai market is not efficient. For these stocks, accordingly, the series in Shanghai is associated with a zero coefficient in the cointegrating space.

Table 2.3 Tests of exclusion[a]

Series	chi-squared	p-value	Decision
Aluminum Corporation of China Ltd (ACC)			
Shanghai	0.207	0.649	F
Hong Kong	724.553	0.000	R
New York	725.599	0.000	R
China Eastern Airlines Co Ltd (CEA)			
Shanghai	4.343	0.037	R
Hong Kong	582.173	0.000	R
New York	582.177	0.000	R
China Life Insurance Co Ltd (CLI)			
Shanghai	13.254	0.000	R
Hong Kong	766.148	0.000	R
New York	766.366	0.000	R
China Petroleum & Chemical Co (CPC)			
Shanghai	2.110	0.146	F
Hong Kong	1095.078	0.000	R
New York	1095.046	0.000	R
China Southern Airlines Co Ltd (CSA)			
Shanghai	7.280	0.007	R
Hong Kong	490.114	0.000	R
New York	490.114	0.000	R
Guangshen Railway Co Ltd (GSR)			
Shanghai	2.479	0.115	F
Hong Kong	250.383	0.000	R
New York	250.859	0.000	R
Huaneng Power International Inc (HNP)			
Shanghai	6.448	0.011	R
Hong Kong	598.491	0.000	R
New York	598.106	0.000	R
PetroChina Co Ltd (PTC)			
Shanghai	13.636	0.001	R
Hong Kong	764.981	0.000	R
New York	764.592	0.000	R
Sinopec Shanghai Petrochemical Co (SSP)			
Shanghai	0.802	0.370	F
Hong Kong	557.392	0.000	R
New York	556.734	0.000	R
Yanzhou Coal Mining Co Ltd (YZC)			
Shanghai	0.619	0.432	F
Hong Kong	406.761	0.000	R
New York	406.616	0.000	R

[a] The tests are performed based on one cointegrating vector (two cointegrating vectors for PTC). The null hypothesis is that the particular series is excluded from the long-run relations. If decision is “R”, it means we reject the null at 5% significance level. If decision is “F”, it means we fail to reject the null at 5% significance level. The test statistic is distributed chi-squared with one degree of freedom (two degrees of freedom for PTC).

2.5.3.2. Test of Weak Exogeneity

The results of weak exogeneity tests are shown in Table 2.4. The null hypothesis is that a particular series does not respond to the disturbances in the long-run relations. For CLI, CPC, CSA, GSR, PTC and YZC, all the three series are rejected under null. For the other four stocks (ACC, CEA, HNP and SSP), each has one series that is weakly exogenous. Particularly, the Shanghai market is weakly exogenous for three stocks. One possible reason may be its ownership restriction hinders the information transmission from other offshore markets. For these weakly exogenous series, corresponding zero coefficients are imposed in the adjustment coefficients.

Table 2.4 Tests of weak exogeneity[a]

Series	chi-squared	p-value	Decision
Aluminum Corporation of China Ltd (ACC)			
Shanghai	10.499	0.001	R
Hong Kong	297.534	0.000	R
New York	1.599	0.206	F
China Eastern Airlines Co Ltd (CEA)			
Shanghai	3.751	0.053	F
Hong Kong	24.725	0.000	R
New York	128.493	0.000	R
China Life Insurance Co Ltd (CLI)			
Shanghai	59.760	0.000	R
Hong Kong	303.456	0.000	R
New York	8.059	0.005	R
China Petroleum & Chemical Co (CPC)			
Shanghai	15.452	0.000	R
Hong Kong	303.833	0.000	R
New York	28.310	0.000	R
China Southern Airlines Co Ltd (CSA)			
Shanghai	5.386	0.020	R
Hong Kong	91.234	0.000	R
New York	14.400	0.000	R
Guangshen Railway Co Ltd (GSR)			
Shanghai	5.803	0.016	R
Hong Kong	74.008	0.000	R
New York	6.218	0.013	R
Huaneng Power International Inc (HNP)			
Shanghai	1.345	0.246	F
Hong Kong	157.642	0.000	R
New York	10.931	0.001	R
PetroChina Co Ltd (PTC)			
Shanghai	54.478	0.000	R
Hong Kong	365.179	0.000	R
New York	11.997	0.002	R
Sinopec Shanghai Petrochemical Co (SSP)			
Shanghai	1.081	0.299	F
Hong Kong	82.976	0.000	R
New York	36.797	0.000	R
Yanzhou Coal Mining Co Ltd (YZC)			
Shanghai	4.590	0.032	R
Hong Kong	49.496	0.000	R
New York	41.852	0.000	R

[a] The tests are performed based on one cointegrating vector (two cointegrating vectors for PTC). The null hypothesis is that the particular series does not respond to the perturbation in the long-run relations. If decision is “R”, it means we reject the null at 5% significance level. If decision is “F”, it means we fail to reject the null at 5% significance level. The test statistic is distributed chi-squared with one degree of freedom (two degrees of freedom for PTC).

2.5.4. Identification of Contemporaneous Structure

2.5.4.1. Normality Test of Innovations

Specific normality tests are conducted to check whether the innovations from VECM estimation are Gaussian distributed (Table 2.5). For test of skewness, six out of the thirty innovation series fail to reject the null of no skewness at 5% significance level, indicating potential symmetrical distribution for these series. However, all series show evidence of excess kurtosis. For both the Jarque-Bera test and the Kolmogorov Smimov test, the null hypothesis of normal distribution is strongly rejected for all stocks in all markets. Figure 2.1 displays the histograms and kernel density curves of the estimated innovations. Gaussian distributions with corresponding means and variances are overlaid on the same graphs. The graphs mirror quite well the results of normality tests. Thus, innovations from all thirty series are non-Gaussian, implying the appropriateness of using LiNGAM algorithm in this chapter.

Table 2.5 Normality tests of innovations

Series	Skewness Test [a]	Kurtosis Test [b]	Jarque-Bera Test [c]	Kolmogorov Smirnov Test [d]
Aluminum Corporation of China Ltd (ACC)				
Innovation_ Shanghai	0.133**	2.801***	635.594***	0.085***
Innovation_ Hong Kong	0.574***	7.764***	4946.193***	0.079***
Innovation_ New York	0.470***	4.202***	1488.474***	0.086***
China Eastern Airlines Co Ltd (CEA)				
Innovation_ Shanghai	0.156***	3.565***	2347.542***	0.090***
Innovation_ Hong Kong	1.298***	22.363***	92897.912***	0.084***
Innovation_ New York	0.803***	9.521***	17089.411***	0.100***
China Life Insurance Co Ltd (CLI)				
Innovation_ Shanghai	-0.045	2.812***	662.010***	0.071***
Innovation_ Hong Kong	-0.684***	10.227***	8902.279***	0.069***
Innovation_ New York	0.372***	4.566***	1789.859***	0.074***
China Petroleum & Chemical Co (CPC)				
Innovation_ Shanghai	0.164***	8.345***	9942.213***	0.091***
Innovation_ Hong Kong	0.015	5.342***	4067.956***	0.068***
Innovation_ New York	0.245***	6.195***	5504.364***	0.084***
China Southern Airlines Co Ltd (CSA)				
Innovation_ Shanghai	0.023	2.697***	881.723***	0.078***
Innovation_ Hong Kong	0.612***	8.117***	8164.882***	0.071***
Innovation_ New York	0.429***	5.017***	3139.012***	0.072***
Guangshen Railway Co Ltd (GSR)				
Innovation_ Shanghai	0.183***	4.097***	1422.080***	0.077***
Innovation_ Hong Kong	0.321***	3.778***	1234.166***	0.063***
Innovation_ New York	0.094*	4.186***	1475.639***	0.069***
Huaneng Power International Inc (HNP)				
Innovation_ Shanghai	0.337***	6.099***	5221.890***	0.078***
Innovation_ Hong Kong	0.075*	3.667***	1867.901***	0.062***
Innovation_ New York	0.137***	4.717***	3096.985***	0.075***
PetroChina Co Ltd (PTC)				
Innovation_ Shanghai	0.234***	5.548***	2316.133***	0.098***
Innovation_ Hong Kong	-0.684***	11.998***	10894.230***	0.064***
Innovation_ New York	-0.132**	6.483***	3144.975***	0.081***
Sinopec Shanghai Petrochemical Co (SSP)				
Innovation_ Shanghai	-0.712***	46.714***	495190.974***	0.100***
Innovation_ Hong Kong	0.099***	6.272***	8928.225***	0.074***
Innovation_ New York	0.382***	4.660***	5056.632***	0.069***
Yanzhou Coal Mining Co Ltd (YZC)				
Innovation_ Shanghai	0.061	2.822***	1405.569***	0.073***
Innovation_ Hong Kong	0.283***	4.439***	3528.157***	0.069***
Innovation_ New York	0.265***	4.892***	4266.624***	0.078***

Notes: *, ** and *** denotes rejection of the null hypothesis at 10%, 5% and 1% significance levels.
[a] The null hypothesis for skewness test is a particular series is not skewed (Sk=0).
[b] The null hypothesis for kurtosis (excess) test is a particular series has no excess kurtosis (Ku=0).
[c] The null hypothesis for Jarque-Bera test is a particular series is normal.
[d] The null hypothesis for Kolmogorov Smirnov test is a particular series is normal.

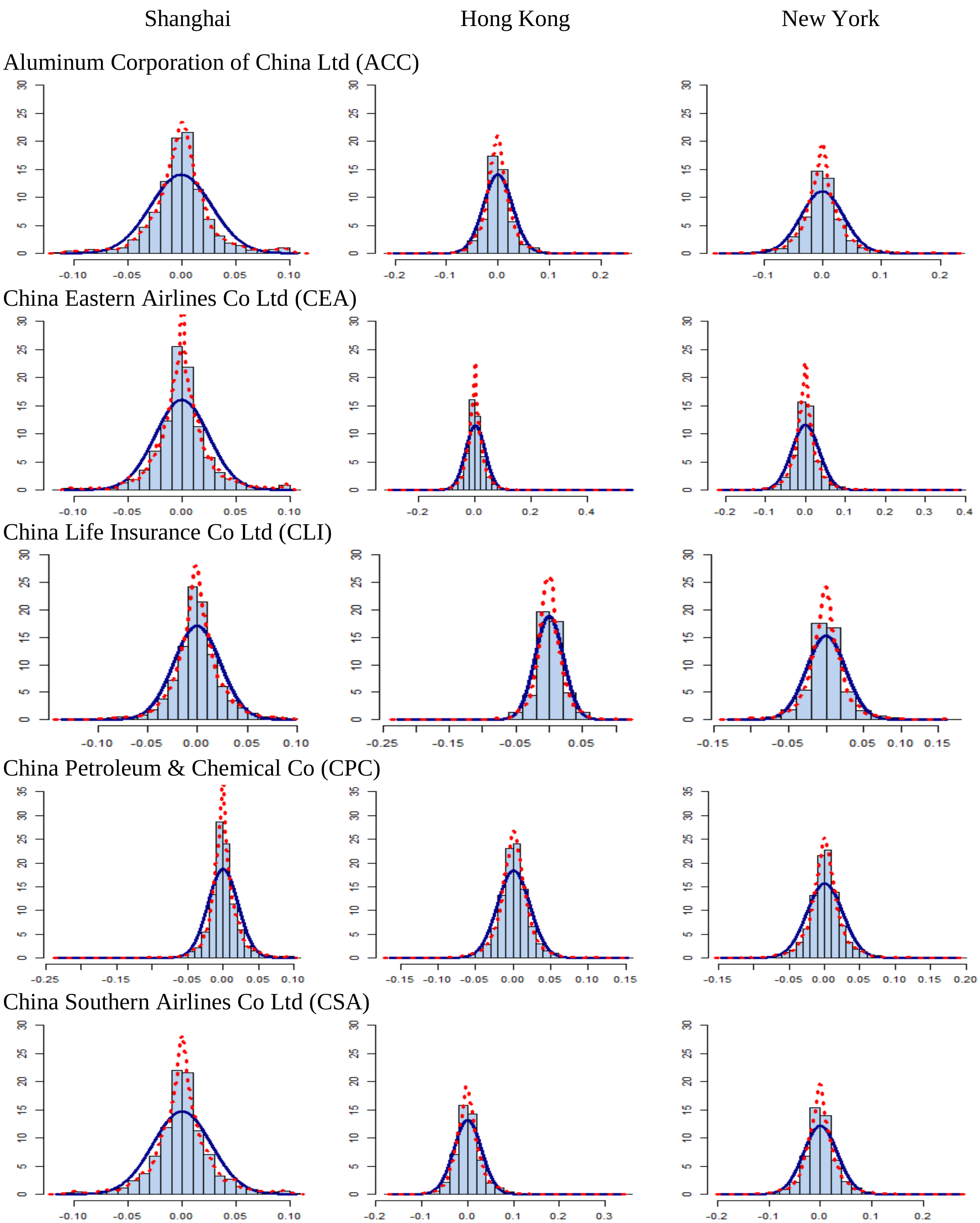


Figure 2.1 Histograms and kernel density curves of the innovations[a]

[a] The solid lines illustrate overlaid normal distributions with corresponding means and variances. The dash lines illustrate the kernel density curves. The choice of algorithm to calculate the bandwidth will affect the smoothness and closeness of the kernel density curves. Different bandwidths are used to generate kernel density curve for each series and the shape of the kernel density curve for each series is not sensitive to the selection of bandwidth in this chapter. The presenting kernel density curves are generated based on the method of Sheather & Jones (1991).

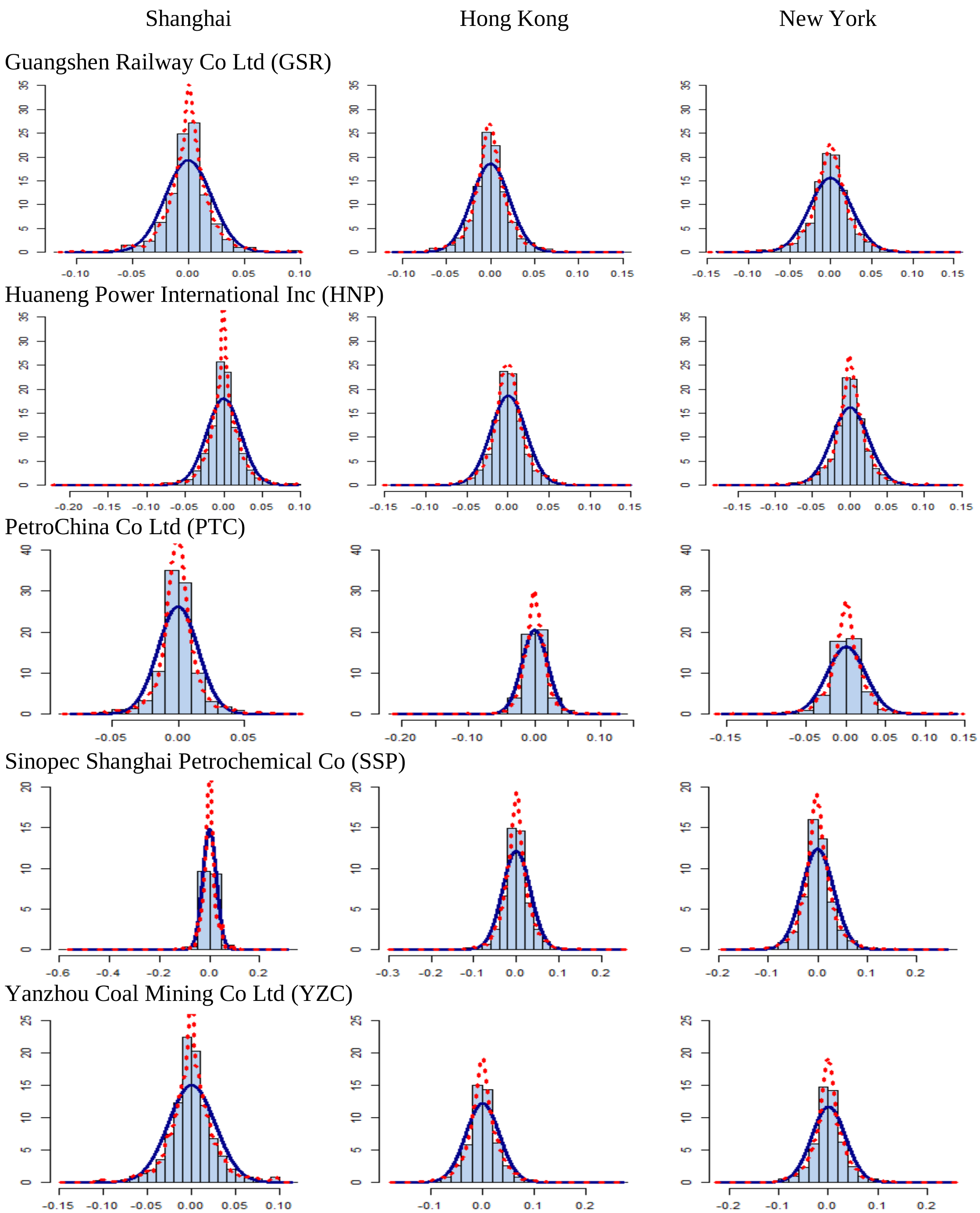


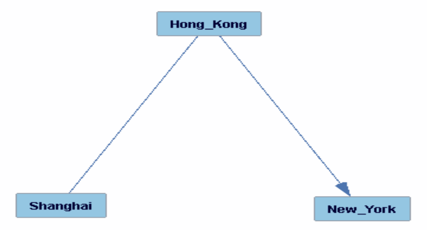
Figure 2.1 Continued

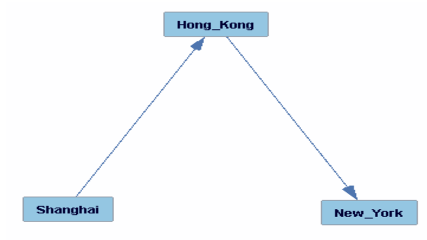
2.5.4.2. Graphical Representations

Graphs generated using both PC and LiNGAM algorithms are presented in Figure 2.2. The restriction that the New York market cannot contemporaneously cause the Shanghai and Hong Kong Markets is imposed since the New York market opens after the other two markets are closed (Figure 2.3). For all ten stocks, LiNGAM algorithm yields stronger causal identifications comparing to PC algorithm. Specifically, the contemporaneous information flow between the Shanghai market and the Hong Kong market is undirected by PC algorithm for all stocks. However, LiNGAM algorithm is capable of sorting out the causal flow from the Shanghai market to the Hong Kong market (Appendix A).

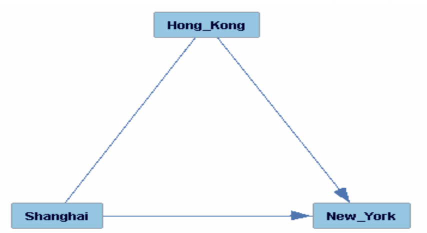
PC | LiNGAM

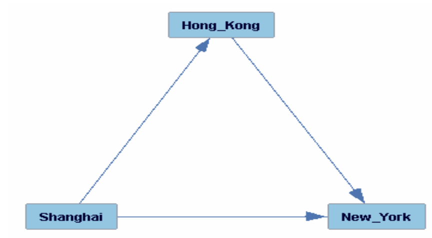
Aluminum Corporation of China Ltd (ACC)



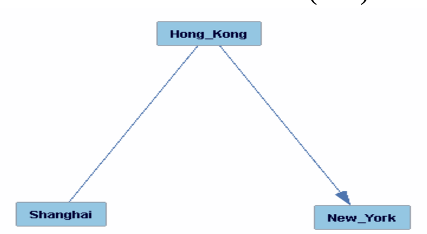


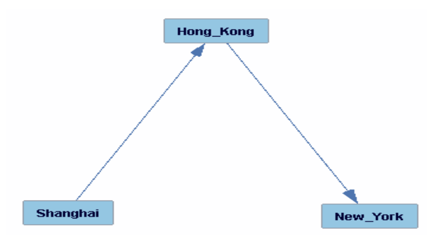
China Eastern Airlines Co Ltd (CEA)



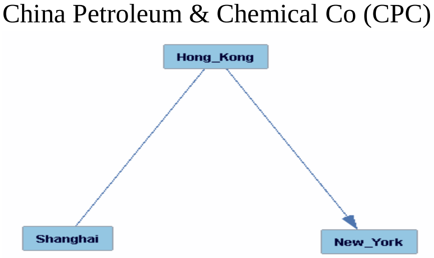


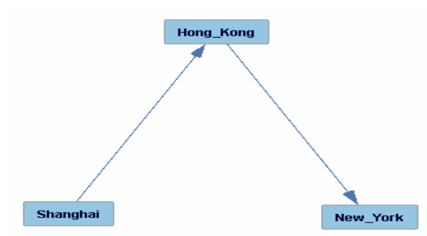
China Life Insurance Co Ltd (CLI)



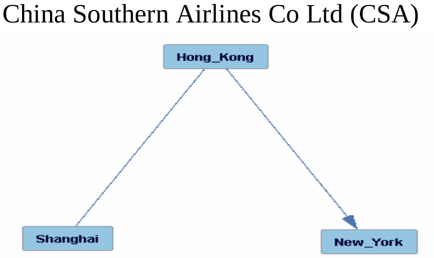


China Petroleum & Chemical Co (CPC)





China Southern Airlines Co Ltd (CSA)



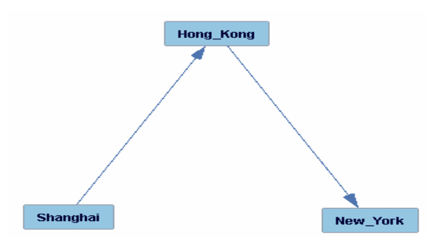


Figure 2.2 Graphs on innovations from the estimated VECM on three price series for each stock[a]

[a] The graphs are generated based on the restriction that the New York market cannot cause the Shanghai and Hong Kong markets, since the New York market opens after the other two markets are closed on the contemporaneous day. The significance level used in PC algorithm is 0.001 and the prune factor used in LiNGAM algorithm is 1.

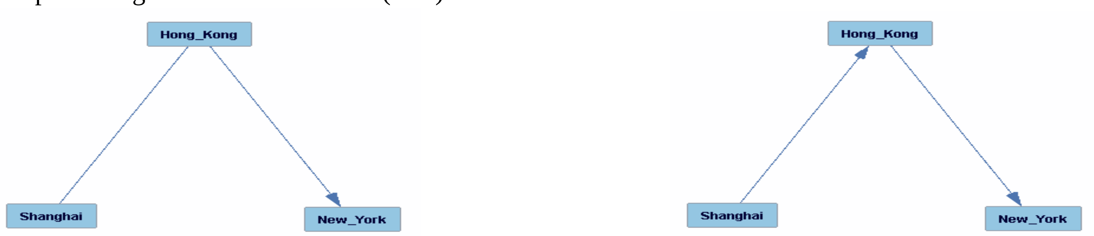
PC LiNGAM

Guangshen Railway Co Ltd (GSR)

Huaneng Power International Inc (HNP)

PetroChina Co Ltd (PTC)

Sinopec Shanghai Petrochemical Co (SSP)



Yanzhou Coal Mining Co Ltd (YZC)

Figure 2.2 Continued

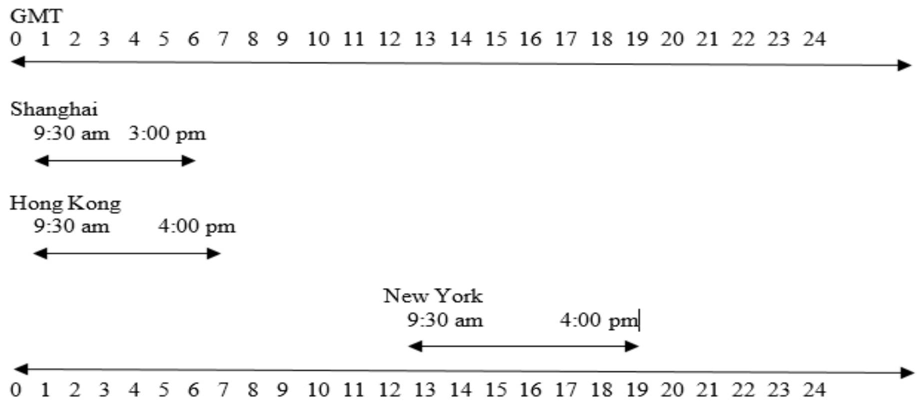


Figure 2.3 Trading hours of the three stock markets

According to the results from LiNGAM, there is a causal chain from the Shanghai market to the Hong Kong market and then to the New York market for all stocks, indicating the Shanghai market leads the Hong Kong market and the Hong Kong market causes the New York market in contemporaneous time. In terms of the market interdependence between Shanghai and New York, the Shanghai market plays a leading role for CEA, GSR and YZC. For the remaining seven stocks (ACC, CLI, CPC, CSA, HNP, PTC and SSP), there is still indirect information flow from the Shanghai market to the New York market mediated by the Hong Kong market.

Overall, the Shanghai market, which acts as the home market for the ten stocks, plays a dominant role in contemporaneous information transmission. The home bias hypothesis is supported in contemporaneous structure in the ten firm equities studied here. This is consistent with the findings of previous literature that the prices of cross-listed stocks will be highly determined by home market factors (Chen et al., 2010; Grammig et al., 2005; Kim et al., 2000; Lau and Diltz, 1994). For the underlying companies of these ten stocks, they are headquartered in mainland China and most of their related information is originated from mainland China. Therefore, the Shanghai market becomes the main source of information flow. On the other hand, the Shanghai market is highly exogenous with no information inflows but just information outflows in contemporaneous time, which may be partially due to its trading exclusivity. The exogenous role of the Shanghai market is also found in Chelley-Steeley and Steeley (2012) and Wang and Jiang (2004).

The Hong Kong market has remarkably more direct influence on the New York market, compared to the Shanghai market. Stronger market integration is also found between the two markets.[6] One reason is that these stocks are issued as ADRs in the New York market based on the underlying stocks traded in the Hong Kong market. Another possible reason is that the higher degree of market openness and information transparency enhances the interaction between the two markets.

2.5.5. Identification of Short-Run Structure

Based on the contemporaneous causal structures identified through LiNGAM algorithm (Figure 2.2), forecast error variance decompositions are given in Table 2.6. Different horizons (steps ahead) are listed with 0 for contemporaneous time, 1 and 2 days for the short horizon, and 40 days for the long horizon.

---

[6] Eun and Shim (1989) suggest the pattern of correlation as an indicator of the degree of market integration. The contemporaneous innovation correlation between the Hong Kong and New York markets is around 0.7 to 0.8 for all ten stocks. The contemporaneous correlations between the Shanghai and Hong Kong markets and between the Shanghai and New York markets are approximately 0.2 to 0.5 for all ten stocks.

