UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Department of Economics

Three Essays in Financial Econometrics: Fractional Cointegration, Nonlinearities and Asynchronicities

by

Chi Wan Cheang

Thesis for the degree of Doctor of Philosophy

February 27, 2018

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES
Department of Economics

Doctor of Philosophy

THREE ESSAYS IN FINANCIAL ECONOMETRICS: FRACTIONAL COINTEGRATION, NONLINEARITIES AND ASYNCHRONICITIES

by Chi Wan Cheang

This thesis develops theoretical tools for the stylised facts of multivariate volatility processes and stock returns in financial markets. The first essay of this thesis contributes to the literature of fractionally cointegrated processes. Threshold adjustment is allowed in the error correction of bivariate fractionally cointegrated systems. Hypothesis testing for the presence of threshold and simulation evidence are provided to support the need of threshold specification in the adjustment dynamics of fractionally cointegrated processes. Empirical application considers the cointegrating relation and adjustment dynamics of S&P500 option implied volatility index spot and futures. Empirical finding shows that investors tend to hedge against volatility by using volatility-tracking products during market turbulence. The next two essays investigate some econometric issues that arise from the use of asynchronous data on modelling the joint dynamics of stock returns. The return correlation is inaccurate if asynchronicity is not taken into consideration. As a result, portfolio risk management can be highly distorted. Aiming to develop an accurate estimation on the return correlation dynamics, several econometric techniques are introduced to tackle this asynchronicity problem that allow financial practitioners to adequately adjust the asynchronous stock return series. This research also attempts to analyse asynchronicity problem as a measurement error problem, parameter estimates from the conventional vector autoregressive models are inconsistent if the vector of multivariate stock returns contains asynchronous returns. A good proxy of measurement error can effectively correct the asynchronous return vector and hence yield consistent parameter estimates.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Chi Wan Cheang , declare that the thesis entitled *Three Essays in Financial Econometrics: Fractional Cointegration, Nonlinearities and Asynchronicities* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:..

Date:..

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Jose Olmo, for his unwavering support, patience, and careful supervision during these years. I thank Maria Kyriacou and Jean-Yves Pitarakis for their helpful support and guidance throughout the thesis and during my stressful job market period. I also thank the remaining faculty who taught and inspired me during my time in University of Southampton.

My special thanks to my best friend Andrea Giovannetti for his intellectual inspiration and for the precious time shared during our PhD experience. I also thank my friends and fellow students in the department especially Liu Liu and Panos Giannarakis for their valuable discussions and encouragement.

This research would not materialise without the financial support from the Economic and Social Science Research Council.

Finally my deepest love to my parents and sisters for their endless encouragement, understanding and support.

Chapter 1

Introduction

Fractional cointegration is a generalised class of cointegrated systems which provides feasibility to estimate the fractional and cofractional memory orders of the time series, rather than fixing the memory parameters to be integer values. Empirical studies found that many macroeconomic and financial variables possess long memory in the long-run; however, not much focus has been given to the short-run adjustment dynamics of the fractional cointegrating relation. The feature of nonlinear adjustments in long-run equilibrium relation of cointegrating variables is separately documented in the strand of $I(1)/I(0)$ cointegration literature.

Nonstationarity and nonlinearity are key features in time series analysis, yet only a few literature attempted to describe long memory and nonlinearity within a time series model. Chapter 2 extends the fractional cointegrated vector autoregressive model (FCVAR) by allowing two regimes in the speed of adjustment parameter in the error-correction term, treating long memory features and cointegrating relation invariant across regimes. Since the threshold parameter is not identified under the null hypothesis of no threshold, supLM test for the presence of a threshold is proposed for the FCVAR model. Bootstrapping test statistic and p-value are derived. Monte Carlo simulation demonstrates the test under finite-sample maintains satisfactory size and power. A DGP from threshold FCVAR model provides simulation evidence about the effects of misspecification in long memory parameter and threshold adjustment from other cointegration models. The threshold FCVAR model is applied to the volatility index spot (VIX) and its related futures. VIX-tracking products are more attractive during market turbulence since investors treat VIX products as a security for hedging market downside risk. It is argued that the adjustment towards the relation of VIX spot and futures could be regime specific. Empirical result shows that in the two-regime case, the co-movement among the VIX spot and futures is insignificant in the contango regime, while the futures contributes significant error correction in the backwardation regime. This result provides a theoretical ground to explain the momentum strategy based on contango and backwardation commonly adopted by volatility traders.

Another focus is given to asynchronicity of stock returns and its related econometric problems. Accurately modelling the joint dynamics of stock returns across different markets is a fundamental prerequisite for understanding how crises propagate, how strongly markets co-move and for quantifying the risk characteristics of international portfolios containing assets from geographically different segments. Multivariate time series modellings have long been used to identify the return correlation and volatility dynamics across financial markets. Vector autoregressive moving average processes with multivariate generalised autoregressive conditional heteroscedastic variances are found to be appropriate to analyse the joint dynamics of return correlation and volatility, in a multivariate return series framework. Nonetheless, when attempting to track return correlations across time and markets, an important concern arises from the fact that assets trade at their local trading time hence causing raw return series based on closing prices to be effectively asynchronous. The multivariate time series models may lead to highly distorted correlation estimates if the presence of asynchronicity in the financial data is not taken into consideration. A biased correlation dynamic does not reflect the true underlying degree of co-movement among the financial assets, portfolio management and risk analyses can be mislead.

Chapter 3 and 4 develop formal econometric techniques that tackle this problem of return asynchronicity and offer a rigorous set of approaches that allow practitioner to construct suitably adjusted series for the purpose of portfolio risk management, dynamics correlation analysis, and many other financial applications on the multivariate financial assets. In particular, Chapter 3 proposes a synchronisation technique which is generalised from the benchmark synchronisation method by Burns et al. (1998). The synchronised correlation estimates by the proposed model are less restrictive than the benchmark method since the assumption of random walk stock prices is relaxed when formulating the synchronised returns. Empirical application considers a seven-market data series including stock price indices located from eastern to western countries. Empirical findings show that the correlation dynamics are larger for the estimated synchronised returns. In a one-day Value-at-Risk (VaR) back-testing exercise, the synchronised returns and synchronised correlations produce more reliable risk measures and superior forecasting performance than the VaR from asynchronous data. The portfolio value calculated from asynchronous data is more volatile since the portfolio recognises profit and loss from the stale market value overnight.

Chapter 4 analytically reviews the use of asynchronous data on the conventional vector autoregressive (VAR) models results in measurement error problem. The measurement error is defined as the difference between synchronous and asynchronous returns. It follows that the coefficient estimates are no longer consistent. Next, the maximum likelihood estimators from the conventional VAR model are proved to be asymptotically consistent if an adequate proxy is used for the measurement error. Several transformed

VAR models are provided as the solutions to this measurement error problem. Assuming the measurement error in asynchronous returns has a linear relation to the true synchronous returns, the unobserved measurement error can be explained in terms of observed asynchronous returns. A transformed VAR model given this assumption is derived from the asynchronous VAR model, in which the maximum likelihood estimators are asymptotically consistent. An additional spatial proxy is also discussed in which the spatial weighted matrix captures the time zone differences correspond to the later markets closing. Using the spatial proxy as a measurement error correction results in a spatio-temporal vector autoregressive model.

Chapter 2

Threshold fractionally cointegrated vector autoregressive model and application to volatility index

2.1 Introduction

Fractional cointegration is a generalisation of a standard $I(1)/I(0)$ cointegration which allows the time series variables and their error correction dynamics following fractionally integrated processes. This research aims to make two contributions to the fractional cointegration literature. The first is an econometric contribution to propose a two-regime threshold adjustment in the equilibrium error in the class of fractionally cointegrated vector autoregressive models (FCVAR) by Johansen (2008). The second is an empirical contribution to analyse the fractionally cointegrated relation of the S&P500 options implied volatility (VIX) and the one-month volatility futures using the proposed threshold model.

The proposed threshold FCVAR model is a bivariate fractionally cointegrated vector autoregressive process on a pair of long memory time series $X_t = (x_{1t}, x_{2t})$, with one cointegrating relation $\beta' X_t$ and a discontinuous two-regime switching in the adjustment coefficient α. The fractional and cofractional orders and the cointegrating relation are presumed to determine the intrinsic long memory characteristic of the time series processes in the long-run. In this sense, the long memory parameters d and b, the cointegrating vector β and the level parameter (if any) μ are unchanged; while the adjustment coefficient α and the short-run dynamic Γ (if any) are allowed to switch across regimes.

A two-step conditional Gaussian maximum likelihood is implemented for the estimation. In the first step, the parameters (d, b, μ) are estimated from the linear model. Given the parameter estimates $(\hat{d}, \hat{b}, \hat{\mu})$, the second step conducts a two-dimensional grid search over the cointegrating vector $\beta \in [\beta_L, \beta_U]$ and the threshold parameter $\gamma \in [\gamma_L, \gamma_U]$. To provide statistical evidence of the nonlinear specification in favour of the linear one, a supLM test for the presence of threshold is derived based on the threshold cointegration literature by Hansen and Seo (2002). The difficulty of a threshold test is that the asymptotic distribution is nonstandard due to the presence of nuisance parameters under null hypothesis. As a result, simulation-based bootstrap sampling null distribution and p-value are derived. According to a Monte-Carlo simulation, the size and power of the test are working well in small sample. Concerning the misspecification issues of a class of (both fractional and non-fractional) cointegration systems with (or without) threshold adjustment, this study also provides simulation evidence for a comparison of coefficient estimates from four different cointegration models. Similar to the theoretical results from Łasak (2010), misspecifying the long memory parameters results in inconsistent estimates of the adjustment coefficient. The adjustment dynamics are distorted in the linear cases.

In recent literature, empirical evidence usually showed that realised volatility possesses a long memory parameter d of 0.4 to 0.5 and implied volatility possesses a parameter d of 0.5 to 0.8, which correspond to long memory stationary and nonstationary, respectively. Semiparametric estimators of fractionally cointegrated systems see e.g. the narrow-band least squares and the exact local Whittle estimation, are found to be useful in modelling bivariate volatility processes. Yet they are developed for assessing long-run features through cointegrating vector and memory parameters.

Thanks to the attractive features of Johansen (2008)'s fractionally cointegrated VAR model, estimation of multiple cointegration relationship is allowed through the reduced-rank approach. The number of cointegrating ranks can be determined via statistical tests, hence the fractional and cofractional orders, cointegrating relations, adjustment coefficients, and the short-run dynamics can be jointly estimated. Each of these features is relevant to the empirical analysis of multiple volatility processes. While the memory parameters and cointegrating vector express the long-run characteristics of the volatility processes, the adjustment process and lead-lag dynamic provide the short-run information dependency which are particularly useful for empirical practice. Considering the cointegrated relation between implied volatility spot and futures, investors are interested in assessing the size of disequilibrium between the spot and futures and the forecast of disequilibrium correction in the next period. Hence, one can design short term investment decisions correspond to the estimate of adjustments.

Due to the empirical importance of the adjustment dynamics, this research investigates the presence of threshold adjustment in the equilibrium relation between the VIX spot and futures through the fractionally cointegrated VAR model. A large number of empirical studies in economic theory demonstrated the presence of stochastic nonlinear dependence in long-run equilibrium relations. For instance, the equilibrium model of prices parity exhibits disequilibrium adjustment only if the arbitrage is larger than transaction costs. However, lack of attention has been given to the nonlinear mean reversion of volatility processes. The intuition of nonlinear adjustment among the volatility spot and futures can be observed from the investment need of VIX futures. The S&P500 implied volatility index is a mathematical measure for how sharply the market moves. Since the VIX prices tend to skyrocket when the stock prices plunge, investors may use the volatility related derivatives as a mean to insure the possible market downturn. In practice, constantly longing a futures is not a sophisticated strategy for portfolio management. The generally negative correlation between the stock prices and VIX says that the cost of buying volatility "insurance" eats up profit from the investment portfolio. Technically speaking, futures contracts on VIX are normally in "contango", that is, the longer-dated futures prices are higher than the near-dated one and the spot. Thus, constantly rolling futures contracts means buying high and selling low which guaranteed loss to the portfolio in long-term. A legitimate use of volatility futures or other related derivatives is to buy volatility securities when market has high uncertainty. Hedging against volatility is more attractive when market is facing a bad time than a good time. This implies investors behave differently in volatility investment given different market conditions. In fact, some volatility traders who adopt momentum investing strategy use the temporary "contango" and "backwardation" conditions of VIX as the enter-and-exit signal. For example, buy when you need - one can long the VIX-tracking ETFs when futures is in backwardation, and short when futures is back to contango.

Several findings are reported in empirical application. First, there is a significant threshold effect in the equilibrium relation $\beta' X_t$ where $X_t = (s_t, f_t)$ denotes implied volatility spot and futures. Assuming a constant cointegrating vector β in the long-run, the adjustment coefficients are regime-dependent. Second, the cointegrating relationship vanished when the disequilibrium does not exceed the threshold. On the other hand, only the futures exhibits significant adjustment when the disequilibrium exceeds the threshold. This result provides empirical regularity to the nonlinear responses towards the disequilibrium of implied volatility spot and futures which is concealed in the linear FCVAR model. Spot has no adjustment effect on the equilibrium relation since the VIX spot is a non-tradable measure. In addition, futures seems not to be attractive until it is significantly below the spot prices.

The remainder of the chapter is organised as follows. Section 2.2 provides an extensive literature review on fractional cointegration, threshold cointegration, and their intersection. Section 2.3 introduces the formulation of threshold FCVAR model, derives the estimation procedure and briefly discusses other forms of nonlinearity. Section 2.4 proposes a supLM test for the presence of threshold in fractional cointegration. The bootstrapping test statistic and p-value are also discussed. Section 2.5 reports Monte

Carlo simulation evidence for the size and power of the test and the misspecification issues by ignoring fractional parameters and threshold nonlinearity. Section 2.6 applies the threshold FCVAR model to the fractional cointegration of the S&P500 implied volatility index and its one-month futures.

2.2 Literature review

Fractional cointegrated systems laid the ground for the long-run relationship of multiple volatility processes. Threshold specification provides feasibility to take nonlinear relations into account in the short-run adjustment dynamics among the variables. This chapter provides a review of the major econometric work and empirical findings on fractional cointegration and threshold models.

2.2.1 Fractional cointegration

Cointegration of nonstationary time series had been studied intensively over the last three decades since the seminal contributions of Granger (1981) and Engle and Granger (1987). The early studies of cointegration focused on the standard $I(1)/I(0)$ type of cointegration in which the linear combinations of the $I(1)$ nonstationary processes are $I(0)$ stationary. Fully parametric inference of cointegration in the error correction mechanism was developed by Johansen (1988, 1991, 1995). In Johansen's framework, cointegration is modelled by the vector autoregressive model for nonstationary variables. The cointegrating vector, the speed of adjustment towards the long run cointegrated relation and the short run dynamics are allowed to be estimated after selecting a cointegrating rank. Empirical applications adopted Johansen's cointegration in the studies of prices parities[1] by rational expectations theory in which the model based expectation restrictions provide testable information on cointegrating relations and short-run adjustments.

However, the premises of the standard $I(1)/I(0)$ cointegration that the time series variables are integrated of order one and co-moved of order zero are somehow restrictive. Substantial evidence in the literature demonstrated that many economic and financial time series possess long range dependence in the autocorrelation function, but not exactly exhibit a unit root process, so called the long memory process. For instance, the volatility of asset prices, forward premium, exchange rates, interest rate term structure, and see Baillie (1996) for a comprehensive review of long memory processes in econometrics. Granger and Joyeux (1980) and Hosking (1981) defined a time series X_t be a fractionally integrated process of order d for non-integer $d \in (0, 1)$, denote $X_t \in I(d)$, if it has a $I(0)$ stationary, invertible autoregressive moving average representation after

[1] For instance, the parity relations include interest rate parity, purchasing power parity, commodity spot and futures (see Hansen et al., 1981).

fractional differencing, i.e. $\Delta^d X_t \in I(0)$. The precise definition is:

$$\Delta^d X_t = \sum_{n=0}^{\infty} \pi_n(-d) X_{t-n} \tag{2.1}$$

where the parameter d determines the memory of the process, and the fractional coefficients, $\pi_n(-d)$, are defined by the binomial expansion[2]

$$\begin{aligned}(1-Z)^{-d} &= \sum_{n=0}^{\infty} (-1)^n \left(\begin{array}{c} -d \\ n \end{array} \right) Z^n \\ &= \sum_{n=0}^{\infty} \frac{d(d+1)...(d+n-1)}{n!} Z^n = \frac{\Gamma(d+n)}{\Gamma(d)\Gamma(n+1)}, \qquad |Z|<1, \qquad d \in \mathbb{R}\end{aligned}$$

in which the fractional coefficients are $O(n^{d-1})$. This shows that, for a valid $d < 1/2$, $\sum_{n=0}^{\infty} \binom{-d}{n}^2 < \infty$. Hence the fractional process X_t is stationary process with finite variance and can be written as

$$\Delta^{-d}\epsilon_t = (1-L)^{-d}\epsilon_t = \sum_{n=0}^{\infty} (-1)^n \left(\begin{array}{c} -d \\ n \end{array} \right) \epsilon_{t-n}$$

with ϵ_t the *iid* variables with zero mean and finite variance. For $d \geq 1/2$, the infinite sum does not exist; yet a nonstationary $I(d)$ process is defined by partial sums by the operator Δ_+^{-d} as

$$\Delta_+^{-d}\epsilon_t = (1-L)_+^{-d}\epsilon_t = \sum_{n=0}^{t-1} (-1)^n \left(\begin{array}{c} -d \\ n \end{array} \right) \epsilon_{t-n}, \qquad t = 1, ..., T.$$

The semiparametric estimators of the long memory parameter d have been developed in the literature include the log-periodogram estimator, see Geweke and Porter-Hudak (1983), Robinson (1995a) and some modified versions by Andrews and Guggenberger (2003) and Sun and Phillips (2003), and the Gaussian semiparametric local Whittle estimator by Künsch (1987), Robinson (1995b), Andrews and Sun (2004) and Shimotsu and Phillips (2005), among others. The standard unit root and stationarity hypothesis tests may result in contradicting evidence if a time series possesses long memory. The semiparametric estimators of the fractional integrated order are helpful for supporting the long memory fashion in the time series process. Noteworthy the fully parametric long memory estimator, e.g. the ARFIMA model, is more efficient, but it is inconsistent if the parametric form such as the number of lags is misspecified.

One of the most important motivations distinguishing the long memory from $I(0)$ and $I(1)$ processes is that long memory processes imply different predictions of long run dynamics and effects of shocks to the macroeconomic variables. The arbitrary restriction

[2]Further definition and intermediate results of the fractional difference operator can be seen in Johansen (2008) and Johansen and Nielsen (2012).

of integrated orders d to integer values in the $I(1)/I(0)$ cointegration framework will result in misspecified likelihood function and incorrect statistical inference. The empirical discussions regarding the fractional cointegrated economic variables can be dated back to 1990s by Baillie and Bollerslev (1994), Cheung and Lai (1993), Dueker and Startz (1998) and Mohanty et al. (1998), among others, which argued that the macroeconomic variables possess long memory may be well described as fractional integrated process, and fitting the fractional integrated variables into the standard case may result in false rejection of the economic hypotheses.

Granger (1986) and Engle and Granger (1987) provided representation theorem for a general case of cointegration that the variables are fractionally integrated of non-integer order d and cofractional of a smaller order $(d-b)$, for $0 \leq d-b < 1/2$. The parameter b determines the degree of cointegration among the variables. In the early studies of long memory data, the autoregressive fractionally integrated moving average (ARFIMA) model is a feasible parametric method for the fractional cointegration which extended the ARMA model to account for hyperbolic rate of decaying autocorrelations. As discussed by Robinson (1991, 1994) and Robinson and Hidalgo (1997), the complication is that, when the cointegrating errors have long memory, they are correlated at long horizons and thus rendering the ordinary least squares estimators inconsistent. The recent study by Haldrup et al. (2017) showed that the standard ARFIMA models fail to fit the dynamics of the fractional processes properly when the long memory is caused by aggregation. Having made these points, several semiparametric approaches had been developed for the fractionally cointegrated processes. Robinson (1994) derived a semi-parametric consistent narrow-band frequency domain least squares estimator (henceforth NBLS) that performs OLS on a degenerating band of frequencies around the origin. The NBLS in the multivariate long memory systems had been developed by Lobato (1997), Marinucci and Robinson (2001). Hence, Robinson and Marinucci (2003) established limiting distribution of the NBLS of the cointegrating relation for the long memory nonstationary case $d > 1/2$, $d-b \geq 0$. Christensen and Nielsen (2006) provided complementary asymptotic distribution theory for the NBLS estimator in the stationary case $d > 0$, $d-b \geq 0$, $2d-b \leq 1/2$. The consistency and asymptotic nomality of the local Whittle estimator in a bivariate stationary cointegrated system were derived by Robinson (2008). A two-step estimation of the exact local Whittle estimator of a bivariate fractionally cointegrated system was proposed by Shimotsu (2012) which accommodates both stationary $(d \leq 1/2)$ and nonstationary $(d > 1/2)$ integrated and/or cointegrated order.

Much attention has been given to the long run cointegrating relation of the fractionally cointegrated systems through the semiparametric approaches, Johansen (2008) generalised the well-known $I(1)/I(0)$ cointegrated vector autoregressive model of Johansen (1995) by allowing for modelling the long memory time series with non-integer order of integration. The fractionally cointegrated vector autoregressive (FCVAR) model permits one to determine the number of equilibrium relations via cointegrating rank test, and to jointly estimate the memory parameters, the long-run cointegrating relations with the adjustment parameters, and the short-run lagged dynamics. The FCVAR model allows for flexibility to estimate the long memory parameters of the cointegrated systems; it reduces to the standard cointegration case when the fractional and cofractional orders equal to unity. Asymptotic theory for likelihood estimation and inference in the fractionally cointegrated VAR model was developed by Johansen and Nielsen (2010, 2012). The role of observed and unobserved initial values in conditional maximum likelihood estimators of the nonstationary fractional systems was analysed by Johansen and Nielsen (2016). The inclusion of a level parameter in the FCVAR formulation is found to have the advantage of reducing the bias of pre-sample observations. Dolatabadi et al. (2016) provided an additional formulation of deterministic term which allows for deterministic linear time trends or drift in the variables.

Empirical relevance of the fractionally cointegrated VAR system has been recognised in the areas of financial markets and political economics. Dolatabadi et al. (2015, 2016) applied the FCVAR model for the analysis of price discovery in commodity spot and futures for five non-ferrous metals (aluminium, copper, lead, nickel and zinc). The discussions of long-run contango or backwardation characteristic and disequilibrium errors are allowed through equilibrium relation. By using the same data from Figuerola-Ferretti and Gonzalo (2010) who instead applied the standard cointegrated VAR model, Dolatabadi et al. (2015, 2016) found more support of price discovery in the spot compared to the result from the non-fractional case. Other applications in empirical finance include a no-arbitrage relation between spot and futures (Rossi and Santucci de Magistris, 2013), and the stock prices predictability from a relation of high and low prices (Caporin et al., 2013). Regarding the studies in social science, Jones et al. (2014) examined the fractionally cointegrated relationship between Canadian political support and macroeconomic conditions.

2.2.2 Threshold in the error correction mechanism

One potential limitation of the cointegration model in relation to empirical application is that, due to the linear combinations of time series variables linked through the long-run cointegrating relationships, the adjustment processes tend to correct any small deviation from the long-run equilibrium. A large number of economic studies[3] showed that key macroeconomic data such as output growth, interest rate term structure, commodity prices with transaction costs, and the stock return volatility exhibit nonlinear adjustment over the business cycle.

Threshold autoregressive (TAR) model was first developed by Tong and Lim (1980). The model allows the autoregression to depend on the state of the transition variable of interest in which the process is piecewise linear in the threshold space. A useful extension

[3]See Hansen (2011) for a comprehensive literature of threshold autoregressive models.

of the univariate threshold autoregresion to a multivariate cointegrated system was made by Balke and Fomby (1997). Their threshold cointegration model allows the equilibrium errors in the cointegrating relation to follow a self-exacting threshold autoregressive process, such that the error correction parameter exhibits regime-switching dynamics. For the AR(1) process as an example, the adjustment z_t takes autoregression depending on the state of the previous value:

$$z_t = \begin{cases} \rho_l z_{t-1} + u_t & \text{if } \; z_{t-1} \leq \theta_l \\ \rho_m z_{t-1} + u_t & \text{if } \; \theta_l \leq z_{t-1} \leq \theta_m \\ \rho_h z_{t-1} + u_t & \text{if } \; \theta_h \leq z_{t-1} \end{cases} \tag{2.2}$$

The coefficients ρ_l, $\rho_h < 1$ are the sufficient condition for stationarity[4]. The inner regime is allowed to have a unit root. This type of discrete adjustment dynamics is appropriate to describe many economic phenomena. In macroeconomics, the exchange rate might be maintained around a certain bound, in such case policy intervention only takes place when the exchange rate is deviated alove or below the range. In asset pricing, sometimes arbitrage can be too small to be profitable with the presence of transaction costs.

There are several other model specifications of the threshold cointegration in literature. Enders and Siklos (2001) generalised the Engle-Granger cointegration test to accommodate the threshold autoregressive and momentum-threshold autoregressive in multivariate content. Hansen and Seo (2002) adopted Johansen (1995)'s reduced rank approach, developed maximum likelihood estimation for a bivariate threshold cointegrated VECM model. A supLM test for testing the presence of a threshold in the adjustment process was also derived. Gonzalo and Pitarakis (2006a) proposed an alternative threshold framework to allow the long-run cointegrating relationship itself to be regime-specific whereas the adjustment coefficient is linear. Gonzalo and Pitarakis (2006b) used an external variable rather than the equilibrium deviation as a transition variable of interest. It is easier to perform model estimation, yet the current threshold tests restrict the external variable to be a stationary. Also, one would need to convince why an influencing variable is included in the threshold switching dynamics but not in the main equation.

Despite the empirical importance of threshold models, there are complications in developing asymptotic theory and statistical inference of threshold estimation. Chan (1993) showed that in the univariate self-exciting model with a single threshold (SETAR), the least squares estimator of the threshold parameter $\hat{\theta}$ is n-consistent (or said to be superconsistent) whereas the estimator of autoregressive slope parameter $\hat{\rho}$ is $\sqrt{n}$-consistent. The limiting distribution of the threshold parameter is a compound Poisson process. The super consistence of the threshold estimator allows the estimated value to take as given asymptotically when conducting inference on $\hat{\rho}$. The inference on threshold parameters was provided by Hansen (1996, 1997, 1999, 2000). The statistical tests for

[4]See Balke and Fomby (1997) for a set of sufficient conditions in a more general TAR model.

nonlinearity has nonstandard distribution since the threshold parameter under null hypothesis remains unidentified. The tabulation of asymptotic distribution of the test is not feasible. The sup-test statistics (Davies, 1987) [5] are employed and the asymptotic null distribution is obtained through bootstrap methods.

The econometric theory of threshold cointegration is more difficult than the univariate case. Allowing for joint estimation of the cointegrating relation β and adjustment dynamic α complicates the asymptotic distributions of the quasi-maximum likelihood estimators and the test statistics. It is because the limiting distributions of the estimators of cointegrating vector and adjustment parameter are asymptotically dependent (Kristensen and Rahbek, 2013). Thus, when testing for linearity in the adjustment dynamics through sup-test inference, the estimation of long-run parameter should not be ignored if it is unknown. The existing literature addressed this issue by assuming the cointegrating vector β is known (see, e.g. Gonzalo and Pitarakis, 2006b; Seo, 2006) or the estimation error of β will not affect the asymptotic behaviour of the test statistic (Hansen and Seo, 2002). When the parameter β is known, all regressors can be treated as stationary, hence the sup-testing is reduced to the testing framework in Hansen (1996). In order to deal with the tests of nonlinearity and nonstationarity simultaneously, Kristensen and Rahbek (2013) provided full asymptotic theory to the case with the estimation of β in the multivariate nonlinear VECMs. The limiting distribution of sup test statistic consists of two components, one is a stationary component from the short-run adjustment dynamics, another is a nonstationary component from the long-run cointegrating parameter. The results are closely related to the previous studies by Caner and Hansen (2001) and Shi and Phillips (2012) in univariate threshold case with unit root under weak identification. Some other studies proposed a Taylor approximation of the nonlinear component for the test of linearity in cointegrating systems, see for instance, the approximation frameworks of Saikkonen and Choi (2004) and Kapetanios et al. (2006).

2.2.3 Nonstationarity and nonlinearity

Nonstationarity and nonlinearity are two of the key features in time series analysis. Interestingly, only few of them attempted to describe long memory and nonlinearity simultaneously within a time series model. Van Dijk et al. (2002) was the first empirical study considered the coexistence of nonlinear and long memory properties in the US unemployment rate. They introduced a fractionally integrated smooth transition (FI-STAR) model to capture the asymmetric behaviour over the business cycle (during expansion and recession), and found that the FI-STAR model outperforms the regular STAR model regarding the measure of fit. Haldrup and Nielsen (2006) developed a regime dependent vector autoregressive model allows the degree of long memory to be different in different regime states. Their model was applied to the electricity prices in

[5] See also Tsay (1998), Lee et al. (2011), and among others for sup-type threshold tests.

Nordic countries, and resulted that the price behaviour is subject to occasional congestion periods in which the dynamic of two long memory electricity prices is bilateral in the congestion state but it exhibits fractional cointegration in the non-congestion state.

The combination of nonstationarity and nonlinearity gives rise to some challenges. The features of these two processes can easily be mistaken for each other. A long memory process has autocorrelation function with hyperbolic decay at long lags; nevertheless, a short memory process with regime switching or a smooth trend has similar autocorrelation feature. The classic work of unit root by Perron (1989) emphasised the standard tests of unit root can misinterpret a stationary series with one-time break as an $I(1)$ process with drift. In the recent long memory literature, Granger and Teräsvirta (1999), Diebold and Inoue (2001) and Granger and Hyung (2004) demonstrated that the stochastic regime switching can be fitted into long memory models and result in spurious conclusion of long memory. Qu (2011) based on the derivative of profile local Whittle likelihood function, proposed a test for the null hypothesis that a stationary time series has long memory against the alternative that it has regime switching. Although some studies argued the possibility of confusing long memory and nonlinearity, some other addressed statistical tests for the coexistence of long memory dynamics and nonlinearity in short memory dynamics, see for instance, Baillie and Kapetanios (2007) and Choi and Saikkonen (2010). This research contributes to this area by jointly estimating the long-run fractionally cointegrated relationship and the short-run threshold adjustment on volatility data, and specifying a fractionally cointegrated models against nonlinear alternatives.

On the other hand, a theoretical work by Dittmann and Granger (2002) derived the properties of various nonlinear transformations of discrete time[6] fractionally intergrated processes. They showed that memory parameter d of the nonlinear transformed fractionally integrated series may or may not reduce depending on whether the initial process is stationary or not. Any nonlinear transformation of an anti-persistent $I(d)$ process with $d < 0$ is $I(0)$. Nonlinear transformation of a stationary long memory process with $d \in (0, 1/2)$ results in either the same value of d or a smaller d depending on the Hermite rank of the transformation. The polynomial transformations of nonstationary process with $d \in (1/2, 1]$ still have very persistence nonstationary long memory dynamics and the long memory parameter d is just slightly smaller than the initial one. Their simulation results provide conjecture in support of the adequacy of nonlinear extension in the error correction dynamics in this study. Referring to equation (2.3), the long-run equilibrium relation in the FCVAR model, $\beta' X_t$ possesses memory of order $d - b \in [0, 1]$. Therefore, introducing nonlinearity in the adjustment process implies a nonlinear transformation to the equilibrium relation which may affect the degree of long memory in the initial process. Dittmann and Granger (2002) showed that the memory parameter

[6]Two similar studies, Taqqu (1979) and Giraitis and Surgailis (1985), prior to Dittmann and Granger (2002) worked out continuous time long-memory properties of nonlinear transformations of fractional Brownian motions.

of the polynomial transformed break process decreases only slightly. More interestingly, a logistic transformation of stationary long memory process retains exactly the same memory parameter as in the initial process. For nonstationary process, logistic transformation reduces the memory order but it is still larger than a half. Those indicate that nonlinear transformations do not seriously distort the memory dynamics of data in the long-run.

2.3 Econometric models

The representation theorem of the fractionally cointegrated vector autoregressive (FCVAR) model (Johansen, 2008) is generalised from the $I(1)/I(0)$ cointegrated vector autoregressive (CVAR) mechanism (Johansen, 1995) to allow for fractional processes of order d that cointegrate to order $d-b$. The recent development of FCVAR framework only accommodates linear and symmetric adjustment in the error correction term. In this chapter, the fractional cointegrated VAR process permits threshold adjustment towards the long-run equilibrium relation. The proposed threshold FCVAR model provides feasibility to assess long memory and nonlinerity which are two important empirical features of time series. With certain restrictions on the parameters, the threshold FCVAR model effectively reduces to other cointegration models include:

- the linear FCVAR model when the regime-specific parameter sets Φ_1 and Φ_2 in regime 1 and 2 respectively are not statistically different from each other;
- the $I(1)/I(0)$ threshold cointegration model (Hansen and Seo, 2002) when $d = b = 1$; and
- the $I(1)/I(0)$ CVAR model when both $d = b = 1$ and $\Phi_1 = \Phi_2$.

2.3.1 The fractionally cointegrated VAR models

To derive the FCVAR model, it is straightforward to begin with the well-known CVAR model with nonfractional integrated order. The CVAR model for a p-dimensional nonstationary time series Z_t is

$$\begin{aligned}\Delta Z_t &= \alpha\beta' Z_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Z_{t-i} + \epsilon_t \\ &= \alpha\beta' L Z_t + \sum_{i=1}^{k} \Gamma_i \Delta L^i Z_t + \epsilon_t, \quad t = 1, ..., T \end{aligned} \tag{2.3}$$

where $\Delta Z_{t-i} = Z_{t-i} - Z_{t-i-1}$ and Z_{t-i} can be written as $L^i Z_t$ by the use of lag operator L. In the standard cointegration framework, the time series Z_t is $I(1)$ and their cofractional $\beta' Z_t$ is $I(0)$.

The FCVAR model allows Z_t to be fractional integrated of order d and $\beta' Z_t$ to be fractional of order $d-b \geq 0$ which can be built from (2.3) in two steps. First, the usual lag operator $L = 1 - \Delta$ and the difference operator Δ are replaced by their fractional counterparts, $L_b = 1 - \Delta^b$ and $\Delta^b = (1-L)^b$, where Δ^b is defined by the binomial expansion $\Delta^b Z_t = \sum_{n=0}^{\infty}(-1)^n \binom{b}{n} Z_{t-n}$. It results:

$$\Delta^b Z_t = \alpha\beta' L_b Z_t + \sum_{i=1}^{k} \Gamma_i \Delta^b L_b^i Z_t + \epsilon_t \tag{2.4}$$

Next, applying model (2.4) to $Z_t = \Delta^{d-b} X_t$. This defines the fractional cointegrated VAR as

$$\Delta^d X_t = \alpha\beta' L_b \Delta^{d-b} X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_{t-i} + \epsilon_t, \quad t = 1, ..., T \tag{2.5}$$

where ϵ_t is p-dimensional $i.i.d.(0, \Omega)$ and Ω is a positive definite. The fractional parameter d determines the fractional integration order of the variables X_t, while the cofractional paramter b determines the degree of cointegration, i.e. the amount of reduction in fractional integration. It appears that both $\Delta^d X_t$ and $\Delta^{d-b}\beta' X_t$ are $I(0)$. The long-run parameter β and the adjustment parameter α are the $p \times r$ matrices where r denote the number of cointegrating ranks and $0 \leq r \leq p$. The parameters $\Gamma_i = (\Gamma_1, ..., \Gamma_k)$ govern the short-run dynamics from lag terms. Denote the product of α and β as $\Pi = \alpha\beta'$ a $p \times p$ matrix. When cointegrating relations exist, the elements of $\beta' X_t$ are the linear combinations of the variables in the system. The adjustment coefficients α determine the speed of adjustment towards the long-run equilibrium.

In some special cases, when $r = p$, the matrix Π is unrestricted; when $r = 0$, the cointegrating relation is not present; and when $r = k = 0$, the model reduces to $\Delta^d X_t = \epsilon_t$. It might be helpful to note that the cointegrating relation $\beta' X_t$ enters into the right-hand-side of (2.5) with time period t but one is only required to provide observations up to time $t-1$ to calculate the fractional difference of $\beta' X_t$. To see this, the fractional process X_t is govern by the fractional lag operator L_b such that $L_b X_t = (1 - \Delta^b) X_t$ is an infinite series starts from X_{t-1} to negative infinity.

Deterministic terms can be assumed in the FCVAR model in several ways. Johansen and Nielsen (2012) considered the inclusion of restricted constant ρ in the long-run cointegrating relation. Dolatabadi et al. (2016) suggested an unrestricted constant ξ as the linear trend of the fractionally integrated processes. The general formulation of the FCVAR with deterministic terms can be written as

$$\Delta^d X_t = \xi + \alpha L_b \Delta^{d-b}(\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_{t-i} + \epsilon_t. \tag{2.6}$$

The restricted constant is interpreted as the mean level of equilibrium relation; on the other hand, the unrestricted constant is the level of the fractionally differenced variables. According to (2.6), if $d = b = 1$, the restricted constant becomes $\alpha\rho'$ and will be reduced to unrestricted constant. Otherwise, the FCVAR model allows both unrestricted and restricted constants exist in the model at the same time.

There is an important issue about sampling fractional processes. Fractional difference (see equation (2.1)) is defined in terms of an infinite series; however, sample data only has finite number of observations. Consequently, the fractional difference calculated from the observed sample is not the same as defined. To reduce sampling bias, one can assume the fractional series X_t has initial value of zero $X_0 = 0$ in sample; albeit economic data seldom initiates from zero. Another solution is to assume a sample length of $N + T$ on X_t where N is the number of initial values for conditioning and T is the number of observations for modelling. Johansen and Nielsen (2016) proposed a simpler model to get rid of the impact of pre-sample observations. The fractionally integrated variables are assumed to have non-zero level parameter μ. The level parameter in model (2.5) shifts the fractional series by a constant:

$$\Delta^d (X_t - \mu) = \xi + \alpha\beta' L_b \Delta^{d-b}(X_t - \mu) + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i (X_{t-i} - \mu) + \epsilon_t, \tag{2.7}$$

which is equivalent to the inclusion of restricted constant by having $\beta'\mu = \rho'$. Johansen and Nielsen (2016) showed that the formulation (2.7) has advantage of reducing sampling bias even when conditioning on no initial values.

The FCVAR models (2.5 and 2.6) are estimated by conditional maximum likelihood with respect to the parameter set $\lambda = (d, b, \alpha, \beta, \xi, \rho, \Gamma_i)$, conditional on N initial values. For model (2.7), the number of initial values can be zero. Johansen and Nielsen (2012) showed that, for a fixed pair of memory parameters (d, b), the estimation of FCVAR model is reduced to reduced rank regression as in Johansen (1995) of $\Delta^d X_t$ on $\Delta^{d-b} L_b X_t$ corrected for $\{\Delta^d L_b^i X_t\}_{i=1}^k$. In this way, the parameters $(\alpha, \beta, \xi, \rho, \Gamma_i)$ can be concentrated out from the likelihood function of model (2.5 or 2.6) and estimation of (d, b) can be done by maximising the profile likelihood function over the memory parameters d and b only. With the presence of level parameter μ in (2.7), the optimisation is conducted over $(d, b,\ \mu)$.

Asymptotic theory of the quasi-maximum likelihood estimators is provided by Johansen and Nielsen (2012). Note that the parameter value $b_0 = 1/2$ is a singular point thus inference is different for $b_0 < 1/2$ or $b_0 > 1/2$. For $0 < b_0 < 1/2$ with b_0 denotes the true value of b, the limit distributions are standard Gaussian for $(\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{\Gamma}_i)$, or mixed Gaussian for $(\hat{\beta}, \hat{\rho})$ with the presence of restricted constant ρ, and chi-squared for likelihood ratio statistic of $\Pi = \alpha\beta'$. For $b_0 > 1/2$, the limit distributions of estimators for $(\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Gamma}_i)$ are Gaussian, while the distribution of $T^{b_0}(\hat{\beta} - \beta_0)$ is mixed Gaussian. The asymptotic distributions for other deterministic terms $\hat{\xi}$ and $\hat{\mu}$ remain unknown in the literature.

2.3.2 FCVAR model with threshold adjustment

The main extension to the fractionally cointegrated vector autoregressive models (2.5-2.7) is to allow the speed of adjustment parameter α to differs across two regimes depending on a threshold parameter. Threshold cointegration has been widely adopted in macroeconomics and finance to analyse nonlinear dynamics in $I(1)$ nonstationary systems. This research takes a further step to provide flexibility in the degree of long memory of the nonstationary processes.

A bivariate two-regime threshold FCVAR model for a pair of fractionally cointegrated processes $X_t = (x_{1t}, x_{2t})$ is formulated as

$$\begin{aligned}\Delta^d(X_t-\mu) =& \Big(\Delta^{d-b}L_b\alpha_1\beta'(X_t-\mu)+\sum_{i=1}^{k}\Gamma_{i,1}\Delta^d L_b^i(X_{t-i}-\mu)\Big)D_{1t}(\beta,\gamma)\\ &+\Big(\Delta^{d-b}L_b\alpha_2\beta'(X_t-\mu)+\sum_{i=1}^{k}\Gamma_{i,2}\Delta^d L_b^i(X_{t-i}-\mu)\Big)D_{2t}(\beta,\gamma)+\epsilon_t \end{aligned} \tag{2.8}$$

where $D_{1t}(\beta,\gamma) = 1(e_{t-1} \le \gamma)$, $D_{2t}(\beta,\gamma) = 1(e_{t-1} > \gamma)$ in which $1(\cdot)$ denotes the indicator function and e_{t-1} equals to the disequilibrium error $\beta'(X_t-\mu)$ in the previous period.

Threshold FCVAR model (2.8) exhibits regime switching depending on the value of the error-correction term e_t in the previous period. The equilibrium deviation is not the only specification for the threshold variables. Other choices of threshold variables can be an external factor (see Gonzalo and Pitarakis, 2006b) or the lagged difference of the dependent variable (see Caner and Hansen, 2001) which are expected to drive the regime-specific adjustment dynamics. In general, it is only necessary to ensure the threshold variable is stationary and ergodic with a continuous distribution function.

Threshold effect has content only if the error correction term e_{t-1} satisfies $0 < P(e_{t-1} \le \gamma) < 1$, otherwise the model reduces to linear FCVAR model. A constraint, $\pi_0 \le P(e_{t-1} \le \gamma) \le 1-\pi_0$, is imposed to guarantee each regime has no less than π_0% of observations of the total sample. The pre-determine trimming parameter, π_0, is chosen by concerning two aspects. One is that the critical values of supLM statistic will increase as π_0 decreases and the distribution of the test statistic diverges to positive infinity as $\pi_0 \to 0$. The power of the test will be reduced if the choice of π_0 is too close to the endpoint. On the other hand, it is desirable to choose a value of π_0 so that the trimming parameter given the true value of threshold lies in the interval $[\pi_0, 1-\pi_0]$. To balance the trade-off among these considerations, Andrews (1993) suggested π_0 may take a value between 0.05 and 0.15 according to his simulation result.

For model (2.8), the long memory parameters d, b and the cointegrating relation $\beta'(X_t-\mu)$ are presumed to determine the intrinsic long memory characteristic of the variables

X_t in the long-run, thus they are assumed to be unchanged across regimes. The short-run dynamics of the time series, include adjustment coefficient α and lagged terms coefficients Γ_i, are assumed to be different across regimes.

The above model specification is somehow similar to Van Dijk et al. (2002) who modelled the nonlinear and long memory dynamics of the US unemployment by a fractionally integrated smooth transition autoregressive (FI-STAR) model. Their model aimed to capture the nonlinear responses of US unemployment to economic shocks during recession and expansion. The long-run properties of the time series are considered as the characteristics of the time series itself, are restricted to be constant; however, the short-run dynamics react to economic shock differently, are allowed to vary depending on the indicator function. The proposed model of this study incorporates the idea of Van Dijk et al. (2002) into a multivariate content, in which the nonlinearity is coming from the disequilibrium among the time series and affecting the short-run dynamics but not deviating the long-run properties. Although it is not the focus of this study, the fractionally cointegrated VAR models can also be extended to have regime-specific long memory parameters; for instance, if the time series of interest are believed to have different long memory characteristics under different states of switching variable.

It is relevant to note that, in model (2.8), the level parameter μ and its implicit form of restricted constant $\rho' = \beta'\mu$ measure the mean level of cointegrating relation, thus it is considered to be constant at different regimes. If one is interested in regime dependent deterministic term, the alternative form with regime dependent unrestricted constant ξ can be taken as

$$\begin{aligned}\Delta^d X_t =& \Big(\xi_1 + \Delta^{d-b} L_b \alpha_1 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,1} \Delta^d L_b^i X_{t-i}\Big) D_{1t}(\beta,\gamma) \\ &+ \Big(\xi_2 + \Delta^{d-b} L_b \alpha_2 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,2} \Delta^d L_b^i X_{t-i}\Big) D_{2t}(\beta,\gamma) + \epsilon_t. \qquad (2.9)\end{aligned}$$

The Matlab programme for threshold FCVAR models (2.8 and 2.9) provided in appendix B is written based on Nielsen and Popiel (2016) and Hansen and Seo (2002). There are options to opt-in level parameter and/or unrestricted constant which are compatible to Nielsen and Popiel (2016)'s Matlab programme for linear FCVAR model.

2.3.3 Estimation

The estimation of a bivariate two-regime threshold FCVAR model can be done by maximum likelihood estimation (MLE) in two steps. Under the assumption that the residuals ϵ_t are $i.i.d.(0,\Omega)$, the Gaussian likelihood function from model (2.8) with $N = 0$ initial values is given as

$$-2T^{-1}\log\mathcal{L}_T(\lambda) = \log\big(\det(\Omega)\big) + \mathrm{tr}\Big(\Omega^{-1}T^{-1}\sum_{t=1}^{T}\epsilon_t(\lambda)\epsilon_t(\lambda)'\Big), \tag{2.10}$$

where the MLE parameter set $\lambda = (d, b, \mu, \Omega, \beta, \alpha_1, \alpha_2, \Gamma_{i,1}, \Gamma_{i,2}, \gamma)$ involves two sets of regime-specific parameters which are denoted $\Phi_1 = (\alpha_1, \Gamma_{i,1})'$ and $\Phi_2 = (\alpha_2, \Gamma_{i,2})'$ henceforth, and the residuals

$$\begin{aligned}\epsilon_t(\lambda) =& \Delta^d(X_t-\mu) - \Big(\Delta^{d-b}L_b\alpha_1\beta'(X_t-\mu) + \sum_{i=1}^{k}\Gamma_{i,1}\Delta^d L_b^i(X_{t-i}-\mu)\Big)D_{1t}(\beta,\gamma)\\ &- \Big(\Delta^{d-b}L_b\alpha_2\beta'(X_t-\mu) + \sum_{i=1}^{k}\Gamma_{i,2}\Delta^d L_b^i(X_{t-i}-\mu)\Big)D_{2t}(\beta,\gamma).\end{aligned}$$

Given a pair of fixed long memory parameters (d, b), the threshold FCVAR model is reduced to the $I(1)/I(0)$ threshold cointegration by Hansen and Seo (2002). In addition, threshold FCVAR model assumes the long-run fractionally cointegrated relationship remains unchanged across regimes, turns out that in the first step of estimation, the maximum likelihood parameters of (d, b, μ) can be estimated by maximising function (2.10) regardless any nonlinearity in the short-run dynamics.

In the second step of estimation, since the MLE estimates $(\hat{d}, \hat{b}, \hat{\mu})$ is given from the previous step, the concentrated likelihood can be computed for the regime-specific parameters (Φ_1, Φ_2, Ω) by holding (β, γ) fixed. The MLE $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are piecewise linear in the threshold space for the sub-samples for which $e_{t-1} \leq \gamma$ and $e_{t-1} > \gamma$, respectively. More specifically, this yields the concentrated likelihood function

$$\begin{aligned}\log\mathcal{L}_T(\beta,\gamma) &= \log\mathcal{L}_T\big(\hat{\Phi}_1(\beta,\gamma), \hat{\Phi}_2(\beta,\gamma), \hat{\Omega}(\beta,\gamma)\big)\\ &= -\frac{T}{2}\log\det\big(\hat{\Omega}(\beta,\gamma)\big) - \frac{Tp}{2}.\end{aligned} \tag{2.11}$$

The MLE $(\hat{\beta}, \hat{\gamma})$ maximise the likelihood function (2.11) subject to the constraint $\pi_0 \leq T^{-1}\sum_{t=1}^{T} 1(e_{t-1} \leq \gamma) \leq 1 - \pi_0$. Since, the likelihood function in (2.11) is not smooth, so the gradient hill climbing optimisation techniques cannot guarantee a global maximum. Therefore, similar to Hansen and Seo (2002), a grid search over the two-dimensional space (β, γ) is adopted for the estimation of β and γ.