Table 2.6 Forecast error variance decompositions[a, b]

Step	Shanghai	Hong Kong	New York	Shanghai	Hong Kong	New York
	Aluminum Corporation of China Ltd (ACC)			*China Eastern Airlines Co Ltd (CEA)*		
Shanghai						
0	100.000	0.000	0.000	100.000	0.000	0.000
1	97.856	0.508	1.635	99.418	0.239	0.342
2	96.817	0.598	2.585	99.021	0.516	0.463
40	93.816	0.825	5.360	98.542	1.330	0.128
Hong Kong						
0	17.222	82.778	0.000	8.065	91.935	0.000
1	14.488	68.459	17.053	8.870	89.370	1.759
2	13.185	62.083	24.733	9.318	87.991	2.691
40	10.406	48.826	40.767	10.335	85.098	4.567
New York						
0	10.000	48.064	41.936	6.656	54.175	39.169
1	10.400	49.370	40.229	7.396	65.081	27.522
2	10.375	49.047	40.578	8.000	69.122	22.878
40	10.227	47.997	41.777	10.016	83.343	6.641
	China Life Insurance Co Ltd (CLI)			*China Petroleum & Chemical Co (CPC)*		
Shanghai						
0	100.000	0.000	0.000	100.000	0.000	0.000
1	96.465	0.015	3.519	98.639	0.411	0.951
2	95.214	0.052	4.734	98.106	0.468	1.427
40	92.516	0.166	7.318	96.680	0.554	2.766
Hong Kong						
0	27.879	72.121	0.000	13.105	86.895	0.000
1	24.997	60.609	14.394	10.535	80.002	9.464
2	23.975	55.338	20.687	9.750	76.859	13.391
40	21.423	44.897	33.680	8.100	67.789	24.111
New York						
0	15.598	40.352	44.050	8.055	53.411	38.534
1	18.304	41.896	39.800	7.968	58.921	33.111
2	19.102	42.554	38.344	7.941	61.140	30.919
40	20.699	44.288	35.013	7.964	66.554	25.482

[a] For the forecast error variance decompositions, a main problem is how to set up the contemporaneous correlation structure. Here, a structural factorization, instead of the Choleski factorization, is applied to allow for more general causal ordering. LiNGAM algorithm is used to provide a data-driven as well as a unique modeling of the contemporaneous innovation structure. Therefore, for all stocks, the forecast error variance decompositions are obtained based on the contemporaneous structures identified through LiNGAM as shown in Figure 2.2.

[b] The sum of elements in each row may not equal to 1 due to rounding to three decimal places.

Table 2.6 Continued

Step	Shanghai	Hong Kong	New York	Shanghai	Hong Kong	New York
	China Southern Airlines Co Ltd (CSA)			*Guangshen Railway Co Ltd (GSR)*		
Shanghai						
0	100.000	0.000	0.000	100.000	0.000	0.000
1	98.710	0.559	0.730	99.425	0.005	0.570
2	97.861	0.883	1.256	98.970	0.013	1.017
40	96.293	1.305	2.403	97.416	0.014	2.571
Hong Kong						
0	15.055	84.945	0.000	11.023	88.977	0.000
1	15.717	80.308	3.976	10.551	80.904	8.545
2	16.856	77.296	5.847	11.256	76.440	12.305
40	16.647	67.544	15.810	14.554	59.119	26.327
New York						
0	10.346	58.378	31.276	7.953	49.458	42.589
1	12.172	63.433	24.395	8.741	55.342	35.916
2	13.532	64.921	21.547	9.786	56.673	33.541
40	16.037	66.653	17.310	14.347	56.791	28.862
	Huaneng Power International Inc (HNP)			*PetroChina Co Ltd (PTC)*		
Shanghai						
0	100.000	0.000	0.000	100.000	0.000	0.000
1	99.397	0.260	0.343	95.847	0.148	4.005
2	98.963	0.578	0.459	94.868	0.174	4.958
40	98.386	1.195	0.419	94.940	0.373	4.687
Hong Kong						
0	5.244	94.756	0.000	12.888	87.112	0.000
1	4.624	87.009	8.367	9.474	71.664	18.863
2	4.903	83.660	11.438	8.382	65.021	26.597
40	4.408	73.092	22.500	4.480	54.922	40.598
New York						
0	3.331	60.190	36.479	6.868	46.422	46.710
1	3.324	64.656	32.019	6.923	50.617	42.459
2	3.560	66.630	29.810	6.733	51.124	42.143
40	4.092	71.615	24.294	4.364	53.851	41.786

Table 2.6 Continued

Step	Shanghai	Hong Kong	New York	Shanghai	Hong Kong	New York
	Sinopec Shanghai Petrochemical Co (SSP)			*Yanzhou Coal Mining Co Ltd (YZC)*		
Shanghai						
0	100.000	0.000	0.000	100.000	0.000	0.000
1	99.907	0.046	0.047	99.351	0.110	0.538
2	99.895	0.036	0.069	98.860	0.190	0.950
40	99.904	0.077	0.019	97.958	0.484	1.558
Hong Kong						
0	3.062	96.938	0.000	5.664	94.336	0.000
1	3.050	93.862	3.088	6.214	90.217	3.569
2	3.277	91.979	4.744	6.727	88.292	4.981
40	4.286	83.636	12.079	7.065	82.268	10.667
New York						
0	1.980	62.662	35.358	5.062	57.025	37.913
1	2.445	70.028	27.527	5.914	65.374	28.712
2	2.830	73.296	23.874	6.336	68.763	24.910
40	4.247	81.241	14.512	7.009	79.807	13.184

The Shanghai market acts as an isolated island in terms of receiving information at all horizons for all stocks. In contemporaneous time, the variation in the Shanghai market is exclusively explained by its own innovations (100%). At both short and long horizons, the Hong Kong and New York markets together explain less than 5% of volatility in the Shanghai market for most stocks. The exogenous Shanghai market is consistent with our findings in Figure 2.2.

In contemporaneous time, volatility in the Hong Kong market is substantially explained by innovations from the Shanghai market for most stocks (approximately 10% to 30%). At the short horizon, the Shanghai market still accounts for a non-trivial share of price movements in the Hong Kong market and is the most influential market other than Hong Kong for most stocks. However, when moving to the long horizon, the importance of the Shanghai market is taken over by the New York market. In particular, at the 40-days-ahead horizon, price information from the New York market explains over 30% variation in the Hong Kong market for ACC, CLI and PTC, which is almost equivalent as Hong Kong's own influence. The New York market, even if it is not the home market, apparently still plays a role in global information transmission. This is consistent with the literature suggesting New York as a global financial center and highlighting its impact on price movements in other markets (Arshanapalli and Doukas, 1993; Bessler and Yang, 2003; Eun and Shim, 1989; Xu and Fung, 2002).

The New York market is dominated by the Hong Kong market at both short and long horizons. For all stocks, the Hong Kong market explains more variations in the New York market than the New York market itself, accounting for almost more than 50% volatility in the New York market. The Shanghai market only has a modest contribution to the price movement in the New York market and its impact is much less relative to the Hong Kong market, especially at the long horizon.

In sum, the importance of the home market, Shanghai, diminishes over time and its leading role (the home bias hypothesis) is only modestly supported at short horizon but not at long horizon. Rather, the Hong Kong and New York markets are observed with non-trivial and growing influences on each other along with the horizon. Consistent with the results found in Figure 2.2, the degree of market integration and market interdependence between Hong Kong and New York markets increases as the horizon increases (similar possible explanations can be found in Section 2.5.4.2).

2.5.6. Robustness Check: Forecasting Performance

As a check of robustness, the estimated VECM is compared with a benchmark model, the autoregressive integrated moving average (ARIMA) model, through recursive forecasting performance. The model is re-estimated and the parameters are updated at each step as the model moves through the forecasting period.

Following Bessler and Wang (2012), point forecasts from the VECM and ARIMA model are evaluated by the idea of d-separation. If one model dominates the other, the information flow from the less preferred model to the actual data will be blocked (d-separated) by the preferred model. Therefore, d-separation can serve as a criterion of model selection.

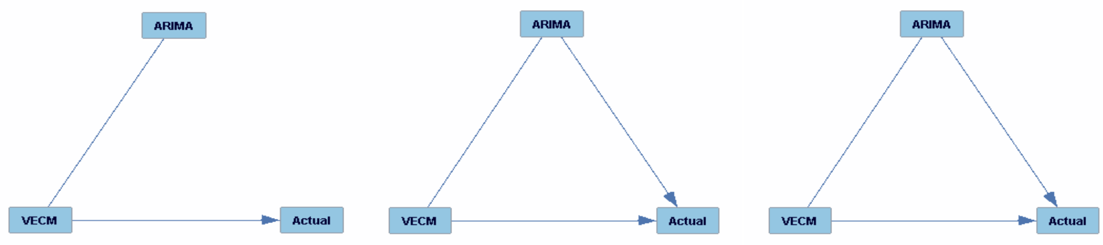
Figure 2.4 shows the causal structures among the forecasts from the two models and the actual realization of stock prices. Take ACC as an example. For the Shanghai market, the information flow from ARIMA forecasts to the actual realization is blocked by VECM forecasts, indicating the VECM is able to capture all the useful information contained in the ARIMA model for the next-step actual data. For the Hong Kong and New York markets, neither model can dominate the other since both models show unblocked paths to the actual realization. As a result, we say VECM forecasts dominant ARIMA forecasts for stock ACC. Similar statement summarizing the superiority of VECM is also found in other seven stocks (CEA, CLI, CPC, CSA, HNP, PTC and SSP). ARIMA forecasts dominate VECM forecasts for only two out of the ten stocks (GSR and YZC). As a result, d-separation results imply the appropriateness of using the estimated VECM in general.

As a supplement to the d-separation evaluation, the adequacy of VECM is reaffirmed by testing point forecasts using root mean squared error (RMSE). The results are listed in Table 2.7. For each stock, if VECM forecasts have lower RMSE than ARIMA forecasts in at least two of the three markets, we say the VECM performs better than the ARIMA model in terms of forecasting for that particular stock. As a result, VECM forecasts dominate ARIMA forecasts for nine out of the ten stocks.[7]

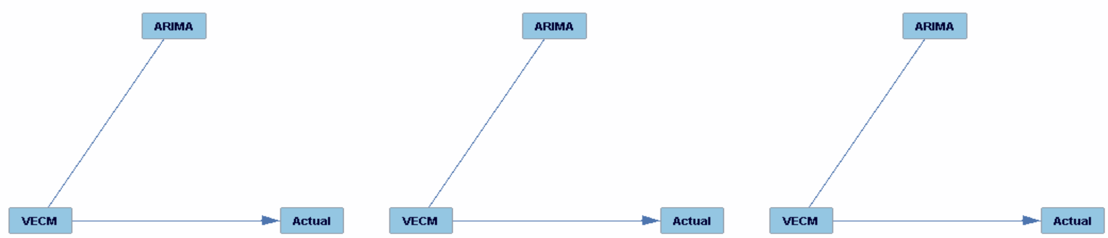
[7] We consider results from a panel specification (Love and Zicchino, 2006) as an alternative to comment on aggregate behavior. Here we group stocks having the same contemporaneous innovation structures together. For firms having a complete directed graph on innovations (CEA, GSR and YZC): Shanghai → Hong Kong, Hong Kong → New York and Shanghai → New York, the panel VAR gives the same causal structure on innovations as shown in their individual models. However, for firms having no edge between Shanghai and New York in their individual results (ACC, CLI, CPC, CSA, HNP, PTC and SSP), the underlying causal structure on innovations from the panel VAR shows an edge between Shanghai and New York in addition to the edges present in their individual models. Therefore, in this case, the panel aggregation does not conflate the underlying qualitative causal structures to a large extent. This "aggregation" issue is left for future study.

Shanghai Hong Kong New York

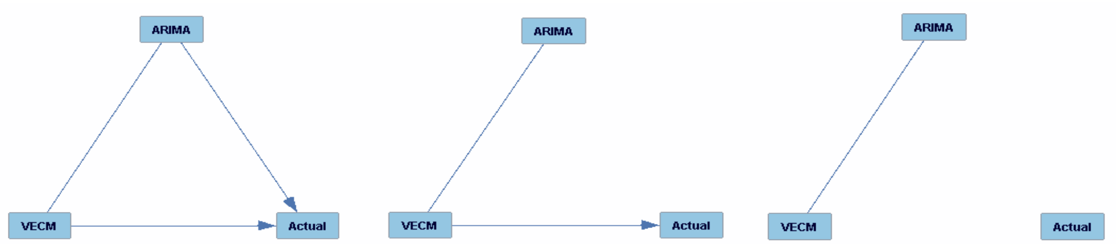
Aluminum Corporation of China Ltd (ACC)



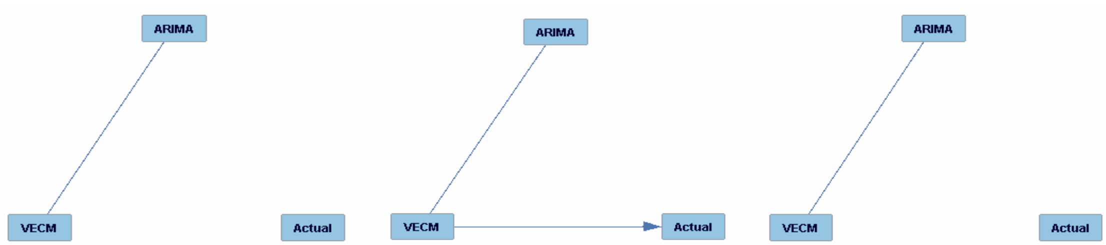
China Eastern Airlines Co Ltd (CEA)



China Life Insurance Co Ltd (CLI)



China Petroleum & Chemical Co (CPC)



China Southern Airlines Co Ltd (CSA)

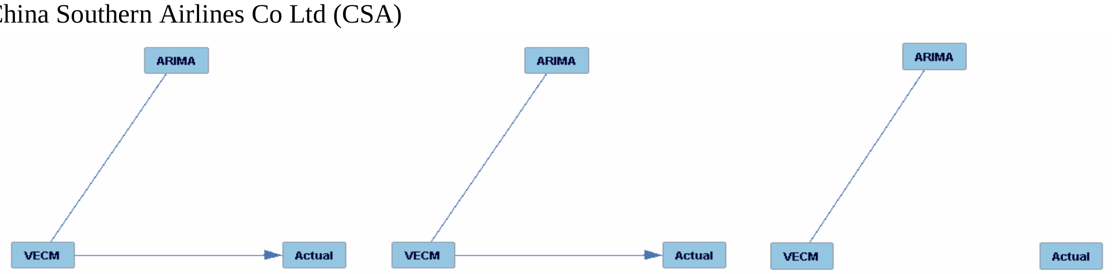
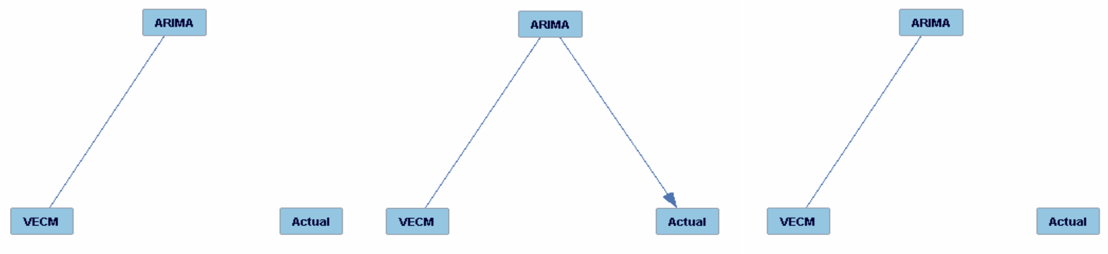


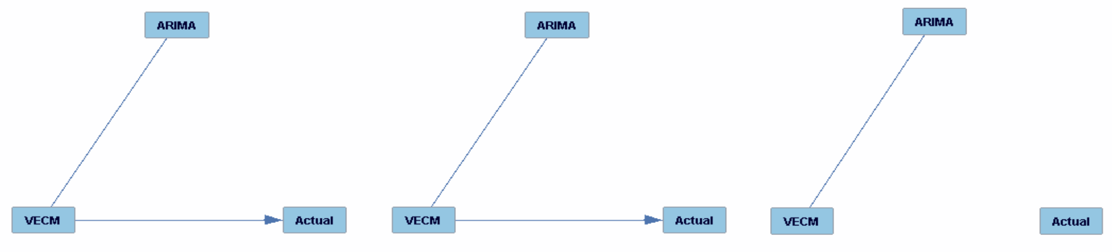
Figure 2.4 Graphs on actual realization and forecasts from the VECM and ARIMA model

Shanghai Hong Kong New York

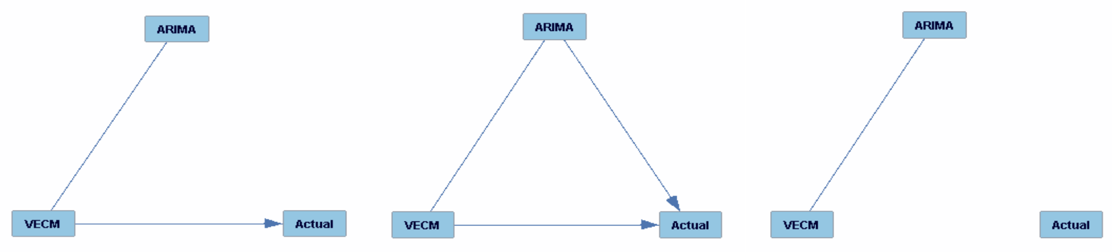
Guangshen Railway Co Ltd (GSR)



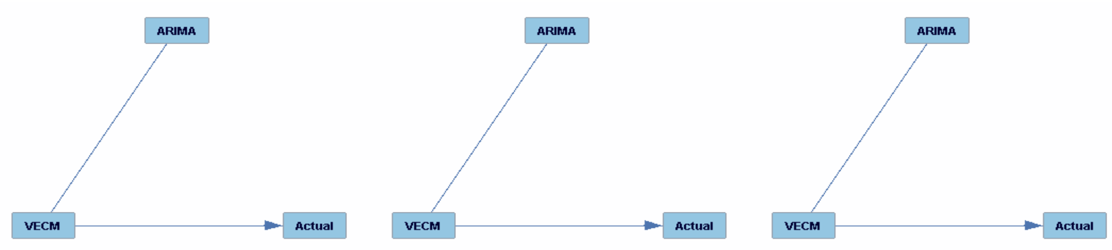
Huaneng Power International Inc (HNP)



PetroChina Co Ltd (PTC)



Sinopec Shanghai Petrochemical Co (SSP)



Yanzhou Coal Mining Co Ltd (YZC)

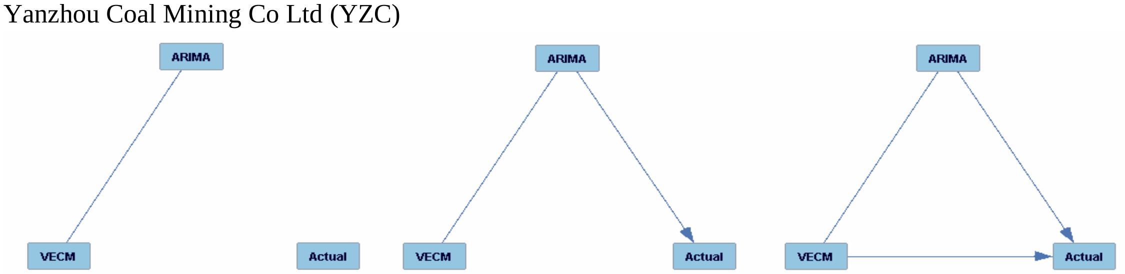


Figure 2.4 Continued

Table 2.7 Root mean squared error

Series	VECM	ARIMA	VECM	ARIMA
	Aluminum Corporation of China Ltd (ACC)		*China Eastern Airlines Co Ltd (CEA)*	
Shanghai	0.03205*	0.03231	0.03753*	0.03781
Hong Kong	0.02229*	0.02231	0.02846*	0.02900
New York	0.02468*	0.02494	0.02932*	0.03060
	China Life Insurance Co Ltd (CLI)		*China Petroleum & Chemical Co (CPC)*	
Shanghai	0.03947*	0.04122	0.02602*	0.02611
Hong Kong	0.01970*	0.02282	0.01343*	0.01499
New York	0.02357*	0.02372	0.01694	0.01677*
	China Southern Airlines Co Ltd (CSA)		*Guangshen Railway Co Ltd (GSR)*	
Shanghai	0.03421*	0.03469	0.03212	0.03198*
Hong Kong	0.02872*	0.02945	0.02011	0.01959*
New York	0.02900*	0.02920	0.01871*	0.01887
	Huaneng Power International Inc (HNP)		*PetroChina Co Ltd (PTC)*	
Shanghai	0.02792*	0.02840	0.02966*	0.03007
Hong Kong	0.02213*	0.02373	0.01614*	0.01797
New York	0.02549*	0.02554	0.02010	0.01996*
	Sinopec Shanghai Petrochemical Co (SSP)		*Yanzhou Coal Mining Co Ltd (YZC)*	
Shanghai	0.02502	0.02492*	0.03529*	0.03536
Hong Kong	0.02145*	0.02225	0.02089	0.02037*
New York	0.02236*	0.02267	0.02225*	0.02341

Notes: * indicates the smaller RMSE value between VECM and ARIMA.

2.6. Conclusion

This chapter aims to investigate the interdependence structure and information transmission among Chinese stocks cross-listed in Shanghai, Hong Kong and New York. Many studies examine cross-listed stocks originating from developed markets with high market quality and find results consistent with the home bias hypothesis, indicating that home market will play a leading role in information transmission of cross-listed stocks (Frijns et al., 2010; Kim et al., 2000; Lau and Diltz, 1994),. We want to explore whether the home bias hypothesis still holds for Chinses cross-listed stocks considering China is an emerging market with inferior market quality.

Several contributions of the chapter are summarized as follows. First, this chapter fills the gap of previous studies on Chinese cross-listed stocks by extending to a multi-market analysis and including all three markets of Shanghai, Hong Kong and New York. Hong Kong and New York are the two most popular offshore stock markets for Chinese firms to cross-list in. Second, this is the first chapter to examine the contemporaneous causal structure of cross-listed stocks using inductive causal graphs. Third, by exploiting the non-Gaussian nature in data, the newly developed machine learning LiNGAM algorithm is capable of offering stronger contemporaneous causal relations that cannot be identified through the traditional PC algorithm. This is also a new attempt in the related area. Fourth, short-run interdependence structure is examined through forecast error variance decompositions. Rather than using the Choleski factorization and determining causal ordering from subjective assumptions or prior economic theories, we obtain the forecast error variance decompositions based on contemporaneous structures identified through structural factorization and LiNGAM algorithm. LiNGAM algorithm is capable of providing a data-determined as well as a unique modeling of the contemporaneous structure, which facilitates a unique measure of each market's contribution to price discovery. Fifth, VECM and cointegration analysis are used to explore the long-run relationship of cross-listed stocks.

As the home market of the selected cross-listed stocks, the Shanghai market plays a dominant role in contemporaneous information transmission. As discovered by

LiNGAM logarithm, there are either direct or indirect information flows transmitting from the Shanghai market to both the Hong Kong and New York markets. The contemporaneous interdependence structure of cross-listed stocks has never been explored by previous literature. Nevertheless, in dynamic (lagged) time, the Shanghai market shows modest impact on other markets at the short horizon and fails to play a nontrivial role at a longer horizon. Instead, the Hong Kong and New York markets, which are suggested with superior level of market quality (Chen et al., 2010), tend to exhibit increasing influences on each other along with the horizon. Thus, the importance of the Shanghai market weakens as horizon increases. The home bias hypothesis is strongly supported for these Chinese cross-listed stocks in contemporaneous time, modestly supported at the short horizon and not supported at the long horizon. Perhaps, the potential deficiencies of the Shanghai market does not impede its leading role in contemporaneous information transmission, but retards its efficiency and information communication with other markets when moving to longer horizon.

In addition, the Shanghai market is highly exogenous at all horizons. The variations of Shanghai-listed stock prices are predominantly explained by the shocks in their own innovations. The exogenous role of the Shanghai market is also backed through the tests of weak exogeneity (Table 2.4) and the market-pair scatter plots of innovations from the estimated VECM (Footnote 3). It seems that the ownership restriction of A-shares has blocked the Shanghai market from receiving outside information to some degree. Chelley-Steeley and Steeley (2012) and Wang and Jiang

(2004) also find evidence indicating the Shanghai market is exogenous in pricing information transmission.

The Hong Kong market causes the New York market in contemporaneous time and shows substantial influence on the New York market at both short and long horizons. The market integration gets stronger between the Hong Kong and New York markets as the horizon increases, with a bidirectional impact. It is worth mentioning that, contrary to the highly exogenous Shanghai market, both the Hong Kong and New York markets have more exposure to other markets. One possible explanation for these results is that the stocks issued in the New York market are ADRs with underlying stocks traded in the Hong Kong market. Another possible explanation is that the market quality in the Hong Kong and New York markets is higher (i.e. higher degree of economic openness and market transparency), which improves the efficiency of information transmission between the two markets.

CHAPTER III

STOCK RETURNS AND INTEREST RATES IN CHINA: THE PREQUENTIAL APPROACH

3.1. Introduction

Interest rates and term structure variables can serve as predictors of real economic activity such as output and consumption (Chen, 1991; Estrella and Hardouvelis, 1991; Harvey, 1988, 1989; Sims, 1980b). Stock returns are found to be positively related to various real economic activities and can be considered as indicators of expected future output (Fama, 1981; Patelis, 1997). Further interest rate related variables, which reflect underlying macroeconomic conditions, can help in forecasting stock returns (Campbell, 1991).

The contribution of interest rates to the predictable movements of stock returns has been extensively discussed in the literature. By regressing expected common stock returns on the one-month treasury bill rate, Fama and Schwert (1977) find the estimated coefficient to be significantly negative. The significant negative relationship is also found between expected returns of common stock and the change of treasury bill rate, as a proxy of the change of expected inflation rate. Campbell (1987) forecasts excess returns of stock using predictors such as the one-month treasury bill rate and spreads between interest rates with different maturities. Based on the Wald tests of estimated coefficients and the R^2 statistics of regression, it is suggested the short-term interest rates and the term structure of interest rates have modest but reliable power in explaining excess stock returns. Campbell (1991) extends the analysis to a vector autoregressive (VAR) model and argues modest but statistically significant result that relative interest rates help to forecast stock returns. By comparing the R^2 values generated from his VAR model with that from Fama and French (1988b)'s univariate autocorrelation model of stock returns, it is shown that multivariate model has stronger explanatory power. The promoting role of short-term interest rates in forecasting stock returns is also supported in other papers (Campbell and Ammer, 1993; Cutler et al., 1991; Fama and French, 1989; Hashemzadeh and Taylor, 1988).