The two-dimensional grid $[\beta, \gamma]$ is constituted by $N_\beta \times N_\gamma$ number of grid points, with N_β and N_γ denote the evenly spaced grids on $[\beta_L, \beta_U]$ and $[\gamma_L, \gamma_U]$, respectively. It computationally convenient to calibrate the grid region $[\beta_L, \beta_U]$ over the slope parameter β_2 from $\beta' = (\beta_1, \beta_2)$ with a fixed value of $\tilde{\beta}_1 = 1$. Hence, the grid of β_2 can be obtained by the confidence interval of consistent estimate $\tilde{\beta}_2$ in the equilibrium relation of the linear FCVAR model. Set $\tilde{e}_{t-1} = e_{t-1}(\tilde{\beta}_2)$, the grid region $[\gamma_L, \gamma_U]$ is given by the empirical support of the sorted distribution of $\tilde{e}_{t-1}$ with a trimming constraint $\pi_0 \leq T^{-1}\sum_{t=1}^{T} 1(\tilde{e}_{t-1} \leq \gamma) \leq 1 - \pi_0$. For each value of (β_2, γ) on the grid, the likelihood parameter estimates $(\hat{\Phi}_1, \hat{\Phi}_2, \hat{\Omega})$ are recalculated. A pair of optimal MLE $(\hat{\beta}, \hat{\gamma})$ yields the largest value of the likelihood value in (2.11). Finally the MLE $(\hat{\Phi}_1, \hat{\Phi}_2, \hat{\Omega})$ are set given the optimal $(\hat{\beta}, \hat{\gamma})$, as $\big(\hat{\Phi}_1(\hat{\beta}, \hat{\gamma}), \hat{\Phi}_2(\hat{\beta}, \hat{\gamma}), \hat{\Omega}(\hat{\beta}, \hat{\gamma})\big)$.

At this stage, the optimisation is only described to implement maximum likelihood estimations. Since the regime-specific parameters are discontinuous and piecewise linear in the threshold space, deriving asymptotic theory for the estimators is challenging. The proof of consistency of the MLE could be a focus for further research. According to the likelihood inference in linear FCVAR model Johansen and Nielsen (2012), the MLE $\hat{\beta}$ converges to normal distribution at rate $T^{1/2}$ for cofractional parameter $b_0 < 1/2$, and to mixed normal at rate T^{b_0} for $b_0 > 1/2$. In threshold stationary models, $\hat{\gamma}$ converges to γ at rate T. It might be reasonable to guess in the threshold FCVAR model, the MLE $(\hat{\beta}, \hat{\gamma})$ may converge to (β, γ) at rate between T^{b_0} and T. With known values of β and γ, the limit distributions of the remaining parameters (d, b, Φ_1, Φ_2) are conjectured to be Gaussian, as they are showed to be Gaussian in the linear FCVAR model.

2.3.4 Other nonlinear extensions

This study examines one of various forms of nonlinearity in the cointegrated time series. The two-regime threshold cointegration has been widely adopted to provide empirical regularity in many macroeconomic problems, for instance, in the business cycles and the equilibrium parities of exchange rates or commodity prices. Nevertheless, other nonlinear specifications are also worth exploring to select an appropriate nonlinear model.

The nonlinear adjustment in the FCVAR model can be extended to other forms of nonlinearity. For instance, the three-regime threshold cointegration has been examined (see Gonzalo and Pitarakis, 2002; Seo, 2003; etc.) in nonfractional case. The 3-regime threshold specification of the fractional case can be formulated as

$$\begin{aligned}\Delta^d X_t =& \Big(\Delta^{d-b} L_b \alpha_1 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,1} \Delta^d L_b^i X_{t-i}\Big) 1(e_{t-1} \leq \gamma_1) \\ &+ \Big(\Delta^{d-b} L_b \alpha_1 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,1} \Delta^d L_b^i X_{t-i}\Big) 1(\gamma_1 < e_{t-1} \leq \gamma_2) \\ &+ \Big(\Delta^{d-b} L_b \alpha_1 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,1} \Delta^d L_b^i X_{t-i}\Big) 1(e_{t-1} > \gamma_2) + \epsilon_t \end{aligned} \tag{2.12}$$

The three-regime case is a straightforward extension from the two-regime one. Error correction depending on three states is also commonly observed in many macroeconomic time series such as the interest rates term structure exhibits different mean-reversion dynamics at good/normal/bad economic conditions. However, the grid search techniques for MLE will be computationally intensive with two thresholds and one slope parameter.

Another nonlinear specification can make use of smooth transition in the adjustment dynamic. Unlike threshold models which are characterised by discrete regime shifts, the smooth transition models allow for continuous and smooth transitions across regimes.

The recent literature e.g. Saikkonen (2008), Seo (2011) and Kristensen and Rahbek (2013) investigate the nonfractional smooth tsransition vector error correction models. The two-regime smooth transition extension in FCVAR model is defined as

$$\Delta^d X_t = \Big(\Delta^{d-b} L_b \alpha_1 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,1} \Delta^d L_b^i X_{t-i}\Big) \mathcal{G}(\beta' X_{t-1}) + \Big(\Delta^{d-b} L_b \alpha_2 \beta' X_t + \sum_{i=1}^{k} \Gamma_{i,2} \Delta^d L_b^i X_{t-i}\Big) \{1 - \mathcal{G}(\beta' X_{t-1})\} + \epsilon_t \tag{2.13}$$

where the transition function $\mathcal{G}(\cdot)$ is commonly the logistic or the exponential.

The logistic smooth transition function is: $\mathcal{G}(\beta' X_t) = (1+\exp\{-\delta(\beta' X_t - c)\})^{-1}$, and the exponential function is: $\mathcal{G}(\beta' X_t) = 1+\exp\{-\delta(\beta' X_t - c)^2\}$, where $\delta > 0$ is the smoothness (velocity) of the transition and c is the location parameter. Smooth transition model is continuous in the smooth transition parameter δ. The model reduces to linear model when $\delta = 0$ and approaches to a discrete two-regime threshold model when $\delta \to 0$.

Asymptotic theory of regime-switching and/or discontinuous cases is difficult to establish due to the lack of uniformity in the convergence over the cointegrating vector space. Given the smooth transition case is a continuous generalisation of threshold models, it may provide convenient properties to derive asymptotic inference in the nonlinear fractionally cointegration. Hopefully this question can be neatly answered in my future research.

2.4 Testing for a threshold

Strictly speaking, there are two testing objectives on the two-regime threshold fractionally cointegrated VAR model. One is testing for the presence of long-run fractional cointegration, and the other is testing for the threshold nonlinearity in short-run adjustment dynamics. This study adopts a two-step approach in which the absence of linear fractional cointegration against the linear fractional cointegration alternative is examined in the first step, while the null of linear fractional cointegration against the threshold nonlinear comes in the second step. The first step of hypothesis testing applies the cointegration rank test of fractionally cointegrated VAR systems by Johansen and Nielsen (2012). This section derives a hypothesis testing for the presence of threshold adjustment in a FCVAR model.

2.4.1 Test statistic

This research focuses on model-based statistical tests for direct model comparisons. The crucial issue in nonlinear testing of this type is the null hypothesis contains a nuisance parameter, i.e. the threshold parameter is not identified, leading to a nonstandard testing problem. Thus, the test follows Hansen and Seo (2002) and employs the sup-type Lagrange Multiplier statistic (henceforth supLM), as it does not require a distribution theory for the parameter estimates in unrestricted model.

Let $\mathcal{H}_0$ denote the class of linear FCVAR models (2.7) and $\mathcal{H}_1$ denote the class of two-regime threshold FCVAR models (2.8). The restricted and unrestricted models in compact form can be written as

$$\mathcal{H}_0 : \qquad \Delta^d(X_t-\mu) = \Phi' z_t(\beta) + \epsilon_t \tag{2.14}$$

$$\mathcal{H}_1 : \qquad \Delta^d(X_t-\mu) = \Phi_1' z_t(\beta) D_{1t}(\beta,\gamma) + \Phi_2' z_t(\beta) D_{2t}(\beta,\gamma) + \epsilon_t \tag{2.15}$$

$$\text{where} \qquad z_t(\beta) = \begin{pmatrix} \Delta^{d-b} L_b \beta'(X_t-\mu) \\ \Delta^d L_b^1 (X_t-\mu) \\ \vdots \\ \Delta^d L_b^k (X_t-\mu) \end{pmatrix}.$$

The $q \times 1$ regressor $z_t(\beta)$ corresponds to the $q \times p$ where $q = pk+1$ parameter matrix $\Phi = (\alpha, \Gamma_i)'$ in $\mathcal{H}_0$ and its regime-specific parameter sets $\Phi_j = (\alpha_j, \Gamma_{i,j})'$ for $j = 1,2$ denotes the state j of regimes in $\mathcal{H}_1$. The linear model is nested in the threshold alternative which satisfies $\Phi_1 - \Phi_2 = 0$.

For fixed memory parameters (d,b) and known values of parameters (β,γ), the alternative of threshold FCVAR model is effectively reduced to piecewise linear reduced rank regressions. Specifically, $\Delta^d(X_t-\mu)$ on $z_t(\beta)D_{1t}(\beta,\gamma)$ in regime 1 and $\Delta^d(X_t-\mu)$ on $z_t(\beta)D_{2t}(\beta,\gamma)$ in regime 2.

The LM statistic has standard expression as

$$LM(\lambda_0,\gamma) = \mathbb{S}(\lambda_0,\gamma)'\mathbb{H}^{-1}(\lambda_0,\gamma)\mathbb{S}(\lambda_0,\gamma) \tag{2.16}$$

where $\lambda_0 = (d,b,\mu,\Omega,\beta,\Phi)$ are the parameters of interest under the null, $\mathbb{S}(\lambda_0,\gamma)$ and $\mathbb{H}(\lambda_0,\gamma)$ are the score and Hessian of the log-likelihood (3.8) evaluated at the parameter set under the null, respectively. As in Hansen and Seo (2002), the asymptotically equivalent version of LM statistic could be given as

$$LM(\beta,\gamma) = \text{vec}\big(\hat{\Phi}_1(\beta,\gamma) - \hat{\Phi}_2(\beta,\gamma)\big)'\big(\hat{V}_1 + \hat{V}_2\big)^{-1}\text{vec}\big(\hat{\Phi}_1(\beta,\gamma) - \hat{\Phi}_2(\beta,\gamma)\big) \tag{2.17}$$

where $\text{vec}(\cdot)$ is the vectorization operator, $\hat{V}_1$ and $\hat{V}_2$ are the covariance estimators for $\text{vec}\big(\hat{\Phi}_1(\beta,\gamma)\big)$ and $\text{vec}\big(\hat{\Phi}_2(\beta,\gamma)\big)$, respectively. They are defined as

$$\hat{V}_1 = \hat{V}_1(\beta,\gamma) = M_1(\beta,\gamma)^{-1}\Xi_1(\beta,\gamma)M_1(\beta,\gamma)^{-1}, \tag{2.18}$$

$$\hat{V}_2 = \hat{V}_2(\beta,\gamma) = M_2(\beta,\gamma)^{-1}\Xi_2(\beta,\gamma)M_2(\beta,\gamma)^{-1} \tag{2.19}$$

with the product matrices

$$M_1(\beta,\gamma) = I_p \otimes z_1(\beta,\gamma)' z_1(\beta,\gamma) \quad \text{and} \quad M_2(\beta,\gamma) = I_p \otimes z_2(\beta,\gamma)' z_2(\beta,\gamma),$$
$$\Xi_1(\beta,\gamma) = I_p \otimes \zeta_1(\beta,\gamma)' \zeta_1(\beta,\gamma) \quad \text{and} \quad \Xi_2(\beta,\gamma) = I_p \otimes \zeta_2(\beta,\gamma)' \zeta_2(\beta,\gamma)$$

Let $z_1(\beta,\gamma)$ and $z_2(\beta,\gamma)$ be the matrices of the staked rows $z_t(\beta)D_{1t}(\beta,\gamma)$ and $z_t(\beta)D_{2t}(\beta,\gamma)$, respectively, in (2.15). Denote $\zeta_1(\beta,\gamma)$ and $\zeta_2(\beta,\gamma)$ as the matrices of the stacked rows $\tilde{\epsilon}_t \otimes z_1(\beta,\gamma)$ and $\tilde{\epsilon}_t \otimes z_2(\beta,\gamma)$, respectively, with $\tilde{\epsilon}_t$ the estimate of residuals in (2.14).

For those who are interested in assuming time-varying conditional variances, the expression in (2.17) can allow for heteroskedasticity-robust covariance estimators in $\hat{V}_1$ and $\hat{V}_2$.

If the threshold parameter γ is identified, LM statistic in (2.16) or (2.17) would be the test statistic. However, if γ is unidentified under the null of $\mathcal{H}_0$, information matrix is singular and the asymptotic distribution of LM statistic is nonstandard due to the presence of nuisance parameters under $\mathcal{H}_0$. A solution to the unidentified parameter in $\mathcal{H}_0$ has been raised by Davies (1987) who proposed to obtain the optimal LM test statistic from a set of LM statistics at each value of the unidentified parameter. Each one of those test statistics would be chi-squared with one degree of freedom, but the maximum of a set of dependent chi-squared distributions is not a chi-squared. The nonstandard limiting distribution of the test can be identified through bootstrap techniques.

The supremum of LM test statistic is defined as

$$\sup LM = \sup_{\gamma \in \Gamma^*} LM_T(\tilde{\beta},\gamma) \tag{2.20}$$

where $\tilde{\beta}$ denotes the null estimate of β, and the threshold parameter γ is given in the search region $\Gamma^* = [\gamma_L, \gamma_U]$ with γ_L is the π_γ percentile of error correction term e_{t-1} and γ_U is the $1-\pi_\gamma$ percentile[7].

For the development of asymptotic properties, the following assumptions are formulated.

Assumption 2.1.

(1) The errors $\{\epsilon_t\}$ are i.i.d.$(0,\Omega)$ with $\Omega > 0$ and $E|\epsilon_t|^8 < \infty$.

(2) The initial values X_{-n} for $n \geq 0$, are uniformly bounded.

(3) Define the parameter set $\mathcal{N} = \{d, b : 0 < b \leq d \leq d_1\}$ for some $d_1 > 0$ which can be arbitrarily large. The true parameter values $(d_0, b_0) \in \mathcal{N}$, $0 \leq d_0 - b_0 < 1/2$, $b_0 \neq 1/2$.

(4) If rank $r < p$, then det$(\Psi(y)) = 0$ has $(p-r)$ unit roots and the remaining roots are outside $C_{max(b_0,1)}$.

[7] Note that for empirical application, the trimming parameter π_0 corresponds to the constraint in the grid search of MLE $\hat{\gamma}$, while the trimming parameter π_γ corresponds the constraint in the search region of LM test statistic.

(5) Under $\mathcal{H}_0$, $LM^{\delta}(\delta,\gamma) = LM(\beta+\delta/T,\gamma)$ has the same asymptotic finite dimensional distribtuions as $LM(\beta_0,\gamma)$.

(6) The estimates of d and b under $\mathcal{H}_0$ do not affect piecewise linear reduced rank regressions under $\mathcal{H}_1$.

Conditions (1)-(4) are assumptions imposed in the fractionally cointegrated VAR data generating process, see section 2.4 of Johansen and Nielsen (2012) for more detail. Assumption 2.1(1) does not impose Gaussian to the errors but only assumes the *i.i.d.* and finite eight moments. Assumption 2.1(2) is needed for nonstationary processes to ensure the fractionally differencing series $\Delta^d X_t$ is defined for any $d \geq 0$. In Assumption 2.1(3), the condition $0 \leq d_0 - b_0 < 1/2$ ensures $\beta_0' X_t$ is asymptotically stationary. Assumption 2.1(4) guarantee the cofractionality when $r > 0$ in which X_t is fractional of order d_0 and $\beta_0' X_t$ is fractional of order $d_0 - b_0$, such that both $\Delta^{d_0} X_t$ and $\Delta^{d_0-b_0}\beta_0' X_t$ are $I(0)$ stationary. Assumption 2.1(5) is suggested by the Theorem 2 of Hansen and Seo (2002) which implies the use of the estimate $\tilde{\beta}$ from linear model, rather than the true β_0, does not affect the asymptotic distribution of the LM test statistic. Similar to 1(5), Assumption 2.1(6) ensures the estimate $\tilde{d}$ and $\tilde{b}$ from linear model does not affect the asymptotic distribution of the test statistic.

Definition 2.1. *Let $\mathcal{F}(\cdot)$ be the marginal distribution of the error correction e_{t-1}. Define $\omega_{t-1} = \mathcal{F}(e_{t-1})$, $M(r) = I_p \otimes E[z_t z_t' 1(\omega_{t-1} \leq r)]$ and $\Xi(r) = E[1(\omega_{t-1} \leq r)(\epsilon_t \epsilon_t' \otimes z_t z_t')]$.*

Theorem 2.1. *Under $\mathcal{H}_0$, Assumption 2.1 and Definition 2.1,*

$$\sup LM \Rightarrow \mathbb{B} = \sup_{r\in\Lambda} B(r)$$

where

$\Lambda = [\pi, 1-\pi]$ *and* $\pi = P(\omega_{t-1} \leq r)$,

$B(r) = S^*(r)'\Xi^*(r)S^*(r)$, *in which*

$S^*(r) = S(r) - M(r)M(1)^{-1}S(1)$,

$\Xi^*(r) = \Xi(r) - M(r)M(1)^{-1}\Xi(r) - \Xi(r)M(1)^{-1}M(r) + M(r)M(1)^{-1}\Xi(1)M(1)^{-1}M(r)$.

Theorem 2.1 is reproduced from Theorem 2.1 of Hansen and Seo (2002) for the case of fractionally cointegrated series. Sketch proof is provided in Appendix A.

2.4.2 Bootstrap statistic and p-value

The LM statistic has nonstandard distribution and hence tabulation is not feasible. As in Hansen and Seo (2002), fixed regressor and residual-based bootstrap inferences are proposed to approximate the distribution of supLM. Refinement of finite sample performance is investigated through simulation experiment in section 2.5.1.

Fixed regressor bootstrap constructs the bootstrap distribution of supLM test using the residuals from the reduced rank regression in $\mathcal{H}_0$ and the parameter estimates $\hat{\Phi}_1(\gamma)$ and $\hat{\Phi}_2(\gamma)$ under the unrestricted model. The fixed regressor bootstrap procedure proceeds as follows.

First, it is noted that Assumption 2.1(5) suggests the estimate of cointegrating vector $\tilde{\beta}$ in linear model does not affect the asymptotic distribution of the supLM test. Hence, the estimate of β does not need to be considered in the inference of threshold estimate and can be held fixed along the bootstrap procedure. Let the dependent variable $\tilde{z}_{0,t} = \Delta^{\tilde{d}}(X_t - \tilde{\mu})$, the regressor $\tilde{z}_t = z_t(\tilde{\beta})$ in which $z_t(\beta)$ is given in (2.14) and the error correction term $\tilde{e}_{t-1} = e_{t-1}(\tilde{\beta})$. The residuals $\tilde{\epsilon}_t$, the estimates of fractional memory parameters $\tilde{d}$, $\tilde{b}$ and the level parameter estimate $\tilde{\mu}$ from the linear model in $\mathcal{H}_0$ are held fixed at their sample values.

Next, let $v_{b,t}$ be an $i.i.d.N(0, I_2)$ and set $u_{b,t} = \tilde{\epsilon}_t \circ v_{b,t}$. Regressing $u_{b,t}$ on $\tilde{z}_t$ yields bootstrap residuals $\tilde{\epsilon}_{b,t}$ in linear model. Regressing u_t on $\tilde{z}_t D_{1t}(\tilde{\beta}, \gamma)$ and $\tilde{z}_t D_{2t}(\tilde{\beta}, \gamma)$ yields bootstrap estimates $\hat{\Phi}_{1,b}(\gamma)$, $\hat{\Phi}_{2,b}(\gamma)$ and residuals $\hat{\epsilon}_{b,t}(\gamma)$ in unrestricted model. Hence, define the bootstrap covariance matrices $\hat{V}_{1,b}(\gamma)$ and $\hat{V}_{2,b}(\gamma)$ as in (2.18) and (2.19) by replacing $\beta = \tilde{\beta}$ and $\tilde{\epsilon}_t = \tilde{\epsilon}_{b,t}(\gamma)$.

Then, the bootstrap supLM test statistic can be set as

$$\sup LM^b = \sup_{\gamma \in \Gamma^*} \text{vec}\big(\hat{\Phi}_{1,b}(\gamma) - \hat{\Phi}_{2,b}(\gamma)\big)' \big(\hat{V}_{1,b}(\gamma) + \hat{V}_{2,b}(\gamma)\big)^{-1} \text{vec}\big(\hat{\Phi}_{1,b}(\gamma) - \hat{\Phi}_{2,b}(\gamma)\big). \quad (2.21)$$

The bootstrap description above generates one draw from the distribution. With a repeated large number of draws (e.g. 1000 times), the simulated distribution of supLM^b is created. The bootstrap p-value can be obtained by counting the percentage of simulated supLM exceeds the sample supLM statistic. If the bootstrap p-value is less than the nominal size chosen, the null hypothesis is rejected in favour of the alternative of nonlinear FCVAR model.

Residual bootstrap constructs the bootstrap sampling distribution of the test using a complete specification of the model in $\mathcal{H}_0$. The disturbance term ϵ_t in (2.14) is assumed to be $i.i.d.$ normal. The resampled residuals $\tilde{\epsilon}_{b,t}$ are randomly drawn from the sample estimated residuals $\tilde{\epsilon}_t$. Then the bootstrap sample $\tilde{z}^b_{0,t}$ can be constructed using the parameter estimates in $\mathcal{H}_0$ and the resampled residuals. The bootstrap supLM^b statistic can be calculated for each resampled data. By repeating sufficient large number of resamples, the bootstrap p-value can be obtained by counting the percentage of simulated supLM exceeds the sample supLM statistic.

2.5 Simulation evidence

This chapter performs a set of simulations to evaluate 1) the size and power of supLM test, and 2) the model misspecification of a threshold fractionally cointegrated VAR model. Simulation evidence is provided in support of the supLM test maintains good empirical size and power in small sample. With regard the incomplete asymptotic theory of Gaussian maximum likelihood estimators in the threshold FCVAR model, a data generating process from the threshold FCVAR is simulated and fitted into the misspecified forms of cointegration models. By doing this, the misspecification of long memory parameters (d, b) and threshold adjustment from other forms of cointegration might be evident.

2.5.1 Finite-sample size and power

The empirical size and power of the supLM test under finite-samples are studied based on linear fractionally cointegrated VAR model in null hypothesis $\mathcal{H}_0$ and the alternative choices $\mathcal{H}_0$ of the two-regime threshold adjustment in the FCVAR model. A bivariate fractionally cointegrated process $X_t = (x_{1,t}, x_{2,t})$ is the variables for data generating processes.

Empirical size is assessed under $\mathcal{H}_0$ which is the fractionally cointegrated VAR model with linear adjustment coefficients. Data is generated based on:

$$\Delta^d(X_t - \mu) = \Delta^{d-b} L_b \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \beta'(X_t - \mu) + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i (X_{t-i} - \mu) + \epsilon_t \tag{2.22}$$

where disturbance term $\epsilon_t \sim i.i.d.N(0, I_2)$. Assume the number of lags k equals 0 and 1. Parameters are fixed at $d = 0.8$, $b = 0.6$, $\mu = (1, 1)'$, $\beta = (1, -1)'$ and $\alpha_1 = -0.05$. The parameter α_2 varies among $\{0, 0.25, 0.5\}$. The lagged term coefficient Γ varies among $\Gamma_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\Gamma_1 = \begin{bmatrix} 0.1 & 0.3 \\ 0 & 0.5 \end{bmatrix}$ The supLM test is calculated setting the trimming parameter $\pi_0 = 0.15$ and number of grid points $N_\gamma = 50$ on $[\gamma_L, \gamma_U]$. The number of simulations is 1000 and bootstrap frequency is 200 for each replication. The samples of sizes 200 and 500 with number of initial values $N_0 = 10$ are considered. Note that the DGP of fractional difference series adopts Jensen and Nielsen (2014)'s fast fractional differencing algorithm and the initial value is necessary to generating fractional difference series.

For each simulated sample, supLM statistic and p-value from fixed regressor bootstrap and residual bootstrap are calculated. Table 2.1 summarises the rejection frequencies [8] from the conventional 5% and 10% tests for supLM statistics. It is shown that the rejection sizes are close to the conventional sizes with the variates of α_2 among $\{0, 0.25, 0.5\}$.

[8] The rejection frequencies of each scenario is the percentage of simulated $p-$values are smaller than the level of significance.

The empirical sizes improve slightly when sample size increased from 200 to 500. The lagged term coefficient Γ_i varies among Γ_0 and Γ_1 does not distort the sizes of the test. Similarly to Hansen and Seo (2002)'s threshold test for nonfractional cointegration, the fixed regressor bootstrap tend to over reject the null hypothesis, while the residual bootstrap gives much better size than the fixed regressor bootstrap. The overall size of the supLM test using fixed-regressor bootstrap range from 0.065 to 0.088 for 5% rate and from 0.119 to 0.132 for 10% rate. The size using residual bootstrap range from 0.051 to 0.067 for 5% rate and from 0.098 to 0.116 for 10% rate. The proposed supLM test using residual bootstrap is found to maintain satisfactory empirical size under small sample.

Table 2.1 about here (see P.46).

Empirical power of the test is assessed against $\mathcal{H}_1$ of the two-regime adjustment coefficients in fractionally cointegrated VAR model. To keep the computation manageable, the short-run lagged dynamics is ignored, i.e. assume $k = 0$. A simple 2-regime threshold FCVAR model in $\mathcal{H}_1$ is as

$$\begin{aligned}\Delta^d(X_t - \mu) =& \Delta^{d-b} L_b \begin{pmatrix}\alpha_1^1 \\ \alpha_2^1\end{pmatrix} \beta'(X_t - \mu) 1(e_{t-1} \leq \gamma) \\ &+ \Delta^{d-b} L_b \begin{pmatrix}\alpha_1^2 \\ \alpha_2^2\end{pmatrix} \beta'(X_t - \mu) 1(e_{t-1} > \gamma) + \epsilon_t \end{aligned} \tag{2.23}$$

with disturbance term $\epsilon_t \sim i.i.d.N(0, I_2)$. Parameters are fixed at $d = 0.8$, $b = 0.6$, $\mu = (1,1)'$, $\beta = (1,-1)'$, $\alpha_1 = (0, 0.05)'$ and the parameter $\alpha_2 = (-0.01, \alpha_2^2)'$ with α_2^2 varies among $\{0.2, 0.5, 0.8\}$. The threshold parameter γ is set to have $P(e_{t-1} \leq \gamma)$ equals to 0.2 and 0.5. Again, sample sizes of 200 and 500 with number of initial values 10 from the fractional difference series are considered. Table 2.2 presents the rejection frequencies of the test at 5% level of significance with different values of adjustment coefficient α_2^2 and different sub-sample sizes π_0 in the first regime. The fixed regressor bootstrap seems to have higher empirical power than residual bootstrap; yet it is likely caused by the artifact of the empirical size distortions. The power performance increases when the values of adjustment coefficient increase, since the threshold adjustment is more obvious with larger α_2^2. When sample size increased from 200 to 500, the power improved a lot at each scenario. Comparing the powers with different values of π_0, the threshold effect is more likely to be identified with larger portion of the sub-sample, thus power is higher with larger value of π_0.

Table 2.2 about here (see P.46).

2.5.2 Misspecification

The goal of this simulation experiment is to show the need of accurate specifications on the long run parameters (d, b) and the regime-specific threshold adjustment α in a class of cointegrated systems. If the true data generating process (DGP) is from a two-regime threshold FCVAR model but it is fitted to different econometric models, the model is misspecified and the coefficient estimates might be inconsistent.

Suppose that a true DGP for $X_t = (x_{1,t}, x_{2,t})'$ is generated from a bivariate two-regime threshold FCVAR model as

$$\Delta^d \begin{pmatrix} x_{1,t} - \mu_1 \\ x_{2,t} - \mu_2 \end{pmatrix} = \Delta^{d-b} L_b \begin{pmatrix} \alpha_1^1 \\ \alpha_2^1 \end{pmatrix} e_t 1(e_{t-1} \leq \gamma) + \Delta^{d-b} L_b \begin{pmatrix} \alpha_1^2 \\ \alpha_2^2 \end{pmatrix} e_t 1(e_{t-1} > \gamma) + \epsilon_t \qquad (2.24)$$

with equilibrium relation $e_t = \beta'(X_t - \mu)$ and innovations $\epsilon_t \sim i.i.d.N(0, I_2)$.

The DGP is generated given the autoregressive order $k = 0$ and remaining parameters: $d = 0.8$, $b = 0.6$, $\mu = (10, 10)'$, $\alpha_1 = (0, 0)'$, $\alpha_2 = (-0.005, 0.25)'$, $\beta = (1, -1)'$ and $\gamma = 1$. This DGP mimics the characteristics of volatility spot-futures relation which will be discussed in empirical application of this study, in which the memory parameters (d, b) are larger than 0.5, with non-zero level parameter μ, and adjustment dynamic α_1 vanished in the first regime. To avoid the possible small sample bias, sample size of one simulated X_t process is $T = 3000$ with number of initial values $N_0 = 10$. Figure 2.1 shows an example of simulated X_t from a bivariate 2-regime threshold FCVAR process.

Figure 2.1 about here (see P.42).

The above DGP is fitted to three different specifications of cointegrated VAR models, they are: (1) linear cointegrated VAR; (2) threshold cointegrated VAR; and (3) linear fractionally cointegrated VAR. For each fitted model, data generating process and model fitting are conducted repeatedly for 1000 times. Table 2.3 summarises the mean and standard error of the sampling distributions of coefficient estimates[9] from three different cointegrated models. There are three simulation evidence that seem relevant to the misspecification of long memory parameters and threshold adjustments.

Table 2.3 about here (see P.47).

First, fitting DGP to an $I(1)/I(0)$ 2-regime threshold CVAR (model 2) forced the long memory parameters $d = b = 1$. Although the cointegrating slope parameter $\hat{\hat{\beta}}_2 = 1.02$ is preserved, the threshold parameter estimate $\hat{\hat{\gamma}} = 0.518$ is distorted. Hence the nonlinear adjustment dynamics from $x_{2,t}$ in regime 2 is diluted by 40%; while the adjustment from $x_{1,t}$ in regime 1 become significant. This result provides partial evidence in support of Łasak (2010)'s Theorem 2, which suggests in linear fractional case that, for any fixed d, $d \neq d_0$ and $d > 0.5$, the MLE $\hat{\hat{\beta}}$ remains consistent with a rate $\hat{\hat{\beta}} - \beta = O_p(T^{1/2-d_0})$ but $\hat{\alpha}$ is not consistent any more. The simulation evidence here further suggests that in

[9] To avoid confusion with the the maximum likelihood estimates in empirical chapter, the MLE from the misspecified cointegration models is denoted with symbol “double hat”.

threshold case, misspecifying long memory parameters may also induce misspecification of threshold parameter γ. It is suspected that the unabsorbed long range dependence in (d, b) is captured by the short-run nonlinear adjustment dynamics and thus misspecified $\hat{\hat{\gamma}}$ and $\hat{\hat{\alpha}}$.

Second, by comparing DGP to the estimation of linear FCVAR (model 3), it can be seen that the long memory parameters d and b are close to the true values in DGP, but the slope parameter of cointegrating relation $\hat{\hat{\beta}}_2 = 1.08$ is slightly bigger than unity. Interestingly, the linear adjustment dynamics have very close magnitudes as in the regime 2 of DGP. This simulation scenario suggests that misspecifying short-run adjustment dynamics may not affect the estimates of memory parameters; however, the possible nonlinear adjustment is concealed.

Lastly, one can see that forcing integer memory parameters and linear adjustment dynamics in linear CVAR (model 1) may affect the estimate of cointegrating slope parameter, i.e. $\hat{\hat{\beta}}_0 = 1.13$, and the adjustment dynamics $\hat{\hat{\alpha}}$ become very weak.

To sum up the simulation evidence from the threshold FCVAR DGP,

- the non-fractional 2-regime threshold cointegration fails to capture the long memory feature of DGP, and hence the unabsorbed long memory may distort the estimates of threshold parameter and adjustment coefficients;
- the linear FCVAR model preserved similar values in long memory parameters;
- cointegrating slope parameter does not have dramatic changes in all the misspecified models.

The simulation experiment demonstrates the importance of accurate specifications in fractional memory parameters and threshold adjustment dynamics. More importantly, it also shows that long memory feature is not distorted in both linear and nonlinear FCVAR systems; thus assuming fixed values of long memory parameters across regimes is reasonable at least in sufficiently large sample.

2.6 Application: volatility spot-futures relation

The appropriate modelling of S&P500 option implied volatility index (VIX) and its related futures prices is of interest for several reasons. First, with regard the empirical literature, the relation between realised and implied volatility of an underlying asset is of interest to identify either option market efficiency or short-run unbiasedness of implied volatility as a forecast of realised volatility, see e.g. Christensen and Prabhala (1998).

Multivariate volatility process is found to be fractionally cointegrated by semi-parametric approaches[10].

Second, because the futures is a tradable asset, the relation between VIX spot and futures is similar to the cointegrated price processes in futures markets and commodity markets. Thus, different to the typical analysis of implied-realised volatility relation, modelling the cointegrated relation among VIX spot and futures provides separate interest to identify the futures price dynamics for hedging and risk management.

Furthermore, volatility traders tend to use VIX related derivatives as volatility security and behave differently in different states of market. When the market has high uncertainty, people are willing to hedge against volatility risk. However, when the market is certain, buying security is not necessary. The logic can be thought of one does not need travel insurance if no travel is planned.

Having make all these points above, this research models the fractionally cointegrated relation between VIX spot and futures by a parametric fractionally cointegrated VAR framework. Especially, it is of interest to investigate the possible regime-specific dynamics of the VIX spot-futures relation and to provide theoretical ground for the nonlinear trading behaviour in volatility derivatives market.

2.6.1 Volatility index and futures

The Chicago Board Options Exchange (CBOE) introduced the S&P500 implied volatility index (VIX)[11] in 2003. The VIX measures a risk neutral 30-day implied volatility derived from a basket of out-of-money S&P500 options which has an average maturity of 30 days. Essentially, the VIX is a current expectation on market volatility over the next 30 days given by investors' attitude on S&P500 options. The higher VIX indicates that investors expect a more volatile market; therefore, it is termed the "fear gauge" (Whaley, 2000).

Volatility index itself is not tradable. To allow for trading opportunities, investors rely on the derivatives products such as futures, options or exchange traded funds (ETF) on VIX. Due to the negative relation between VIX and S&P500 equity index, volatility tracking products are mainly used for hedging the downside risk of S&P500. Since the second moment of volatility measure is extremely high, VIX products provide a huge leverage of hedge on S&P500 which means one could secure for a crash in the S&P500 by giving up just a small amount of gain in the portfolio.

[10] Bandi and Perron (2006) used the narrow-band least squares estimators and Nielsen (2007) used the local Whittle estimator to to analyse the long-run implied-realised volatility relation.

[11] The S&P500 implied volatility index is denoted as "spot" in order to distinguish from the VIX futures, but it is not a direct underlying of the futures.

2.6.2 Stylised facts about VIX futures market

The VIX futures can be seen as a current bet on some future values of VIX. It is noteworthy that the settlement price of futures is not calculated by the "cost of carry relationship" as in other futures markets. In fact, the fair price of VIX futures expiring at day j is calculated by the current value of a basket of S&P500 options expiring at day $j + 30$. For example, a short-term VIX futures with $j = 5$ days of expiration is projecting a current bet on the S&P500 option implied volatility in 35 days. This implies an imperfect correlation between VIX spot and futures because futures is not tracking spot directly.

In addition, VIX futures has normal contango state, or noted as "long-run" contango in econometric literature (see Figuerola-Ferretti and Gonzalo, 2010). Contango is a state of which the futures price is higher than the spot price, in turn investors are paying at a premium for the asset at some point in the future. The opposite market state to contango is known as backwardation. The long-run contango of VIX futures means that rolling futures contracts forces to buy the long-date contract at high price and to sell the short-date contract at low price continuously. In long term, volatility hedge effectively eats up potential gain in the portfolio. Instead, it makes more practical sense to buy volatility securities only when market is facing downside pressure. Brexit referendum is a good example to show the volatility derivatives as a protective bet against market uncertainty. At the day of referendum result announcement, the S&P500 index spot reported a plunge of 3.6% while the one-month VIX futures had a spike of 35%.

With regard the short-term futures price dynamics, contango and backwardation of VIX futures tend to depend on the market conditions. The scatter plot in figure 2.2 shows that, when the VIX spot is relatively low [12] which indicates market is calm, on average, the futures price is at contango state. It is because market expects the future volatility to rise with respect to the current VIX level. On the other hand, when the level of VIX is high which implies market has high uncertainty, the futures price tends to be at backwardation. It means that market expects the future volatility may drop below the current spot level.

Figure 2.2 about here (see P.42).

To illustrate the nonlinear response by investors upon the states of contango and backwardation in volatility futures market, a preliminary evidence is shown by an OLS regression for the period 2009-2016.

[12] In practice, the level of VIX spot is considered to be low when it is below 20. Note that the sample average of VIX spot from 2003 to 2016 is 19. The VIX level was above 35 during most of the time of 07-08 financial crisis and of the European debt crisis in 2011, and it is above 25 during the Brexit referendum.

The percentage change in trading volume of VIX futures, $\Delta log(\text{vol})_t$, is regressed on its lag term and the logarithmic difference of spot and futures in the previous period, $vd_{t-1} = log(s_{t-1}) - log(f_{t-1})$, with the regime switch depends on contango $(s - f \leq 0)$ and backwardation $(s - f > 0)$.

$$\Delta log(\text{vol})_t = \underset{(0.010)}{0.031} - \underset{(0.024)}{0.323} \Delta log(\text{vol})_{t-1} - \underset{(0.159)}{0.412} vd_{t-1} \, 1(s - f \leq 0) + \underset{(0.368)}{1.229} vd_{t-1} \, 1(s - f > 0) \tag{2.25}$$

where $1(\cdot)$ denotes the indicator function and standard errors are in parentheses.

The regression result in (2.25) shows that, on average, when it is at contango state, a 100% increase in volatility difference increases the volume change by 41.2%; however, when it is at backwardation, a 100% increase in volatility difference increases the volume change by 122.9%, which is three times of the change at contango. As can be seen, market is more nervous during the volatile period and is urging to hedge the downside risk by trading volatility futures, while there is less incentive to trade volatility futures during the calm period. Moreover, the difference between spot and futures is more sensitive to volume change at backwardation than at contango.

Having seen investors' willingness to hedge against volatility may depend on market condition, a question is posed to ask whether the possible regime dependent relation between VIX spot and futures can be an "enter-and-exit" indicator for volatility investment. Using the states of contango and backwardation to derive investment strategies is related to a method called "momentum strategies" adopted by many financial practitioners. This study aims to provide a theoretical ground for the empirical regularity of momentum strategies.

2.6.3 Data description

This research uses a balanced dataset of the Chicago Board Options Exchange S&P500 option implied volatility index and the corresponding one-month futures contract from 24 March 2004[13] to 30 December 2016. There are 3215 observations in the sample. Along the study, the VIX index and VIX futures are simply denoted as spot (s_t) and futures (f_t). The one-month futures prices are the weighted prices of the first and the second month futures contract which give futures prices with an average 30 days of expiration. Using one-month futures price has the advantage that the price difference between spot and futures implies the change of market expectation on market volatility over a constant time window between day 30 and day 60, that is, similar to the idea of term structure. Instead of taking natural logarithm as in many macroeconomic and asset prices data, the original values of VIX spot and futures are used in this study due to several reasons. First of all, this application aims to verify the "momentum strategy" which uses contango and

[13]The first day of the sample is the date of introduction of VIX futures.

backwardation as enter-and-exit signal. The financial practitioners always look at the original values of spot and futures rather than the logarithmic values. Secondly, since fractionally cointegrated processes are sensitive to initial values (Johansen and Nielsen, 2016), transforming the data series may initiate the risk of inconsistency in parameter estimates.

The time series plot of VIX spot, futures and their price difference, spot minus futures, is shown in figure 2.3. From a preliminary inspection on spot and futures, they seem to co-move with each other across the sample period. They have a level around 15 to 20 over time. Noticeably, there are several spikes match the periods of market crashes, including the global financial crisis in 2008, the European sovereign debt crises around 2010 and 2011 and the China market flash crash in 2015. The strong persistence of volatility around the crash periods may indicate the presence of long memory feature. Next, considering the difference series of spot minus futures, $s_t - f_t$, it has a slightly negative level over the sample which matches the property of long-run contango ($s_t < f_t$) in volatility futures market. Although the difference price series seems to be more stationary than the original volatility processes, the spikes around market downturn periods are still very obvious. Moreover, the significant positive spikes indicate strong backwardation ($s_t > f_t$) state. Those imply that the long-run relation of volatility spot and futures may have informative switch from contango to backwardation when market has risk to turn from normal to downside.

Figure 2.3 about here (see P.43).

To show the presence of long memory in volatility processes and their volatility difference, table 2.4 reports the estimators of long memory parameter d using three semi-parametric estimators. Given different Fourier frequencies for various estimators, the long memory parameter estimates, $\hat{d}$, of VIX spot range from 0.75 to 0.92, while the $\hat{d}$ of futures price is slightly higher than but very close to those of spot, range from 0.84 to 0.97. With $\hat{d} > 1/2$, the volatility spot and futures are long memory nonstationary processes. Regarding the difference of spot and futures, the estimates of d range from 0.31 to 0.62. It indicates that the difference series possesses long memory, yet the stationarity is inconclusive.

Table 2.4 about here. (see P.48)

2.6.4 Model specification

The relation of volatility spot (s_t) and futures (f_t) is modelled by a bivariate 2-regime threshold FCVAR model as the framework in model (2.8). A pair of fractionally cointegrated processes $X_t = (s_t, f_t)'$ has a fractionally integrated order of d. The long-run cointegrating relationship $\beta' X_t = (1, -\beta_2)(s_t, f_t)'$ can be written in a single equation, $s_t - \beta_2 f_t = e_t$. A lot of literature focused on assessing the unity of β_2 and the $I(0)$ stationarity of disequilibrium error e_t for cointegration analysis. In fractionally cointegrated VAR framework, the cointegrating vector $\beta' X_t$ has an order $d - b \in [0, 1)$ where b determines the degree of fractional cointegration.

It is relevant to distinguish between long-run and short-run contango/ backwardation from the cointegrating vector $s_t - \beta_2 f_t = e_t$. The long-run contango or backwardation refers to the "normal" state of a futures market. It is about the discussion of the slope of cointegrating vector β_2, see the econometric studies in commodity futures by Figuerola-Ferretti and Gonzalo (2010) and Dolatabadi et al. (2015). When the slope parameter $\beta_2 > 1$ ($\beta_2 < 1$), the market has long-run backwardation (contango). On the other hand, the short-run contango or backwardation used by financial practitioners refers to the temporary discrepancy from long-run equilibrium, that is, the disequilibrium error e_t, and the error correction is picked up by adjustment parameter α in the error correction mechanism. When $e_{t-1} > 0$ ($e_{t-1} < 0$), the market is at backwardation (contango) state. Since this study aims to test for nonlinear adjustment towards the long-run relation, the short-run "contango" and "backwardation" are of interest. In addition, when the idea of temporary contango and backwardation is framed into the regime-switching adjustment dynamics, the disequilibrium error e_t is said to be deviated from a threshold value γ rather than a fixed value 0. The threshold parameter γ is optimised through the maximum likelihood value of the threshold model.

The initial value of fractional series is a concern in fractionally integrated model specification. As mentioned previously, fractional difference is an infinite series by definition, yet observed sample cropped the fractional difference with finite terms. It is unrealistic to impose the initial value of volatility processes to be zero which implies S&P500 equity market has zero volatility, and also, this assumption induces bias to parameter estimation. Therefore, a level parameter $\mu = (\mu_s, \mu_f)'$ is included within the fractional integrated process X_t which shifts each of the fractional difference series by a constant, that is, becomes $\Delta^d(X_t - \mu)$. The cointegrating vector takes a slightly different form as $(s_t - \mu_s) - \beta_2(f_t - \mu_f) = e_t$. Actually, it does not affect the definition of contango and backwardation regime states, since the elements of level parameter is entered into the equilibrium relation as a constant term which will be absorbed by the threshold parameter estimate.

The model selection choices in terms of the number of lag length k is specified in the linear FCVAR model. Several information criteria include the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) the Likelihood-Ratio (LR) test statistic for significance of Γ_k are reported. For each criterion, the model specification with level parameter μ and full rank $r = p$ is considered up to the maximum $k = 3$. The BIC suggests $k = 0$, LR statistic suggests $k = 2$ and AIC suggests $k = 3$ for lag length selection. Since our focus is more on the presence of threshold adjustment in the error correction term, the model with no lags $k = 0$ is slected based on BIC for the simplification of the threshold model. In addition, the cointegration rank test determines the number of cointegrating relation based on a sequential test of null hypothesis $\mathcal{H}(r)$: $rank = r$ against $\mathcal{H}(p) : rank = p$ for r = 0, 1, 2. The estimated rank is selected by the first non-rejected value of the LR test statistic. The asymptotic distribution of the LR statistic is nonstandard. Asymptotic p-value can be calculated using the algorithm by Johansen and Nielsen (2016). The null hypothesis that $r = 0$ is rejected but $r = 1$ is not rejected.

Have made the model specifications from the above, the relation of VIX spot and futures is modelled by a bivariate 2-regime threshold FCVAR model with level parameter μ, number of lags $k = 0$ and rank $r = 1$ as the following formulation.

Unrestricted model:

$$\Delta^d \begin{pmatrix} s_t - \mu_s \\ f_t - \mu_f \end{pmatrix} = \Delta^{d-b} L_b \begin{pmatrix} \alpha_s^1 \\ \alpha_f^1 \end{pmatrix} e_t D_{1t}(\beta, \gamma) + \Delta^{d-b} L_b \begin{pmatrix} \alpha_s^2 \\ \alpha_f^2 \end{pmatrix} e_t D_{2t}(\beta, \gamma) + \epsilon_t \tag{2.26}$$

Equilibrium relation:

$$s_t = \mu_s + \beta_2 (f_t - \mu_f) + e_t \tag{2.27}$$

States of deviation from equilibrium relation:

$D_{1t}(\beta, \gamma) = 1(e_{t-1} \leq \gamma)$ indicates market is at short-run contango at $t - 1$, and
$D_{2t}(\beta, \gamma) = 1(e_{t-1} > \gamma)$ indicates market is at stort-run backwardation at $t - 1$,
where $1(\cdot)$ denotes the indicator function.

2.6.5 Empirical results

To facilitate comparison with a general class of cointegration models, the estimations from the standard $I(1)/I(0)$ cointegrated VAR (Johansen, 1995), the 2-regime threshold cointegration (Hansen and Seo, 2002) and the linear FCVAR (Johansen, 2008) also presented as benchmark for the proposed threshold fractionally cointegrated VAR model. It is relevant to figure out common and different specifications of model 1 to 4. The fractional cointegration (model 3 and 4) has similar representation theorem as the usual $I(1)/I(0)$ cointegration (model 1 and 2), apart from the long memory parameters in the standard cointegration are restricted as unity. The threshold model (2 and 4) is estimated after the estimation of linear model (1 and 3), in which those parameters assumed to be fixed across regimes can be concentrated out from the maximum likelihood function for threshold estimation. In model 2, the estimate of cointegrating parameter $\hat{\beta}$ is extracted from linear case and undertaken a 2-dimensional grid search for threshold estimate. Similarly, the proposed threshold FCVAR (model 4) of this study also extracts the long memory parameter estimates $\hat{d}$, $\hat{b}$ and the estimates $\hat{\beta}$ and $\hat{\mu}$ in the long-run cointegrating relation from the linear model. Note that unlike in model 2 where the unrestricted constant is changed across regimes, the level parameter in model 4 is assumed to be fixed as it is inside the long-run relation as a constant.

Before proceeding to threshold estimation, a supLM test for the presence of threshold effect is conducted on the unrestricted threshold FCVAR model. The result of the test shown in table 2.5 supports a 2-regime threshold effect for the nonlinear adjustment dynamics in the long-run relation of volatility spot and futures. The supLM statistic is calculated at 19.050 with a bootstrap p-value of 0.009, it is statistically significant at 1% level. The nonstandard distribution is calculated by 500 replications and with a trimming parameter $\pi_\gamma = 0.15$.

Table 2.5 about here (see P.48).

The estimation results are presented in table 2.6.

Table 2.6 about here (see P.49).

First, the estimate of fractional parameter $\hat{d}$ for the fractionally integrated volatility spot and futures is 0.848 indicates a pair of long memory nonstationary processes, which is close the range of d estimates by various semi-parametric estimators (see table 2.4). The cofractional parameter $\hat{b}$ is 0.771 which indicates the degree of cointegration between spot and futures is fairly strong. In addition, the memory of long-run relation $\beta'(X_t - \mu)$ is $d - b = 0.077$ shows that the long memory feature in disequilibrium error is mild. It is different from the semi-parametric estimates of the memory of volatility difference series $(s_t - f_t)$ which are ranging from 0.31 to 0.62. It might be due to the slope parameter β is estimated rather than taken as one, or due to the possible misspecification of d and b; yet further investigation should be given.

Second, recall that the cointegrating parameter β provides empirical evidence on the normal state of contango or backwardation in long-run. By the fact that VIX futures market has long-run contango $(s_t \leq f_t)$, the estimate $\hat{\beta}_2$ is expected to be less than one. Based on the estimations reported, only the threshold FCVAR (model 4) successfully obtained an estimate as expected, which is $\hat{\beta}_2 = 0.990$, while all other slope parameter estimates from model 1 to 3 are larger than unity. Moreover, the estimates of the mean level of volatility processes $\hat{\mu} = (\hat{\mu}_s, \hat{\mu}_f)'$ from all models reflect the long-run contango with $\hat{\mu}_f > \hat{\mu}_s$.