One characteristic of this literature is that most studies examine the influence of short-term interest rates on forecasting stock returns based on in-sample tests of model fit. A common method is to regress stock returns on potential forecasting variables such as interest rates and dividend and price ratio. The predictive capability of the forecasting variables is evaluated based on the significance tests of estimated coefficients and the marginal R^2 statistics (Campbell and Shiller, 1988). However, as suggested by Stock and Watson (2001), the in-sample significance tests may fail to provide substantial information to identify stable predictors. They take advantage of out-of-sample forecasts to assess predictors of economic output and inflation. Granger (1980) also argues that causality should be examined based on out-of-sample forecasts rather than on in-sample tests of model fit.

Another characteristic of this literature is that it focuses on developed economies such as the U.S. market. Very few studies have examined interest rates' predictive role in stock returns in China. Different from the developed economies, interest rates in

China have been highly controlled and regulated by the central bank (the People's Bank of China) (Porter and Xu, 2009). This may lead to inconsistent results of the interrelation between stock returns and interest rates, comparing to the literature listed above (Fang et al., 2016). However, China started the liberalization of interest rates by allowing the interbank offered rates to be determined on a market basis in 1996. Since 2013, the liberalization of interest rates in China has experienced momentous development by fully eliminating the regulations on loan rates and lifting the upper bound on deposit rates of financial institutions. In late 2015, the liberalization of interest rates mainly completed. Particularly, as an important implementation in interest rates liberalization, the Shanghai Interbank Offered Rate (Shibor) was introduced, which is targeted as the benchmark interest rate in Chinese monetary market. Now the Shibor has become one of the most market-based interest rates in China. It is expected that the market-based Shibor should well reflect the changes in macro-economy and improve the forecasts of asset prices.

Motivated by these two characteristics, we explore whether interest rates help in predicting stock returns in China by emphasizing the role of out-of-sample forecasts instead of in-sample tests of model fit. Following Bessler and Ruffley (2004), prequential analysis is applied to study the issue from both calibration (reliability) and sorting (resolution) perspectives. A two variable VAR model on stock returns and interest rates and a univariate autoregressive (AR) model on each series are examined. If the VAR model outperforms the AR model for stock returns, then including interest rates improves the forecasts of stock returns.

This chapter contributes to the existing literature in several ways. First, this is the first attempt to use prequential analysis to study the role of interest rates in predicting stock returns in China. Prequential analysis refers to a system where probability forecasts for future observations are issued sequentially given a set of currently known data (Dawid, 1984, 1985). Suppose for a $m\times1$ vector time series X_t, a sequence of observed data are $x_t' = (x_{1t}, \dots, x_{mt}), t = 1, \dots, T$. At time *T*, a set of probability distributions $P_{T,h} = (P_{T+j}; j = 1, \dots, h)$ are assigned for unknown obervations $X_{T+j}, j = 1, \dots, h$, based on the known observations $x_t, t = 1, \dots, T$. A 'prequential forecasting system' (PFS) is defined as a rule *P*, where a choice of $P_{T,h}$ is matched with each *T* value and each set of realized outcomes $x_t, t = T + 1, \dots, T + h$ (Dawid, 1984).

A PFS is judged based on its forecasting ability. Dawid (1984) introduces the prequential principal: a rule *P* can only be assessed through the sequence of its issued probabilities. Therefore, the sequence of out-of-sample predictive distributions and the subsequent actual realizations are evaluated to judge the 'goodness' of a PFS system. Such system is not assessed through prior considerations, such as goodness of model fit or agreement with economic theories. This principal echoes the instrumentalism idea of Friedman (1953) and the idea of Granger (1980) which asserts the role of out-of-sample forecasts in causality tests.

The second contribution is that both a bootstrap-like simulation method and a nonparametric kernel-based simulation method are used in the generating process of probability forecasts. Third, most literature using prequential analysis classify subintervals and events based on arbitrary selection. In this chapter, several data-driven methods, such as the rule of Sturges (1926) and the normal scale rule of Scott (1979), are considered for settings of subintervals and events when evaluating probability forecasts. Fourth, Yates' (1988) idea of slope, as a symbol of a model's sorting ability, is extended to a multiple-event case for the prequential analysis of continuous data. In addition, covariance graphs are presented to graphically illustrate the idea of slope.

The organization of the chapter is as follows. Section 2 presents the methods used to evaluate the prequential forecasting system. Section 3 describes the empirical estimation procedures and data. The empirical results are discussed in Section 4 and Section 5 is the conclusion of the chapter.

3.2. Assessment of Probability Forecasts

3.2.1. Calibration

Calibration is suggested as a measure of the adequacy of probabilities by Lichtenstein et al. (1982). Dawid (1984) proposes the calibration criterion to assess 'goodness' of probability forecasts by comparing the issued probabilities for a sequence of future observations with their relative frequency. If the *ex post* relative frequency of all events which are issued with a probability of p^*, is in fact p^*, the PFS is treated as well-calibrated. For example, for a group of events, the occurrence probability of each event is issued as 0.6 by a model. If 60% of these events actually occur, the model is said to be well-calibrated at 60%. Calibration is similar to the relative frequency definition of probability. However, the latter idea looks at repeated events under identical conditions, while the former idea requires no such conditions. Rather, it requires agreement of relative frequencies with ex ante probabilities.

Dawid (1984) applies the probability integral transform to evaluate the adequacy of a PFS. For continuous quantities $X_{i,t+h}$ with continuous distribution functions $F_{i,t+h}$, the random fractiles $U_{i,t+h} = F_{i,t+h}(X_{i,t+h})$ are uniformly distributed random variables at [0,1]. For discrete quantities $X_{i,t+h}$, the random fractiles also follow uniform distribution with discontinuous functions. As a result, for both cases, if the PFS is well-calibrated, the associated observed fractiles $u_{i,t+h} = F_{i,t+h}(x_{i,t+h})$ should follow uniform distribution U[0,1] with cumulative distribution functions $G(u_{i,t+h}) = u_{i,t+h}$.

The same assessment idea can be achieved through the graphical representation of calibration. The calibration plot has been widely used in the analysis of subjective probabilities (Fischohoff et al., 1977; Lichtenstein et al., 1982). It graphs realized fractile (*x*-axis) against the relative frequency (*y*-axis). The fractile (quantile) is determined by fitting the actual outcome in the corresponding cumulative distribution function for a forecasting data point. It represents the realization under the estimated distribution. For a PFS to be well-calibrated, its calibration plot should be an approximate 45-degree line. The deviation of calibration plot from the 45-degree line is the miscalibration of a PFS.

The calibration plot is attained from the estimated cumulative distribution function, $\hat{G}(U_{i,t+h})$, which is referred to as the 'calibration function' (Bunn, 1984). A sequence of realized fractiles $u_{i,t+h} = F_{i,t+h}\left(x_{i,t+h}\right), t = 1, \dots, T$ are sorted in ascending order to formulate a T-element sequence: $u_{i,h}(1), \dots, u_{i,h}(T)$. The calibration function of these observed fractiles are calculated as follows:

$$\hat{G}\left[u_{i,h}(j)\right] = (j/T),\ j = 1, \dots, T \quad (12)$$

The judgement of well-calibration based on calibration plot is somewhat subjective since there is no formal definition of how close a PFS should be to the 45-degree line in order to be treated as well-calibrated. Dawid (1984) suggests a formal test of calibration. If there is a sequence of *T* ascending observed fractiles, the sequence can be mapped into a unit interval, which is exhausted into *J* nonoverlapping subintervals with length of L_j $(0 \leq L \leq 1)$ for each subinterval *j*. If a PFS is well-calibrated (null hypothesis), it is expected that $(T * L_j)$ observed fractiles will fall into subinterval *j* for each *j*. The test can be constructed by comparing the actual number of realized fractiles in subinterval *j*, y_j, and the expected number of realized fractiles in subinterval *j*, $T * L_j$. Eq. (13) is the goodness-of-fit statistic.

$$\chi^2 = \sum_{j=1}^{J}\left[\left(y_j - T * L_j\right)^2 / T * L_j\right] \sim \chi^2(J-1) \quad (13)$$

The test statistic is compared to the chi-squared statistic with *J*-1 degrees of freedom. If it is larger than the chi-squared critical value, the null hypothesis of well-calibration is rejected. Kling and Bessler (1989) appear as the early users of these ideas in economics.

3.2.2. Brier Score and Yates Partition

Calibration focuses on the forecasting model's reliability by checking the consistency of issued probability forecasts and the event relative frequency. Besides calibration, maximizing a model's sorting ability in distinguishing events that occur from events that do not occur is also of interest (Murphy and Winkler, 1987). For this purpose, the mean probability score, also known as the Brier score (Brier, 1950) is considered, which contains components representing both calibration (reliability) and sorting (resolution).

The Brier score of a single event takes a quadratic form as

$$\overline{PS}(p,d) = 1/T \sum_{t=1}^{T} (p_t - d_t)^2 \quad (14)$$

where *t* is the occasion index with $t = 1, \ldots, T$, p_t is the issued probability forecast for occasion *t*, and d_t is an indicator of occurrence at occasion *t* ($d_t = 1$, if the event occurs at occasion *t*; $d_t = 0$, if the event does not occur at occasion *t*). The minimum value of Brier score is 0 when probability 1 is issued to all occasions for which the event actually occurs ($p_t = d_t = 1$) and probability 0 is issued to all occasions for which the event does not occur ($p_t = d_t = 0$). If the model issues probability 0 to occasions where the event actually occurs ($p_t = 0,\ d_t = 1$) and issues probability 1 to occasions where the event turns out to not occur ($p_t = 1,\ d_t = 0$), the Brier score finds its maximum value of 1. Hence, we expect the dominant model to have a smaller Brier score.

Yates (1988) partitions the Brier score into various components for an in-depth evaluation of probability forecasts. His 'covariance decomposition' can be applied to both discrete and continuous probability forecasts (Bessler and Ruffley, 2004). The Yates partition is written as

$$\overline{PS}(p,d) = Var(d) + Scat(p) + MinVar(p) + Bias^2 - 2 * Cov(p,d) \quad (15)$$

$Var(d)$ is the variance of the occurrence indicator d_t and describes what the base rate is for the target event to occur. Consequently, it reflects the variability beyond the forecasting model's control in most cases. The formula of $Var(d)$ is defined as

$$Var(d) = \bar{d}(1 - \bar{d}) \quad (16)$$

in which $\bar{d}$ is the mean of occurrence indicator over T occasions given as

$$\bar{d} = 1/T \sum_{t=1}^{T} d_t \qquad (17)$$

The term of Bias is calculated as

$$Bias = \bar{p} - \bar{d} \qquad (18)$$

where

$$\bar{p} = 1/T \sum_{t=1}^{T} p_t \qquad (19)$$

$Bias$ measures to which extent the issued probabilities are biased. Thus, the $Bias^2$ represents the overall calibration error no matter the error is positive or negative.

The covariance of probabilistic forecast and occurrence indicator is calculated as

$$Cov(p, d) = [Slope(p)][Var(d)] \qquad (20)$$

The $Slope(p)$ term is given as

$$Slope(p) = \overline{p_1} - \overline{p_0} \qquad (21)$$

in which

$$\overline{p_1} = 1/T_1 \sum_{t=1}^{T_1} p_{1t} \qquad (22)$$

$$\overline{p_0} = 1/T_0 \sum_{t=1}^{T_0} p_{0t} \qquad (23)$$

$\overline{p_1}$ is the mean of probabilities issued to the T_1 occasions for which the event actually obtains. $\overline{p_0}$ is the mean of probabilities issued to the T_0 occasions for which the event does not obtain. We have $T = T_1 + T_0$. The term $Slope(p)$ is the difference between $\overline{p_1}$ and $\overline{p_0}$. It reaches its maximum value of 1 when perfect forecasts are made ($p_t = d_t = 1$ for the T_1 occasions, $p_t = d_t = 0$ for the T_0 occasions). The $Cov(p, d)$ term is the most essential element for evaluating probability forecasts in the sense that it reflects a model's sorting ability because of $Slope(p)$. We want the forecasting model to generate a large $Cov(p, d)$ value, by assigning high probabilities to occasions for which the event actually occurs and assigning low probabilities to occasions for which the event does not occur.

The $Scat(p)$ is defined as

$$Scat(p) = (1/T)[T_1 Var(p_1) + T_0 Var(p_0)] \tag{24}$$

where

$$Var(p_1) = 1/T_1 \sum_{t=1}^{T_1} (p_{1t} - \overline{p_1})^2 \tag{25}$$

$$Var(p_0) = 1/T_0 \sum_{t=1}^{T_0} (p_{0t} - \overline{p_0})^2 \tag{26}$$

$Var(p_1)$ is the variance of probabilities assigned to the T_1 occasions where the event actually occurs and $Var(p_0)$ is the variance of probabilities assigned to the T_0 occasions where the event does not occur. The term $Scat(p)$ reflects the total noise in a PFS that is extraneous to the occurrence of target event.

$MinVar(p)$ is calculated as

$$MinVar(p) = Var(p) - Scat(p) \tag{27}$$

where $Var(p)$ is the forecast variance. $MinVar(p)$ acts as the variance of effect variables, reflecting the minimum forecast variance that must be tolerated if the model makes its fundamental forecasts. $Scat(p)$ corresponds to the variance of errors.

Overall, all components other than $Var(d)$ in Brier score are controllable under the forecasting model. In order to gain a minimized Brier score, we want to minimize $MinVar(p)$, $Scat(p)$ and $Bias^2$. However, the $Cov(p, d)$ term, which is of the most importance in probability forecasting, needs to be maximized.

The Brier score and Yates partition can be extended to multiple events. Following Murphy (1972), the multiple-event Brier score for an *N*-event case is defined as

$$\overline{PSN}(p,d) = \sum_{n=1}^{N} \overline{PS}(p,d)_n \tag{28}$$

Therefore, the *N*-event Brier score is equal to the summation of the mean probability scores of all *N* events. The Yates partition for the *N*-event case is given by Yates (1988) as well:

$$\overline{PSN}(p,d) = \sum_{n=1}^{N} Var(d_n) + \sum_{n=1}^{N} Scat(p_n) + \sum_{n=1}^{N} MinVar(p_n) + \sum_{n=1}^{N} (Bias_n)^2 - 2\sum_{n=1}^{N} Cov(p_n, d_n) \tag{29}$$

where the subscript *n* indexes one of the *N* events.

3.3. Data and Empirical Estimation

Here prequential analysis is used to study the relationship between stock returns and interest rates in China. The log return of stock is calculated using Shanghai Stock Exchange Composite Index (SSECI) and the interest rate is measured by the 3-month Shibor. Daily data from 2013 to 2015 are collected. Based on the augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) and the Phillips–Perron test (Phillips and Perron, 1988), the stock return series is stationary while the interest rate series presents a unit root. After detrending daily interest rates through a first difference transformation, the series appear stationary. Thus, the following estimations are applied on stock returns and relative interest rates (the changes in interest rates).

The lag length is selected based on Schwarz Information Criterion for both VAR and AR models. As a result, the VAR model is generated as a first-order process. For the AR model of stock returns, zero lag is suggested, indicating that here the stock index fits in a random walk model. One lag is selected for the AR model of interest rates.

The first half of the data set is used for tests of model specification (230 observations). The remaining sample data are considered as the forecast period for generating one-step-ahead out-of-sample probability forecasts (229 observations).

Following Fair (1986) and Kling and Bessler (1989), two aspects of uncertainty are considered to generate the cumulative distribution functions for each out-of-sample time point – uncertainty in estimated parameters and uncertainty in estimated innovation. We can take the VAR model as an example:

$$X_t = A(B)_t X_{t-1} + u_t \tag{30}$$

where $A(B)_t$ is the parameter matrix of the model at time *t*. Given the estimated parameter matrix $\hat{A}(B)_t$ and its variance-covariance matrix $V_t = P_t P_t'$, the elements of $A(B)_t$ are assumed to be normally distributed as $A(B)_t \sim N(\hat{A}(B)_t, V_t)$. The draws of the parameter matrix $A(B)_t^*$ can be attained through the random generations from the following equation:

$$A(B)_t^* = \hat{A}(B)_t + P_t e_t \tag{31}$$

where e_t is a vector of draws from the standard normal distribution.[8]

With respect to the simulation of innovation u_{t+1}^*, two methods are considered. The first is a bootstrap-like method to draw from the reshuffled historical innovations.

[8] The actual program used is an updated version of the RATS program used in Kling and Bessler (1989).

The second is to simulate from the kernel distribution function of the historical innovations.[9] As a nonparametric method, kernel estimation makes statistical inference on population of data based on a finite sample. It is able to estimate the probability density function of a random variable without prior assumption of its distribution. Here, U_k is a random variable of innovations for variable *k* (*k*=1 or 2 in our case) and $\left(u_{k,1}, \dots, u_{k,t}\right)$ are the iid historical innovations at time *t*. The kernel estimator (Parzen, 1962) for the density function of U_k is

$$\widehat{f_h}(u_k) = 1/t \sum_{i=1}^{t} K_h\left(u_k - u_{k,i}\right) = 1/th \sum_{i=1}^{t} K\left(u_k - u_{k,i}/h\right) \quad (32)$$

where $K_h(x) = 1/h * K(x/h)$ is the kernel function and *h* is the bandwidth for data smoothing. Accordingly, the kernel estimator for the distribution function of U_k is

$$\widehat{F_h}(u_k) = \int_{-\infty}^{u_k} \widehat{f_h}(y)dy \quad (33)$$

The Altman and Leger (1995)'s plug-in method with an unbiased cross-validation is implemented to select bandwidth in kernel distribution estimation. Therefore, in the rest of the chapter, simulation 1 refers to the method where innovations are simulated from the reshuffled historical innovations, while simulation 2 refers to the method where innovations are simulated from the kernel distribution estimation of historical innovations.

[9] The Portmanteau test by Box and Pierce (1970) and the modified Portmanteau test by Ljung and Box (1978) are used to examine the autocorrelation in the historical innovations. As a result, the historical innovations from the estimated VAR model of stock returns show no autocorrelation for 196 out of the 229 forecasting dates. The historical innovations from the estimated AR model of stock returns show no autocorrelation for 197 out of the 229 forecasting dates. In terms of interest rate, historical innovations from both estimated models exhibit no autocorrelation for all 229 forecasting dates.

If the draws of parameter matrix and innovation are both repeated for 1000 times at each *t*, we can get 1000 one-step-ahead forecasts X^*_{t+1} at each *t+1* as

$$X^*_{t+1} = A(B)^*_t X_t + u^*_{t+1} \tag{34}$$

For each forecast date *t+1*, the associated cumulative distribution function is estimated based on the 1000 simulated forecasts X^*_{t+1}. The forecasts are generated recursively in the sense that the model is re-estimated and the estimated parameter matrix and the historical innovations are updated as the model moves forward one data point at each out-of-sample time horizon.

3.4. Empirical Results

3.4.1. Results of Calibration

Calibration plots from both the VAR and AR specifications using different simulation methods are presented in Figure 3.1.[10] To evaluate whether the models are well-calibrated, we investigate the calibration plots visually to see how close they are to the 45-degree line. Both simulation methods generate very similar results. However, the plots from simulation 2 (Figure 3.1: Panel B) are closer to the 45-degree line compared to those from simulation 1 (Figure 3.1: Panel A) for each series and each model.

For stock returns, plots of the VAR and AR models are in line with the 45-degree line with modest amount of deviation. The plots of the VAR model are closer to the 45-

[10] We also consider performance of probability forecasts from the 'linear opinion pool', as suggested by Stone (1961). For calibration plot and calibration test, we average the densities of the VAR and AR models at each forecasting time point. For Brier score and Yates partition in the following Section 4.2, probability forecasts are an average of the densities of the VAR and AR models in each event (mutually exclusive and exhaustive events) at each forecasting time point. The linear opinion pool probability forecasts do not improve the forecasting performance relative to the best model described below.

degree line relative to those of the AR model, indicating the probability forecasts from the VAR model are better calibrated. Both models show over confidence on their forecasts since their realized fractiles lie below the 45-degree line. In terms of interest rates, the first half of the observed fractiles shows larger deviation from the 45-degree line and appears not to be as well-calibrated as the second half for both models. Except the fractile range of 0.45 to 0.70, the plots of the VAR model tend to be closer to the 45-degree line.

In sum, consistent results in terms of model superiority are obtained under both simulations. The VAR model appears to outperform the AR model for stock returns, since its probability forecasts seem to have higher degree of well-calibration. For interest rates, it is difficult to tell the dominance between the two models. All these judgments, however, are obtained through visual inspection of the calibration plots, which may fail to draw objective conclusions on the performance of probabilistic forecasts.

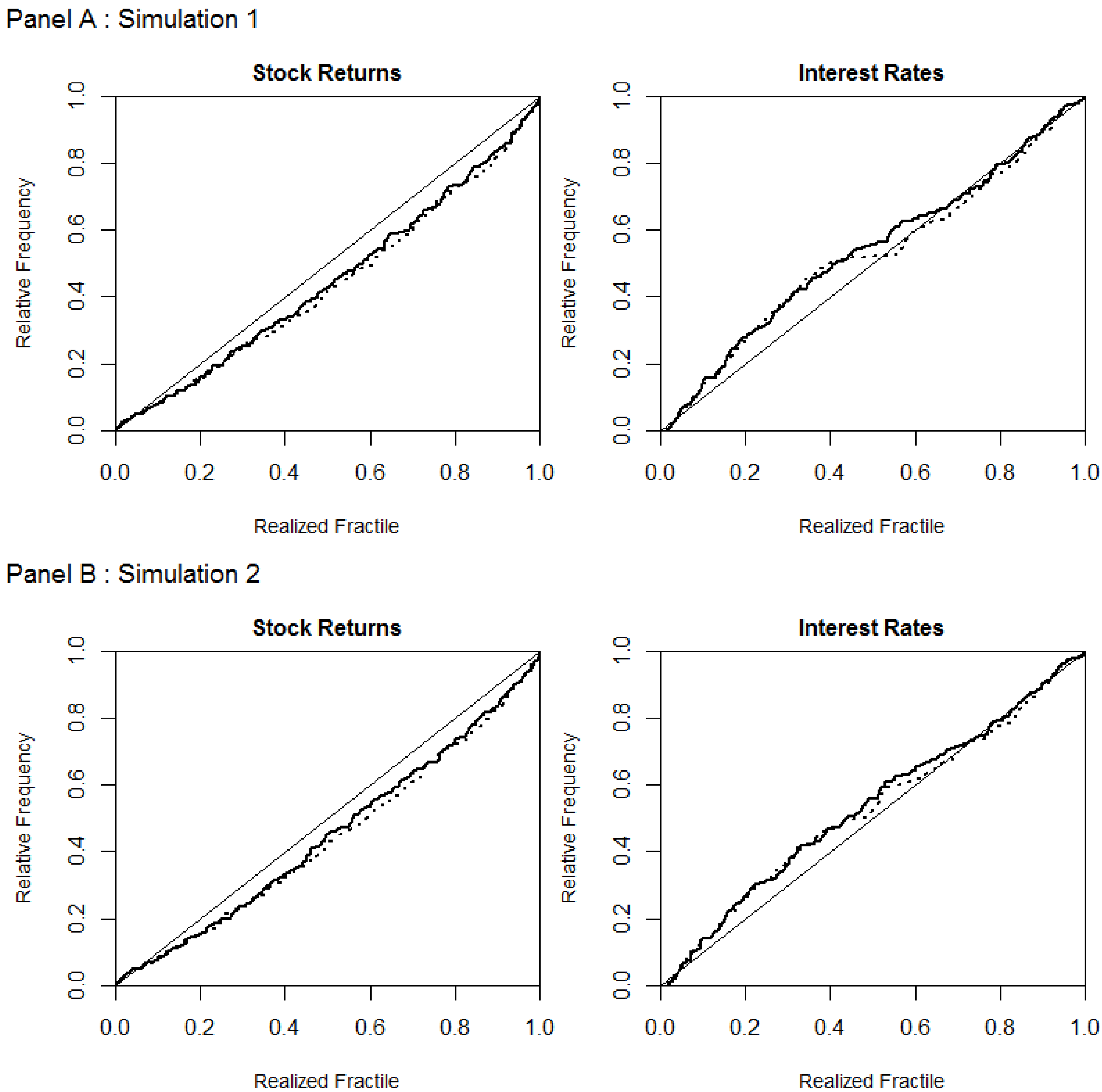


Figure 3.1 Calibration plots on probability forecasts from simulation 1 (Panel A) and simulation 2 (Panel B).

Notes:

1) The solid line represents the plot of VAR model and the dash line represents the plot of AR model.

2) In simulation 1, the innovations are simulated from the reshuffled historical innovations. In simulation 2, the innovations are simulated from the kernel distribution of the historical innovations.

The chi-squared goodness of fit test is carried out as a formal test of calibration. For the calibration test, an unspecified issue is how many subintervals (J) to select.

Bessler et al. (2015) argue that the classification of subintervals should consider the subjective judgements of field experts. However, the choice of the number of subintervals in most prequential analysis studies is somewhat arbitrary. Following Bessler and Kling (1990), we select 10 subintervals.

Referring to the extensive literature discussing the optimal number of bins (or optimal bin width) for a histogram, more data-based methods are considered to determine the optimal number of subintervals (bins) in this chapter. Sturges (1926) proposes one of the earliest rules to choose the number of bins for histogram (Scott, 2015), which is given as

$$\hat{k} = 1 + log_2(n) \quad (35)$$

where $\hat{k}$ is the estimated bin number, log_2 is to take logarithms at base 2, and n is the size of sample data. Here the sample data refers to the observed data over the forecast period and n equals to 229 which is the number of dates in forecast period. Sturges's rule suggests 9 subintervals (bins) for both series, which supports our subjective selection of 10 subintervals.[11]

The normal scale rule of Scott (1979), which is based on the asymptotic minimization of integrated mean squared error (IMSE), is also considered. First, the optimal bin width is calculated as:

[11] We also consider the chi-squared goodness of fit test under 9 subintervals. In the following Section 4.2, Brier score and Yates partition are conducted under 9 events as well. The conclusions concerning whether interest rates help to forecast stock returns in China are the same to those obtained under 10 subintervals/events. Therefore, the results corresponding to 9 subintervals/events are not reported to save space.

$$\hat{h} = 3.49\hat{\sigma}n^{-\frac{1}{3}} \tag{36}$$

in which $\hat{h}$ is the estimated bin width, $\hat{\sigma}$ is the estimated standard deviation of the sample data and n is the sample size. The choices of sample data and n are the same to Eq. (35). The corresponding optimal number of bins is calculated as

$$\hat{k} = (Range\ of\ data)/\hat{h} \tag{37}$$

Accordingly, the optimal number of subintervals (bins) is 17 for stock return series and 16 for interest rate series respectively. As a result, we conduct the calibration tests under two settings. In setting 1, 10 subintervals are used for both series. In setting 2, different numbers of subintervals are used for stock returns (J=17) and interest rates (J=16).[12]

The results of calibration tests are given in Table 3.1. Under simulation 1, for both settings and both series, the null hypothesis of well-calibration cannot be rejected for forecasts from the VAR model, but is rejected for forecasts from the AR model. Under simulation 2, forecasts from both models are well-calibrated for both settings and both series. Particularly, simulation 2 generates smaller chi-squared statistics than simulation 1 in most cases. Therefore, simulation 2 which takes advantage of kernel distribution estimation leads to better calibration results.