Next, we consider the threshold estimate γ and the sample split between regimes in the threshold models. This application takes a trimming parameter $\pi_0 = 0.1$ which implies at least 10% of observations will be in either one of the regimes. In model 4, the estimate of threshold parameter $\hat{\gamma} = 1.156$ is optimised through a 2-dimensional grid search over $[\beta, \gamma]$ with $[100 \times 100]$ grids. The bound $[\beta_L, \beta_U] = [0.98, 1.11]$ with increment $\Delta_\beta = 0.0013$, is given by the confidence interval of the consistent estimate $\tilde{\beta}_2$ in the equilibrium relation in the linear FCVAR model, in which $\tilde{\beta}_2 = 1.05$ ($s.e. = 0.01$) and 6 standard error apart from mean are used for calculation. The bound $[\gamma_L, \gamma_U]$ is given by the sorted distribution of the disequilibrium error $\tilde{e}_{t-1}$ with a trimming constraint $0.10 \leq P(e_t \leq \gamma) \leq 0.90$. The log-likelihood plot over the 2-dimensional grid $[\beta_2, \gamma]$ is shown in figure 2.4. Both MLE $\hat{\beta}_2$ and $\hat{\gamma}$ are precise and strongly convexified around a small region where the lowest negative likelihood lies on. An optimal pair of $[\hat{\beta}_2, \hat{\gamma}]$ yields the lowest value of negative log-likelihood. The percentage of observations are 89% in regime 1 and 11% in regime 2, which is realistic to suggest market turbulence accounts for 11% of the total sample period. However, the standard threshold cointegration (model 2) tells a different story regarding the threshold estimate and the sample split. The threshold estimate is negative, $\hat{\gamma} = -1.431$, and it divides the sample into 68% and 32% of observations in regime 1 and 2, respectively. Yet the observations in regime 2 still account for smaller proportion, it is surprising to say the market has 32% bad time over the sample period.

Figure 2.4 about here (see P.44).

The adjustment parameter $\alpha = (\alpha_s, \alpha_f)'$ is of interest in this empirical application. Threshold cointegration and threshold FCVAR models found different evidence in adjustment dynamics. First of all, the threshold FCVAR reports zero adjustment in the first regime which implies that the cointegrating relation of VIX spot and futures dies out when the market is calm. The adjustment parameter estimate in the second regime is $(-0.006, 0.213)'$ with $s.e. = (0.080, 0.029)'$, which indicates during the turbulent time, there is an error correction dynamics by VIX futures but the adjustment by spot remains minimal. Note that $\alpha_f^2 = 0.213$ means that volatility futures in period $t + 1$ will adjust on average 21.3% of the disequilibrium error e_t in period t. The result of adjustment dynamics from threshold FCVAR mdoel successfully provides economic meaning to the asymmetric hedging behaviour by using VIX products. When market is highly uncertain, i.e. when volatility spot is much higher than futures, people tend to hedge against the downside risk. Hence, futures price has upward tendency and reverts the volatility difference towards the long-run equilibrium relation. On the other hand, when market is normal, volatility products are less attractive, thus there is no noticeable adjustment by VIX futures. Comparing to threshold FCVAR, standard threshold cointegration found opposite results in adjustment parameters across regimes. Futures exhibits mean-reverting behaviour in the first regime with $\alpha_f^1 = 0.072$ (significant at 5%); however, the cointegrating relation is statistically insignificant in the second regime.

Lastly, the linear fractional and nonfractional cointegration models conceal the heterogeneous error correction dynamics across different regimes due to the linear model specification; yet both model 1 and 3 found significant adjustment behaviour from VIX futures, standard cointegration also found significant adjustment by spot.

2.6.6 Momentum strategy

Empirical results demonstrate that the disequilibrium error e_t from the long-run relation of VIX spot and futures exhibits regime dependent responses given different economic states. Financial practitioners refer the states as (short-run) contango and backwardation and use them as an enter-and-exit directional signal for frequent trading. For example, from the estimation of 2-regime threshold FCVAR model, one may enter into VIX futures contract when the error $e_t = (s_t - \mu_s) - \beta_2(f_t - \mu_f)$ is larger than $\hat{\gamma} = 1.156$, i.e. in regime 2. As illustrated in figure 2.5, the sequence of switches between regime 1 and regime 2 seems to be able to detect the market turbulence, such as the Chinese stock market crash in 2015 and Brexit referendum in 2016.

Figure 2.5 about here (see P.45).

This section attempts to examine the profitability of such strategy based on the information arrival of contango and backwardation of the VIX futures market. The portfolio performance is compared to those from the benchmark $I(1)/I(0)$ threshold cointegration and the simple $s_t - f_t$ definition for contango and backwardation adopted by traders. A simple strategy is constructed by:

> *longing* VIX futures when it is in backwardation, holding a long position until it enters into contango.

This trading strategy takes the advantage of contract rolling, because futures is bought at discount in backwardation and sold at premium in contango. It is noteworthy that this strategy may expose to risks associated with adverse moves in VIX futures, other researchers (see. Simon and Campasano, 2014) suggested a more sophisticated strategy by hedging the adverse moves of VIX futures with opposite direction of mini S&P500 futures. This study would rather keep the trading application straightforward to compare the nonlinear trading signals generated from different models. Both threshold FCVAR and threshold cointegration use the threshold estimate $\hat{\gamma}$ as the trigger point to buy VIX futures, the decision rule is longing futures when $e_t > \hat{\gamma}$. The commonly adopted contango-backwardation strategy fixed the threshold γ at zero and β at one, hence traders long futures when $s_t - f_t > 0$.

An abstract portfolio is constructed as follows. Assuming there is an initial investment of \$1000 in VIX one-month futures[14]. The sample period from 4 January to 30 December 2016 is tested. It is important to note that VIX futures settlement is calculated with a Special Opening Quotation using opening prices of actual traded S&P500 options that expire 30 days from the day of settlement. Accumulated profit is calculated as the sum of profit(or loss) in futures contract times principal over the sample period.

Table 2.7 about here (see P.50).

Table 2.7 shows the profit performance using different sequences of enter-and-exit information for trading VIX futures. Regarding the total profit, the $s_t - f_t$ contango-backwardation strategy harvested the largest amount of profit, it is \$962.70, while threshold cointegration gained the least, it is \$399.98; the proposed threshold FCVAR model earned \$454.01. However, if we consider the number of trading signals each model are given, threshold FCVAR model provided only 7 directions for the entire year and only 2 out of 7 (29%) made a loss. The simple contango-backwardation gave 14 signals and 5 out of 14 (36%) made negative profit. The threshold cointegration was the noisiest one which provided 17 directions but 10 out of 17 (59%) generated negative profit. If other transaction costs and slippage are taking effect in the portfolio, the more number of transactions implied the more profit can be eaten up.

Threshold FCVAR provides the most conservative signal to enter a long position in VIX futures, while threshold cointegration seems providing an aggressive trading signal. Given the similar representation theory of threshold models, one may concern the signalling difference is a consequence of the misspecification of long memory parameters. Any unabsorbed long range dependence in the nonfractional threshold cointegration might be wrongly assigned to threshold adjustment dynamics, and ends up generated lots of noise in adjustment dynamics. As shown in section 2.5, misspecifying the memory parameters distorted the estimates of threshold and adjustment parameters.

From this illustrative example, regime dependent adjustment dynamic is found to be useful to provide investment direction for VIX futures. Due to the simplicity of the portfolio construction, there is no enough evidence to conclude which model provides the most profitable signal for VIX trading. If one is interested in VIX trading, more effort should be given to decide sophisticated strategies in a realistic portfolio.

[14] There are various VIX-tracking derivatives products available on financial market, e.g. iPath S&P500 VIX short-term futures ETN, ProShares Ultra VIX short-term futures ETF, Velocity Shares Daily 2X VIX short-term ETN, etc. The same strategy can be applied to different VIX products, though one should beware the imperfect correlation between the underlying futures contracts and the ETF/ETN products.

2.7 Conclusion

This research contributes to the coexistence of long memory and nonlinearity in time series analysis. The recently developed fractionally cointegrated vector autoregressive (FCVAR) model is extended to accommodate regime dependent adjustment dynamics in the error correction mechanism. A model-based supLM test is derived to test for the presence of threshold. Since the distribution of test statistic is nonstandard given by the presence of nuisance parameter in the restricted model in null hypothesis, Bootstrapping test statistic and p-value are deriving by simulation. Size and power of the test are verified. Simulation evidence regarding the model misspecification is also presented in support of the need of accurate estimates on memory parameters and threshold.

With the presence of long memory in bivariate volatility processes, their long-run relation is found to exhibit regime switching, in which the states of switch can be explained by different situations - contango and backwardation - of futures markets. Previous theoretical studies mainly paid attention to the equilibrium relation among commodity futures market and investigate the long-run contango and backwardation. However in financial practice, the nonlinear short-run deviations from normal relation effectively provide investment direction.

Still, there are many interesting questions worth for further research. The asymptotic theory in the context of threshold model is challenging. The main problem is due to the discontinuity in threshold. One extension to this recent research is to investigate different nonlinear forms in the FCVAR model. A smooth transition FCVAR model could be a promising attempt, in which asymptotic theory is more straightforward to be derived given its continuity. Furthermore, a more comprehensive hypothesis testing for the presence of threshold in FCVAR in which the null of linear no fractional cointegration against nonlinear fractional cointegrated alternative could be an improvement to the current two-step testing approach. It is of an empirical interest to examine the coexistence of long memory and short-run regime dependent adjustment in other macroeconomic and financial time series.

Figures of Chapter 2

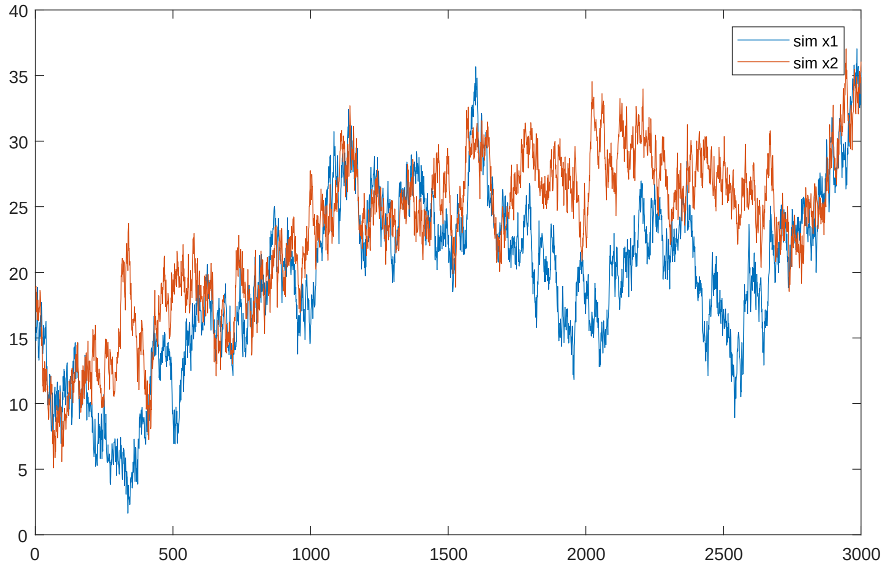


Figure 2.1: X_t generated by a 2-regime threshold FCVAR model.

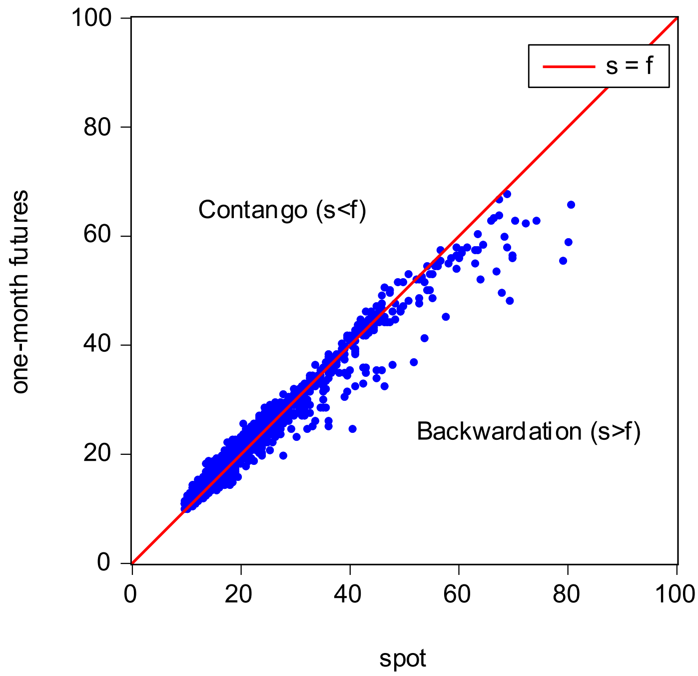


Figure 2.2: Scatter plot of VIX index spot and one-month futures for the observation from 26 March 2004 to 30 December 2016.

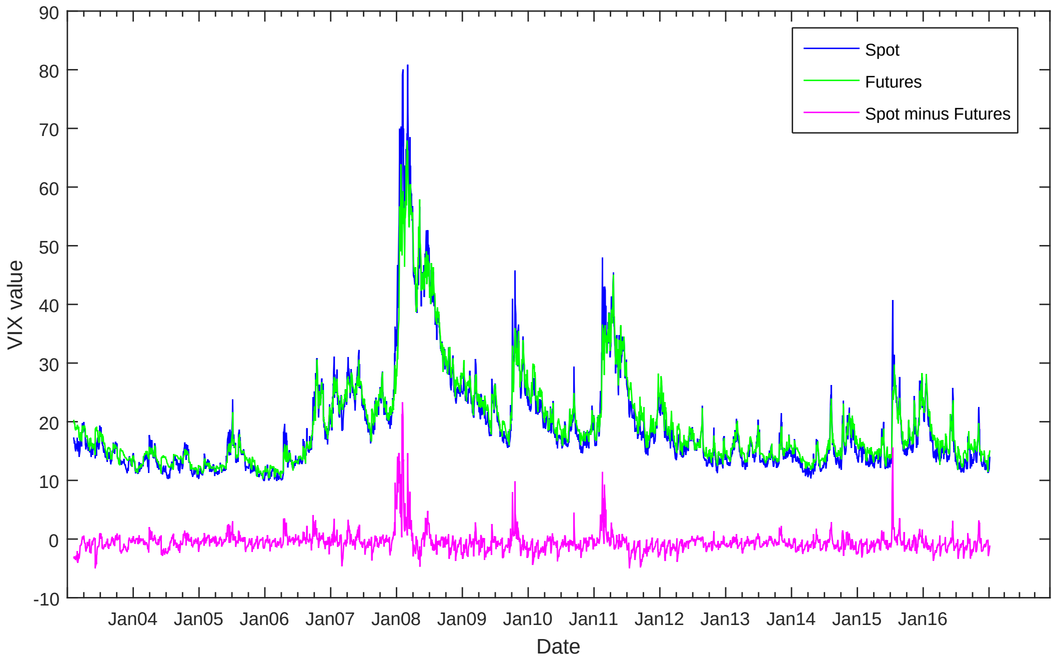


Figure 2.3: Time series plot of VIX index spot, one-month futures and the difference $(s_t - f_t)$ from 26 March 2004 to 30 December 2016.

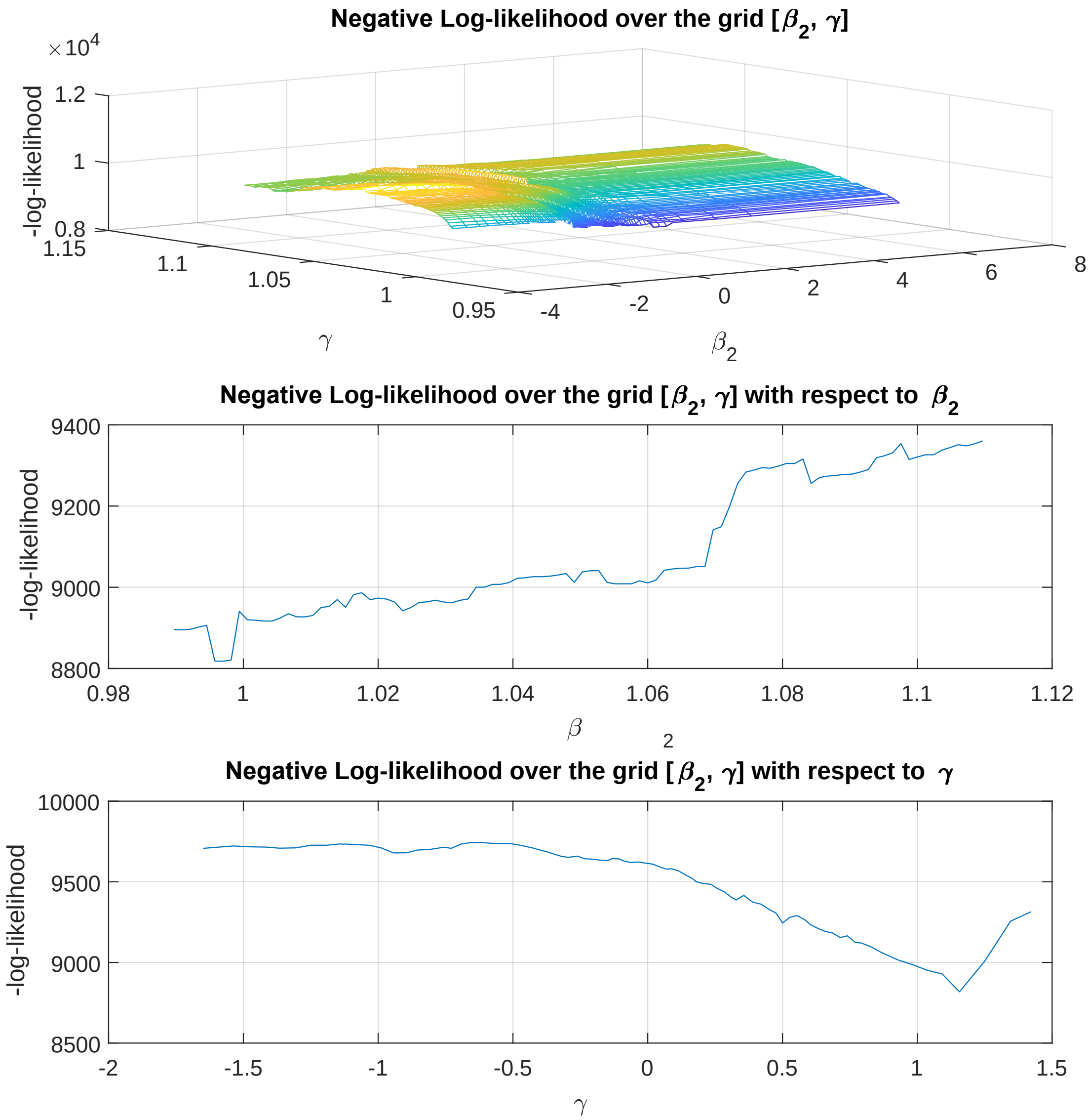


Figure 2.4: Maximum log-likelihood over the two-dimensional grid search $[\beta, \gamma]$.

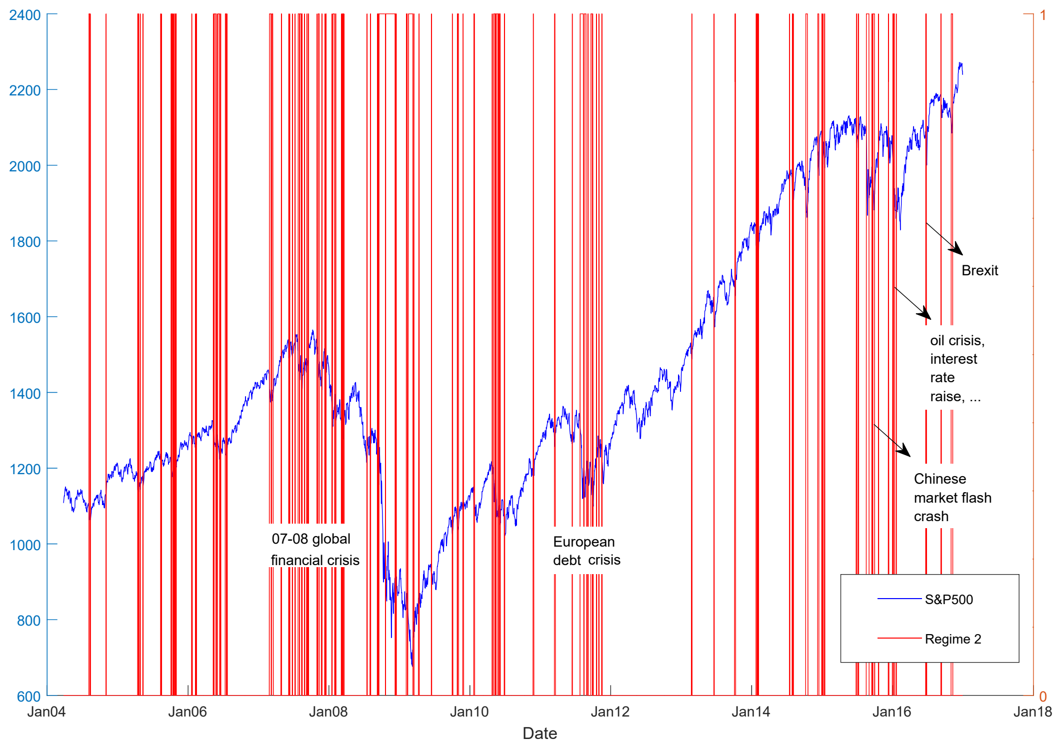


Figure 2.5: Time series plot of threshold error correction dynamic in Regime 2 ($e_{t-1} > \gamma$).

Tables of Chapter 2

level of significance		**5%**			**10%**	
α_2	0	0.25	0.5	0	0.25	0.5
$n = 200$, $\Gamma_i = \Gamma_0$						
Fixed regressor bootstrap	0.088	0.084	0.079	0.130	0.128	0.132
Residual bootstrap	0.067	0.056	0.062	0.116	0.109	0.111
$n = 500$, $\Gamma_i = \Gamma_0$						
Fixed regressor bootstrap	0.073	0.065	0.069	0.125	0.121	0.119
Residual bootstrap	0.056	0.051	0.052	0.107	0.098	0.101
$n = 200$, $\Gamma_i = \Gamma_1$						
Fixed regressor bootstrap	0.079	0.081	0.082	0.126	0.121	0.130
Residual bootstrap	0.072	0.062	0.060	0.113	0.106	0.120
$n = 500$, $\Gamma_i = \Gamma_1$						
Fixed regressor bootstrap	0.071	0.068	0.070	0.125	0.121	0.119
Residual bootstrap	0.058	0.053	0.055	0.111	0.105	0.108

Table 2.1: Size of supLM Test.

$\pi_0 = P(e_{t-1} \leq \gamma)$		**0.2**			**0.5**	
α_2^2	0.2	0.5	0.8	0.2	0.5	0.8
n = 200						
Fixed regressor bootstrap	0.317	0.495	0.649	0.330	0.508	0.714
Residual bootstrap	0.297	0.392	0.594	0.308	0.476	0.687
n = 500						
Fixed regressor bootstrap	0.531	0.875	0.939	0.558	0.896	0.940
Residual bootstrap	0.486	0.819	0.923	0.591	0.824	0.931

Table 2.2: Power of supLM Test at 5% size.

	DGP **2-Regime FCVAR**	*Model (1)* Linear CVAR	*Model (2)* 2-Regime CVAR	*Model (3)* Linear FCVAR
d	**0.8**	1	1	0.802 (0.001)
b	**0.6**	1	1	0.573 (0.008)
β	**[1, −1]**	[1, −1.128]	[1, −1.020]	[1, -1.084]
μ_1	**10**	−0.089 (0.030)	-	15.240 (0.030)
μ_2	**10**	0.183 (0.029)	-	18.062 (0.028)
γ	**1**	-	0.518	-
	regime 1	linear	regime 1	linear
α_1^1	**0**	−0.015 (0.003)	−0.012 (0.004)	−0.0158 (0.0118)
α_2^1	**0**	0.027 (0.003)	0.004 (0.003)	0.2360 (0.0291)
μ_1^1	-	-	−0.032 (0.032)	-
μ_2^1	-	-	0.068 (0.034)	-
	regime 2		regime 2	
α_1^2	−0.005	-	−0.089 (0.039)	-
α_2^2	**0.250**	-	0.155 (0.040)	-
μ_1^2	-	-	0.038 (0.090)	-
μ_2^2	-	-	0.164 (0.093)	-

Notes:
(1) The coefficients of DGP in the first column assume similar values as in the process of VIX spot-futures relation.
(2) All other cointegration models assume one cointegrating relationship between $x_{1,t}$ and $x_{2,t}$ and the number of lags $k = 0$.

Table 2.3: Coefficient estimates of the threshold FCVAR DGP from three different cointegration models.