Regardless of the subinterval settings and simulation methods, the VAR model always has a smaller chi-squared test statistic than the AR model for both series. These

[12] We consider the alternative to select number of subinterval based on the simulated data for the first forecasting date. For different models and simulation methods, Sturges' rule agrees with 11 subintervals (bins) for both series, which is also closely consistent with our subjective selection of 10. As regards Scott's normal scale rule, different subintervals (bins) numbers are selected for different models and simulation methods. For stock returns, 24 to 26 subintervals (bins) is suggested. For interest rates, the subintervals (bins) number ranges from 43 to 50. The advantage of selecting subinterval number based on the observed data over forecast period is to provide a consistent standard for assessment of probability forecasts, which will not vary by models or simulation methods.

results seem to characterize well the situation observed in calibration plots (Figure 3.1). For both series, the VAR model issues better probability forecasts, which are also well-calibrated, relative to the AR model, implying that stock returns and short-term interest rates help to forecast each other.

Table 3.1 Chi-squared goodness of fit tests for calibration

	Setting 1[a]		Setting 2[b]	
	VAR	AR	VAR	AR
Simulation 1[c]				
Stock Returns	12.0568*	20.7031	19.7467*	27.1703
Interest Rates	16.0218*	25.4542	24.4149*	40.4847
Simulation 2[d]				
Stock Returns	12.1441*	14.3275*	12.6201*	22.1223*
Interest Rates	10.7817*	13.5764*	11.6987*	22.0393*

Notes: * Indicates well-calibration. The null hypothesis is that the observed fractiles are well-calibrated. The null hypothesis is rejected if the calculated chi-squared statistic is larger than the 5% critical value.

[a] In setting 1, the number of subintervals (*J*) is 10 for both stock return series and interest rate series. This results in 22.9 fractiles in each subinterval for both series. The corresponding test statistic is chi-squared distributed with 9 (*J*-1) degrees of freedom and the 5% critical value is 16.9190 for both series.

[b] In setting 2, the number of subintervals (*J*) is 17 for stock return series and the number of subintervals (*J*) is 16 for interest rate series. This leads to 13.47 fractiles in each subinterval for stock return series and 14.31 fractiles in each subinterval for interest rate series. For stock returns, the corresponding test statistic is chi-squared distributed with 16 (*J*-1) degrees of freedom and the 5% critical value is 26.2962. For interest rates, the corresponding test statistic is chi-squared distributed with 15 (*J*-1) degrees of freedom and the 5% critical value is 24.9958.

[c] In Simulation 1, the innovation is simulated from the reshuffled historical innovations.

[d] In simulation 2, the innovation is simulated from the kernel distribution of the historical innovations.

3.4.2. Results of Brier Score and Yates Partition

The Brier score and its decompositions are used to provide information of the model's sorting ability. For each series, the multiple events are defined based on the histogram distribution of observed data over the forecast period. As in the case of the calibration tests discussed above, two settings of the number of events are considered.

Setting 1 has 10 events for both series and setting 2 includes 17 events for stock return series and 16 events for interest rate series.[13]

The results of Brier score and Yates partition under each setting are shown in Table 3.2 (setting 1) and Table 3.3 (setting 2). In setting 1, forecasts from the VAR model have a smaller Brier score (Brier) than those from the AR model for both series and simulations, indicating that the VAR model outperforms the AR model on this measure. The dominance of the VAR model in terms of Brier score is also obtained in setting 2.

For all cases, the VAR and AR models provide the same variance of indicator variable (Dvar). This is not a surprising result. Dvar stands for the forecasting factors that cannot be controlled by the models. Hence, the same Dvar is obtained regardless of forecasting models.

The noise of forecast (Scat) measures how responsive the model is to the irrelevant information in forecasting. It is smaller for forecasts of stock returns from the AR model in both settings and simulations. For interest rates, however, the VAR model is superior on this measure in all cases. It is suggested that when forecasting stock returns, the VAR model contains more noise and performs worse at ignoring irrelevant information. While predicting interest rates, the VAR model is less responsive to extraneous information than the AR model.

[13] For a specific setting and a specific series, the range of each bin in the histogram defines the range of each event, with the exception that the first and last events are extended to include infinity in both directions.

Except for the interest rates under setting 2, the minimum forecast variance (MinVar) that must be accepted is smaller for the AR model in all cases. The sum of Scat and MinVar is the variance of forecast. Thus, the results of Scat and MinVar show that the AR model has lower forecast variance for stock returns and the VAR model has lower forecast variance for interest rates under both settings and simulations.

In terms of the squared absolute calibration error (Bias_squared), the preference of the two models varies by series. For stock returns, forecasts from the VAR model have less overall miscalibration, which is consistent with the results from calibration tests. For interest rates, however, forecasts from the AR model have lower overall miscalibration.

As indicated by Yates (1988), covariance (2Cov) is at the heart of probability forecasting. We want a large 2Cov since it represents to what extent the model is capable of sorting events into subgroups such that events that occur are issued with high probabilities and events that do not occur are issued with low probabilities. For stock returns, the VAR model has a larger 2Cov than the AR model, indicating the VAR model is more responsive to information related to the occurrence of event when predicting stock returns. This result is robust under different event settings and simulation methods. The dominance of the VAR model is particularly strong in the case of setting 1 and simulation 1, where the AR model presents a negative 2Cov, implying it makes poor forecasts by assigning low probabilities to events that actually obtain and high probabilities to events that do not obtain. The stronger sorting ability of the VAR model is also found for interest rates in setting 1. In setting 2, however, the VAR model gives lower 2Cov for interest rates under both simulations.

Scat and 2Cov reflect how responsive the model is to the irrelevant and relevant information, respectively. The VAR model gives two-sided results of these two measures for stock returns. It forecasts stock returns with both larger Scat and 2Cov in all cases. One possible reason is that when forecasting stock returns, the VAR model is better at capturing related information, at a cost of also incorporating more extraneous information.

Table 3.2 Brier score and Yates partition (setting 1[a])

	Stock Returns		Interest Rates	
	VAR	AR	VAR	AR
Simulation 1				
Brier	0.649206*	0.657323	0.249891*	0.255579
Dvar	0.647432	0.647432	0.350337	0.350337
Scat	0.003019	0.001165*	0.105328*	0.108148
MinVar	0.000018	0.000003*	0.066206	0.063873*
Bias_squared	0.004833*	0.007211	0.011021	0.008807*
2Cov	0.006096*	-0.001512	0.283000*	0.275586
Simulation 2				
Brier	0.651469*	0.654364	0.249325*	0.251764
Dvar	0.647432	0.647432	0.350337	0.350337
Scat	0.003249	0.001085*	0.104406*	0.106130
MinVar	0.000010	0.000001*	0.066135	0.064790*
Bias_squared	0.005135*	0.006027	0.011499	0.008964*
2Cov	0.004358*	0.000181	0.283052*	0.278457

Notes: * Indicates the better-performed model for that particular measure. For Brier score (Brier) and the three Yates decomposition components (MinVar, Scat, Bias_square), a smaller value indicates better forecasting performance of the underlying model. For the covariance term (2Cov), a larger value indicates better forecasting performance of the underlying model. The magnitude of Dvar term is out of the model's control.

[a] In setting 1, the number of events is 10 for both stock return series and interest rate series.

Table 3.3 Brier score and Yates partition (setting 2 [a])

	Stock Returns		Interest Rates	
	VAR	AR	VAR	AR
Simulation 1				
Brier	0.783750*	0.784998	0.519846*	0.521098
Dvar	0.779543	0.779543	0.622871	0.622871
Scat	0.003794	0.001198*	0.171461*	0.181579
MinVar	0.000021	0.000009*	0.049432*	0.054100
Bias_squared	0.005210*	0.005751	0.002826	0.002399*
2Cov	0.004818*	0.001503	0.326744	0.339851*
Simulation 2				
Brier	0.781528*	0.785890	0.514366*	0.519251
Dvar	0.779543	0.779543	0.622871	0.622871
Scat	0.003625	0.001148*	0.164392*	0.175432
MinVar	0.000029	0.000006*	0.048596*	0.052270
Bias_squared	0.005307*	0.005596	0.003194	0.002328*
2Cov	0.006976*	0.000402	0.324688	0.333650*

Notes: * Indicates the better-performed model for that particular measure. For Brier score (Brier) and the three Yates decomposition components (MinVar, Scat, Bias_square), a smaller value indicates better forecasting performance of the underlying model. For the covariance term (2Cov), a larger value indicates better forecasting performance of the underlying model. The magnitude of Dvar term is out of the model's control.

[a] In setting 2, the number of events is 17 for stock return series and the number of events is 16 for interest rate series.

As a complementary to the 2Cov, the slope, which is the difference between mean probabilities issued to occasions where the event actually obtains and mean probabilities issued to occasions where the event does not obtain, is calculated based on Eq. (21). For an *N*-event case, we can get the slope by summarizing individual slopes of all events:

$$Slope_{N(p)} = \sum_{n=1}^{N} Slope(p)_n = \sum_{n=1}^{N} (\overline{p_1} - \overline{p_0})_n \tag{38}$$

A larger value of $Slope\_N(p)$ is wanted as a sign of model's stronger ability in distinguishing events that actually occur from events that do not occur. In Table 3.4, the results of multiple-event slope (Slope) reconfirm what we find under the measure of 2Cov in Table 3.2 and Table 3.3. The Slope from the VAR model is much larger than that from the AR model for stock returns under both settings and both simulations, indicating that including interest rates enhances model's overall sorting ability when predicting stock returns. The 2Cov and Slope values associated with stock returns are much smaller than those of interest rates. This suggests interest rates are easier to forecast than stock returns, consistent with the finding in Bessler and Ruffley (2004).

Table 3.4 Results of Slope[a]

	Setting 1		Setting 2	
	VAR	AR	VAR	AR
Simulation 1				
Stock Returns	0.017502*	-0.009945	0.014064*	0.000740
Interest Rates	2.242403*	2.004940	2.877623	3.102813*
Simulation 2				
Stock Returns	0.009970*	-0.004636	0.014309*	-0.004510
Interest Rates	2.247070*	2.078880	2.826906	3.054535*

Notes: * Indicates the model with stronger sorting ability.

[a] The term of Slope is calculated based on Eq. (27): $Slope\_N(p) = \sum_{n=1}^{N} Slope(p)_n = \sum_{n=1}^{N}\left(\overline{p_{1,n}} - \overline{p_{0,n}}\right)$, which is a summation of slopes for all events. For each single event, slope is the difference between mean probabilities issued to occasions where the event obtains and mean probabilities issued to occasions where the event does not obtain. The larger the Slope is, the stronger the underlying model's overall sorting ability is.

According to Yates (1988), for a single event, the slope can be statistically obtained as the slope from regressing issued probabilities on occurrence indicators. The covariance graphs are shown in Figure 3.2 as an illustration of Yates' idea of single-event slope. The *x*-axis represents occurrence indicators, with *x*=0 meaning the event does not occur and *x*=1 meaning the event actually occurs. The *y*-axis indicates the issued probabilities for all the occasions for a specific event. The dash line represents the regression of issued probabilities on occurrence indicators. The larger the slope of the dash line is, the stronger the underlying model's sorting ability is for the corresponding single event.

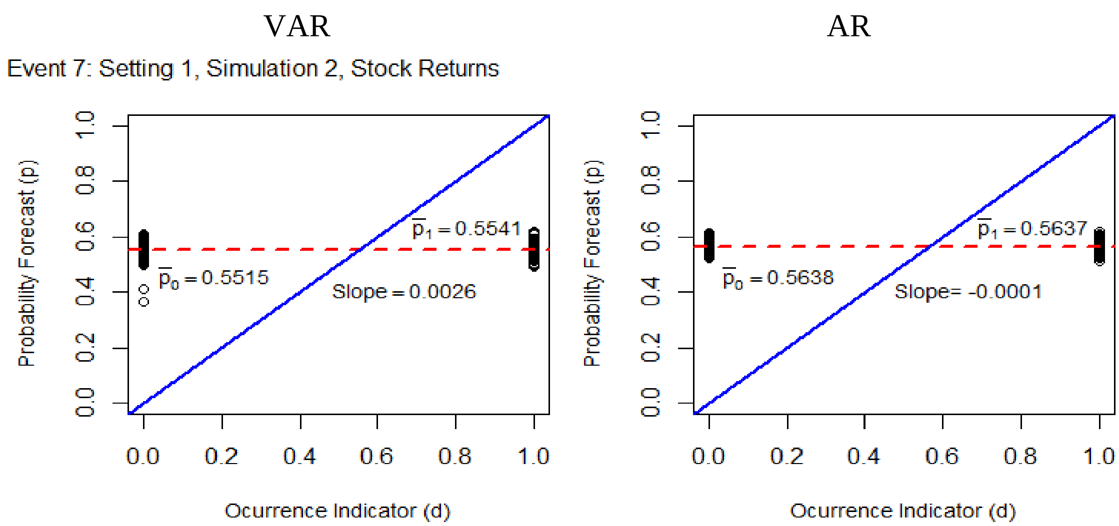


Figure 3.2 Covariance graphs of probability forecasts for the VAR and AR models.

Notes:

1) The dash line generates from the regression of issued probabilities on occurrence indicators. The solid 45-degree line illustrate the situation when the model has perfect sorting ability (issuing 0 probabilities to occasions where the event does not occur and issuing 1 probabilities to occasions where the event does occur). The slope of the dash line represents the sorting ability of the estimated model for a single event. The larger the illustrated slope is, the stronger the model's sorting ability is for that particular event.
2) For each series and each setting, only one event is selected to present the covariance graphs in the consideration of simplicity. The selected event is the event that actually occurs for the most times under T occasions.
3) Under each setting, the summation of the slopes for each event (Table 3.4) is used to evaluate the model's overall sorting ability.
4) Since the covariance graph results from simulation 1 and simulation 2 are very similar, only results using simulation 2 are illustrated here.

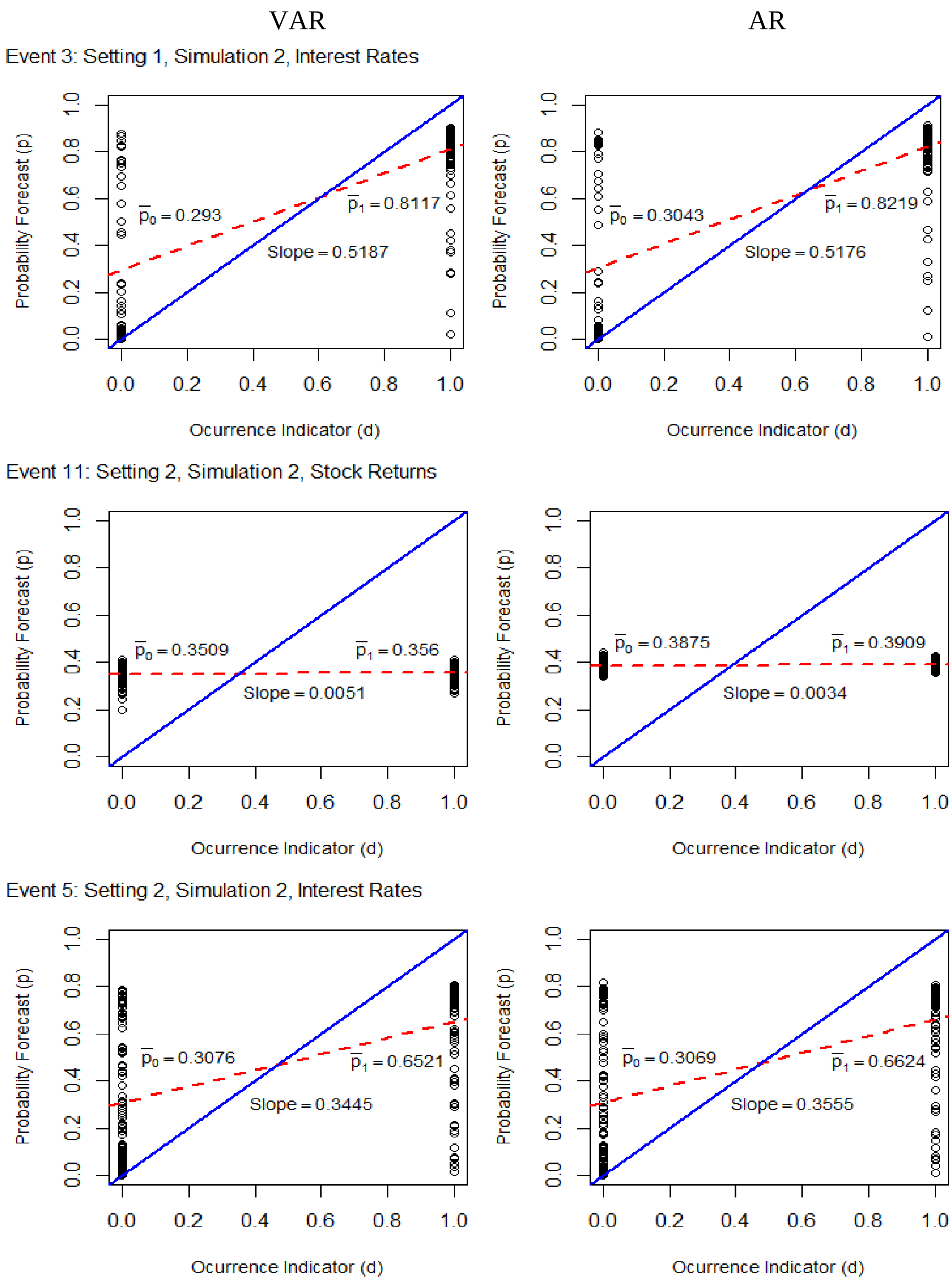


Figure 3.2 Continued

3.4.3. Overall Evaluations

The objective of this chapter is to explore whether interest rates help to forecast stock returns through both calibration (reliability) perspective and sorting (resolution) perspective. We want the dominant model to be able to issue honest forecasts (be well-calibrated) as well as make distinctions between events that obtain and events that do not obtain. Table 3.5 summarizes the models' predictive performance based on chi-squared goodness of fit test (Chi-sq), Brier score (Brier), the covariance component (2Cov) and the overall slope (Slope).

Table 3.5 Summarization of models' predictive performance

	Setting 1				Setting 2			
	Chi-sq[a]	Brier[b]	2Cov[c]	Slope[d]	Chi-sq	Brier	2Cov	Slope
Simulation 1								
Stock Returns	Yes[e]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interest Rates	Yes	Yes	Yes	Yes	Yes	Yes	No[f]	No
Simulation 2								
Stock Returns	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interest Rates	Yes	Yes	Yes	Yes	Yes	Yes	No	No

[a] Chi-sq is the Chi-squared goodness of fit test statistics.
[b] Brier is the Brier score.
[c] 2Cov is the covariance component in the decompositions of Brier score.
[d] Slope is the overall slope, obtained by summarizing slopes of all single events.
[e] The label of 'Yes' indicates the VAR model outperforms the AR model for that particular measure.
[f] The label of 'No' indicates the AR model outperforms the VAR model for that particular measure.

For stock returns, the VAR model performs better than the AR model in terms of all four measures. The robustness of this result is confirmed under both settings and both simulations. The VAR model not only issues forecasts that are better calibrated but also has a stronger sorting ability. The clear dominance of the VAR model suggests that incorporating interest rates helps in explaining the predictable pattern of stock returns in China. However, another point that cannot be overlooked is that our VAR model is still not particularly satisfying at forecasting stock returns, given the slight values of 2Cov and Slope in Table 3.2 – 3.4. According to Patelis (1997), interest rates can only partially contribute to the forecasting of stock returns. Other variables like various term structure of interest rates and dividend price ratio also have predictive power for stock returns (Campbell, 1987; Campbell and Shiller, 1988; Fama and French, 1988a; Lin et al., 2009). Therefore, more additional information should be incorporated when predicting stock returns.

Focusing on interest rates, the VAR model is superior than the AR model at all cases. The only exception is in setting 2, where the VAR model exhibits weaker sorting ability than the AR model for both simulations. Thus, the inclusion of stock returns helps to generate better calibrated forecasts of interest rates. However, there are no consistent results concerning whether it improves the model's performance in sorting.

Although the 2Cov (Slope) measure tells a different story of model' sorting ability for interest rates under different event settings, evaluations of stock return forecasts are robust to both settings as well as both simulation methods. Considering the core of this chapter is to study interest rates' role in forecasting stock returns, we do not provide deep discussion of the varying results for interest rates equations. The sensitivity of probability forecasting results to event settings is an issue needs further exploration.

3.5. Conclusion

This chapter aims to investigate whether interest rates help to explain the predictable movements of stock returns in China through prequential analysis. A two variable VAR model containing stock returns and interest rates and a univariate AR model on each series are estimated. For both model specifications, out-of-sample probability forecasts from two different simulation methods are assessed based on both calibration (reliability) and sorting (resolution) perspectives. Several data-driven methods are considered to select the number of subintervals and events when evaluating probability forecasts. We expect the dominant model to have the ability to generate honest probability forecasts (be well-calibrated) and distinguish between events that obtain and events that do not obtain.

For stock returns, the results of calibration tests suggest that the VAR model provides probability forecasts that are better calibrated relative to those from the AR model, regardless of subinterval settings and simulation methods. Thus, including interest rates in the model improves the reliability of the forecasted stock returns. Brier score and its decompositions under Yates partition are considered to further examine the two models with respect to sorting. The key element among the various decomposed components is the covariance (slope) term. As a result, the VAR model outperforms the AR model in terms of Brier score and covariance (slope) measure under both event settings and both simulation methods. Therefore, interest rates help in forecasting stock returns in China from both calibration and sorting perspectives. The significant effect of interest rates on stock returns in China is also obtained in Li (2015) and Fang et al.

(2016). Since this chapter mainly focuses on exploring whether interest rates help to forecast stock returns, we only include interest rates in the VAR model. To improve the forecasts of stock returns, it is suggested that more other predictors should be considered as well.

With respect to the interest rates, the VAR model shows better performance than the AR model in calibration, indicating stock returns help to generate realistic forecasts of interest rates (be well-calibrated). However, the results with regarding to models' sorting ability are inconsistent under different event settings. Further research is suggested to study the influence of event classifications on evaluation of probability forecasts.

CHAPTER IV

FACTOR ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES: FORECAST EVALUATIONS BASED ON FACTOR-AUGMENTED VECTOR AUTOREGRESSION (FAVAR) APPROACH

4.1. Introduction

The development of information technology has led to increasing availability of high-dimensional data sets consist of thousands of economic and financial time series. One of the main problems in high-dimensional data analysis is the "curse of dimensionality". For example, as the data dimensionality increases, the number of model parameters in parametric estimation may increase exponentially. For kernel estimation, when the dimensionality of the variable space rises, the estimator converges to its asymptotic distribution with an exponentially deteriorating rate (Li and Racine, 2007; Racine, 2008). In addition, the increase of dimensionality may cause problems such as expanding noise and complex computation (Fan et al., 2014). One way of achieving dimension-reduction is through the use of factor models, aiming to estimate a small number of factors capable of summarizing the main variation and common structure of the high-dimensional data.

The theories of classical factor models are well discussed by Lawley and Maxwell (1971), including multiple issues such as the methods of factor estimation, the selection of number of factors and the interpretation of estimated factors. Geweke (1977) and Sargent and Sims (1977) model the latent factors following a vector autoregression (VAR) process and extend the classical factor models to dynamic factor models.

In order to obtain consistent estimation of factors, principle components method is used when the time series data has fixed cross-section dimension (N) and large time dimension (T); that is $N < T$ (Lawley and Maxwell, 1971; Anderson, 1958, 1963). The method of asymptotic principal components is developed by Conner and Korajczyk (1986, 1988) to fit situations of fixed time dimension and large cross-section dimension ($N > T$). They provide asymptotic results of factor estimation in a classical factor model when $N \to \infty$. The method of asymptotic principal components is extended to dynamic factor models and demonstrated to provide consistent estimation of factors as well (Frorni and Reichlin, 1996; Forni et al., 2000, 2004, 2005).

With respect to the estimation of the number of factors, an informal but common procedure is to determine the number of factors based on the screen plot, which graphs all eigenvalues of the covariance structure of examined data in a descending order. Some literature just use a presumed number of factors (Stock and Watson, 1989; Ghysels and Ng, 1998). Bai and Ng (2002) propose a formal procedure to consistently estimate the number of factors. Their proposed criteria are developed for data with both large cross-section dimension and time dimension, and therefore are appropriate for many practical economic analyses.

One type of applications of factor analysis is to obtain a small set of estimated factors and use these factors in forecasting. Stock and Watson (1998, 2002a, 2002b) compare the forecast performances of factor-augmented regression with some benchmark models like univariate autoregression (AR) model, small vector autoregression (VAR) model and a leading indicator model. Based on an approximate dynamic factor model, the diffusion indexes (factors) are constructed from a large set of macroeconomic predictors using principal components method. The estimated factors are used to forecast a single macroeconomic time series. Their factor-augmented regression outperforms other models in terms of forecasting accuracy. It is indicated that factor-augmented regression may have improved forecast performance relative to other models incorporating no factors, since the factors are assumed to be able to capture the main useful information from a large panel of macroeconomic predictors. A similar result is found in Stock and Watson (1999), where a composite index (factor) representing aggregate activities is constructed from a large number of macroeconomic series and is employed to forecast inflation. It is shown that models including a single factor generate better forecasts than other forecasting models.

Bernanke et al. (2005) extend the factor-augmented regression to a multivariate VAR framework by incorporating estimated factors into a standard VAR analysis. The factors are identified using both principal components method and Bayesian method. The VAR framework has obtained substantial popularity and it emphasizes to rely mainly on data rather than *a priori* restrictions (Motena et al., 2013; Sims, 1980a). An important limitation of the conventional VAR methodology is that it generally just fits a small set of variables no greater than eight, which may fail to cover sufficient information contained in the massive economic activities. The factor-augmented VAR (FAVAR) approach is developed to address this issue. The FAVAR preserves the statistical efficiency of VAR estimation as having just a small number of variables. On the other hand, it also enables analysts to exploit the large amount of information summarized in a few factors.