	$\hat{d}$ for Spot (s_t)			$\hat{d}$ for futures (f_t)			$\hat{d}$ for $s_t - f_t$		
Fourier frequencies	GPH	LW	ELW	GPH	LW	ELW	GPH	LW	ELW
m = $\lfloor n^{0.5} \rfloor = 56$	0.836	0.757	0.761	0.918	0.838	0.849	0.388	0.314	0.310
$\lfloor n^{0.6} \rfloor = 127$	0.923	0.906	0.908	0.966	0.962	0.968	0.553	0.574	0.622
$\lfloor n^{0.7} \rfloor = 285$	0.879	0.836	0.842	0.936	0.889	0.895	0.465	0.479	0.484

Notes:
(1) The table reports semi-parametric estimators of long memory parameter d for S&P500 option implied volatility spot s_t, one-month futures f_t and their difference $s_t - f_t$. Three semi-parameter estimators are considered, they are: Geweke-Porter-Hudak (GPH) estimator, local Whittle (LW) estimator and the 2-step exact local Whittle (ELW) estimator. The estimations use the computation algorithms by Kanzler et al. (1998) for the GPH estimator and Shimotsu and Phillips (2005) for LW and ELW estimators.
(2) The number of Fourier frequencies equals to $m = \lfloor n^w \rfloor$ where m is the floor integer part, n is the number of observations and w is the bandwidth size. According to literature, the choices of bandwidth are usually ranging from 0.25 to 0.8. This study reports three bandwidth choices, 0.5, 0.6 and 0.7.

Table 2.4: Estimates of d for VIX spot, futures, and the difference between spot and futures

	2-Regime FCVAR (spot, futures)
No of replications	500
SupLM Test Statistic	19.050
Critical Value at 5%	15.134
p-value	0.009

Table 2.5: supLM test for the presence of a threshold in the FCVAR model.

	Model 1 $I(1)/I(0)$	*Model 2* 2-Regime $I(1)/I(0)$	*Model 3* Linear FCVAR	*Model 4* 2-Regime FCVAR
d	1	1	0.848	0.848
b	1	1	0.771	0.771
β	$[1, -1.071]$	$[1, -1.069]$	$[1, -1.050]$	$[1, -0.990]$
μ_s	-0.216	-	20.419	20.419
μ_f	0.140	-	20.912	20.912
γ	-	-1.431	-	1.156
		regime 1		regime 1
	(*linear*)	(*contango*)	(*linear*)	(*contango*)
% of obs.	-	68%	-	89%
α_s^1	-0.111 (0.054)***	-0.052 (0.044)	-0.007 (0.030)	-0.000 (0.035)
α_f^1	0.074 (0.026)***	0.072 (0.037)**	0.156 (0.021)***	0.000 (0.029)
μ_s^1	-	-0.061	-	-
μ_f^1	-	0.108	-	-
		regime 2		regime 2
		(*backwardation*)		(*backwardation*)
% of obs.	-	32%	-	11%
α_s^2	-	-0.143 (0.101)	-	-0.006 (0.080)
α_f^2	-	0.045 (0.053)	-	0.213 (0.029)***
μ_s^2	-	-0.220	-	-
μ_f^2	-	0.179	-	-
likelihood	-609.635	-600.600	-9661.439	-8817.840

Notes:
(1) The table shows estimation results for the cointegrating relationship between VIX spot and futures from four different cointegration models for comparison. Model 1 and 2 are the standard I(1)/I(0) and threshold cointegration, respectively which are estimated using the Matlab programme of Hansen and Seo (2002). Model 3 is the fractionally cointegrated VAR model estimated using the Matlab programme of Nielsen and Popiel (2016). Model 4 is the 2-regime threshold fractionally cointegrated VAR model proposed by this study. The main Matlab codes are provided in appendix B.
(2) The likelihood value for each model is the maximum log-likelihood value resulted by the optimal parameter estimates. The larger value, the better measure of fit to the data.
(3) one, two and three asterisk(s) indicate 10%, 5% and 1% level of significance, respectively.

Table 2.6: Estimation results from four different cointegration models.

	Enter Date	Exit Date	Return	Profit of $1,000 investment
Threshold FCVAR	04 Jan, 16	05 Jan, 16	4.7%	46.7
	07 Jan, 16	12 Jan, 16	−2.4%	−23.6
	21 Jan, 16	22 Jun, 16	−3.4%	−34.5
	22 Jun, 16	23 Jun, 16	0.6%	5.9
	24 Jun, 16	27 Jun, 16	7.6%	76.4
	09 Sep, 16	12 Sep, 16	22.2%	221.9
	01 Nov, 16	09 Nov, 16	16.1%	161.3
	Total profit			$454.01
Threshold Cointegration	04 Jan, 16	19 Jan, 16	28.1%	281.2
	21 Jan, 16	22 Jan, 16	−3.4%	−34.5
	08 Feb, 16	10 Feb, 16	1.3%	12.8
	15 Mar, 16	16 Mar, 16	−5.6%	−55.6
	18 Apr, 16	21 Apr, 16	−4.5%	−45.2
	17 May, 16	19 May, 16	6.1%	61.3
	15 Jun, 16	16 Jun, 16	4.5%	45.2
	20 Jun, 16	21 Jun, 16	−4.4%	−43.6
	22 Jun, 16	27 Jun, 16	25.5%	255.1
	28 Jun, 16	29 Jun, 16	−4.0%	−40.0
	19 Jul, 16	21 Jul, 16	−3.5%	−34.8
	15 Aug, 16	18 Aug, 16	−0.2%	−2.0
	09 Sep, 16	14 Sep, 16	7.0%	70.1
	19 Sep, 16	21 Sep, 16	−3.9%	−38.6
	31 Oct, 16	09 Nov, 16	20.2%	201.7
	14 Nov, 16	17 Nov, 16	−7.6%	−76.3
	14 Dec, 16	22 Dec, 16	−15.7%	−156.8
	Total profit			$399.98
$s_t - f_t$ ***Cotango-Backwardation Strategy***	04 Jan, 16	05 Jan, 16	4.7%	46.7
	06 Jan, 16	20 Jan, 16	36.7%	366.8
	21 Jan, 16	22 Jan, 16	−3.4%	−34.5
	25 Jan, 16	26 Jan, 16	−1.1%	−11.4
	08 Feb, 16	12 Feb, 16	6.8%	68.2
	20 Apr, 16	21 Apr, 16	4.3%	42.7
	17 May, 16	19 May, 16	6.1%	61.3
	15 Jun, 16	16 Jun, 16	4.5%	45.2
	22 Jun, 16	28 Jun, 16	12.0%	120.2
	09 Sep, 16	12 Sep, 16	22.2%	221.9
	13 Sep, 16	14 Sep, 16	−0.8%	−8.2
	20 Sep, 16	21 Sep, 16	−5.1%	−50.7
	01 Nov, 16	09 Nov, 16	16.1%	161.3
	14 Nov, 16	16 Nov, 16	−6.7%	−66.7
	Total profit			$962.70

Table 2.7: Profit performance using different sequences of enter-and-exit signal for buying VIX futures.

Chapter 3

Deriving synchronised daily correlations from asynchronous stock returns

3.1 Introduction

Accurately modelling the joint dynamics of asset returns across different markets is a fundamental requirement for understanding how strongly markets co-move and for quantifying the risk characteristics of portfolios containing assets from different geographical segments. An important concern when measuring return correlations across international markets arisen from the fact that assets trade at their local time hence causing the daily return series based on closing prices to be asynchronous at the time point of data collection. The asynchronicity of returns may lead to highly distorted correlation dynamics if the modellers do not take this issue into consideration.

This research produces reliable synchronised correlation estimates that allow financial practitioner to construct suitably adjusted series for the purposes of correlation analysis and portfolio risk management. Since synchronous returns are generally unobserved, an assumption is imposed by formulating the structure of synchronous returns as a function of some observables. The benchmark synchronisation model in Burns et al. (1998) (henceforth BEM) imposed a random walk assumption on synchronous prices. In their case, any lead-lag correlations among the stock returns of different markets are considered as misspecified correlation. This assumption is somehow restrictive because it rules out any lead-lag movements between markets. Instead, the synchronised correlations proposed in this study are derived from a vector autoregressive process of asynchronous stock returns with less restrictive assumption on price processes. The formulation of synchronised returns assumes a fraction of asynchronous returns from the later close market contains information to explain part of the synchronised returns from the earlier close market. This is a more intuitive and loose assumption comparing to the assumption of random walk. More importantly, the existing asynchronous GARCH model of BEM is nested in the proposed synchronised model of this study. This model is identical to BEM's model with certain restriction imposed to the parameter. In other words, the synchronisation model in this study may provide implications about the degree of market efficiency among the stock markets with asynchronous trading hours.

Correlation dynamics play an important role in many financial applications. One may be interested in estimating how the assets move in relation to each other before constructing a multi-asset portfolio or setting up their hedging strategies. The multivariate time series models, e.g vector autoregressive moving average process, had long been adopted to explore correlation dynamics and transmission mechanism between the financial assets. Yet, significant bias on the correlation dynamics may result from the assumption that multiple time series are sampled simultaneously but in fact the sampling is non-synchronous (Lo and MacKinlay, 1990).

When Information flows continuously across international stock markets, stock prices change in response to the relevant information. However, assets at different markets trade at their local trading time and hence the prices are only recorded in a discrete time basis. Stock prices are stale at closing time and are no longer reflecting the current market values upon new information. The closing prices observed from the different time of measurement are known to be asynchronous.

To illustrate the phenomenon of return asynchronicity, think of an internationally diversified portfolio contains equity investment in Japanese NIKKEI 225, the UK FTSE 100 and the US S&P 500 (see figure 3.1 for a graphical illustration). Suppose at the time of the US market closes, the S&P 500 reports a drop by 1% of stock price, the Japanese and the UK markets had already closed thus cannot respond to the innovations from US. If both of the Japanese and the UK stock markets are positively correlated with the US market, the true market values of the Japanese and the UK should have declined at the same trading day in response to the drop from the US; however, their closing prices are stale but their next day opening prices would assimilate the overnight information and show decline in values. Consequently, the use of asynchronous closing prices underestimates the contemporaneous correlations and leads to spurious lag-1 cross correlations. The systematic error on correlation is bigger when the asynchronous trading difference between markets is larger, as there is larger portion of asynchronous return incorporated to the next day's return. In our example, the systematic error on correlation between the US and the Japanese stocks is expected to be bigger than the US and the UK one. Scherer (2013) found empirical evidence that the unreliable correlations affect risk models arrive at too low (high) VaR forecasts for long (short) position in portfolios, and risk decisions arrive at too small hedge ratios. Therefore, the phenomenon of asynchronicity should be taken into consideration when modelling the stock returns of global financial markets.

Figure 3.1 about here (see P.72).

By looking at some international stock markets which are trading partially overlap with others, we could have a comparative view of how the degree of systematic error on correlation depends on the asynchronous timing. Figure 3.2 compares the unconditional correlations between the UK FTSE 100 and the US S&P 500 using synchronous (16:30 GMT) and asynchronous (closing) data[1], in which the time horizon moves from 1-day to 20-day return intervals. As we can observe the systematic error on correlations diminishes when the time horizon becomes longer, by the reason that the degree of asynchronicity is less sensitive for the longer sampling time interval[2]. The daily correlation calculated from asynchronous data is significantly lower than the one from synchronous data (0.52 vs. 0.82). Once past the 5-day (weekly) interval, the effect of asynchronicity on correlation is minor.

Figure 3.2 about here (see P.73).

Among the existing empirical studies related to the price co-movement and return correlation, some studies (Theodossiou et al., 1997; Ramchand and Susmel, 1998; Chow et al., 2003) simply by-pass the asynchronicity problem using weekly or monthly data. However, the low frequency data is relatively small sample which may lead to inefficient parameter estimates in multivariate time series, and in practice the potential investors or decision makers especially treasure the correlation measures from shorter horizons. It is because the daily or even the higher frequency data allow the exploration of market microstructures at the same time acquiring more relative information. Moreover, the daily synchronous prices can only be observed for limited markets (e.g. UK with US) which have common trading within 24 hours. For other stock markets (e.g. Japan with US) which have no overlapping trading, the asynchronicity issue remains unsolved.

The issues of asynchronous data take on greater importance in today's global financial applications. Various asynchronous problems have been studied for many years. Perhaps Scholes and Williams (1977) is the first literature that considered the effects of asynchronous trading on asset modelling. They proposed a statistical method to estimate a consistent Beta at the Capital Asset Pricing Model. Dimson (1979) and Cohen et al. (1980; 1983) considered the asynchronous returns caused by infrequent trading or other frictions in trading process can bias the beta estimates of the asset pricing models, hence provided an analytical expression to the relationship between observed returns and true

[1]The UK FTSE 100 is trading from 8:00 to 16:30 GMT, while the US S&P 500 is trading from 9:30 to 16:00 EST which corresponds to 14:30 - 21:00 GMT; thus the US and the UK stock markets have 2 hours (14:30 - 16:30 GMT) of contemporaneous trading. Then the synchronous prices for both UK and US markets can be observed between 14:30 and 16:30 GMT. In this illustration, the synchronous data are collected from Bloomberg at 16:30 GMT using 30-minute frequency.

[2]The idea is that the 5-hour trading difference between the UK and the US markets has greater impact on the correlations at daily 24-hour interval than at weekly (5 trading days) 120-hour interval.

returns. Another milestone of study, Lo and MacKinlay (1990), focused on the spurious correlations induced by asynchronous data, analysed the market microstructure using high-frequency data by a stochastic model with some probabilities of trading (or non-trading) assigned to the data at each time interval. Riskmetrics^{TM} provided an explicit expression for synchronised correlations as a function of asynchronous correlations. The drawback of Riskmetrics is that the syncrhonised covariance matrix is not guaranteed to be a positive definite hence the synchronised correlations are not bounded on $[-1, 1]$.

More related to this study, Burns et al. (1998) proposed a data synchronisation method particular for daily return series whose stock markets are located at different time regions. They used a first-order vector moving average process as a linear projection of the asynchronous part of the returns. Their synchronising technique can be adopted as either the preliminary synchronisation step on asynchronous data before applying the multivariate time series models, or the mean process of the multivariate-GARCH type procedure. Martens and Poon (2001) applied BEM and compared it with Riskmetrics on their Value-at-Risk measures. The BEM synchronisation approach produces better VaR estimates than the Riskmetrics. Audrino and Buhlmann (2004), Scherer (2013) and Bell et al. (2013) improved the BEM method with a first-order vector autoregressive process, which is somehow simpler due to the Markov structure, because the conditional expectation of a Markov process depends only on finite number of previous term(s).

The synchronisation methodology of this study mainly relaxes random walk stock prices assumption in BEM's technique. It is an important consideration because the efficient market hypothesis applied to asynchronous data can be less reasonable. Unlike synchronous markets have mutual information for every participant, asynchronous markets have asymmetric information where investors may behave differently. Given asymmetric information from the global financial markets, investors can only assess the best guess (expectation) of other's market values, instead of the contemporaneous observations from others.

The remainder of this study is organised as follows. Section 3.2 discusses the existing synchronisation methodology introduced by Burns et al. (1998). Section 3.3 discusses the econometric techniques adopted in this study. The assumptions and propositions imposed are also explained. Section 3.4 conducts an empirical analysis on seven international stock markets located at different time regions. Section 3.5 presents a Value-at-Risk back-testing exercise on both asynchronous and synchronised returns. The performance of Value-at-Risk one-step ahead forecast from different returns series are discussed. Section 3.6 concludes the study.

3.2 The existing synchronisation model

Burns et al. (1998) considered the trading time difference of international stock markets underestimated daily return correlations and volatility measures. Their study proposed a solution to estimate a synchronised return series and from that the correlation estimates are free of systemic error. The idea is to recognise that asset values may change even when markets are closed, these unrecorded asset prices can be estimated for use before the markets reopen. When a market closes before synchronised time t, there are past values of this market $P^i_{t_i-1}$, $P^i_{t_i-2}$, ... and new information from other subsequently close markets $P^j_{t_j}$, $P^k_{t_k}$, ... for $i < j < k$ that can be used to predict what the market price of this earlier close market would be if it was open.

A random walk process is assumed in the stock prices series such that any future changes of stock prices from the synchronised time are unpredictable. In this way synchronised returns R_t can be formulated by an innovation with mean zero and time-varying covariance matrix H_t. Assuming zero mean in the synchronised stock returns, then it is given by

$$R_t = \epsilon_t.$$

Given the assumption of random walk prices, the synchronised prices are also unbiased estimates of the next recorded prices. In other words, the conditional expected future price movements beyond the synchronised time t given the complete information at synchronised time t is zero. Then synchronised returns can be written in terms of asynchronous returns plus the next day's expected asynchronous returns given information at synchronised time t (which is the missing part) minus today's expected asynchronous returns given information at $t-1$ (which is the extra part):

$$R_t = r_t + E[r_{t+1}|\mathcal{F}_t] - E[r_t|\mathcal{F}_{t-1}], \tag{3.1}$$

denote the logarithmic asynchronous close-to-close return as r_t.

The random walk stock prices assure synchronised returns do not depend on the past stock prices. However, asynchronous returns show serial cross-correlations in practice. It is because different markets measure close-to-close returns at different time and there are time shifts forward for the later close markets. As a result, next day's asynchronous returns of the earlier close markets are predictable from the current asynchronous returns of the later close markets. The vector of asynchronous returns is modelled by a vector first-order moving average process with a GARCH covariance matrix to capture the return predictability for one day in the future as

$$\begin{aligned} r_t &= M u_{t-1} + u_t \\ Var_{t-1}[u_t] &= h_t, \end{aligned} \tag{3.2}$$

where the error term u_t is assumed to be serially uncorrelated and $E[u_t|\mathcal{F}_{t-1}] = 0$, h_t is a time-varying conditional covariance matrix.

The formulation in (3.2) is named asynchronous GARCH model. The first-order moving average coefficient M is a $J \times J$ matrix. The random walk stock prices suggest that the coefficient matrix should be zero if the stock prices are synchronised. Therefore, any predictability captured by M is entirely attributed to time asynchronicity of returns vector. This predictability resulted from asynchronicity is spurious and doesn't rule out the random walk assumption on stock prices. With elements ordered by the closing time of J markets, the diagonal and below-diagonal elements of M should be zero, since there are no overlaps in trading time. This implied that coefficient M should be in this form if the innovations u_t are serially uncorrelated. BEM only restricted the last row of matrix M must be zero and empirically resulted in some non-zero estimates on the diagonal and below-diagonal elements.

From model (3.2), the conditional expectation of r_t is $E[r_t|\mathcal{F}_{t-1}] = Mu_{t-1}$, synchronised returns R_t in equation (3.1) can be constructed using the VMA(1) parameter as

$$\begin{aligned} R_t &= r_t + E[r_{t+1}|\mathcal{F}_t] - E[r_t|\mathcal{F}_{t-1}] \\ R_t &= \left(Mu_{t-1} + u_t\right) + Mu_t - Mu_{t-1} \\ &= (I + M)u_t. \end{aligned} \tag{3.3}$$

This synchronising procedure brings forward the fraction of daily return which is occurred but not yet recorded by the synchronised time. The covariance matrix of synchronised returns H_t is

$$Var[R_t|\mathcal{F}_{t-1}] = (I + M)h_t(I + M)'. \tag{3.4}$$

The synchronised covariance matrix H_t is positive-definite since the asynchronous covariance matrix h_t is positive-definite. The asynchronous variances and covariances are typically smaller as some of the variability of asynchronous returns are spread across days.

Their asynchronous GARCH is applied to the G-7 equity markets[3] which obviously have asynchronicity problem. Empirical findings yielded the unconditional correlation estimates from asynchronous data are too small for the "high-asynchronous" markets. Although there is no reason to believe correlations are always large for each pair of markets, the weekly correlations are much higher than the daily correlations for the high-asynchronous markets. The weekly data are time-aggregated thus have less degree of asynchronicity, yet they are not perfectly synchronised. Empirically the unconditional correlation estimates from synchronised returns are slightly larger than the weekly correlations.

The conditional covariance, correlations and forecasts of these for both asynchronous and synchronised returns can be computed at each particular time by their proposed asynchronous GARCH model. In their study, a BEKK type multivariate GARCH is assumed in the asynchronous covariance matrix. Recall that this synchronisation method based on the assumption of random walk stock prices required serially uncorrelated error

[3] The G-7 equity markets include France CAC40, Germany DAX30, the U.K. FTSE 100, Italy MIL, Japan NIKKEI225, the U.S. S&P500 and Canada TSX. Some markets are trading perfectly synchronous and some other are completely asynchronous.

terms and the diagonal and below diagonal elements of the moving average matrix be zero. However, BEM only restricted the last row of the moving average matrix must be zero and empirically resulted in some non-zero estimates on the diagonal and below-diagonal elements. Moreover, the post-estimation diagnostics by Ljung-box tests for testing the dependence of standardised residuals showed that the standardised residuals are serially correlated; the diagnostics tests indicated that the error terms from the asynchronous GARCH model still have some non-captured patterns and a richer model may be needed.

Since the synchronisation model proposed by Burns et al. (1998) is an important contribution to the applications in risk management, other literatures adopted the asynchronous GARCH focused on testing the forecasting performance of the synchronised data, but they put less focus on the model specification and diagnostic tests. Martens and Poon (2001) compared BEM with RiskMetrics$^{\text{TM}}$ on their Value-at-Risk measures using both synchronised and synchronous[4] returns. Their key findings are that correlation dynamics are highly sensitive to the model chosen and the data used, and both synchronisation models add noise to the correlation dynamics. Regarding the VaR estimates, the RiskMetrics VaR provisions have fewer violations than the Asynchronous GARCH by using both data types; while the synchronised data resulted in fewer violations than the synchronous data on both synchronisation methods. However, those results should not be interpreted as one is superior to another. The number of VaR violations depends on the size of conditional covariance estimates. In fact, the RiskMetrics covariance is higher than the covariance from asynchronous GARCH, and the covariance from synchronised returns are expected to be higher.

It is noteworthy that Martens and Poon (2001) strictly followed the assumption of zero serial correlation and zero serial cross-correlations in the efficient market for both synchronisation models; therefore, the vector of asynchronous returns is assumed a VMA process the diagonal and the below-diagonal elements of moving average matrix in asynchronous GARCH are assumed to be zero. Other studies including Audrino and Bühlmann (2002), Scherer (2013) and Bell et al. (2013) assumed a first order autoregressive structure to the mean process of the asynchronous GARCH model since the first-order autoregressive Markov structure is simpler for estimation. However, those studies ignored the fact that the VAR(1) process on asynchronous returns violated the assumption of random walk stock prices, such that the formulation (3.1) for synchronised returns is invalid. This proposition can be expressed mathematically as follows.

Given the first order autoregressive asynchronous returns $r_t = Mr_{t-1} + u_t$ where $E[u_t|\mathcal{F}_{t-1}] = 0$, its conditional expectation $E[r_t|\mathcal{F}_{t-1}] = Mr_{t-1}$ forms the estimated

[4]The synchronised data are obtained from the synchronisation method proposed by BEM. Taking the advantage of partially trading overlap between the UK FTSE 100 and the US S&P 500 stock markets, the synchronous data are collected at 16:00 London time for both stock exchanges.

synchronised returns as

$$\begin{aligned} R_t &= r_t + Mr_t - Mr_{t-1} \\ &= Mr_t + u_t. \end{aligned} \tag{3.5}$$

Taking the conditional expectation on the estimated synchronised returns on the above, yields

$$E[R_t|\mathcal{F}_{t-1}] = M^2 r_{t-1}, \tag{3.6}$$

where the conditional expected synchronised returns are time-varying and depending on the past stock prices unless the first order autoregressive matrix $M = 0$. Recall that the formulation of estimated synchronised returns at (3.1) was built on top of the assumption that all the future stock price changes beyond the synchronised time t are unpredictable, which required a constant expected synchronised returns.

The assumption of random walk stock prices in BEM can be too restrictive. Symmetric information is observed by every market in the case of contemporaneous trading, hence informationally efficient market condition is easier to be achieved. Regarding the world of asynchronous trading, information is asymmetric - investors in markets that closed earlier are not able to response further innovations; even investors in the latest close market who have the most recent information do not know the true response from those markets have already been closed. Investors may respond to innovations slower than they should, due to the uncertainty on other markets' true values, and thus the information available at synchronised time t can possibly be used to predict at least some morning transactions in the next day. This may be the reason why there are non-zero elements in the coefficient matrix of asynchronous GARCH model, bringing forward all the predictability from the past to the current returns may over-allocate correlation estimates. Therefore, this study aims to modify BEM to provide a less restrictive synchronisation method by relaxing the assumption of random walk stock prices when formulating synchronised returns.

3.3 The synchronising methodology

3.3.1 Asynchronous and synchronised returns

Asynchronous returns are formulated in line with BEM. Since daily returns are measured from one time to 24 hours later next day at the same time, different markets may have different time of measurement.

Time. Concerning the time notation along this chapter, denote t_j as the closing time of each individual market j in an order of closing time from the earliest to the latest.