Following the approach of Bernanke et al. (2005), Monch (2008) studies the movement of short-term interest rate and highlights the role of FAVAR approach in improving forecasting accuracy. A subsequent literature has also discussed the merits of the FAVAR model for providing reasonable empirical results in macroeconomic analysis (Boivin et al., 2009; Forni and Gambetti, 2010).

In this chapter, factors are estimated from 288 pricing time series of Chinese equities. According to the arbitrage pricing theory (APT) proposed by Ross (1976), a large number of asset returns could be modeled by a small number of factors. We attempt to search for a few factors reflecting the common forces that drive the movements and dynamics in the Chinese equity market.

The main objective and key contribution of this chapter is that two methodologies of factor estimation are applied and the two methodologies are evaluated through forecast performances. The first methodology has been extensively used in existing literature. It is to estimate factors using principal components analysis (PCA) and determine the optimal number of factors by the method of Bai and Ng (2002), which is treated as the benchmark methodology. The second methodology, recently developed by Lam and Yao (2012), is to identify both common factors and serial correlated idiosyncratic components as factors. They also propose a two-step procedure to extract factors with different degrees of strength. To the best of our knowledge, the comparison between the commonly followed PCA methodology and the newly developed methodology has rarely been discussed in the literature.

In order to evaluate these two methodologies, the factors estimated through each methodology are employed to forecast a single macroeconomic time series. Particularly, the corresponding estimated factors are fitted with Shanghai Interbank Offered Rate (Shibor) in FAVAR models respectively. Forecasting exercises are performed based on both FAVAR models. The corresponding interest rate forecasts are evaluated through a variety of statistical measures and tests. Our purpose is to assess the ability of Lam and Yao (2012)'s method, which discovers both common factors and serially correlated idiosyncratic components, in providing reasonable forecasts of the Shibor. For convenience, the factors estimated through both methodologies will be called as PCA factors and LY factors respectively in the following chapter. The corresponding interest rate forecasts generated using PCA factors are referred to as PCA forecasts, and interest rate forecasts generated using LY factors are labelled as LY forecasts.

The rest of the chapter is organized as follows. Section 2 introduces the empirical methodology of estimation. The data and a series of preliminary tests are discussed in Section 3. The empirical results including factor estimation and evaluation of forecast performances are presented in Section 4. Section 5 concludes the chapter.

4.2. Empirical Methodology

Factor models have been discussed extensively in the literature (Lawley and Maxwell, 1971; Geweke, 1977; Pena and Box, 1987; Stock and Watson, 2006; Stock and Watson, 2011). The representation of a standard factor model is as follows:

$$X_t = \Lambda F_t + \varepsilon_t, \quad t = 1, 2, \ldots, T \tag{39}$$

where X_t is a $N\times1$ vector of observed time series, F_t is a $r\times1$ vector of unobserved factors (r is the number of factors), Λ is a $N\times r$ unobserved factor loading matrix, and ε_t is a $N\times1$ vector of idiosyncratic disturbances. The observed time series are explained jointly by the common components ΛF_t and the idiosyncratic disturbances.

A factor model is aimed to achieve dimension-reduction by identifying a low-dimensional vector ($r\times1$) of latent factors F_t from a high-dimensional vector ($N\times1$) of observed data X_t. If r is much smaller than N, a dimension-reduction is realized effectively. The latent factors summarize the most useful information in the original data and can reflect underlying economic forces that cannot be described by just one or two specific variables but can be represented by a combination of multiple variables (Bernanke et al., 2005). For example, the factors can represent the major market fundamentals or the leading aggregate economic activities.

4.2.1. Method of Principal Components

The factor models can be estimated through the method of principal components or principal components analysis (PCA). According to Stock and Watson (2006), the factor loading matrix Λ and the common factors F_t are estimated by solving the minimization problem

$$\min_{F_t,\Lambda} \; T^{-1}\sum_{t=1}^{T}(X_t - \Lambda F_t)'(X_t - \Lambda F_t) \tag{40}$$

subject to the constraint of $\Lambda'\Lambda = I_r$. This is equivalent to the following maximization problem with the same constraint

$$\max_{\Lambda} \Lambda'\hat{\Sigma}_X\Lambda \tag{41}$$

where $\hat{\Sigma}_X = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$ is the $N{\times}N$ covariance matrix of X_t. As a result, the columns of the estimated factor loading matrix $\widehat{\Lambda}$ are the *r* eigenvectors of $\widehat{\Sigma}_X$ corresponding to its *r* largest eigenvalues. The common factors are estimated according to $\widehat{F_t} = \widehat{\Lambda}'X_t$. The resulting $\widehat{F_t}$ is referred to as PCA estimator of static factors. Practically, method of principal components can be performed though eigenvalue and eigenvector analysis based on the covariance or correlation structure of the high-dimensional data. It uses a few linear combinations of the original series to account for the major variations in the data.

4.2.2. Factor Estimation by Lam and Yao (2012)

According to Lam and Yao (2012), much of the previous literature on factor analysis focuses on identifying the common factors that can explain the dynamics in most of the original series. However, besides the strong common factors, there are also the idiosyncratic components, which may not generally explain most of the original series, but still account for variations in some of the original series. The exploration of these weak idiosyncratic components are often neglected. They recently developed an eigenanalysis method to estimate both the common factors and the serially correlated idiosyncratic components. They emphasize the serial dependence in data and attempt to identify factors motivating the data dependence structure. Moreover, a two-step estimation procedure is introduced to estimate factors with different strength levels. Specifically, the factors are treated as strong factors and weak factors.

The first-step is to estimate strong factors by conducting eigenanalysis on

$$\widehat{M} = \sum_{k=1}^{k_0} \hat{\Sigma}_X(k)\hat{\Sigma}_X(k)' \quad (42)$$

where $\widehat{M}$ is a $N \times N$ nonnegative definite matrix, k is the time lag, and $\hat{\Sigma}_X(k) = \frac{1}{T-k}\sum_{t=1}^{T-k} X_{t+k}X_t'$ is the $N \times N$ sample autocovariance matrix of X_t at nonzero time lag k. Since serial correlation deteriorates along with the increase of time lags, small k_0 is suggested. Lam et al. (2011) conducted simulation analysis and indicated that both the estimation of factor loading matrix and the estimation of number of factors are not sensitive to the magnitude of k_0. In this chapter, two different k_0 ($k_0 = 5$ *and* $k_0 = 2$) are considered to discover the sensitivity of the choice of k_0.

The matrix $\widehat{M}$ is one of the improvements that distinguishes the method by Lam and Yao (2012) from the method of principal components. Instead of focusing on just contemporaneous variance-covariance structure $\hat{\Sigma}_X$, this method is based on the autocovariance matrices at nonzero lags $\hat{\Sigma}_X(k)$. The summation structure of $\widehat{M}$ is to account for accumulative information from different lags. The reason for using $\hat{\Sigma}_X(k)\hat{\Sigma}_X(k)'$ is to prevent the cancellation of information at different time lags.

Assume the estimated number of strong factors is known as $\hat{r}_1$, the estimated strong factor loading matrix $\widehat{\Lambda}_1$ is constructed as the $\hat{r}_1$ eigenvectors corresponding to the $\hat{r}_1$ largest eigenvalues of $\widehat{M}$.

The second-step is to estimate weak factors. After obtaining $\widehat{\Lambda}_1$ from the first-step, the residuals denoted as X_t^* can be obtained by

$$X_t^* = X_t - \widehat{\Lambda}_1\widehat{\Lambda}_1' X_t \quad (43)$$

Based on X_t^*, the similar summation structure as (4) could be established as follows

$$\widehat{M}^* = \sum_{k=1}^{k_0} \hat{\Sigma}_{X^*}(k)\hat{\Sigma}_{X^*}(k)' \quad (44)$$

where $\hat{\Sigma}_{X^*}(k)$ is the $N\times N$ sample autocovariance matrix of X_t^* at time lag *k*.

If the estimated number of weak factors is known as $\hat{r}_2$, we can get the estimated weak factor loading matrix $\widetilde{\Lambda}_2$ as the $\hat{r}_2$ orthonormal eigenvectors of $\widehat{M}^*$ corresponding to its $\hat{r}_2$ largest eigenvalues.

Combining $\widehat{\Lambda}_1$ and $\widetilde{\Lambda}_2$, the two-step estimated factor loading matrix is as

$$\widetilde{\Lambda} = (\widehat{\Lambda}_1, \widetilde{\Lambda}_2) \quad (45)$$

The two-step estimated factors are then obtained through projecting $\widetilde{\Lambda}$ on the original data as

$$\hat{F}_t = \widetilde{\Lambda}' X_t \quad (46)$$

4.2.3. Estimation of the Number of Factors

Bai and Ng (2002) propose some panel information criteria to consistently estimate the number of factors. By treating factors as random variables and factor loadings as parameters, the problem of selecting number of factors resembles the problem of model selection. There is a trade-off between model fit and parsimony (Wang and Bessler, 2005). The model with more factors will exhibit improved model fit at the cost of less efficiency of model estimation. Therefore, it is of importance to select appropriate number of factors to maintain the balance between model fit and estimation efficiency.

In this chapter, both IC and PC criteria with different penalty functions are considered. The number of factors which leads to the minimum value of the corresponding criteria is selected. The six panel information criteria are constructed as

$$IC_{p1}(r) = \ln\left(V\left(r,\hat{F}\right)\right) + rg_1(N,T) = \ln\left(V\left(r,\hat{F}\right)\right) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$IC_{p2}(r) = \ln\left(V\left(r,\hat{F}\right)\right) + rg_2(N,T) = \ln\left(V\left(r,\hat{F}\right)\right) + r\left(\frac{N+T}{NT}\right)\ln C_{NT}^2,$$

$$IC_{p3}(r) = \ln\left(V\left(r,\hat{F}\right)\right) + rg_3(N,T) = \ln\left(V\left(r,\hat{F}\right)\right) + r\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right),$$

$$PC_{p1}(r) = V\left(r,\hat{F}\right) + r\hat{\sigma}^2 g_1(N,T) = V\left(r,\hat{F}\right) + r\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$PC_{p2}(r) = V\left(r,\hat{F}\right) + r\hat{\sigma}^2 g_2(N,T) = V\left(r,\hat{F}\right) + r\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln C_{NT}^2,$$

$$PC_{p3}(r) = V(r,\hat{F}) + r\hat{\sigma}^2 g_3(N,T) = V(r,\hat{F}) + r\hat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) \quad (47)$$

where $V(r,\hat{F}) = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \Lambda'\hat{F}_t\right)^2$ is the scaled sum of squared residuals when there are r factors, $\hat{\sigma}^2 = V\left(r_{max},\hat{F}_{max}\right)$ is estimated at the upper bound $r = r_{max}$[14], and $C_{NT}^2 = \min(N,T)$ selects the smaller dimension of original time series. $g_i(N,T)$, i = 1, 2

[14] Following Bai and Ng (2002), $8 * int[(min\{N,T\}/100)^{1/4}]$ is used as a rule to choose r_{max}, which is similar as the rule suggested by Schwert (2002).

and 3, stands for the different formulations of penalty functions which varies along the changes of both *N* and *T*. The penalty functions find their minimum values at certain rate indicating that the selected model is neither over-fitted nor under-fitted.

Another method of estimating the number of factors is suggested by Lam and Yao (2012), denoted as the ratio-based estimator of r. The basic idea is to find a cut-off value $\hat{r}$ so that the subsequent eigenvalues are considerably smaller than the $\hat{r}$th eigenvalue when ordering all eigenvalues from the largest to the smallest. The ratio-based estimator is defined as follows

$$\hat{r} = \underset{1 \le i \le R}{\operatorname{argmin}} \hat{\lambda}_{i+1} / \hat{\lambda}_i \tag{48}$$

where $\hat{\lambda}_i$ are the eigenvalues of structures as (42) or (44) and are sorted in descending order. Following Lam and Yao (2002), the choice of R is determined by $R = N/2$.

4.2.4. Factor-Augmented Vector Autoregression (FAVAR) Model

Following Bernanke et al. (2005), the FAVAR model is given by

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + u_t \tag{49}$$

where F_t is the vector of unobserved factors, Y_t represents the vector of observable series, $\Phi(\mathrm{L})$ is a conformable lag polynomial with finite order *p*, and u_t is the error term. Equation (11) describes the joint dynamics of F_t and Y_t. If the coefficients of factor terms in Y_t equation are not all zero, then equation (11) represents the FAVAR. Otherwise, it becomes a standard VAR of Y_t. In this chapter, Y_t only includes one variable, which is the interest rate (Shibor).

Since the factors F_t are unobserved, equation (49) cannot be estimated directly. Instead, we need to estimate F_t based on equation (39). After obtaining the estimated factors $\widehat{F}_t$, F_t can be substituted by $\widehat{F}_t$ and equation (49) can be estimated through standard VAR approach.

4.3. Data Description

The factor estimation is applied to the major securities in Chinese equity market with the expectation that the extracted factors will summarize the main information of market dynamics. The securities selected in this chapter are the 300 securities used to develop the CSI 300 index by China Securities Index Company (CSI). All of the 300 securities are selected from the two stock exchanges in mainland China (Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE)). They are actively traded and account for approximately 60 percent of the total market capitalization of the two exchanges. The 300 securities cover all 10 sector classifications, as given by the Global Industry Classification Standard (GICS) developed by MSCI and Standard & Poor's. The 10 sectors are Energy, Financials, Industrials, Materials, Utilities, Consumer Discretionary, Consumer Staples, Health Care, Information Technology, and Telecom Services. The 300 securities are believed to well represent the general dynamics and movements in Chinese equity market. The interest rate used in this chapter is the Shibor, which is targeted to be the potential benchmark interest rate in China (Gang, 2009).

Daily closing prices of the selected securities are obtained from the Datastream database and the overnight Shibor is collected from its official website[15]. Due to data availability issue, 288 out of the 300 securities are used in our study for further estimation. The time period of the original data is from Jun 11, 2012 to May 11, 2015, yielding 761 time observations.

A series of data transformations are applied to the data before performing factor estimation. First, the 288 security series are transformed to logarithms. For interest rate, no logarithm transformation is applied since it is already in percentage (Stock and Watson, 2002b). Second, the stationarity of the data series are examined through two unit root tests (the augmented Dickey-Fuller (ADF) test by Dickey and Fuller (1979) and the Phillips–Perron test by Phillips and Perron (1988)). All the security series are nonstationary in level, while the interest rate is stationary in level. Accordingly, the 288 security series are transformed through first differencing and the resulting series are all stationary at significance level of 1%. Third, following Stock and Watson (2002a, 2002b, 2005), the security series are standardized to have sample mean zero and sample variance one after the first two steps of transformation.

4.4. Empirical Results

4.4.1. Optimal Number of Factors by Bai and Ng (2002)

For the first methodology, six criteria developed by Bai and Ng (2002) are used to select the optimal number of factors. Corresponding results are shown in Table 4.1.

[15] The overnight Shibor is collected from http://www.shibor.org/shibor/web/html/index_e.html.

Consistent with many previous studies, no consistent results concerning the optimal number of factors are suggested by different criteria. Criteria IC_{p1}, IC_{p2} and PC_{p2} all agree with 5 factors, PC_{p1} indicates 6 factors and IC_{p3} and PC_{p3} suggest 8 factors. Comparing to the *PC* criteria, the three types of *IC* information criteria have the advantage of not depending on the choice of r_{max}. According to Bai and Ng (2002), their simulation results suggest that all the six criteria could provide precise estimates of the number of factors when $min\{N, T\}$ is 40 or larger. Particularly, criteria IC_{p1}, IC_{p2}, PC_{p1} and PC_{p2} perform better than other criteria in Bai and Ng's work. Therefore, the optimal number of factors is selected as 5 using criteria IC_{p1} and IC_{p2} in this chapter.

The relative importance of the factors are explained by the proportion and cumulative proportion of factors in Table 4.2. The first factor plays a leading role since it accounts for 29.97% of the total variability in data, which is much larger than the contribution of other factors in explaining the data variations. The remaining factors after the 5th factor have very limited contributions. The cumulative proportion of the total variance in data explained by the first 5 factors is 43.83%, which supports the choice of optimal number of factors as 5.

Table 4.1 Optimal number of factors

Number of Factors	IC_{p1}	IC_{p2}	IC_{p3}	PC_{p1}	PC_{p2}	PC_{p3}
1	-0.3319	-0.3304	-0.3379	0.7128	0.7136	0.7097
2	-0.4111	-0.4080	-0.4229	0.6566	0.6582	0.6504
3	-0.4368	-0.4322	-0.4545	0.6385	0.6409	0.6292
4	-0.4447	-0.4386	-0.4684	0.6322	0.6354	0.6198
5	-0.4502[a]	-0.4425 [a]	-0.4798	0.6278	0.6319 [a]	0.6124
6	-0.4500	-0.4407	-0.4854	0.6272 [a]	0.6320	0.6086
7	-0.4481	-0.4373	-0.4895	0.6277	0.6334	0.6061
8	-0.4437	-0.4314	-0.4910 [a]	0.6299	0.6364	0.6052 [a]

Notes: IC_{pi} and PC_{pi} refer to the six information criteria in (9) estimated with different penalty function $g_i(N, T)$. The r_{max} is calculated according to $r_{max} = 8 * int[(min\{N, T\}/100)^{1/4}] = 8$ in this chapter.
[a] The optimal number of factors.

Table 4.2 Proportion and cumulative proportion of factors

Number of Factors	Proportion of Factors	Cumulative Proportion of Factors
1	0.2997	0.2997
2	0.0696	0.3693
3	0.0315	0.4008
4	0.0197	0.4206
5	0.0177	0.4383
6	0.0140	0.4523
7	0.0128	0.4652
8	0.0112	0.4764

Notes: The proportion of factors is calculated based on $PF_i = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_{288}}$, and the cumulative proportion of factors is calculated based on $CPF_i = \frac{\lambda_1 + \cdots + \lambda_i}{\lambda_1 + \cdots + \lambda_{288}}$, where λ_i is the *i*th largest eigenvalue and *i=1, ..., 288*.

4.4.2. Estimation of Factors by the Method of Principal Components

Based on the results from Table 4.1, 5 factors are estimated using the method of principal components. To better interpret the factors, simple linear regressions of each security series on each factor are performed following Stock and Watson (2002b). As a result, there are 1440 ($288 * 5$) simple regressions and the R^2 values of each regression are studied. For each factor, bar charts are plotted with the R^2 values as y-axis and the

288 securities as x-axis (The 288 securities are classified into 10 sectors following the GICS and ordered according to the ordering listed in the Appendix B). Therefore, each bar chart illustrates each factor's capability in explaining the variations of the 288 securities.

Figure 4.1 shows the R^2 values associated with each of the 5 factors estimated by the method of principal components. The first PCA factor captures fluctuations of most securities across all sectors with high R^2 values, indicating that the first factor may represent the general movement of the stock market. Accordingly, the first factor can be treated as a market factor. The other four PCA factors are more like sector factors, focusing on explaining the variations in several sectors. The second PCA factor has relatively stronger explaining power in sectors of Financials, Consumer Discretionary, Health Care, Telecom Services and Information Technology. The third PCA factor loads more on Financials, Industrials and Materials sectors, with relatively higher R^2 values comparing to other sectors. The fourth PCA factor reflects more variations in Energy, Financials, Materials and Utilities sectors. The fifth PCA factor focuses on sectors of Consumer Staples and Health Care.

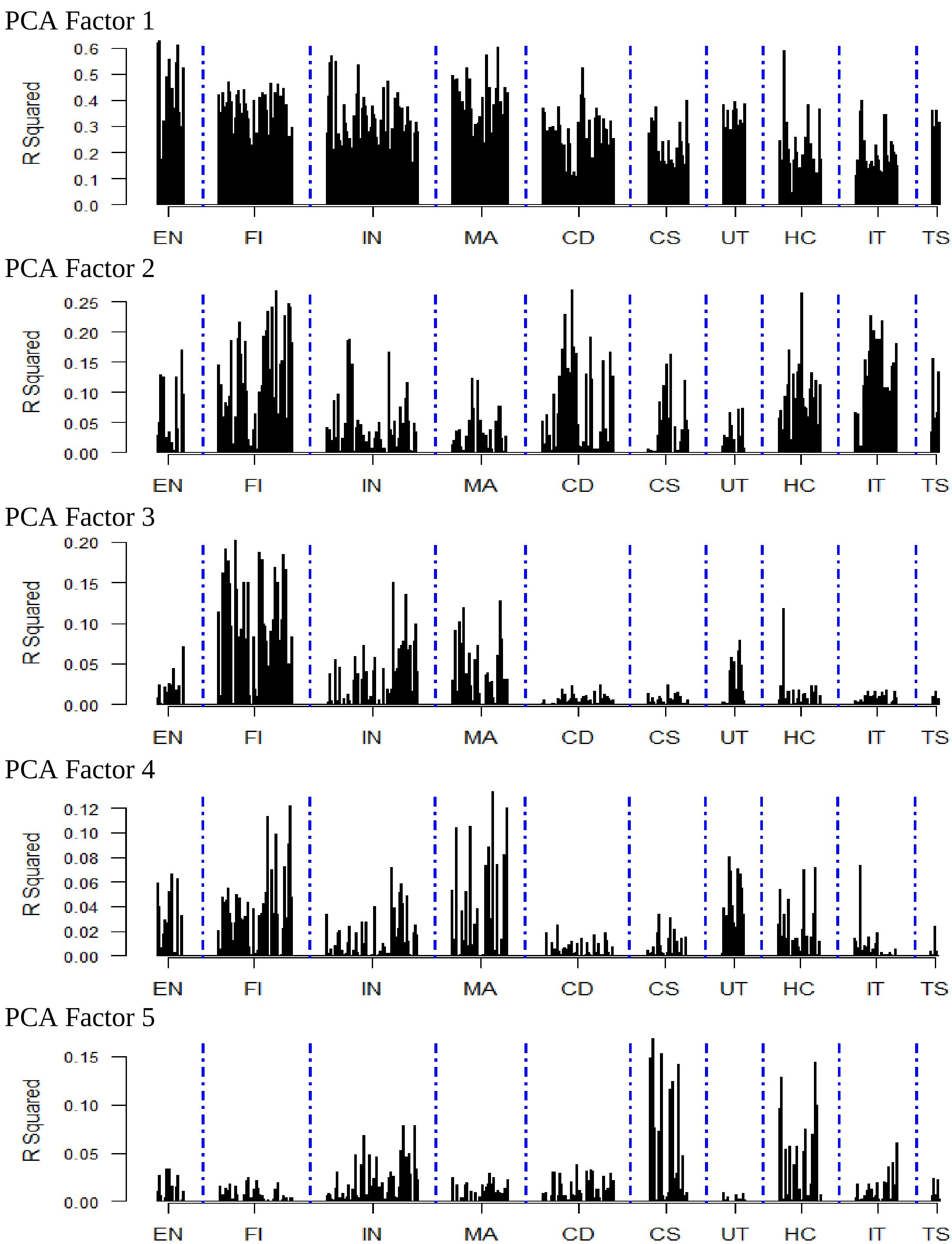


Figure 4.1 R^2 values for 5 PCA factors [a]

[a] The labels in x-axis stand for the sectors. “EN” denotes Energy sector, “FI” denotes Financials sector, “IN” denotes Industrials sector, “MA” denotes Materials sector, “CD” denotes Consumer Discretionary sector, “CS” denotes Consumer Staples sector, “UT” denotes Utilities sector, “HC” denotes Health Care sector, “IT” denotes Information Technology sector, and “TS” denotes Telecom Services sector.

4.4.3. Optimal Number of Factors by Lam and Yao (2012) ($k_0 = 5$)

Following Lam and Yao (2012), $k_0 = 5$ is chose to estimate $\widehat{M}$. The estimated eigenvalues of $\widehat{M}$ are calculated in order to form the ratio-based estimator. The results are illustrated in Figure 4.2 (Panel A). There is a remarkable decrease of the eigenvalue from *i*=1 to *i*=2. The remaining eigenvalues, especially those at *i*>=8, are all quite small with approximately the same size. The ratio-based estimator for the number of strong factors are plotted in Figure 4.2 (Panel B). According to equation (10), the ratio-based estimator in Figure 4.2 (Panel B) is minimized at *i*=1, indicating the sharpest decrease in proportion when moving from the first eigenvalue to the second eigenvalue. Thus, the optimal number of strong factor is 1. Since the ratio-based estimator gets to its second U-shape at *i*=3, there may be several weak factors that need to be explored from the idiosyncratic components.

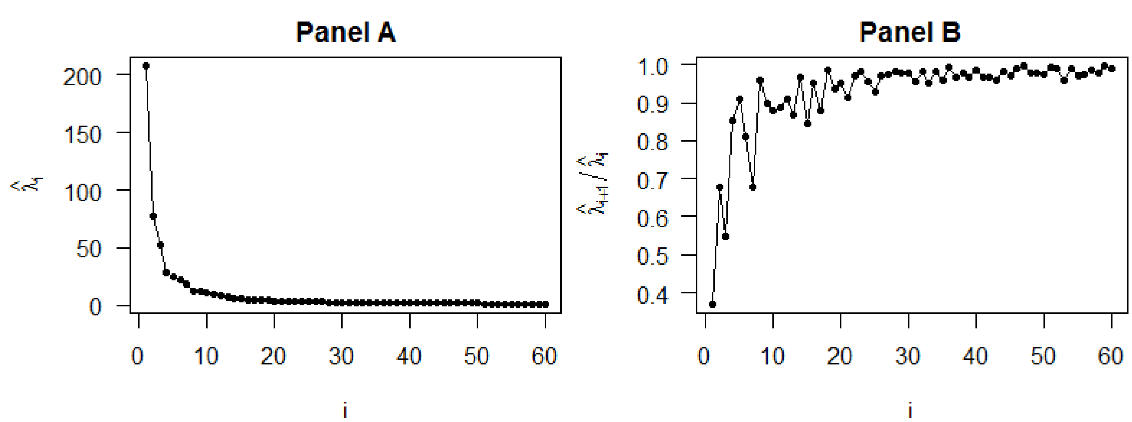


Figure 4.2 Plot of the estimated eigenvalues of $\widehat{M}$ (Panel A) and plot of the ratio-based estimator of the number of strong factors (Panel B) ($k_0 = 5$)[a].

[a] Only the first 60 eigenvalues and ratio-based estimators are plotted respectively, where $\hat{\lambda}_i$ denotes the *i*th largest estimated eigenvalue, $i = 1, \dots, 60$.

After removing the one strong factor, the eigenanalysis is conducted on $\widehat{M}^*$ to obtain the corresponding estimated eigenvalues and the ratio-based estimator for the number of weak factors. The results are plotted in Figure 4.3 (Panel A) and Figure 4.3 (Panel B) respectively. The eigenvalue of $\widehat{M}^*$ decreases strongly to around 10 when *i*=3. The following eigenvalues all diminish at a slight rate and have a relatively small size. In Figure 4.3 (Panel B), the ratio-based estimator based on $\widehat{M}^*$ is minimized at *i*=2, implying there are 2 weak factors.

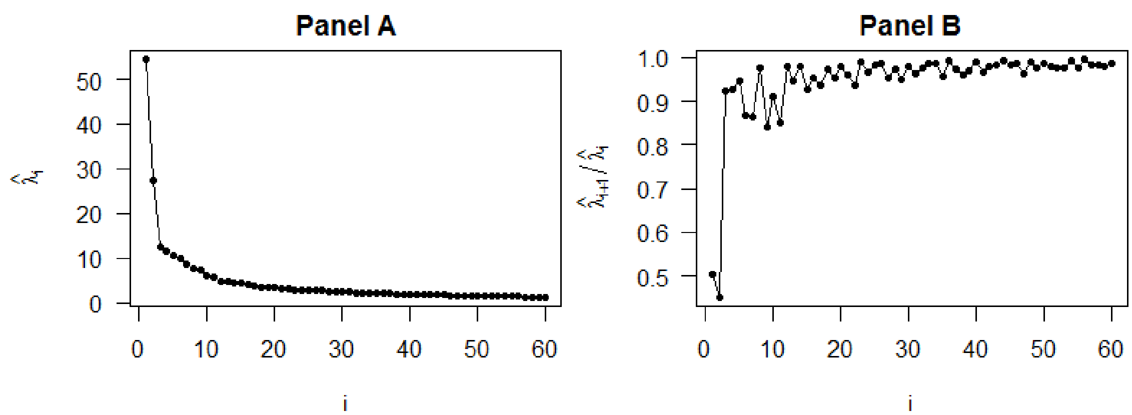


Figure 4.3 Plot of the estimated eigenvalues of $\widehat{M}^*$ (Panel A) and plot of the ratio-based estimator of the number of weak factors (Panel B) ($k_0 = 5$)[a].

[a] Only the first 60 eigenvalues and ratio-based estimators are plotted respectively, where $\hat{\lambda}_i$ denotes the *i*th largest estimated eigenvalue, $i = 1, \dots, 60$..

When $k_0 = 5$, 1 strong factor and 2 weak factors are found through the two-step procedure, leading to 3 factors in total. The first eigenvector of $\widehat{M}$, corresponds to its largest eigenvalue and the first two eigenvectors of $\widehat{M}^*$, correspond to its two largest eigenvalues, are combined to form the factor loading matrix $\widetilde{\Lambda}$.

4.4.4. Estimation of Factors by Lam and Yao (2012) ($k_0 = 5$)

Based on the estimated factor loading matrix $\widetilde{\Lambda}$, the 3 LY factors under $k_0 = 5$ are estimated through equation (8). The bar plots of R^2 values for the 3 LY factors are presented as Figure 4.4. The first LY factor ($k_0 = 5$) tends to be a market factor associated with high R^2 values for all sectors. The second LY factor ($k_0 = 5$) has relatively high R^2 values in sectors of Industrials, Consumer Discretionary, Consumer Staples, Health Care, Information Technology and Telecom Services. The third LY factor ($k_0 = 5$) explains more variations in sectors of Energy, Industrials, Materials and Utilities.

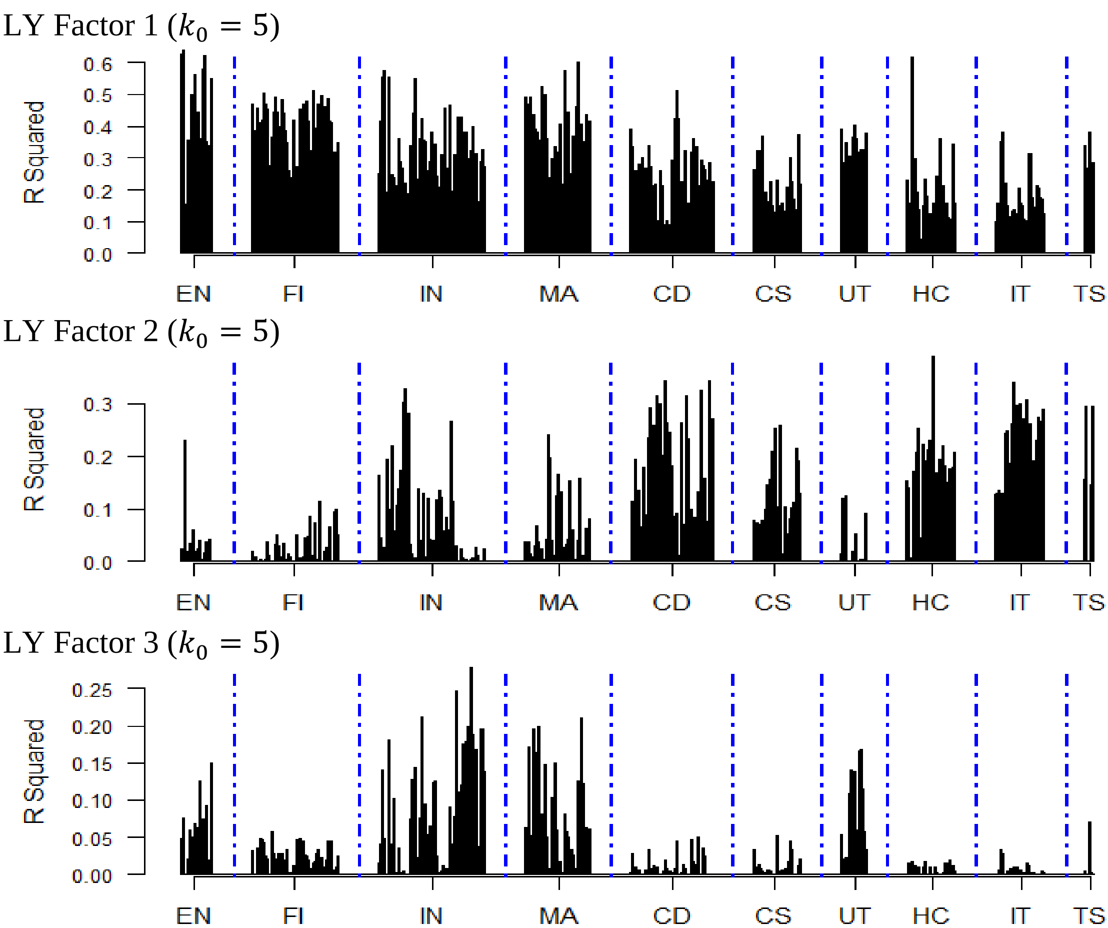


Figure 4.4 R^2 values for 3 LY factors (k_0 = 5)[a]

[a] The labels in x-axis stand for the sectors. "EN" denotes Energy sector, "FI" denotes Financials sector, "IN" denotes Industrials sector, "MA" denotes Materials sector, "CD" denotes Consumer Discretionary sector, "CS" denotes Consumer Staples sector, "UT" denotes Utilities sector, "HC" denotes Health Care sector, "IT" denotes Information Technology sector, and "TS" denotes Telecom Services sector.

4.4.5. Optimal Number of Factors by Lam and Yao (2012) (k_0 = 2)

One of the motivations of Lam and Yao (2012)'s method is to explore factors promoting the serial dependence of the data. This may provide some guidance of the choice of k_0. Although the simulation results by Lam et al. (2011) suggest that estimations of factor loading matrix and number of factors are insensitive to the choice of k_0, we still want to examine whether there is a different value of k_0 driven by the substantial serial dependence in data. To detect the autocorrelation of the data, all the 288 time series are fitted into autoregressive integrated moving average (ARIMA) model. The results using both BIC and AIC criteria are shown in Table 4.3. Based on BIC, 278 out of 288 series exhibit up to 2 autoregression (AR) orders. When using AIC, 262 out of 288 series show up to 2 AR order. The results indicate that all series show no serial correlations or low order serial correlations (no more than 5). As a result, we also perform the method by Lam and Yao (2012) using $k_0 = 2$ since most of the serial correlations in the data could be captured at AR order of 2.

Table 4.3 Summary of selection of AR order on univariate representation on each series

AR order	BIC	AIC
0	239	175
1	23	36
2	16	51
3	2	14
4	7	9
5	1	3

Under $k_0 = 2$, the estimated eigenvalues of $\widehat{M}$ and the ratio-based estimator for the number of strong factors are plotted in Figure 4.5 (Panel A) and Figure 4.5 (Panel B). According to Figure 4.5 (Panel B), 1 strong factor is selected since the ratio-based estimator is minimized at *i*=1.

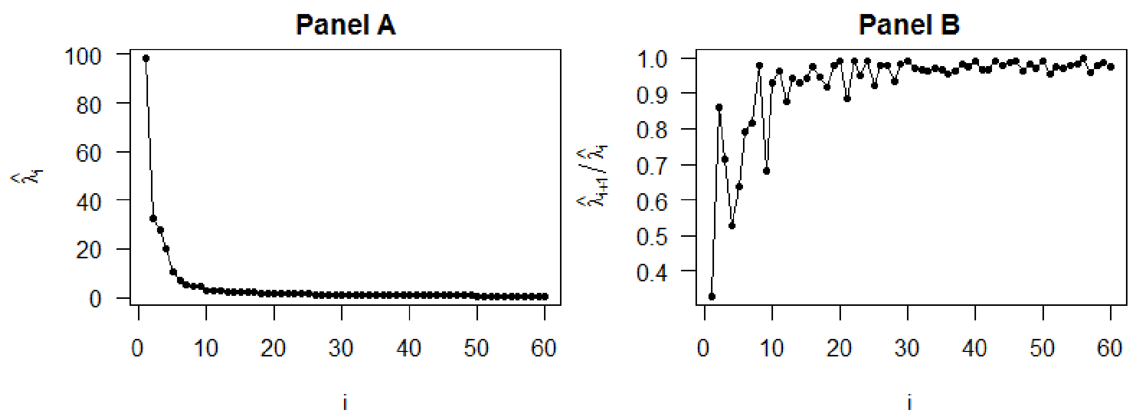


Figure 4.5 Plot of the estimated eigenvalues of $\widehat{M}$ (Panel A) and plot of the ratio-based estimator of the number of strong factors (Panel B) ($k_0 = 2$)[a].

[a] Only the first 60 eigenvalues and ratio-based estimators are plotted respectively, where $\hat{\lambda}_i$ denotes the *i*th largest estimated eigenvalue, $i = 1, \dots, 60$.

After finding 1 strong factor, we continue to examine the number of weak factors. Figure 4.6 (Panel A and Panel B) exhibit the results of estimated eigenvalues of $\widehat{M}^*$and the ratio-based estimator for the number of weak factors. In Figure 4.6 (Panel B), the ratio-based estimator based on $\widehat{M}^*$ is minimized at *i*=1, indicating 1 weak factor under $k_0 = 2$.

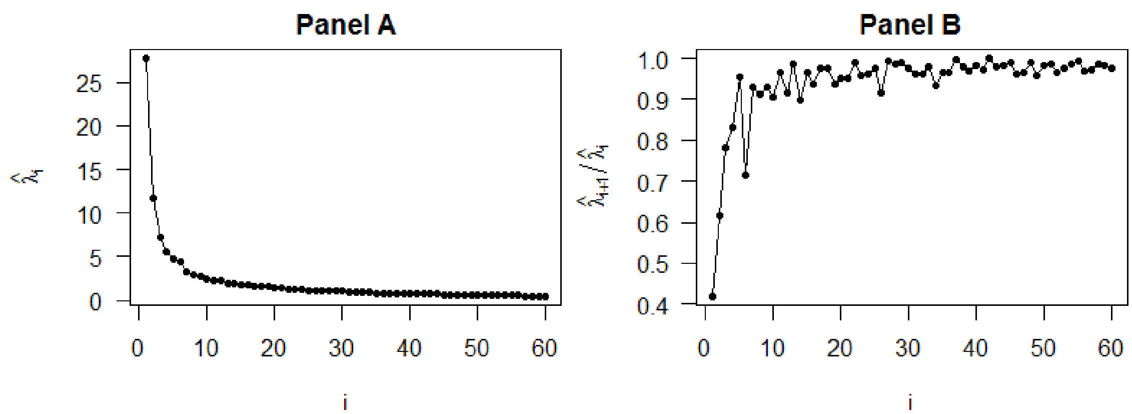


Figure 4.6 Plot of the estimated eigenvalues of $\widehat{M}^*$ (Panel A) and plot of the ratio-based estimator of the number of weak factors (Panel B) ($k_0 = 2$)[a].

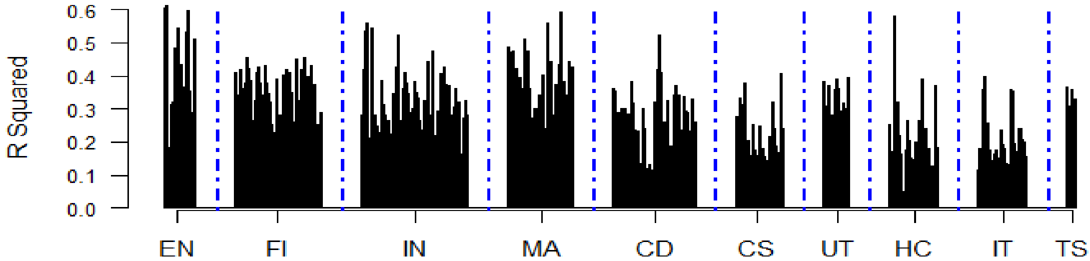
[a] Only the first 60 eigenvalues and ratio-based estimators are plotted respectively, where $\hat{\lambda}_i$ denotes the ith largest estimated eigenvalue, $i = 1, \dots, 60$.

In sum, when $k_0 = 2$, there are 1 strong factor and 1 weak factor, leading to 2 factors in total. Inconsistent to the findings of Lam et al. (2011) and Lam and Yao (2012), the estimation of the number of factors is sensitive to the choice of k_0 in our case.

4.4.6. Estimation of Factors by Lam and Yao (2012) ($k_0 = 2$)

Under $k_0 = 2$, the 2 LY factors are estimated through equation (8). Figure 4.7 presents the bar plots of R^2 values for each of the 2 LY factors. The LY factor 1 ($k_0 = 2$) is a market factor with high R^2 values in all sectors. The second LY factor ($k_0 = 2$) loads more on sectors of Financials, Consumer Discretionary, Health Care and Information Technology.

LY Factor 1 ($k_0 = 2$)



LY Factor 2 ($k_0 = 2$)

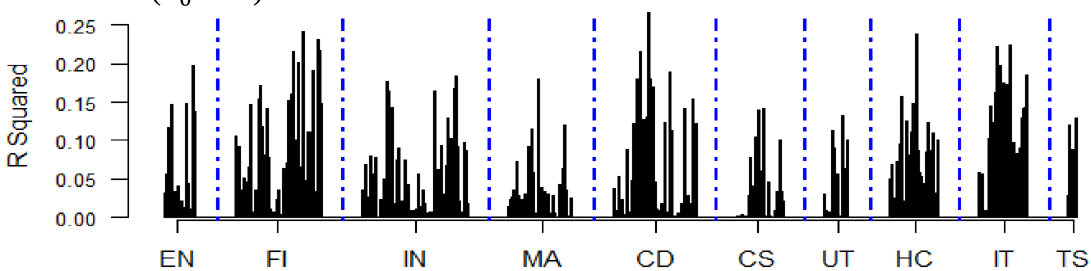


Figure 4.7 R^2 values for 2 LY factors ($k_0 = 2$)[a]

[a] The labels in x-axis stand for the sectors. "EN" denotes Energy sector, "FI" denotes Financials sector, "IN" denotes Industrials sector, "MA" denotes Materials sector, "CD" denotes Consumer Discretionary sector, "CS" denotes Consumer Staples sector, "UT" denotes Utilities sector, "HC" denotes Health Care sector, "IT" denotes Information Technology sector, and "TS" denotes Telecom Services sector.

4.4.7. FAVAR and Evaluation of Forecast Performances[16]

The PCA factors and LY factors are fitted with interest rate in FAVAR models respectively. The first FAVAR model includes 6 variables (1 interest rate variable plus 5 PCA factors), the second FAVAR model has 4 variables (1 interest rate variable plus 3 LY factors under $k_0 = 5$), and the third FAVAR model has 3 variables (1 interest rate variable plus 2 LY factors under $k_0 = 2$). Schwarz-loss and Hannan and Quinn-loss metrics are used to determine the number of lag length. Optimal lag length of 1 is

[16] The economic returns from forecasting have been highlighted in many studies (Granger and Newbold, 1973; Granger, 1992; Brandt and Bessler, 1983). A set of forecasting techniques, such as using various forecasting models, considering non-linear models with a switching regime and applying disaggregated data are expected to provide merits of economic analysis.

selected for all models. For each FAVAR model, out of sample forecasts of interest rate are recursively generated for one-, two-, three-, five-, ten- and fifteen-step-ahead horizons. As a result, there are 100 one-step-ahead forecasts, 99 two-step-ahead forecasts, 98 three-step-ahead forecasts, 96 five-step-ahead forecasts, 91 ten-step-ahead forecasts, and 86 fifteen-step-ahead forecasts.

4.4.7.1 Statistical Measures of Forecast Performances

A variety of statistical measures are applied in this chapter to evaluate forecast accuracy. The root mean squared error (RMSE) measures the average value of forecast error using a quadratic scoring rule. The mean absolute error (MAE) is equal to the average absolute value of forecast error. The mean absolute percentage error (MAPE) calculates the average absolute percentage of forecast error in actual value. RMSE is a commonly used measure of forecast accuracy. Comparing to RMSE, MAE is less sensitive to outliers and MAPE does not rely on the scale of the data (Hyndman and Koehler, 2006).

The results of the three statistic measures for PCA forecasts and LY forecasts under $k_0 = 5$ are listed in Table 4.4. Forecasts using LY factors ($k_0 = 5$) exhibit smaller RMSE for the one- to three-step-ahead forecasts. If measuring in MAE, the LY forecasts ($k_0 = 5$) is more accurate than PCA forecasts at all horizons except for fifteen-step-ahead horizon. With respect to MAPE, forecasts using LY factors ($k_0 = 5$) exhibit lower value for all horizons except ten- and fifteen-step-ahead horizons. Overall, LY factors under $k_0 = 5$ forecast interest rate better than PCA factors, especially at the short horizons (one, two and three steps).

Table 4.4 Results of statistical measures for PCA forecasts and LY forecasts ($k_0 = 5$)[17]

	RMSE		MAE		MAPE	
Steps	PCA	LY ($k_0 = 5$)	PCA	LY ($k_0 = 5$)	PCA	LY ($k_0 = 5$)
One	0.1142	0.1034 [a]	0.0881	0.0787 [a]	3.6676	3.3826 [a]
Two	0.1875	0.1819 [a]	0.1455	0.1400 [a]	6.2848	6.1017 [a]
Three	0.2508	0.2482 [a]	0.1975	0.1938 [a]	8.5928	8.4962 [a]
Five	0.3520 [a]	0.3530	0.2766	0.2746 [a]	12.2618	12.2416 [a]
Ten	0.5234 [a]	0.5252	0.4077	0.4072 [a]	18.6358 [a]	18.6443
Fifteen	0.6199 [a]	0.6214	0.4828 [a]	0.4840	22.3892 [a]	22.4442

Notes: RMSE = $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - x_t^f)^2}$, MAE = $\frac{1}{T}\sum_{t=1}^{T}\left|x_t - x_t^f\right|$ MAPE = $\frac{100}{T}\sum_{t=1}^{T}\left|\frac{x_t - x_t^f}{x_t}\right|$, where x_t is the actual value of interest rate, x_t^f is the forecasted value of interest rate. All the MAPE values are in percentage units.

[a] The smaller statistical measure.

In order to further explore whether the choice of k_0 will affect the prediction ability of the estimated LY factors, the LY factors estimated under $k_0 = 2$ are also compared with PCA factors using the three statistical measures (Table 4.5). LY factors under $k_0 = 2$ have the same prediction ability as LY factors under $k_0 = 5$ when comparing to PCA factors. For example, LY forecasts ($k_0 = 2$) have lower RMSE than PCA forecast for one-, two- and three-step-ahead horizons. In terms of MAE, LY forecasts ($k_0 = 2$) perform better than PCA forecasts at all steps except fifteen steps.

[17] The three statistical measures are also calculated for forecasts generated from a random walk model of interest rate. Accordingly, the forecasts from the random walk model outperform both PCA and LY forecasts for one, two, three, five and ten steps. For the fifteen-step-ahead horizon, the two groups of factor forecasts perform better than the forecasts from the random walk. The same results are also obtained as using $k_0 = 2$.

However, when focusing on only LY forecasts under $k_0 = 2$ and LY forecasts under $k_0 = 5$, we find that the former have lower value than the latter for all three statistical measures and for most of the horizons. This corresponds with findings of prior research that complex models usually forecast not as well as simple models (Armstrong, 1984; Brandt and Bessler, 1984; Zellner, 2001). The FAVAR model fitted with the 5 LY factors under $k_0 = 5$ is more complicated than that fitted with just 2 LY factors since it contains more regressors.

Table 4.5 Results of statistical measures for PCA forecasts and LY forecasts ($k_0 = 2$)

	RMSE		MAE		MAPE	
Steps	PCA	LY ($k_0 = 2$)	PCA	LY ($k_0 = 2$)	PCA	LY ($k_0 = 2$)
One	0.1142	0.1026 [a]	0.0881	0.0783 [a]	3.6676	3.3625 [a]
Two	0.1875	0.1814 [a]	0.1455	0.1398 [a]	6.2848	6.0905 [a]
Three	0.2508	0.2477 [a]	0.1975	0.1936 [a]	8.5928	8.4865 [a]
Five	0.3520 [a]	0.3528	0.2766	0.2748 [a]	12.2618	12.2439 [a]
Ten	0.5234 [a]	0.5250	0.4077	0.4071 [a]	18.6358 [a]	18.6391
Fifteen	0.6199 [a]	0.6213	0.4828 [a]	0.4840	22.3892 [a]	22.4427

Notes: RMSE $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - x_t^f)^2}$, MAE $= \frac{1}{T}\sum_{t=1}^{T}\left|x_t - x_t^f\right|$ MAPE $= \frac{100}{T}\sum_{t=1}^{T}\left|\frac{x_t - x_t^f}{x_t}\right|$, where x_t is the actual value of interest rate, x_t^f is the forecasted value of interest rate. All the MAPE values are in percentage units.

[a] The smaller statistical measure.

Therefore, it is shown that the forecast performance of LY factors relative to PCA factors is not sensitive to the choice of k_0. However, the choice of k_0 may influence the prediction ability of different groups of LY factors. In our case, $k_0 = 2$ yeilds more accurate LY forecasts than $k_0 = 5$. Since the focus of this chapter is to make comparison between PCA factors and LY factors, we won't pay too much attention to the within-group differences of LY factors under different values of k_0. In addition, the result that LY factors have higher prediction accuracy than PCA factors at the short horizons is consistent with the findings in Lam et al. (2011), which only focuses on one-step-ahead forecasts. In this chapter, the long-run out-of-sample forecasts are examined as well and it is indicated that PCA factors show better forecast performances than LY factors mostly at the long horizons in terms of the three statistical measures. One possible explanation for the better performance of LY forecasts only at the short horizons is that the autocorrelations in our data series are of short order (Table 4.3) and the method by Lam and Yao (2012) is able to capture the serial correlations in the data.

4.4.7.2 DM Equality Test

The DM equality test developed by Diebold and Mariano (2012) is used to test forecast accuracy. If the forecast errors generated from PCA factors and LY factors are denoted as $e_{PCA,t}$ and $e_{LY,t}$, the forecast accuracy could be measured by certain loss functions $g(e_{PCA,t})$ and $g(e_{LY,t})$. Typically, the loss functions can take the form of squared error loss as equation (50) or the absolute error loss as equation (51). In this chapter, the loss functions are calculated based on squared error loss.

$$g\left(e_{PCA,t}\right) = e_{PCA,t}^2, \quad g\left(e_{LY,t}\right) = e_{LY,t}^2 \tag{50}$$

$$g\left(e_{PCA,t}\right) = \left|e_{PCA,t}\right|, \quad g\left(e_{LY,t}\right) = \left|e_{LY,t}\right| \tag{51}$$

The null hypothesis of DM equality test is

$$H_0: E[d_t] = 0 \tag{52}$$

where $d_t = g(e_{PCA,t}) - g(e_{LY,t})$. The null hypothesis means the forecasts errors using PCA factors and LY factors are not statistically different. The test statistics of DM test is

$$\text{DM} = \frac{\bar{d}}{(\hat{V}_{\bar{d}}/T)^{1/2}} \quad (53)$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$ is the sample mean of d_t, and $\hat{V}_{\bar{d}}$ is the sample variance of $\sqrt{T}\bar{d}$. $V_{\bar{d}}$ can be represented as$V_{\bar{d}} = \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k$, where $\gamma_k = cov(d_t, d_{t-k})$ is the *k*th autocovariance of d_t. It is shown that under the null hypothesis,

$$DM \overset{Asymptotic}{\sim} N(0,1) \quad (54)$$

and we estimate $\hat{V}_{\bar{d}}$ by $\hat{V}_{\bar{d}} = \hat{\gamma}_0 + 2\sum_{l=1}^{\infty} \hat{\gamma}_k$.

The results of DM equality test for PCA forecasts and LY forecasts ($k_0 = 5$) are listed in Table 4.6. Two alternatives of the test are considered. For alternative 1, the null hypothesis of equality is rejected at 5% significance level for one-step-ahead forecasts. Thus, the forecast accuracy of one-step-ahead forecasts using PCA factors and LY factors ($k_0 = 5$) are different. However, for other steps, the two sets of forecasts are not statistically different. In order to further check which factors yield higher forecast accuracy, alternative 2 is applied. Alternative 2 indicates that LY forecasts ($k_0 = 5$) have higher predictive accuracy. As a result, for one-step-ahead forecasts, the null hypothesis is rejected at 1% significance level and for two-step-ahead forecasts, the null hypothesis is rejected at 10% significance level. Thus, LY factors ($k_0 = 5$) outperform PCA factors in forecasting for the first two steps. The same test is performed for PCA forecasts and LY forecasts ($k_0 = 2$) as well (Table 4.7). The rejection results are exactly the same for each alternative and each horizon when using the two different sets of LY forecasts ($k_0 = 5$ and $k_0 = 2$).

Table 4.6 DM equality test ($k_0 = 5$)

	Alternative 1[a]		Alternative 2[b]	
Steps	DM statistics	P-value	DM statistics	P-value
One	2.3713	0.0197**	2.3713	0.0098***
Two	1.4103	0.1616	1.4103	0.0808*
Three	0.7909	0.4309	0.7909	0.2155
Five	-0.2942	0.7692	-0.2942	0.6154
Ten	-1.2202	0.2256	-1.2202	0.8872
Fifteen	-1.0727	0.2864	-1.0727	0.8568

Notes: *, ** and *** denotes rejection of the null hypothesis at 10%, 5% and 1% significance levels.
[a] Alternative hypothesis 1 is forecasts using PCA factors and LY factors ($k_0 = 5$) have different levels of accuracy.
[b] Alternative hypothesis 2 is forecasts using LY factors ($k_0 = 5$) are more accurate than forecasts using PCA factors.