Denote t as the daily synchronised time vector where $t = \{1, 2, ..., T\} \in \mathbb{N}$. Let the closing time as a fraction of a day such that $t_{j-1} \leq t_j \leq t$.

Asychronous prices. The logarithmic closing price of market j is denoted as

$$P^j_{t_j} \quad \text{for} \quad j = 1, ..., J \quad \text{and} \quad t_j = \{t_1, t_2, ..., t_J\} \in \mathbb{R}^+. \tag{3.7}$$

Typically, market with the latest closing time of the day refers to the time of synchronisation. As an illustration in Figure 3.3, assume the closing price of the US S&P500 to be the last element of the closing price vector denoted P^{US}_1, the Japanese NIKKEI225 price is denoted $P^{\text{JPN}}_{0.375}$, and the UK FTSE 100 price is denoted $P^{\text{UK}}_{0.8125}$.

Figure 3.3 about here (see P.74).

If the closing prices of a market are observed at synchronising time $t \in \mathbb{N}$, the synchronised prices are just its closing prices, i.e. $P^{s,j}_t$; for instance the case for the US stocks. However, most of the international stock markets record their closing prices before time t, the synchronised prices are unobserved for those markets.

Synchronised prices. The logarithmic synchronised price $P^{s,j}_t$ are unobserved at time t and have to be estimated when market j is closed, given by information from markets that are open. Define the conditional expectation of the synchronised price of market j by

$$S^j_t = E[P^{s,j}_t | \mathcal{F}_t], \quad \textit{where } \mathcal{F}_t = \{P^j_{t_j} | t_j \leq t, j = 1, ..., J\}. \tag{3.8}$$

The complete information set $\mathcal{F}_t$, contains all recorded prices up to time t.

Let r_{t_j} be the $J \times 1$ vector of asynchronous close-to-close returns measured at different closing time t_j for j different markets as

$$r_{t_j} = \left(\Delta P^1_{t_1}, ..., \Delta P^J_{t_J}\right)' = \Delta P_{t_j}, \tag{3.9}$$

where $t_j = \{t_1, t_2, ..., t_J\}$ is a multi-index.

Let R_t be the $J \times 1$ vector of estimated synchronised returns as

$$R_t = \left(\Delta S^1_t, ..., \Delta S^J_t\right)' = \Delta S_t. \tag{3.10}$$

where $t = \{1, 2, ..., T\} \in \mathbb{N}$.

By (3.7)-(3.10), synchronised returns can be expressed in terms of closing prices and synchronised prices as

$$\begin{aligned} R_t &= S_t - S_{t-1} \\ &= (P_{t_j} - P_{t_j-1}) + (S_t - P_{t_j}) - (S_{t-1} - P_{t_j-1}) \\ &= r_{t_j} + E[P^s_t - P_{t_j} | \mathcal{F}_t] - E[P^s_{t-1} - P_{t_j-1} | \mathcal{F}_{t-1}] \end{aligned} \tag{3.11}$$

As shown in (3.11), synchronised returns can be seen as asynchronous returns with adjustments from the expectation of unobserved returns given the most updated information available.

Recall that BEM assumed random walk stock prices, any price change beyond the synchronised time point t is unpredictable given the most recent information. It follows that $S_t^j = E[P_t^{s,j}|\mathcal{F}_t] = E[P_{t_j+1}^j|\mathcal{F}_t]$, $t_j \leq t < t_j + 1$. It implies the conditional expected future price movements beyond the synchronised time t given the complete information at synchronised time t is zero $E[P_{t_j+1}^j - P_t^{s,j}|\mathcal{F}_t] = 0$. BEM defined the adjustment terms in (3.11) as $E[P_t^s - P_{t_j}|\mathcal{F}_t] = E[r_{t_j+1}|\mathcal{F}_t]$ and $E[P_{t-1}^s - P_{t_j-1}|\mathcal{F}_{t-1}] = E[r_{t_j}|\mathcal{F}_{t-1}]$.

This study is distinctive from BEM by relaxing the random walk assumption on stock prices at the definition of synchronised returns in (3.5).

Denote the stock return from the closing to the synchronising time of $P_t^s - P_{t_j}$ as the corrected return R_t^* and obtain

$$R_t = r_{t_j} + E[R_t^*|\mathcal{F}_t] - E[R_{t-1}^*|\mathcal{F}_{t-1}]. \tag{3.12}$$

If a market' synchronising time is just its closing time, its R_t^* vanished. However, if a market closes before synchronising time t, there is no record for its synchronised price and hence R_t^* $\forall t$ are unobserved.

It is easy to understand that the correction return R_t^* composes part of the next day's asynchronous close-to-close return which is observable. Therefore, we make the following assumption:

Assumption 3.1. *The $(J \times 1)$ vector of correction returns R_t^* for market j at time t makes up of a $(J \times J)$ matrix fraction A of the next day's asynchronous close-to-close returns r_{t_j+1}, denote*

$$R_t^* = Ar_{t_j+1}. \tag{A1}$$

The $(J \times 1)$ fraction parameter matrix A has elements α_{ij} for the observed market $i, j = 1, ..., J$. The fraction parameter projects the unobserved R_t^* from the asynchronous return r_{t_j}, thus it can be seen as a correlation matrix between R_t^* and r_{t_j}. To maintain a good property of fraction parameter, each element of matrix A is bounded on the closed interval $[-1, 1]$. The parameter A can either be constant or dynamic. In this study, we assume it is constant and unchanged overtime, then synchronised returns is re-written by substituting (A3.1) into (3.12) in terms of asynchronous returns by

$$\begin{aligned} R_t &= r_{t_j} + E[R_t^*|\mathcal{F}_t] - E[R_{t-1}^*|\mathcal{F}_{t-1}] \\ &= r_{t_j} + A(E[r_{t_j+1}|\mathcal{F}_t] - E[r_{t_j}|\mathcal{F}_{t-1}]). \end{aligned} \tag{3.13}$$

Synchronised returns equal asynchronous close-to-close returns plus a correction, which consists of linear combination of return increments from time $t + 1$ to t, representing some degree of dynamics adjusted to the multivariate synchronised return process given asynchronous returns.

Unlike Burns, et al.(1998) who imposed random walk hypothesis and made an approximation that $E[R_t^*|\mathcal{F}_t] = E[r_{t_j+1}|\mathcal{F}_t]$, our assumption only borrows the simple fact of asynchronicity and expresses the correction return as a fraction of the asynchronous return is less restrictive. Nevertheless, the random walk assumption can also be viewed as a special case within our Assumption 3.1 when parameter A equals an identity matrix.

Saying matrix A is a fraction parameter of $E[R_t^*|\mathcal{F}_t]$ on $E[r_{t+1}|\mathcal{F}_t]$ may be abstract, yet it has some economic expressions that i) when any diagonal element of A closes to one, the corresponding market may be more efficient and its own $E[R_{t+1}^{*,j}|\mathcal{F}_t]$ cannot predict its next day's return; ii) when the off-diagonal elements of A are statistically different from zero, the correction return $E[R_{t+1}^{*,j}|\mathcal{F}_t]$ can be explained by the next day's returns of other markets i for $i \neq j$.

3.3.2 The model

The formulation of synchronised returns in (3.13) depends on conditional expectation of asynchronous returns which should be modelled. As this model does not assume random walk stock prices, assuming the first-order autoregressive process to asynchronous returns will not contradict the formulation of synchronised returns.

Assumption 3.2. *Asynchronous returns for market j at day t follow a first order autoregressive process, i.e.*

$$r_{t_j} = c_0 + Br_{t_j-1} + e_t, \tag{A2}$$

with a $(J \times 1)$ vector of serially uncorrelated errors e_t such that $E[e_t|\mathcal{F}_{t-1}] = 0$, $E[e_t e_s'|\mathcal{F}_{t-1}] = 0 \; \forall t \neq s$, $E[e_t e_t'] = \Sigma^e$, a $(J \times J)$ first order autoregressive coefficient matrix B, and a $(J \times 1)$ vector of constants c_0. Assume that the VAR(1) of asynchronous returns satisfies the stationary condition, i.e. the root of $\det(I_J - B) = 0$ lies outside the complex unit circle, or $|B| < 1$.

From Assumption 3.2, synchronised returns in (3.13) is as

$$\begin{aligned} R_t &= r_{t_j} + A(E[r_{t_j+1}|\mathcal{F}_t] - E[r_{t_j}|\mathcal{F}_{t-1}]) \\ &= r_{t_j} + A(c_0 + Br_{t_j} - c_0 - Br_{t_j-1}) \\ &= (I_J + AB)r_{t_j} - ABr_{t_j-1}. \end{aligned} \tag{3.14}$$

Synchronised returns in (3.14) are expressed as a linear combination of two consecutive asynchronous returns. Next, a more specific stochastic process for synchronised returns is needed to model the left-hand side of equation (3.14).

Assumption 3.3. *The $(J \times 1)$ vector of synchronised returns is framed by a location-scale model in terms of its mean and variance, i.e.*

$$R_t = \mu_t + \epsilon_t \quad for\, t \in \mathbb{N}, \tag{A3}$$

with a $(J \times 1)$ vector of conditional mean $\mu_t = E[R_t|\mathcal{F}_{t-1}]$, a $(J \times 1)$ vector of innovation process $\epsilon_t = \Sigma_t Z_t$ where Z_t is a sequence of i.i.d. multivariate innovation variables with zero mean and unit variance. The conditional variance-covariance matrix of the innovation is $E[\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}] = \Sigma_t$. We can further consider GARCH-type structure for the error term ϵ_t.

Alternative parameter identification is also provided in Appendix B, in which the synchronised returns are not derived from Assumption 3.3 but from an auxiliary regression. Note that adopting the parameter identification from an auxiliaryy regression may suffer from efficiency loss of parameter estimates, since the parameter of interest A and B are identified from a set of structural parameters.

Using the derivation of synchronised returns in (3.14), the conditional mean of synchronised returns are:

$$\begin{aligned}\mu_t = E[R_t|\mathcal{F}_{t-1}] &= E[(I_J + AB)r_{t_j} - ABr_{t_j-1}|\mathcal{F}_{t-1}] \\ &= (I_J + AB)(c_0 + Br_{t_j-1}) - ABr_{t_j-1} \\ &= (I_J + AB)c_0 + (I_J + AB - A)Br_{t_j-1} \end{aligned} \tag{3.15}$$

Hence, the location-scale model of synchronised returns can be transformed by replacing R_t with the derivation in (3.14) and the conditional mean with the derivation in (3.15), to obtain

$$\begin{aligned}(I_J + AB)r_{t_j} - ABr_{t_j-1} &= (I_J + AB)c_0 + (I_J + AB - A)Br_{t_j-1} + \epsilon_t \\ \Leftrightarrow (I_J + AB)r_{t_j} &= (I_J + AB)c_0 + (I_J + AB)Br_{t_j-1} + \epsilon_t \end{aligned} \tag{3.16}$$

The matrix $(I_J + AB)$ is invertible. Pre-multiplying the matrix $(I_J + AB)$ the equation (3.16) on both sides, obtaining

$$r_{t_j} = c_0 + Br_{t_j-1} + (I_J + AB)^{-1}\epsilon_t. \tag{3.17}$$

The model structure in (3.17) is exactly the same as the asynchronous VAR(1) model in (A3.2), except the expression of asynchronous error term e_t is as a transformation of synchronised error term $(I + AB)^{-1}\epsilon_t$. It implied that the synchronised conditional variance-covariance matrix $E[\epsilon_t \epsilon_t' | \mathcal{F}_{t-1}] = \Sigma_t$ can be obtained from the asynchronous conditional variance-covariance matrix $E[e_t e_t' | \mathcal{F}_{t-1}] = \Sigma_t^e$ by

$$\Sigma_t = (I_J + AB)\Sigma_t^e (I_J + AB)'. \tag{3.18}$$

The derivation in (3.18) implied that the volatility structure of synchronised returns $Var[R_t|\mathcal{F}_{t-1}]$ can be produced from the volatility of asynchronous returns h_t, with a correction metric $(I + AB)$ derived from the mean equation of stock returns. The proposed conditional variance of synchronised returns in (3.18) can be modelled by some kinds of multivariate volatility models and taken trading asynchronicity into consideration.

It is interesting to see that the conditional variance in equation (3.18) has the same structure of the conditional variance equation by Burns et al. (1998)'s synchronisation in (3.4). The only discrimination between these two variance matrices is that their return correction $(I + M)$ is derived from a VMA(1) process with random walk assumption on stock prices; while the methodology proposed here uses a return correction matrix $(I + AB)$ which is derived from a VAR(1) process with Assumption 3.1 imposed in the relation of asynchronous and synchronous returns. Random walk stock prices is not necessarily be assumed in (3.18), but the critical implication is that the proposed return correction $(I + AB)$ provides flexibility to let the data shows the degree of co-movement among different asynchronous stock markets. For some extreme cases, let the fraction matrix $A = 0$, the synchronised variance-covariance matrix reduced to asynchronous variance-covariance matrix which implied no asynchronicity. Let $A = I$, the synchronised variance matrix reduced to the form of synchronised variance by BEM in (3.4).

3.3.3 The estimation

The model (3.17) involves two parameters to be estimated: the coefficient matrix B and the fraction parameter A. The model suffers from under-identification problem by estimating the process directly using OLS or maximum likelihood estimations, since there are two unknowns ϵ_t and A in one equation $e_t = (I+AB)^{-1}\epsilon_t$. Therefore, this study proceeds with a "two-step" maximum-likelihood procedure (see Greene, 2003 p.576-582, and Heckman, 1977).

To perform the two-step estimation, consider the following system of models constructed by asynchronous VAR in (A3.2) and synchronised VAR in (3.16), assuming the error terms in both models are normally distributed:

Model 1: $r_{t_j} = c_0 + Br_{t_j-1} + e_t$
Model 2: $(I_J + AB)r_{t_j} = (I_J + AB)c_0 + (I_J + AB)Br_{t_j-1} + \epsilon_t$ with
$\epsilon_t = (I + AB)e_t$.

There are two parameter matrices, B and A, to be estimated in the system of models. The coefficient B appears in both model 1 and 2, but the fraction matrix A only appears in model 2. See model 1 as a reduced model whereas model 2 as a full model, a two-step estimation procedure estimates the parameters in the reduced model, and then estimates the full model by embedding the consistent estimators from the reduced model. The two steps of maximum-likelihood estimations are conducted as follows.

Step 1: Estimating the coefficient matrix B of model 1 by maximising the log-likelihood function $\ln\mathcal{L}_1$ as

$$\begin{aligned}\ln\mathcal{L}_1(B,\Sigma^e) = &-\frac{TJ}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma^e| \\ &- \frac{1}{2}\sum_{t=1}^{T}[(r_{t_j} - c_0 - Br_{t_j-1})'(\Sigma^e)^{-1}(r_{t_j} - c_0 - Br_{t_j-1})]\end{aligned} \tag{3.19}$$

Step 2: Maximising the concentrated log-likelihood function $\ln\mathcal{L}_2$ in terms of parameter A embedded with the consistent maximum-likelihood estimators of B and Σ^e from step 1 as

$$\ln\mathcal{L}_2(A,\hat{B},\hat{\Sigma^e}) = C + \ln|I + A\hat{B}| - \frac{T}{2}\ln|(I + A\hat{B})\hat{e}_t\hat{e}_t'(I + A\hat{B})'| \tag{3.20}$$

Finally the estimate of synchronised variance-covariance matrix is obtained by

$$\hat{\Sigma} = (I_J + \hat{A}\hat{B})\hat{\Sigma}^e(I_J + \hat{A}\hat{B})' \tag{3.21}$$

where $\hat{B}$ and $\hat{\Sigma^e}$ are the maximum-likelihood estimates from model 1 and $\hat{A}$ is the maximum-likelihood estimate from model 2 given the consistent estimators of B and Σ^e from model 1. The time-varying synchronised variance Σ_t can also be obtained by further assuming heteroskecdastic structure for asynchronous variance-covariance matrix Σ_t^e.

3.4 Empirical analysis

This section applied the proposed synchronisation method on seven international stock markets from the eastern to western time zones including Japan, Australia, Hong Kong, Germany, the United Kingdom, Canada and the United States. The use of their daily close-to-close returns is obviously suffering from the problems of asynchronicity. This application assumed a dynamic conditional correlation (DCC) structure on the asynchronous variance-covariance matrix to allow time-varying correlations.

3.4.1 Data description

The data consists of seven stock market closing price series collected at the local closing time of each market, including Nikkei Stock Average (NIKKEI225) of Japan, Australian Stock Exchange (ASX) of Australia, Heng Seng Index (HSI) of Hong Kong, German stock index (DAX) of Germany, Financial Times Stock Exchange 100 index (FTSE 100) of the United Kingdom, Toronto Stock Exchange (TSE) of Canada and Standard & Poor 500 index (S&P500) of the United States. The data is extracted from Bloomberg Database for the period 1 January 2005 - 30 September 2015. After dropping the observations with holidays/non-trading days in the time series data panel (i.e. the whole observation for each market j at day t is dropped if day t is the holiday for at least one of the investigated stock markets), there are 2360 common trading days for each stock price series.

The data set is divided into two sub-samples, the first sample is ranging from 1 January 2005 to 31 December 2014, with 2190 trading days in total, for the purpose of model estimation. The second sample is ranging 1 January - 30 September 2015, with 170 trading days, for the purpose of out-of-sample forecast. The vector of asynchronous daily close-to-close returns r_t for the seven market indices ($J = 7$) are calculated by taking the natural logarithmic difference of the closing prices p_t for each market j at the local closing time t_j in a time order of the earliest close to the latest close as

$$r_{t_j} = \{r_{t_1}^{\text{ASX}}, r_{t_2}^{\text{NIK}}, r_{t_3}^{\text{HSI}}, r_{t_4}^{\text{DAX}}, r_{t_5}^{\text{FTSE}}, r_{t_6}^{\text{TSE}}, r_{t_7}^{\text{S\&P}}\}' = \begin{pmatrix} \Delta P_{t_1}^{\text{ASX}} \\ \vdots \\ \Delta P_{t_7}^{\text{S\&P}} \end{pmatrix} = \Delta P_{t_j}$$

These markets are located at different time zones from eastern to the western hemisphere globe, some of them are trading contemporaneously (e.g. DAX and FTSE 100, TSE and S&P500), some of them are trading partially overlap (e.g. NIKKEI225 and HSI), and some of them are completely out of phase. The opening and closing times at local time and at US time, trading time differences correspond to the US closing time, and the trading overlap corresponds to the US market are presented for each data series in Table 3.1. The first three indices, NIKKEI225, ASX and HSI, are the Pacific markets located at the eastern hemisphere with no trading overlap with the North American markets TSE and S&P500; and there are also a few trading differences among themselves. The second two indices, DAX and FTSE 100 are the European markets located in Europe which are trading partially overlap with the North American markets but have no trading overlap with the Pacific stock markets. The last two indices, TSE and S&P500 are the North American markets located at the western hemisphere which are the latest close stock markets every day. The closing time of the TSE and S&P500 is considered as the synchronised time relative to this application.

Table 3.1 about here (see P.75).

The daily contemporaneous correlations and the lagged correlations of the seven international stock markets using asynchronous close-to-close returns are presented in Table 3.2. In panel A of Table 3.2, the contemporaneous correlations are the highest with markets at the same region, for instance, 0.69 between ASX and NIKKEI225, 0.87 between DAX and FTSE 100, 0.79 between TSE and S&P500. The correlations are small when the markets are highly asynchronous, such as 0.21 between NIKKEI225 and S&P500, 0.38 between HSI and TSE. Although there is no reason to believe that all the contemporaneous correlations should be high, the significant lagged correlations may indicate the effect of asynchronicity. Panel B of Table 3.2 presented the lagged-1 correlations among the seven stock returns. The Pacific stock markets have minimal lag-1 cross correlations within the same Pacific region, but the lag-1 correlations are getting larger when the trading asynchronicity is getting higher for the markets close later, the lag-1 correlation between ASX and lag-1 S&P500 0.52 is the largest among all the lag-1 correlations. The lagged correlations between the European stock markets and the North American markets also have noticeable values between 0.07 to 0.22. Regarding the latest close stock markets in North American region, all the lagged correlations with the earlier close markets are negligible, while the TS S&P500 has noticeable negative lag-1 correlations with TSE -0.15 and with itself -0.11. These results implied that the close-to-close returns of the latest markets predict the next day's returns of the earlier markets as the stock prices of the latest markets involve much more information. However, this predictability is spurious because the true contemporaneous correlations are diluted by the lagged correlations with the presence of trading asynchronicity.

Table 3.2 about here (see P.75).

3.4.2 Estimating the synchronised model

The (7×1) vector of asynchronous close-to-close returns of the seven stock markets is modelled by the proposed synchronisation method. Additionally, the time-varying dynamics of the conditional variance-covariance is captured through a Multivariate Dynamic Conditional Correlation (DCC) model (Engle, 2002). The synchronised VAR-DCC is specified as follows.

$$\begin{aligned} r_{t_j} &= c_0 + B r_{t_j-1} + e_t \\ e_t &= (I + AB)^{-1} \epsilon_t \end{aligned} \tag{3.22}$$

where

$$\begin{aligned} & e_t | \mathcal{F}_{t-1} \sim N(0, h_t) \\ & h_t = D_t \tilde{R}_t D_t \\ & D_t = diag(h_t^{\frac{1}{2}}) \\ & h_{i,i,t} = \omega_i + \alpha_i' e_{t-1} e_{t-1}' + \beta_i H_{i,i,t-1} \qquad \text{for } \; i = 1, 2, ..., 7 \\ & \tilde{R}_t = diag(Q_t^{-\frac{1}{2}}) Q_t diag(Q_t^{-\frac{1}{2}}) \\ & Q_t = \Omega + \alpha e_{t-1} e_{t-1}' + \beta Q_{t-1} \end{aligned}$$

This study considers the DCC(1,1) model. The conditional covariance matrix in the DCC model is decomposed into a relation between the estimated univariate GARCH variances D_t and the conditional correlation matrix $\tilde{R}_t$. h_t and $\tilde{R}_t$ are positive definite, D_t is a diagonal matrix with the elements of the estimated univariate GARCH variances,

Q_t is the quasi-correlation matrix which is rescaled by $\rho_{i,j,t} = Q_{i,i,t}^{-\frac{1}{2}} Q_{i,j,t} Q_{j,j,t}^{-\frac{1}{2}}$ to ensure the correlation estimate is $[0,1]$ bounded. The parameters (α_i, β_i) in the GARCH process are positive $\forall i$ and has a sum less than unity to ensure the stationary condition. The maximum-likelihood estimators $\theta \equiv (\omega_i, \alpha_i, ..., \Omega, \alpha, \beta)$ of h_t is estimated by maximising the log-likelihood function (since the asynchronous error e_t are multivariate normal):

$$\ln\mathcal{L}(r_{t_j}, \theta) = -\frac{TJ}{2}\ln(2\pi) - \frac{T}{2}\ln|h_t| - \frac{1}{2}\sum_{t=1}^{T}[e_t' h_t^{-1} e_t] \tag{3.23}$$

The synchronised conditional covariance matrix is then given by

$$E_{t-1}[\epsilon_t \epsilon_t'] = H_t = (I + AB)h_t(I + AB)'. \tag{3.24}$$

The synchronised conditional covariance H_t is positive definite since the asynchronous conditional covariance h_t is positive definite.

The two-step maximum likelihood estimation produces the maximum likelihood estimators B and h_t in the first step and the estimator A in the second step. To reduce the number of parameters to be estimated in the second step, the matrix elements of $\hat{B}$ are set to zero if they are found to be statistically insignificant. The estimation results are reported in Table 3.3. The non-zero diagonal and lower-diagonal elements resulted in matrix B implied that the assumption of random walk stock prices may be invalid.

Table 3.3 about here (see P.76).

3.4.3 Synchronised correlations

The (7×1) vector of estimated synchronised returns is given by

$$\hat{R}_t = \{\hat{R}_t^{\text{ASX}}, \hat{R}_t^{\text{NIK}}, \hat{R}_t^{\text{HSI}}, \hat{R}_t^{\text{DAX}}, \hat{R}_t^{\text{FTSE}}, \hat{R}_t^{\text{TSE}}, \hat{R}_t^{\text{S\&P}}\}' = (I + \hat{A}\hat{B})r_{t_j} - \hat{A}\hat{B}r_{t_j-1}$$

The unconditional correlations of synchronised returns can be directly calculated, they are presented in Table 3.4. As expected, the contemporaneous correlations among the seven stock indices after the synchronisation adjustment are higher than the correlations from asynchronous data. The largest increase are the correlations between the Pacific stock markets and the North American markets from 0.2 to 0.7. The correlations between European markets and the North American markets also raise from 0.6 to 0.8. The correlations among the stock indices in the same regions, such as DAX and FTSE 100, TSE and S&P500 have no noticeable changes because they are relatively synchronised.

Table 3.4 about here (see P.77).