Table 4.7 DM equality test ($k_0 = 2$)

	Alternative 1[a]		Alternative 2[b]	
Steps	DM statistics	P-value	DM statistics	P-value
One	2.5202	0.0133**	2.5202	0.0067***
Two	1.5531	0.1236	1.5531	0.0618*
Three	0.9779	0.3305	0.9779	0.1653
Five	-0.2408	0.8103	-0.2408	0.5949
Ten	-1.5342	0.1285	-1.5342	0.9358
Fifteen	-1.1873	0.2384	-1.1873	0.8808

Notes: *, ** and *** denotes rejection of the null hypothesis at 10%, 5% and 1% significance levels.
[a] Alternative hypothesis 1 is forecasts using PCA factors and LY factors ($k_0 = 2$) have different levels of accuracy.
[b] Alternative hypothesis 2 is forecasts using LY factors ($k_0 = 2$) are more accurate than forecasts using PCA factors.

4.4.7.3 Forecast Encompassing Test

The encompassing test proposed by Chong and Hendry (1986) is also employed to evaluate the forecast accuracy. In this chapter, the test is performed based on the OLS regressions as follows

$$e_{PCA,t} = \lambda_1(e_{PCA,t} - e_{LY,t}) + \varepsilon_t \qquad (55)$$

$$e_{LY,t} = \lambda_2(e_{LY,t} - e_{PCA,t}) + \mu_t \qquad (56)$$

where ε_t and μ_t are the composite forecast errors. The first regression (17) is to test whether PCA factors encompass LY factors in terms of forecasting. The null hypothesis is $\lambda_1 = 0$. If one cannot reject this hypothesis, then forecasts using PCA factors encompass those using LY factors, suggesting that LY forecasts do not capture useful information absent in PCA forecasts. Similarly, the second regression (18) is to test whether forecasts generated from LY factors encompass those from PCA factors. If $\lambda_2 = 0$ in equation (18) (null hypothesis), then LY forecasts encompass PCA forecasts since no additional useful information missing from LY forecasts is captured by PCA forecasts.

Table 4.8 shows the results of encompassing test between PCA forecasts and LY forecasts ($k_0 = 5$) and Table 4.9 presents encompassing test results of PCA forecasts and LY forecasts ($k_0 = 2$). The rejection results are the same regardless of the choice of k_0. For one- and two-step-ahead forecasts, the null of $\lambda_1 = 0$ is rejected at 1% and 5% significance levels for both sets of LY forecasts ($k_0 = 5$ and $k_0 = 2$). Therefore, PCA forecasts does not encompass LY forecasts for these two horizons. However, the null of $\lambda_2 = 0$ fails to be rejected at all horizons, indicating LY forecasts do encompass PCA forecasts. For three-, five-, ten and fifteen-step-ahead horizons, both PCA and LY forecasts can encompass each other, indicating that neither includes useful information absent from the other. Thus, for these horizons, neither forecasts dominate the other. But for one- and two-step-ahead horizons, LY forecasts encompass PCA forecasts.

Table 4.8 Encompassing test ($k_0 = 5$)

	Dependent variable: $e_{PCA,t}$		Dependent variable: $e_{LY,t}$	
Steps	λ_1	P-value	λ_2	P-value
One	1.2526	5.77e-06***	-0.2526	0.3360
Two	1.2946	0.0132**	-0.2946	0.5670
Three	1.1499	0.1480	-0.1499	0.8500
Five	0.0017	0.9990	0.9983	0.4730
Ten	-3.3400	0.3480	4.3400	0.2230
Fifteen	-10.3570	0.1510	11.3570	0.116

Table 4.9 Encompassing test ($k_0 = 2$)

	Dependent variable: $e_{PCA,t}$		Dependent variable: $e_{LY,t}$	
Steps	λ_1	P-value	λ_2	P-value
One	1.3051	2.18e-06***	-0.3051	0.2420
Two	1.3997	0.0081**	-0.3997	0.4420
Three	1.2940	0.1100	-0.2938	0.7150
Five	0.1056	0.9410	0.8944	0.5290
Ten	-3.0440	0.4020	4.0440	0.2660
Fifteen	-9.8830	0.1810	10.8830	0.1410

4.5. Conclusion

The objective of this chapter is to empirically examine the newly developed factor analysis approach by Lam and Yao (2012). Factor analysis is an effective way of dimension-reduction. It attempts to find a small number of factors capable of summarizing the main useful information contained in a large number of data series. The conventional and popular methods in factor analysis are the method of principal components and the method by Bai and Ng (2002). The former is extensively used for estimating factors and the latter is usually applied to determine the optimal number of factors. In this chapter, these two methods are formed as a benchmark methodology to be compared with the methodology by Lam and Yao (2012).

Under these two methodologies, 288 price series of Chinese equities are employed to extract factors. The factors obtained through each methodology are respectively fitted with interest rate variable (Shibor) in FAVAR models to generate forecasts of interest rate. The resulting different sets of interest rate forecasts are evaluated in terms of forecasting ability.

In this chapter, different lags of autocorrelation (k_0) are used when estimating the LY factors using Lam and Yao (2012)'s method. It is found that the forecast performances of LY factors relative to PCA factors are not sensitive to the choice of k_0. In sum, LY factors outperform PCA factors in terms of forecasting accuracy, especially at the short horizons. More specifically, LY forecasts exhibit smaller RMSE, MAPE and MAE statistics for at least one-, two- and three-step-ahead horizons. In DM equality tests, LY factors provide more accurate forecasts than PCA factors for one and two steps. Encompassing tests indicate that LY forecasts encompass PCA forecasts for the first two steps. However, for the long horizons, LY forecasts have very limited predominance comparing to PCA forecasts.

For the future research, a more formal way of selecting k_0 may need to be explored since the choice of k_0 may influence the performances of factors estimated by Lam and Yao (2012)'s methodology. In this chapter, different values of k_0 lead to different results of number of factors. Although the prediction ability of LY factors relative to PCA factors is not sensitive to the choice of k_0 in our case, different sets of LY factors ($k_0 = 5$ and $k_0 = 2$) do exhibit different forecast performances.

CHAPTER V

CONCLUSIONS

This dissertation studies three major issues related to Chinese financial markets: (1) the interdependence structure and information transmission among Chinese stocks cross-listed in Shanghai, Hong Kong and New York; (2) the prequential analysis of stock returns and interest rates in China; (3) forecast performance evaluation of two factor estimation methods based on FAVAR model.

In Chapter II, ten Chinese stocks cross-listed in Shanghai, Hong Kong and New York are examined using VECM and inductive causal graphs. To study the contemporaneous causal structure among these Chinese cross-listed stocks, both PC algorithm and LiNGAM algorithm are applied. Due to the non-Gaussianity of the data, LiNGAM algorithm yields stronger causal identifications comparing to PC algorithm. According to the results from LiNGAM algorithm, the Shanghai market transmits information flows to both the Hong Kong and New York markets, either directly or indirectly. In contemporaneous time, the Shanghai market, as the home market, plays a dominant role in information transmission. Facilitated by the LiNGAM generated contemporaneous causal structure, forecast error variance decomposition is used to study the short-run interdependence of these cross-listed stocks. It is found that in dynamic (lagged) time, the Shanghai market shows modest influence on other markets at the short horizon and fails to have leading impacts at a longer horizon. Therefore, the importance of the Shanghai market weakens as horizon increases. For these Chinese cross-listed stocks, the home bias hypothesis, which indicates the leading role of the home market in information transmission of cross-listed stocks, is strongly supported in contemporaneous time, modestly supported at the short horizon and not supported at the long horizon.

In addition, the Shanghai market is highly exogenous at all horizons, perhaps due to its participation restriction in trading. The Shanghai market acts like an isolated island in terms of receiving information from other markets. The Hong Kong market causes the New York market in contemporaneous time and shows substantial influence on the New York market at both short and long horizons. The market integration and market interdependence gets stronger between the Hong Kong and New York markets as the horizon lengthens.

In Chapter III, the prequential analysis is applied to study whether interest rates in China help to forecast the stock returns. Both a bootstrap-like simulation method and a nonparametric kernel-based simulation method are used in the generating process of probability forecasts. Besides arbitrary selection, several data-driven methods are also considered for settings of subintervals and events when evaluating probability forecasts. With respect to calibration (reliability), including interest rates in the model improves the reliability of the forecasted stock returns. In terms of sorting (resolution), considering interest rates in the model enhances the model's ability to distinguish stock returns that actually occur and stock returns that do not occur. Therefore, interest rates help to forecast stock returns in China from both calibration and sorting perspectives. These results are robust under both simulation methods and both subinterval settings.

In Chapter IV, the traditional factor estimation method of PCA, accompanied by the method of Bai and Ng (2002), is compared with the newly developed method by Lam and Yao (2012). According to Bai and Ng (2002), 5 factors are selected from 288 pricing series of Chinese equities. The Lam and Yao (2012)'s method selects 1 strong factor and 2 weak factors with autocorrelation lag of 5, and 1 strong factor and 1 weak factor under autocorrelation lag of 2. The corresponding estimated factors are fitted with Chinese interest rate in a FAVAR to generate forecasts of interest rate with different horizons (one-, two-, three-, five-, ten- and fifteen-step-ahead horizons). Statistical measures like root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), and statistical tests such as the DM equality test developed by Diebold and Mariano (2012) and the encompassing test proposed by Chong and Hendry (1986) are used to evaluate the performance of the forecasted interest rate under different factor estimation methods. Overall, the factors estimated by Lam and Yao (2012)'s method, regardless of the selection of autocorrelation lag, outperforms factors estimated through PCA in terms of forecasting accuracy, especially at the short horizons. However, for the long horizons, the factors by Lam and Yao (2012) have very limited predominance compared to the factors by PCA.

Further research is needed to address some of the limitations and problems in each chapter. In Chapter II, the estimation is based on daily closing prices of three equity markets. However, since the Hong Kong market closes one hour after the Shanghai market, the identification of contemporaneous structure between these two markets using daily data may be confounded. For a more in-depth exploration, high-frequency intraday data (e.g. data with five-minute spread) should be used to examine the contemporaneous interdependence structure between the Shanghai and Hong Kong markets.

In Chapter III, only short-term interest rate variable is included in the VAR model for the major objective of the Chapter is to investigate whether interest rates help to forecasting stock returns in China. However, the VAR model shows limited sorting ability. Following the suggestions by previous literature, more variables, such as the price dividend ratio, different term structures of interest rates and aggregate returns on other important stock exchanges such as the NYSE, should be incorporated in the model to improve the forecasts of stock returns.

In Chapter IV, both forecasts using PCA method and Lam and Yao (2012)'s method only outperform the forecasts of the random walk at the long horizon. Thus, the long-term forecast performance of these two factor estimation methods needs to be further investigated.

REFERENCES

Altman, N., C. Léger, 1995. Bandwidth selection for kernel distribution function estimation. Journal of Statistical Planning and Inference 46, 195-214.

Anderson, T.W., 1958. An introduction to multivariate statistical analysis. New York: Wiley.

—— 1963. The use of factor analysis in the statistical analysis of multiple time series. Psychometrika 28, 1-25.

Armstrong, J.S., 1984. Forecasting by extrapolation: Conclusions from 25 years of research. Interfaces 14, 52-66.

Arshanapalli, B., J. Doukas, 1993. International stock market linkages: Evidence from the pre-and post-October 1987 period. Journal of Banking & Finance 17, 193-208.

Awokuse, T.O., 2006. Export-led growth and the Japanese economy: Evidence from VAR and directed acyclic graphs. Applied Economics 38, 593-602.

Awokuse, T.O., A. Chopra, D.A. Bessler, 2009. Structural change and international stock market interdependence: Evidence from Asian emerging markets. Economic Modelling 26, 549-559.

Bai, J., S. Ng, 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Bai, J., P. Perron, 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1-22.

Bernanke, B.S., 1986. Alternative explanations of the money-income correlation. In Carnegie-Rochester Conference Series on Public Policy. North-Holland, 25, 49-99.

Bernanke, B.S., J. Boivin, P. Eliasz, 2005. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. The Quarterly Journal of Economics 120, 387-422.

Bessler, D.A., S. Kibriya, J. Chen, and E. Price. 2015. On forecasting conflict in the Sudan: 2009–2012. Journal of Forecasting 35, 179-188.

Bessler, D.A., J.L. Kling. 1990. Prequential analysis of cattle prices. Applied Statistics 39, 95-106.

Bessler, D.A., J.W. Kolari, T.A. Maung, 2011. Dynamic linkages among equity markets: Local versus basket currencies. Applied Economics 43, 1703-1719.

Bessler, D.A., R. Ruffley. 2004. Prequential analysis of stock market returns. Applied Economics 36, 399-412.

Bessler, D.A., Z. Wang, 2012. D-separation, forecasting, and economic science: a conjecture. Theory and Decision 73, 295-314.

Bessler, D.A., J. Yang, 2003. The structure of interdependence in international stock markets. Journal of International Money and Finance 22, 261-287.

Boivin, J., M.P. Giannoni, I. Mihov, 2009. Sticky prices and monetary policy: Evidence from disaggregated US data. The American Economic Review 99, 350-384.

Box, G. E., D. A. Pierce, 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American statistical Association 65, 1509-1526.

Brandt, J. A., D.A. Bessler, 1983. Price forecasting and evaluation: An application in agriculture. Journal of Forecasting 2, 237-248.

Brandt, J.A., D.A. Bessler, 1984. Forecasting with vector autoregressions versus a univariate ARIMA process: An empirical example with US hog prices. North Central Journal of Agricultural Economics, 29-36.

Brier, G.W. 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review 78, 1-3.

Bunn, D.W. 1984. Applied decision analysis. New York: McGraw-Hill.

Cai, C.X., P.B. McGuinness, Q. Zhang, 2011. The pricing dynamics of cross-listed securities: The case of Chinese A-and H-shares. Journal of Banking & Finance 35, 2123-2136.

Campbell, J.Y. 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373-399.

—— 1991. A variance decomposition for stock returns. Economic Journal 101, 157-179.

Campbell, J.Y., J. Ammer. 1993. What moves the stock and bond markets? A variance decomposition for long-term asset returns. The Journal of Finance 48, 3-37.

Campbell, J.Y., R.J. Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195-228.

Chelley-Steeley, P.L., J.M. Steeley, 2012. Price discovery for Chinese shares cross-listed in multiple markets. Applied Financial Economics 22, 1587-1601.

Chen, N.F. 1991. Financial investment opportunities and the macroeconomy. The Journal of Finance 46, 529-554.

Chen, K.C., G. Li, L. Wu, 2010. Price discovery for segmented US-listed Chinese stocks: location or market quality?. Journal of Business Finance & Accounting 37, 242-269.

Chong, Y.Y., D.F. Hendry, 1986. Econometric evaluation of linear macro-economic models. The Review of Economic Studies 53, 671-690.

Connor, G., R.A. Korajczyk, 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of Financial Economics 15, 373-394.

—— 1988. Risk and return in an equilibrium APT: Application of a new test methodology. Journal of Financial Economics 21, 255-289.

Cootner, P.H., 1964. The random character of stock market prices. MIT Press.

Coval, J.D., T.J. Moskowitz, 1999. Home bias at home: Local equity preference in domestic portfolios. The Journal of Finance 54, 2045-2073.

Cutler, D.M., J.M. Poterba, L.H. Summers. 1991. Speculative dynamics. The Review of Economic Studies 58, 529-546.

Dawid, A.P. 1984. Statistical theory: The prequential approach. Journal of the Royal Statistical Society, Series A 147, 278-292.

—— 1985. Calibration-based empirical probability. The Annals of Statistics 13, 1251-1274.

Dickey, D.A., W.A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427-431.

Diebold, F.X., R.S. Mariano, 2012. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134-144.

Doan, T., 1992. RATS: User's Manual 4.0, Estima, Evanston, IL.

Dodge, Y., V. Rousson, 2001. On asymmetric properties of the correlation coefficient in the regression setting. The American Statistician 55, 51-54.

Engle, R.F., C.W. Granger, 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica. 55, 251-276.

Estrella, A., G.A. Hardouvelis, 1991. The term structure as a predictor of real economic activity. The Journal of Finance 46, 555-576.

Eun, C.S., S. Shim, 1989. International transmission of stock market movements. Journal of financial and quantitative Analysis 24, 241-256.

Fair, R.C. 1986. Evaluating the predictive accuracy of models. In Handbook of Econometrics, Volume 3, Eds. Griliches, Z. and M. D. Intriligator, 1979-1995. Amsterdam: North-Holland.

Fama, E.F., 1965. The behavior of stock-market prices. Journal of Business 38, 34-105.

—— 1981. Stock returns, real activity, inflation, and money. The American Economic Review 71, 545-565.

—— 1995. Random walks in stock market prices. Financial Analysts Journal 51, 75-80.

Fama, E.F., K.R. French. 1988a. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3-25.

—— 1988b. Permanent and temporary components of stock prices. The Journal of Political Economy 96, 246-273.

—— 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

Fama, E.F., G.W. Schwert. 1977. Asset returns and inflation. Journal of Financial Economics 5, 115-146.

Fan, J., F. Han, H. Liu, 2014. Challenges of big data analysis. National Science Review 1, 293-314.

Fang, F., W. Dong, X. Lv. 2016. Asymmetric reactions of China's stock market to short-term interest rates. International Journal of Economics and Finance 8, 260.

Fischhoff, B., P. Slovic, S. Lichtenstein. 1977. Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance 3, 552.

Forni, M., L. Gambetti, 2010. The dynamic effects of monetary policy: A structural factor model approach. Journal of Monetary Economics 57, 203-216.

Forni, M., M. Hallin, M. Lippi, L. Reichlin, 2000. The generalized dynamic-factor model: Identification and estimation. Review of Economics and Statistics 82, 540-554.

—— 2004. The generalized dynamic factor model consistency and rates. Journal of Econometrics 119, 231-255.

—— 2005. The generalized dynamic factor model. Journal of the American Statistical Association 100.

Frorni, M., L. Reichlin, 1996. Dynamic common factors in large cross-sections. Empirical Economics 21, 27-42.

French, K.R., J.M. Poterba, 1991. Investor diversification and international equity markets. American Economic Review 81, 222–226.

Friedman, M. 1953. Essays in positive economics. University of Chicago Press.

Frijns, B., A. Gilbert, A. Tourani-Rad, 2010. The dynamics of price discovery for cross-listed shares: Evidence from Australia and New Zealand. Journal of Banking & Finance 34, 498-508.

Frijns, B., A. Gilbert, A. Tourani-Rad, 2015. The determinants of price discovery: Evidence from US-Canadian cross-listed shares. Journal of Banking & Finance 59, 457-468.

Gang, Y.I., 2009. On the market-based interest rate reform. Journal of Financial Research 1: 003.

Garbade, K.D., W.L. Silber, 1979. Dominant and satellite markets: A study of dually-traded securities. The Review of Economics and Statistics, 455-460.

Geweke, J., 1977. The dynamic factor analysis of economic time series. In Latent Variables in Socio-Economic Models, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland, Ch. 19.

Geweke, J., R. Meese, 1981. Estimating regression models of finite but unknown order. International Economic Review 23, 55-70.

Ghysels, E., S. Ng, 1998. A semiparametric factor model of interest rates and tests of the affine term structure. Review of Economics and Statistics 80, 535-548.

Gonzalo, J., C. Granger, 1995. Estimation of common long-memory components in cointegrated systems. Journal of Business & Economic Statistics 13, 27-35.

Grammig, J., M. Melvin, C. Schlag, 2005. Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. Journal of Empirical Finance 12, 139-164.

Grammig, J., F.J. Peter, 2013. Telltale tails: A new approach to estimating unique market information shares. Journal of Financial and Quantitative Analysis 48, 459-488.

Granger, C.W., 1980. Testing for causality: A personal viewpoint. Journal of Economic Dynamics and Control 2, 329-352.

Granger, C. W., 1992. Forecasting stock market prices: Lessons for forecasters. International Journal of Forecasting 8, 3-13.

Granger, C. W., P. Newbold, 1973. Some comments on the evaluation of economic forecasts. Applied Economics 5, 35-47.

Granger, C.W., P. Newbold, 1974. Spurious regressions in dconometrics. Journal of econometrics 2, 111-120.

Harvey, C.R. 1988. The real term structure and consumption growth. Journal of Financial Economics 22, 305-333.

—— 1989. Forecasts of economic growth from the bond and stock markets. Financial Analysts Journal 45, 38-45.

Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. The Journal of Finance 50, 1175-1199.

Hashemzadeh, N., P. Taylor, 1988. Stock prices, money supply, and interest rates: The question of causality. Applied economics 20, 1603-1611.

Hauser, S., Y. Tanchuma, U. Yaari, 1998. International transfer of pricing information between dually listed stocks. Journal of Financial Research 21, 139-157.

Hyndman, R.J., A.B. Koehler, 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679-688.

Hyvärinen, A., J. Karhunen, E. Oja, 2004. Independent component analysis. John Wiley & Sons.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica 59, 1551-1580.

—— 1995. Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. Journal of Econometrics 69, 111-132.

Johansen, S., K. Juselius, 1990. Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. Oxford Bulletin of Economics and statistics 52, 169-210.

—— 1994. Identification of the long-run and the short-run structure an application to the ISLM model. Journal of Econometrics 63, 7-36.

Jorion, P., E. Schwartz, 1986. Integration vs. segmentation in the Canadian stock market. The Journal of Finance 41, 603-614.

Jutten, C., J. Herault, 1991. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. Signal processing 24, 1-10.

Karolyi, G.A., 2006. The world of cross-listings and cross-listings of the world: Challenging conventional wisdom. Review of Finance 10, 99-152.

Kendall, M.G., A.B. Hill, 1953. The analysis of economic time-series-part i: Prices. Journal of the Royal Statistical Society. Series A (General) 116, 11-34.

Kim, M., A.C. Szakmary, I. Mathur, 2000. Price transmission dynamics between ADRs and their underlying foreign securities. Journal of Banking & Finance 24, 1359-1382.

Kling, J.L., D.A. Bessler. 1989. Calibration-based predictive distributions: An application of prequential analysis to interest rates, money, prices, and output. Journal of Business 62, 477-499.

Kwon, D.H., D.A. Bessler, 2011. Graphical methods, inductive causal inference, and econometrics: A literature review. Computational Economics 38, 85-106.

Lam, C., Q. Yao, 2012. Factor modeling for high-dimensional time series: inference for the number of factors. The Annals of Statistics, 694-726.

Lam, C., Q. Yao, N. Bathia, 2011. Estimation of latent factors for high-dimensional time series. Biometrika, asr048.

Lanne, M., H. Lütkepohl, 2012. Structural vector autoregressions with nonnormal residuals. Journal of Business & Economic Statistics 28, 159-168.

Lawley, D.N., A.E. Maxwell, 1971. Factor analysis as a statistical method. New York: American Elsevier Publishing.

Lau, S.T., J.D. Diltz, 1994. Stock returns and the transfer of information between the New York and Tokyo Stock Exchanges. Journal of International Money and Finance 13, 211-222.

Li, J., 2015. The asymmetric effects of investor sentiment and monetary policy on stock prices. Applied Economics 47, 2514-2522.

Li, Q., J.S. Racine, 2007. Nonparametric econometrics: theory and practice. Princeton University Press.

Li, Y., D. Yan, J. Greco, 2006. Market segmentation and price differentials between A shares and H shares in the Chinese stock markets. Journal of Multinational Financial Management 16, 232-248.

Lichtenstein, S., B. Fischhoff, L. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. Eds. Kahneman, D., P. Slovic, and A. Tverski, 306-334. Judgement under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press.

Lieberman, O., U. Ben-Zion, S. Hauser, 1999. A characterization of the price behavior of international dual stocks: an error correction approach. Journal of International Money and Finance 18, 289-304.

Lin, Y.M., Y.S. Hsu, S.L. Chen. 2009. Cash-flow news, market liquidity and liquidity risk. Applied Economics 41, 1137-1156.

Ljung, G. M., G. E. Box, 1978. On a measure of lack of fit in time series models. Biometrika 65, 297-303.

Love, I., L. Zicchino, 2006. Financial development and dynamic investment behavior: Evidence from panel VAR. The Quarterly Review of Economics and Finance 46, 190-210.

Lütkepohl, H., H.E. Reimers, 1992. Impulse response analysis of cointegrated systems. Journal of economic dynamics and control 16, 53-78.

Mönch, E., 2008. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. Journal of Econometrics 146, 26-43.

Moneta, A., D. Entner, P.O. Hoyer, A. Coad, 2013. Causal inference by independent component analysis: theory and applications*. Oxford Bulletin of Economics and Statistics 75, 705-730.

Murphy, A.H. 1972. Scalar and vector partitions of the probability score: Part II. N-state situation. Journal of Applied Meteorology 11, 1183-1192.

Murphy, A.H., R.L. Winkler. 1987. A general framework for forecast verification. Monthly Weather Review 115, 1330-1338.

Parzen, E., 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33, 1065-1076.

Pascual, R., B. Pascual-Fuster, F. Climent, 2006. Cross-listing, price discovery and the informativeness of the trading process. Journal of Financial Markets 9, 144-161.

Patelis, A.D. 1997. Stock return predictability and the role of monetary policy. The Journal of Finance 52, 1951-1972.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82, 669-710.

—— 2000. Causality: models, reasoning and inference, MIT press, Cambridge.

Pena, D., G.E. Box, 1987. Identifying a simplifying structure in time series. Journal of the American Statistical Association 82, 836.

Phillips, P.C., 1996. Econometric model determination. Econometrica 64, 763–812.

Phillips, P.C., P. Perron, 1988. Testing for a unit root in time series regression. Biometrika 75, 335-346.

Porter, N., T. Xu. 2009. What drives China's interbank market? IMF working paper No. 9-189.

Racine, J.S., 2008. Nonparametric econometrics: A primer. Now Publishers Inc.

Roosenboom, P., M.A. Van Dijk, 2009. The market reaction to cross-listings: Does the destination market matter? Journal of Banking & Finance 33, 1898-1908.

Ross, S.A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341-360.

Sargent, T.J., C.A. Sims, 1977. Business cycle modeling without pretending to have too much a priori economic theory. New methods in business cycle research 1, 145-168.

Schwert, G.W., 2002. Tests for unit roots: A Monte Carlo investigation. Journal of Business & Economic Statistics 20, 5-17.

Scott, D.W. 1979. On optimal and data-based histograms. *Biometrika* 66, 605-610.

—— 2015. Multivariate density estimation: Theory, practice, and visualization. New York: John Wiley & Sons.

Sheather, S.J., M.C. Jones, 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society. Series B 53, 683-690.

Shimizu, S., P.O. Hoyer, A. Hyvärinen, A. Kerminen, 2006. A linear non-gaussian acyclic model for causal discovery. Journal of Machine Learning Research 7, 2003-2030.

Shimizu, S., A. Hyvärinen, P.O. Hoyer, Y. Kano, 2006. Finding a causal ordering via independent component analysis. Computational Statistics & Data Analysis 50, 3278-3293.

Shimizu, S., A. Hyvarinen, Y. Kano, P.O. Hoyer, 2012. Discovery of non-gaussian linear causal models using ICA. arXiv preprint arXiv, 1207-1413.

Shimizu, S., Y. Kano, 2008. Use of non-normality in structural equation modeling: application to direction of causation. Journal of Statistical Planning and Inference 138, 3483-3491.

Sims, C., 1980a. Macroeconomics and reality. Econometrica 48, 1-48.

—— 1980b. Comparison of interwar and postwar business cycles: Monetarism reconsidered. American Economic Review 70, 250-57.

—— 1986. Are forecasting models usable for policy analysis? Quarterly Review of Federal Reserve Bank in Minneapolis 10, 2-16.

Sims, C.A., J.H. Stock, M.W. Watson, 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113-144.

Spirtes, P., C.N. Glymour, R. Scheines, 2000. Causation, prediction, and search, MIT press, Cambridge.

Stock, J.H., M.W. Watson, 1989. New indexes of coincident and leading economic indicators. In NBER Macroeconomics Annual 1989, 4, 351-409, MIT press.

—— 1998. Diffusion indexes (No. w6702). National Bureau of Economic Research.

—— 1999. Forecasting inflation. Journal of Monetary Economics 44, 293-335.

—— 2001. Forecasting output and inflation: The role of asset prices. NBER Working Paper No. 8180.

—— 2002a. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167-1179.

—— 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147-162.

—— 2005. Implications of dynamic factor models for VAR analysis (No. w11467). National Bureau of Economic Research.

—— 2006. Forecasting with many predictors. Handbook of Economic Forecasting 1, 515-554.

—— 2011. Dynamic factor models. Oxford Handbook of Economic Forecasting 1, 35-59.

Stone, M., 1961. The opinion pool. The Annals of Mathematical Statistics 32, 1339-1342.

Sturges, H.A. 1926. The choice of a class interval. Journal of the American Statistical Association 21, 65-66.

Su, Q., T.T.L. Chong, 2007. Determining the contributions to price discovery for Chinese cross-listed stocks. Pacific-Basin Finance Journal 15, 140-153.

Swanson, N.R., C.W. Granger, 1997. Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. Journal of the American Statistical Association 92, 357-367.

Tan, L., T.C. Chiang, J.R. Mason, E. Nelling, 2008. Herding behavior in Chinese stock markets: An examination of A and B shares. Pacific-Basin Finance Journal 16, 61-77.

Tesar, L.L., I.M. Werner, 1995. Home bias and high turnover. Journal of International Money and Finance 14, 467-492.

Wang, Z., D.A. Bessler, 2005. A Monte Carlo study on the selection of cointegrating rank using information criteria. Econometric Theory 21, 593-620.

Wang, S.S., L. Jiang, 2004. Location of trade, ownership restrictions, and market illiquidity: examining Chinese A-and H-shares. Journal of Banking & Finance 28, 1273-1297.

Xu, X.E., H.G. Fung, 2002. Information flows across markets: evidence from China–backed stocks dual–listed in Hong Kong and New York. Financial Review 37, 563-588.

Yang, J., D.A. Bessler, 2004. The international price transmission in stock index futures markets. Economic Inquiry 42, 370-386.

Yates, J.F. 1988. Analyzing the accuracy of probability judgments for multiple events: An extension of the covariance decomposition. Organizational Behavior and Human Decision Processes 41, 281-299.

Zellner, A., H.A. Keuzenkamp, M. McAleer, M., 2001. Simplicity, inference and modelling: Keeping it sophisticatedly simple. Cambridge University Press.

APPENDIX A

LiNGAM is applied in this chapter to help specify a linear relationship with causal interpretations between innovations from an estimated VECM. Scatter plots on these innovations for CSA are given as Figure A-1. Others are not reported to save space.

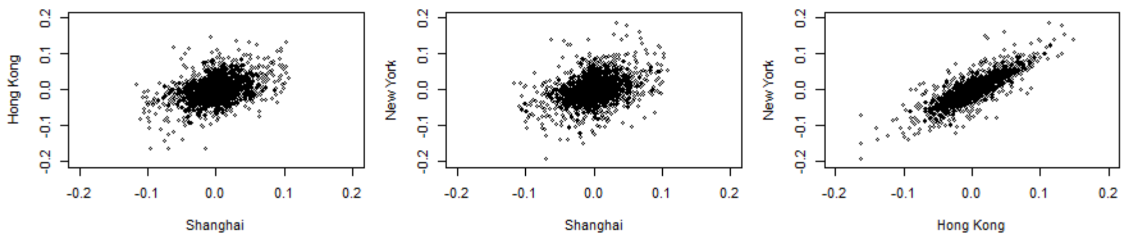


Figure A-1 Scatter plots on innovations for CSA

The plots in Shanghai versus Hong Kong and Shanghai versus New York show the linear relationship, while plausible, are less strong, relative to the Hong Kong versus New York innovations. R^2 from innovations in Hong Kong regressed on innovations in Shanghai is 0.151; for innovations in New York regressed on innovations in Shanghai, the R^2 is 0.118 and for New York innovations regressed on Hong Kong innovations, the R^2 is 0.686.

APPENDIX B

Sector	Symbol	Company	Exchange
Energy			
1	000937	Jizhong Energy Resources Co Ltd	Shenzhen
2	000983	Shanxi Xishan Coal And Electricity Power Co Ltd	Shenzhen
3	002353	Yantai Jereh Oilfield Services Group Co Ltd	Shenzhen
4	600028	China Petroleum & Chemical Corporation	Shanghai
5	600157	Wintime Energy Co Ltd	Shanghai
6	600188	Yanzhou Coal Mining Co Ltd	Shanghai
7	600256	Guanghui Energy Co Ltd	Shanghai
8	600348	Yang Quan Coal Industry (Group) Co., Ltd	Shanghai
9	600395	Guizhou Panjiang Refined Coal Co Ltd	Shanghai
10	600583	Offshore Oil Engineering Co Ltd	Shanghai
11	600688	Sinopec Shanghai Petrochemical Co Ltd	Shanghai
12	601088	China Shenhua Energy Co Ltd	Shanghai
13	601699	Shanxi Lu'an Environmental Energy Development Co Ltd	Shanghai
14	601808	China Oilfield Services Limited	Shanghai
15	601857	PetroChina Co Ltd	Shanghai
16	601898	China Coal Energy Co Ltd	Shanghai
Financials			
17	000001	Ping An Bank Co., Ltd.	Shenzhen
18	000002	China Vanke Co Ltd	Shenzhen
19	000402	Financial Street Holding Co Ltd	Shenzhen
20	000686	Northeast Securities Co Ltd	Shenzhen
21	000728	Guoyuan Securities Company Limited	Shenzhen
22	000750	Sealand Securities Co., Ltd.	Shenzhen
23	000776	GF Securities Co., Ltd.	Shenzhen
24	000783	Changjiang Securities Company Limited	Shenzhen
25	002142	Bank of Ningbo Co Ltd	Shenzhen
26	002146	Risesun Real Estate Development Co Ltd	Shenzhen
27	002500	Shanxi Securities Co Ltd	Shenzhen
28	002673	Western Securities Co Ltd	Shenzhen
29	600000	Shanghai Pudong Development Bank Co Ltd	Shanghai
30	600015	Hua Xia Bank Co Ltd	Shanghai
31	600016	China Minsheng Banking Corp Ltd	Shanghai
32	600030	CITIC Securities Co Ltd	Shanghai
33	600036	China Merchants Bank Co Ltd	Shanghai
34	600048	Poly Real Estate Group Co Ltd	Shanghai

Sector	Symbol	Company	Exchange
35	600109	Sinolink Securities Co. Ltd.	Shanghai
36	600208	Xinhu Zhongbao Co Ltd	Shanghai
37	600340	China Fortune Land Development Co., Ltd.	Shanghai
38	600369	Southwest Securities Co Ltd	Shanghai
39	600383	Gemdale Corporation	Shanghai
40	600663	Shanghai Lujiazui Finance and Trade Zone Development Co Ltd	Shanghai
41	600705	AVIC Capital Co Ltd	Shanghai
42	600837	Haitong Securities Company Limited	Shanghai
43	600999	China Merchants Securities Co Ltd	Shanghai
44	601009	Bank of Nanjing Co Ltd	Shanghai
45	601166	Industrial Bank	Shanghai
46	601169	Bank of Beijing Co Ltd	Shanghai
47	601288	Agricultural Bank of China Co Ltd	Shanghai
48	601318	Ping An Insurance (Group) Company of China Ltd	Shanghai
49	601328	Bank of Communications Co LTD	Shanghai
50	601336	New China Life Insurance Co Ltd	Shanghai
51	601377	Industrial Securities Co Ltd	Shanghai
52	601398	Industrial and Commercial Bank of China Ltd	Shanghai
53	601555	Soochow Securities Co Ltd	Shanghai
54	601601	China Pacific Insurance (Group) Co Ltd	Shanghai
55	601628	China Life Insurance Company Limited	Shanghai
56	601688	Huatai Securities Co Ltd	Shanghai
57	601818	China Everbright Bank Co Ltd	Shanghai
58	601901	Founder Securities Co Ltd	Shanghai
59	601939	China Construction Bank	Shanghai
60	601988	Bank of China Ltd	Shanghai
61	601998	China Citic Bank Corporation Limited	Shanghai
Industrials			
62	000009	China Baoan Group Co.,Ltd.	Shenzhen
63	000039	China International Marine Containers (Group) Co Ltd	Shenzhen
64	000157	Zoomlion Heavy Industry Science & Technology Co Ltd	Shenzhen
65	000338	Wei Chai Power Co Ltd	Shenzhen
66	000400	XJ Electric Co Ltd	Shenzhen
67	000425	XCMG Construction Machinery Co Ltd	Shenzhen
68	000768	Avic Aircraft Co.,Ltd.	Shenzhen
69	000826	Sound Environmental Resources Co Ltd	Shenzhen
70	002051	China CAMC Engineering Co Ltd	Shenzhen
71	002081	Suzhou Gold Mantis Construction Decoration Co Ltd	Shenzhen
72	002202	Xinjiang Goldwind Science & Technology Co Ltd	Shenzhen

Sector	Symbol	Company	Exchange
73	002310	Beijing Orient Landscape Co Ltd	Shenzhen
74	002375	Zhejiang Yasha Decoration Co Ltd	Shenzhen
75	300024	Siasun Robot & Automation Co Ltd	Shenzhen
76	300070	Beijing Originwater Technology Co Ltd	Shenzhen
77	300124	Shenzhen Inovance Technology Co Ltd	Shenzhen
78	600009	Shanghai International Airport Co Ltd	Shanghai
79	600018	Shanghai International Port (Group) Co Ltd	Shanghai
80	600029	China Southern Airlines Co Ltd	Shanghai
81	600031	Sany Heavy Industry Co Ltd	Shanghai
82	600038	AVIC Helicopter Co.,Ltd.	Shanghai
83	600058	Minmetals Development Co Ltd	Shanghai
84	600068	China Gezhouba Group Co Ltd	Shanghai
85	600089	TBEA Co Ltd	Shanghai
86	600115	China Eastern Airlines Corp Ltd	Shanghai
87	600118	China Spacesat Co Ltd	Shanghai
88	600150	China CSSC Holdings Limited	Shanghai
89	600153	Xiamen C&D Inc	Shanghai
90	600170	Shanghai Construction Co Ltd	Shanghai
91	600221	Hainan Airlines Co Ltd	Shanghai
92	600316	Jiangxi Hongdu Aviation Industry Co Ltd	Shanghai
93	600372	China Avic Electronics Co.,Ltd.	Shanghai
94	600406	NARI Technology Co., Ltd.	Shanghai
95	600648	Shanghai Wai Gaoqiao Free Trade Zone Development Co Ltd	Shanghai
96	600739	Liaoning Cheng Da Co Ltd	Shanghai
97	600783	Luxin Venture Capital Group Co.,Ltd.	Shanghai
98	600875	Dongfang Electric Corporation Limited	Shanghai
99	600880	Chengdu B-ray Media Co Ltd	Shanghai
100	600893	AVIC Aviation Engine Corporation PLC	Shanghai
101	601006	Daqin Railway Co Ltd	Shanghai
102	601018	Ningbo Port Co Ltd	Shanghai
103	601111	Air China Ltd	Shanghai
104	601117	China National Chemical Engineering Co Ltd	Shanghai
105	601179	China XD Electric Co Ltd	Shanghai
106	601186	China Railway Construction Co Ltd	Shanghai
107	601333	Guangshen Railway Company Limited	Shanghai
108	601390	China Railway Group Limited	Shanghai
109	601618	Metallurgical Corporation of China Co Ltd	Shanghai
110	601668	China State Construction Engineering Co Ltd	Shanghai
111	601669	Power Construction Corporation of China,Ltd	Shanghai

Sector	Symbol	Company	Exchange
112	601727	Shanghai Electric Group Co Ltd	Shanghai
113	601766	CSR Co Ltd	Shanghai
114	601800	China Communications Construction Company Limited	Shanghai
115	601866	China Shipping Container Lines Co Ltd	Shanghai
116	601989	China Shipbuilding Industry Co Ltd	Shanghai
Materials			
117	000060	Shenzhen Zhongjin Lingnan Nonfemet Co Ltd	Shenzhen
118	000401	Tangshan Jidong Cement Co Ltd	Shenzhen
119	000629	Pangang Group Vanadium Titanium & Resources Co., Ltd.	Shenzhen
120	000630	Tongling Nonferrous Metals Group Co. Ltd	Shenzhen
121	000709	Hebei Iron & Steel Co., Ltd	Shenzhen
122	000778	Xinxing Ductile Iron Pipes Co Ltd	Shenzhen
123	000792	Qinghai Salt Lake Industry Co Ltd	Shenzhen
124	000825	Shanxi Taigang Stainless Steel Co Ltd	Shenzhen
125	000831	China Minmetals Rare Earth Co., Ltd.	Shenzhen
126	000878	Yunnan Copper Co Ltd	Shenzhen
127	000898	Angang Steel Co Ltd	Shenzhen
128	000960	Yunnan Tin Co Ltd	Shenzhen
129	002450	Jiangsu Kangde Xin Composite Material Co.,Ltd.	Shenzhen
130	002470	Kingenta Ecological Engineering Group Co., Ltd	Shenzhen
131	600010	Inner Mongolia Baotou Steel Union Co Ltd	Shanghai
132	600019	Baoshan Iron &Steel Co Ltd	Shanghai
133	600143	Kingfa Sci&Tech Co Ltd	Shanghai
134	600277	Inner Mongolia Yili Energy Company Limited	Shanghai
135	600309	Wanhua Chemical Group Co., Ltd.	Shanghai
136	600352	Zhejiang Longsheng Group Co Ltd	Shanghai
137	600362	Jiangxi Copper Co Ltd	Shanghai
138	600489	Zhongjin Gold Co Ltd	Shanghai
139	600497	Yunnan Chihong Zinc&Germanium Co Ltd	Shanghai
140	600516	Fangda Carbon New Material Co.,Ltd	Shanghai
141	600547	Shandong Gold-Mining Co Ltd	Shanghai
142	600549	Xiamen Tungsten Co Ltd	Shanghai
143	600585	Anhui Conch Cement Co Ltd	Shanghai
144	601168	Western Mining Co Ltd	Shanghai
145	601216	Inner Mongolia Junzheng Energy & Chemical Industry Co Ltd	Shanghai
146	601600	Aluminum Corporation of China Limited	Shanghai
147	601899	Zijin Mining Group Co Ltd	Shanghai
148	601958	Jinduicheng Molybdenum Co Ltd	Shanghai
149	601992	BBMG Corporation	Shanghai

Sector	Symbol	Company	Exchange
150	600111	China Northern Rare Earth (Group) High-Tech Co.,Ltd	Shanghai
Consumer Discretionary			
151	000069	Shenzhen Overseas Chinese Town Co Ltd	Shenzhen
152	000100	TCL Corporation	Shenzhen
153	000156	Wasu Media Holding Co Ltd	Shenzhen
154	000559	Wanxiang Qianchao Co Ltd	Shenzhen
155	000581	Weifu High-Technology Group Co Ltd	Shenzhen
156	000625	Chongqing Changan Automobile Co Ltd	Shenzhen
157	000651	Gree Electric Appliances,Inc. of Zhuhai	Shenzhen
158	000793	Huawen Media Investment Corp	Shenzhen
159	000800	FAW Car Co Ltd	Shenzhen
160	000839	CITIC Guoan Information Industry Co Ltd	Shenzhen
161	000917	Hunan TV & Broadcast Intermediary Co Ltd	Shenzhen
162	002024	Suning Commerce Group Co Ltd	Shenzhen
163	002292	Guangdong Alpha Animation and Culture Co Ltd	Shenzhen
164	002344	Haining China Leather Market Co Ltd	Shenzhen
165	002400	Guangdong Advertising Co Ltd	Shenzhen
166	002429	Shenzhen MTC Co Ltd	Shenzhen
167	002594	BYD Co Ltd	Shenzhen
168	300027	Huayi Brothers Media Co Ltd	Shenzhen
169	300058	BlueFocus Communication Group Co Ltd	Shenzhen
170	300133	Zhejiang Huace Film & TV Co Ltd	Shenzhen
171	300251	Beijing Enlight Media Co Ltd	Shenzhen
172	600060	Hisense Electric Co Ltd	Shanghai
173	600066	Zhengzhou Yutong Bus Co Ltd	Shanghai
174	600104	SAIC Motor Co Ltd	Shanghai
175	600166	Beiqi Foton Motor Co Ltd	Shanghai
176	600177	Youngor Group Co Ltd	Shanghai
177	600373	Chinese Universe Publishing And Media Co Ltd	Shanghai
178	600398	Heilan Home Co.,Ltd	Shanghai
179	600415	Zhejiang China Commodities City Group Co Ltd	Shanghai
180	600633	Zhe Jiang Daily Media Group Co.,Ltd	Shanghai
181	600637	BesTV New Media Co., Ltd.	Shanghai
182	600655	Shanghai Yuyuan Tourist Mart Co Ltd	Shanghai
183	600660	Fuyao Glass Industry Group Co.,Ltd	Shanghai
184	600690	Qingdao Haier Co Ltd	Shanghai
185	600741	HUAYU Automotive Systems Company Limited	Shanghai
186	600839	Sichuan Changhong Electric Co Ltd	Shanghai
187	601098	China South Publishing & Media Group Co Ltd	Shanghai

Sector	Symbol	Company	Exchange
188	601118	China Hainan Rubber Industry Group Co Ltd	Shanghai
189	601258	Pangda Automobile Trade Co Ltd	Shanghai
190	601633	Great Wall Motor Co Ltd	Shanghai
191	601888	China International Travel Service Co Ltd	Shanghai
192	601928	Jiangsu Phoenix Publishing & Media Co Ltd	Shanghai
193	601929	Jishi Media Co Ltd	Shanghai
Consumer Staples			
194	000061	Shenzhen Agricultural Products Co Ltd	Shenzhen
195	000568	Luzhou Lao Jiao Co Ltd	Shenzhen
196	000729	Beijing Yanjing Brewery Co Ltd	Shenzhen
197	000858	Wuliangye Yibin Co Ltd	Shenzhen
198	000869	Yantai Changyu Pioneer Wine Co Ltd	Shenzhen
199	000876	New Hope Liuhe Co Ltd	Shenzhen
200	000895	Henan Shuanghui Investment & Development Co Ltd	Shenzhen
201	000963	Huadong Medicine Co Ltd	Shenzhen
202	002304	Jiangsu Yanghe Brewery Joint-Stock Co Ltd	Shenzhen
203	002385	Beijing Dabeinong Technology Group Co Ltd	Shenzhen
204	002570	Beingmate Baby & Child Food Co., Ltd.	Shenzhen
205	300146	By-Health Co Ltd	Shenzhen
206	600108	Gansu Yasheng Industrial (Group) Co Ltd	Shanghai
207	600315	Shanghai Jahwa United Co Ltd	Shanghai
208	600518	Kangmei Pharmaceutical Co Ltd	Shanghai
209	600519	Kweichow Moutai Co Ltd	Shanghai
210	600597	Bright Dairy & Food Co Ltd	Shanghai
211	600600	Tsingtao Brewery Co Ltd	Shanghai
212	600809	Shanxi Xinghuacun Fen Wine Factory Co Ltd	Shanghai
213	600827	Shanghai Bailian Group Co.,Ltd.	Shanghai
214	600873	Meihua Holdings Group Co., Ltd	Shanghai
215	600887	Inner Mongolia Yili Industrial Group Co Ltd	Shanghai
216	600998	Jointown Pharmaceutical Group Co Ltd	Shanghai
217	601607	Shanghai Pharmaceuticals Holding Co.,Ltd	Shanghai
218	601933	Yonghui Superstores Co Ltd	Shanghai
Utilities			
219	000027	Shenzhen Energy Group Co Ltd	Shenzhen
220	000598	Chengdu Xingrong Investment Co Ltd	Shenzhen
221	000883	Hubei Energy Group Co Ltd	Shenzhen
222	600008	Beijing Capital Co Ltd	Shanghai
223	600011	Huaneng Power International Inc	Shanghai
224	600027	Huadian Power International Corporation Ltd	Shanghai

Sector	Symbol	Company	Exchange
225	600578	Beijing Jingneng Power Co Ltd	Shanghai
226	600642	Shenergy Co Ltd	Shanghai
227	600674	Sichuan Chuantou Energy Co Ltd	Shanghai
228	600795	GD Power Development Co Ltd	Shanghai
229	600863	Inner Mongolia Mengdian Huaneng Thermal Power Corp Ltd	Shanghai
230	600886	SDIC Power Holdings Co.,Ltd.	Shanghai
231	600900	China Yangtze Power Co Ltd	Shanghai
232	601158	Chongqing Water Group Co Ltd	Shanghai
Health Care			
233	000423	Shandong Dong-Ee Jiao Co Ltd	Shenzhen
234	000538	Yunnan Baiyao Group Co., Ltd.	Shenzhen
235	000623	Jilin Aodong Pharmaceutical Group Co., Ltd.	Shenzhen
236	000999	China Resources Sanjiu Medical & Pharmaceutical Co Ltd	Shenzhen
237	002001	Zhejiang NHU Co Ltd	Shenzhen
238	002007	Hualan Biological Engineering INC	Shenzhen
239	002038	Beijing SL Pharmaceutical Co Ltd	Shenzhen
240	002252	Shanghai RAAS Blood Products Co Ltd	Shenzhen
241	002294	Shenzhen Salubris Pharmaceuticals Co Ltd	Shenzhen
242	002399	Shenzhen Hepalink Pharmaceutical Co Ltd	Shenzhen
243	002422	Sichuan Kelun Pharmaceutical Co Ltd	Shenzhen
244	002603	Shijiazhuang Yiling Pharmaceutical Co Ltd	Shenzhen
245	002653	Xizang Haisco Pharmaceutical Group CO., LTD	Shenzhen
246	300015	Aier Eye Hospital Group Co Ltd	Shenzhen
247	600079	Humanwell Healthcare (Group) Co., Ltd.	Shanghai
248	600085	Beijing Tongrentang Co Ltd	Shanghai
249	600196	Shanghai Fosun Pharmaceutical (Group) Co Ltd	Shanghai
250	600252	Guangxi Wuzhou Zhongheng Group Co Ltd	Shanghai
251	600267	Zhejiang Hisun Pharmaceutical Co Ltd	Shanghai
252	600276	Jiangsu Hengrui Medicine Co Ltd	Shanghai
253	600332	Guangzhou Baiyunshan Pharmaceutical Holdings Co Ltd	Shanghai
254	600436	Zhangzhou Pientzehuang Pharmaceutical Co Ltd	Shanghai
255	600535	Tasly Pharmaceutical Group Co Ltd	Shanghai
256	600664	Harbin Pharmaceutical Group Co Ltd	Shanghai
257	600867	Tonghua Dongbao Pharmaceutical Co.,Ltd.	Shanghai
Information Technology			
258	000413	Dongxu Optoelectronic Technology Co., Ltd.	Shenzhen
259	000503	Searainbow Holding Corp	Shenzhen
260	000536	CPT Technology (Group) Co Ltd	Shenzhen
261	000725	BOE Technology Group Co Ltd	Shenzhen

Sector	Symbol	Company	Exchange
262	000970	Beijing Zhong Ke San Huan High-Tech Co Ltd	Shenzhen
263	002008	Han's Laser Technology Industry Group Co., Ltd.	Shenzhen
264	002065	DHC Software Co.,Ltd.	Shenzhen
265	002129	Tianjin Zhonghuan Semiconductor Co., Ltd.	Shenzhen
266	002153	Beijing Shiji Information Technology Co Ltd	Shenzhen
267	002230	Iflytek Co.,Ltd.	Shenzhen
268	002236	Zhejiang Dahua Technology Co Ltd	Shenzhen
269	002241	GoerTek Inc	Shenzhen
270	002410	Glodon Software Co Ltd	Shenzhen
271	002415	Hangzhou Hikvision Digital Technology Co Ltd	Shenzhen
272	002456	Shenzhen O-film Tech Co Ltd	Shenzhen
273	002475	Luxshare Precision Industry Co., Ltd.	Shenzhen
274	300017	Wangsu Science and Technology Co.,Ltd.	Shenzhen
275	600100	Tsinghua Tongfang Co Ltd	Shanghai
276	600271	Aisino Co.,Ltd	Shanghai
277	600570	Hundsun Technologies Inc.	Shanghai
278	600588	Yonyou Network Technology Co., Ltd.	Shanghai
279	600703	Sanan Optoelectronics Co.,Ltd	Shanghai
280	600718	Neusoft Corporation	Shanghai
281	600804	Dr. Peng Telecom&Media Group Co Ltd	Shanghai
282	601231	Universal Scientific Industrial (Shanghai) Co.,Ltd.	Shanghai
283	603000	People.cn CO.,LTD	Shanghai
Telecom Services			
284	000063	ZTE Corporation	Shenzhen
285	002465	Guangzhou Haige Communications Group Incorporated Company	Shenzhen
286	600050	China United Network Communications Co Ltd	Shanghai
287	600485	Beijing Xinwei Telecom Technology Group Co., Ltd.	Shanghai
288	600498	Fiberhome Telecommunication Technologies Co Ltd	Shanghai