3.4.4 Diagnostic tests

The diagnostic tests are performed for the the standardised residuals $z_t = h^{-\frac{1}{2}}e_t$. If the asynchronous VAR model in (A3.2) is correctly specified, the innovations e_t are serially uncorrelated with zero mean and a conditional variance-covariance matrix h_t. Then the covariance matrix of the standardised residuals should be approximately an identity matrix.

The first test is a χ^2 test of whether the estimated covariance matrix of the standardised residuals is an identity matrix. The degrees of freedom is 28 as there are $\frac{1}{2}(7\times6)+7 = 28$ elements to be estimated in the covariance matrix. The χ^2 test statistic is 40.69 with $d.f. = 28$ and a p-value of 0.0573. Therefore, the estimated covariance matrix of the standardised residuals is approximately an identity matrix at 5% level of significance.

The second test is a Lagrange multiplier χ^2 test of the serial correlations of the estimated standardised residuals. The test statistics are computed using lags up to 30, the asymptotc 5% critical value is 66.34 with 49 (7×7) degrees of freedom. All of the χ^2 test statistics for lags up to 30 are exceed the critical value. The serially correlated standardised residuals indicated that there exists some dependences are not captured by the asynchronous VAR thus a richer autoregressive model may be needed.

3.5 An application to Value-at-Risk measurement under asynchronicity

This section performs an out-of-sample analysis to compare the forecasting performance of the synchronised returns derived from the proposed synchronisation method with the asynchronous returns. The economic significance of their differences are illustrated by a Value-at-Risk application. The forecasting exercise uses the second part of the data set from 1 January to 30 September 2015. The synchronised VAR-DCC(1,1) model is not re-estimated, but the one-step-ahead conditional measures are updated with the new observations available.

Value-at-Risk measure is a popular financial risk management tool defined as the maximum estimated loss that is expected to occur given a specified probaility in the market value of a portfolio. That is, mathematically, the prob($\Delta Z_{\Delta t} \leq -\text{VaR}) = \alpha\%$, where ΔZ_t is the change of the market value of the portfolio P within the period of Δt, $\alpha\%$ is the probability of loss.

Consider a portfolio of seven equally weighted investments in the seven stock assets $Z = \{\frac{1}{7}(Z^{\text{ASX}} + Z^{\text{NIK}} + Z^{\text{HSI}} + Z^{\text{DAX}} + Z^{\text{FTSE}} + Z^{\text{TSE}} + Z^{\text{S\&P}})\}'$. The DCC structure of variance-covariance constructed a VaR measure as:

$$VaR_{t+1} = -k \times \sqrt{w_1^2 H_{11,t} + w_2^2 H_{22,t} + ... + 2w_1 w_2 \rho_{12,t} H_{11,t}^{1/2} H_{22,t}^{1/2} + ...$$

where k is the critical value corresponds to the confidence interval of the targeted $\alpha\%$. w_i for $i = 1, 2, ..., 7$ is the weights of the total investments assigned to each asset, $\rho_{ij,t}$ for $i, j = 1, 2, ..., 7$ is the time-varying dynamic correlation coefficient between asset i and j. This exercise assumed the VaR follows the normal distribution with mean zero and 99% confidence interval. The time period $\Delta t = 1$ day. The critical value $-k$ is approximately -2.25 with $\alpha\% = 1\%$ on the left tail.

The performance of Value-at-Risk using different different stock returns series are investigated. The performance is evaluated by a back-testing of an investment portfolio, which is the frequency count of the number of the loss in a day exceeded the VaR measure of 1% provision. This exercise is inspired in a similar study of asynchornous daily correlations by Martens and Poon (2001). To compare the forecasting performance of asynchronous and synchronised returns, the one-day-ahead forecast of the portfolio returns and the VaR measures are calculated and plotted in Figure 3.4. The incidents of violations (the larger dots) when the value of market portfolio losses more than the VaR provision (the black lines). With 170 one-step ahead forecasts, asynchronous returns have 10 (5.9%) violations to the VaR provision; while estimated synchronised returns have only 3 (1.8%)violations to the VaR provision. In addition, the violations in the synchronised VaR measures lie about on the VaR provision line. The performance of VaR using asynchronous data for measuring portfolio risk is more volatile; however, synchronised returns estimated from the proposed model provide a more conservative measure of portfolio risk.

Figure 3.4 about here (see P.74).

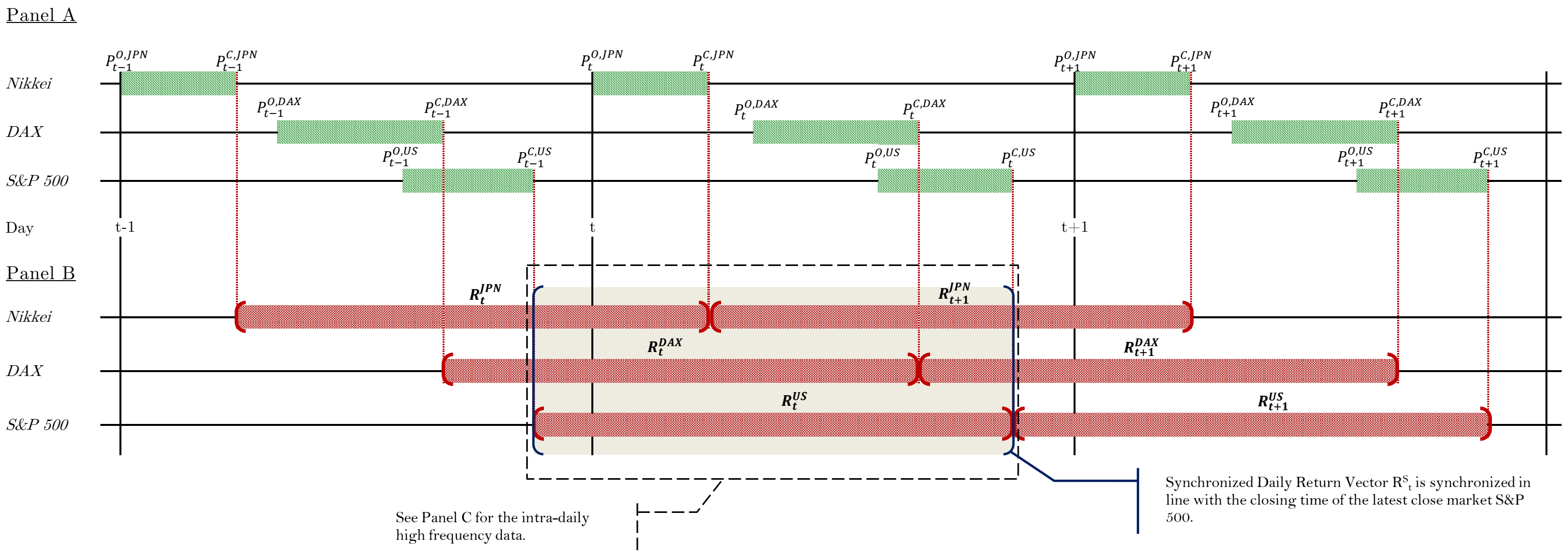
The empirical result agreed with our expectation that the portfolio value calculated from asynchronous returns is more volatile by recognising unnecessary profit and loss overnight. This can be explained by an illustrative example. Imagine there are two perfectly hedged assets in a portfolio, one is trading in the UK (short position) and another one is trading in the US (long position). After the UK market closed, the US asset price dropped by 1%. This perfectly hedged portfolio should not report loss because the value of the UK asset is expected to offset this 1% decrease. However, the UK asset price is stale after the UK market close, any adjustment is reported until the next open of UK market. Then the portfolio recognised this unnecessary loss overnight until the short position responds. Therefore, the VaR forecasting performance of the proposed synchronised stock returns is superior over the use of asynchronous returns.

3.6 Conclusion

Asynchronous data are resulted by the fact that information flows continuously across markets, yet the asset prices are only recorded within their local trading time. The need of synchronising multivariate daily stock returns is strongly motivated by the distorted correlation estimates of asynchronous data. In particular, the daily correlation dynamics of the multivariate financial time series are significantly understated. The underestimation of correlations leads to the distortion of portfolio values, the volatile VaR measures and many other financial applications.

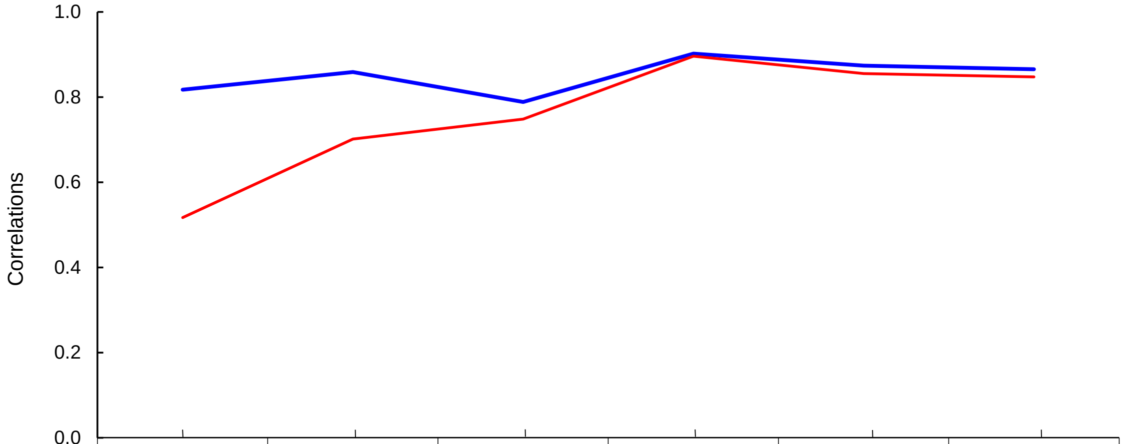
This study proposed a synchronisation model to estimate synchronised returns and their contemporaneous correlations from the observed asynchronous returns. The fundamental set up of this synchronisation algorithm is inspired by the benchmark paper Burns, Engle, and Mezrich (1998) which is the first paper discussed about the importance of time zone differences among the international time series. However, their paper assumed random walk stock price series to construct synchronised returns and hence the synchronised model. Under the random walk hypothesis on stock prices, any future price change is unpredictable; in other words, any predictability from asynchronous returns are due to trading asynchronicity. The empirical results did not strongly support the random walk assumption in their model, and once the random walk assumption is rejected, the entire formulation of synchronised returns is invalid. Therefore, this study relaxed the random walk assumption on stock prices and constructed a more generalised synchronisation model by claiming a less restrictive assumption, that is, the unrecorded returns for the earlier markets make up a fraction of the next day's asynchronous returns. This proposed model is a generalised class of Burns, et al.'s (1998) (when $A = I_J$) and asynchronous VAR (when $A = 0$).

The empirical results show that asynchronous correlations are too low as some of the contemporaneous correlations are spread to the lagged correlations. After synchronising the stock returns, the correlations are brought back to the same synchronised time point. Regarding the diagnostic tests associated with the standardised residuals, the unconditional covariance matrix of the standardised residuals is approximately an identity matrix; yet the LM tests of serial correlated residuals indicated that a rich autoregressive model may be needed to capture the serial dependence. The VaR Back-testing analysis is supporting evidence that the proposed synchronised VAR model leads to better risk measure than those from asynchronous model.



Panel A: The open-to-close daily stock prices for Nikkei, DAX and S&P 500 stock indices.
Panel B: The close-to-close asynchronous returns and synchronizaed returns for Mikkei, DAX and S&P 500 stock indices.

Figure 3.1: The World of Asynchronous Trading



Return Interval (No. of Trading Days)	1-Day	2-Day	3-Day	5-Day	10-Day	20-Day
Synchronous returns at 16:30 GMT	0.817	0.859	0.789	0.902	0.874	0.865
Asynchronous close-to-close returns	0.517	0.701	0.749	0.896	0.855	0.848

Figure 3.2: Correlations between FTSE 100 and S&P 500 using synchronous vs. asynchronous data evolve from 1-day to 20-day return intervals (sample period: 2 Jun - 21 Dec 2014)

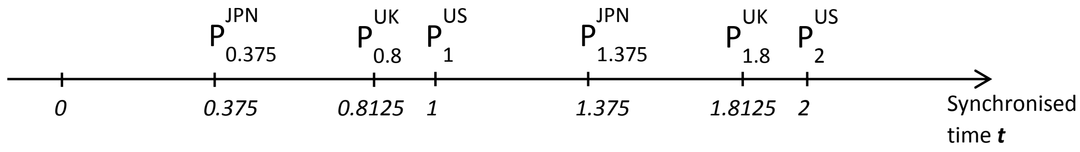


Figure 3.3: The closing prices of Japanese, the UK and the US stock markets corresponds to synchronised time t.

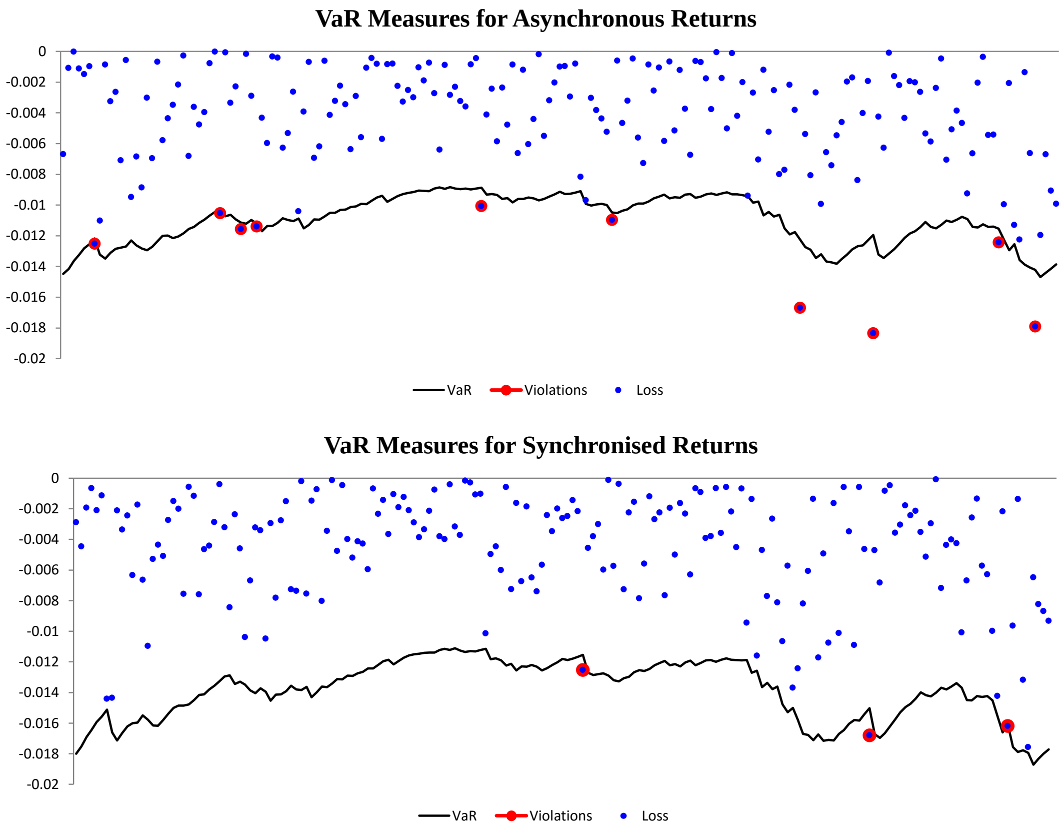


Figure 3.4: The 1% one-day VaR forecast using asynchronous returns and synchronised returns for the period 1 January - 30 September 2015.

Tables of Chapter 3

Index	Local Time Opening	Local Time Closing	US Time Opening	US Time Closing	Overlap with US?	Close time diff. with US (hour)
ASX	10:00	16:00	18:00($-1t$)	00:00	no	16
NIKKEI225	09:00	15:00	19:00($-1t$)	01:00	no	15
HSI	09:30	16:00	20:30($-1t$)	03:00	no	13
DAX	09:00	17:30	03:00	11:30	partially	4.5
FTSE 100	08:00	16:30	03:00	11:30	partially	4.5
TSE	09:30	16:00	09:30	16:00	overlap	0
S&P500	09:30	16:00	09:30	16:00	overlap	0

Table 3.1: Opening and closing times, overlapping and closing time differneces corresponds to the US markets for seven international stock indices.

Panel A: Contemporaneous Correlations

Index	ASX	NIKKEI225	HSI	DAX	FTSE 100	TSE	S&P500
ASX	1.000	0.691	0.676	0.414	0.462	0.315	0.218
NIK 225	0.691	1.000	0.650	0.397	0.419	0.298	0.214
HSI	0.676	0.650	1.000	0.441	0.472	0.382	0.300
DAX	0.414	0.397	0.441	1.000	0.868	0.599	0.664
FTSE 100	0.462	0.419	0.472	0.868	1.000	0.638	0.636
TSE	0.315	0.298	0.382	0.599	0.638	1.000	0.785
S&P500	0.218	0.214	0.300	0.664	0.636	0.785	1.000

Panel B: Lagged-1 Correlations

Index	ASX	NIKKEI225	HSI	DAX	FTSE 100	TSE	S&P500
L1.ASX	−0.041	−0.014	−0.032	−0.030	−0.056	−0.044	−0.065
L1.NIK 225	−0.047	−0.048	−0.052	−0.050	−0.061	−0.006	−0.038
L1.HSI	0.021	0.036	−0.058	−0.014	0.004	−0.021	−0.059
L1.DAX	0.332	0.342	0.239	−0.010	−0.027	0.019	−0.069
L1.FTSE 100	0.336	0.317	0.243	−0.030	−0.052	−0.004	−0.079
L1.TSE	0.428	0.338	0.278	0.069	0.121	−0.055	−0.153
L1.S&P500	0.522	0.461	0.375	0.185	0.223	0.039	−0.110

Table 3.2: Daily contemporaneous correlations (Panel A) and lagged correlations (Panel B) of asynchronous close-to-close stock returns for seven markets for the period 1 January 2005 - 31 December 2014.

Step 1 maximum likelihood

Parameter	ASX	NIKKEI225	HSI	DAX	FTSE 100	TSE	S&P500
$B_{ASX,i}$	−0.097	−0.047	−0.067	0	0	0.094	0.348
$B_{NIK,i}$	−0.095	−0.082	0	0.187	0	0	0.374
$B_{HSI,i}$	0	−0.079	−0.101	0	0	0	0.406
$B_{DAX,i}$	0	0	0	0	−0.189	0	0.355
$B_{FTSE,i}$	0.058	0	0	−0.094	−0.172	0	0.321
$B_{TSE,i}$	0	0	0	0	0	−0.106	0.077
$B_{S\&P,i}$	0	0	0	0	0	−0.084	−0.071
$c_{0,i}$	0.001	0.001	0.001	0.001	0.001	0.001	0.001
ω_i	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_i	0.065	0.112	0.065	0.073	0.075	0.064	0.090
β_i	0.916	0.859	0.918	0.901	0.898	0.918	0.880

Step 2 maximum likelihood

Parameter	ASX	NIKKEI225	HSI	DAX	FTSE 100	TSE	S&P500
$A_{ASX,i}$	0.304	0.272	0.574	0.115	−0.256	−0.024	−0.199
$A_{NIK,i}$	0.284	0.308	0.276	0.703	0.426	−0.019	−0.281
$A_{HSI,i}$	0.276	0.654	0.305	0.648	0.355	−0.019	−0.240
$A_{DAX,i}$	−0.498	−0.267	−0.417	0.186	0.262	0.014	0.132
$A_{FTSE,i}$	0.989	0.081	1.000	0.272	0.305	−0.003	−0.412
$A_{TSE,i}$	−0.269	−0.338	−0.346	−0.326	−0.310	0.297	0.264
$A_{S\&P,i}$	−0.298	−0.295	−0.294	−0.294	−0.290	−0.260	0.325

Note: All the non-zero parameter estimates are statistically significant at 1% level of significance.

Table 3.3: The estimation results of synchronised VAR(1)-DCC(1,1) from the 2-step maximum likelihood proceduce.

Index	ASX	NIKKEI225	HSI	DAX	FTSE 100	TSE	S&P500
ASX	1.000	0.650	0.650	0.629	0.655	0.590	0.619
NIKKEI225	0.650	1.000	0.681	0.732	0.740	0.641	0.724
HSI	0.650	0.681	1.000	0.599	0.634	0.579	0.608
DAX	0.629	0.732	0.599	1.000	0.843	0.662	0.820
FTSE 100	0.655	0.740	0.634	0.843	1.000	0.755	0.827
TSE	0.590	0.641	0.579	0.662	0.755	1.000	0.786
S&P500	0.619	0.724	0.608	0.820	0.827	0.786	1.000

Table 3.4: Daily contemporaneous correlations of estimated synchronised stock returns for seven markets for the period 3 January 2005 - 31 December 2014.

Chapter 4

Vector autoregressive models with measurement errors for asynchronous data and a spatially synchronised correlation

4.1 Introduction

Multivariate time series modelling has crucial implications for the quantitative assessments of the variables of interests in many financial applications. Specifically, the vector autoregressive model is widely used by its attractivenesses of estimation simplicity, and of the identifications of Granger causal relationships (Granger, 1969) between financial assets. The analysis of Granger causality on stock market returns can help identifying information flow between markets and explaining the hypothesis of informational market efficiency. Nevertheless, it is important to highlight that the conventional VAR model would not identify correct relationships between the true variables if the data used is subject to error. When the measurement errors are ignored in the estimation process and the observed variables are assumed to be the true variable of interest, the maximum-likelihood estimators are inconsistent. Some special inferences may carry out for model parameters in order to correct for the inconsistency. The measurement equations should be added to the model to capture the measurement error effect, then the maximum-likelihood estimators are consistent by having known information about measurement error.

As investigated in the previous chapter, the daily stock returns are collected at a discrete basis at the stock markets' local closing time. Asynchronicity is not an issue for the univariate asset return modelling as the time of close is absolute for each market alone;

however, the point of time when each data series is recorded is a material issue for multivariate return modelling because a point of close for one market may be relative to a point of trading for another. Information continuously flows across markets but markets respond only within their trading hours. As a result, for some earlier closed markets, a piece of information from closing to a time of analysis is missed and does not reflect on the recorded stock returns. If the synchronous returns are the true variable of interest in the multivariate time series analysis, the observed asynchronous returns are measured with errors because there are some information missing. Intuitively, the larger the degree of asynchronicity, the more information between true returns and asynchronous returns is missing, as a result, the bigger the measurement error appears.

The main focus of this research is to analyse the inconsistent estimators from the usual VAR models using observed close-to-close stock returns by the reason that the observed stock returns are subject to measurement bias attributed to trading asynchronicity. Next, the maximum-likelihood estimators are proved to be asymptotically consistent if a consistent proxy is used to replace the measurement error. Since the amount of the underestimation on the contemporaneous correlations are subject to the degree of trading asynchronicity. This chapter proposes a predetermined spatial structure to the unrecorded returns, such that the asynchronous returns are adjusted according to the time zone differences between the earlier close markets and the synchronous market. Then the multivariate stock return analysis can still be implemented by a spatio-temporal vector autoregressive models with an alternative inference.

This is the first study considers measurement error problem on asynchronous data. According to the benchmark paper Burns et al. (1998) who modelled the asynchronous returns by a first-order vector moving average process, the synchronised returns are computed by shifting the lagged dependence to the current time, in which the lagged dependence is captured by the first-order moving average matrix. The validation of their synchronisation model is based on the model specification of the asynchronous returns, that is, the error terms from the VMA(1) model are required to be serially uncorrelated. In case the measurement error exhibits on the asynchronous returns, their model is no longer correctly specified. The empirical results and diagnostic tests in their study also suspected that the error terms may not be serially uncorrelated. Patriota et al. (2010) discussed about the observed data with measurement errors lead to inconsistent estimate from the conventional vector autoregressive models, if the measurement errors are not negligible. The alternative inference is to recognise the possible measurement errors and model the true variable of interest instead of the observed variable. Then the estimators are asymptotically consistent by taking the measurement error equation into consideration.

Another scope of literature related to this study is the spatial analysis applied in financial time series. Spatial analysis deals with the measurements of a particular phenomenon associated with specific locations or regions. Generally, the spatial correlations detect whether a event happened in one place affects the same measure of another place. Spatial dependence has been studied in many research areas including economic geography, environmental sciences, and urban economics, etc. However, there are relatively fewer studies consider the spatial dependence in the field of finance. Even there are a few research (Asgharian et al., 2013; Durante and Foscolo, 2013; Fernandez, 2011) focus on the spatial linkages in the international financial markets, the spatial dependencies are usually referred the geographic closeness between firms or markets, bilateral trade, or the size of markets measured by market capitalisation, book-to market-ratio, and many other financial indicators. This study introduces a new spatial dependence among the international financial markets, which is motivated by the absolute time zone differences between the stock markets located in different time regions.

The remainder of this chapter is organised as follows. Section 4.2 discusses the conventional VAR model. Section 4.3 discusses the inconsistent estimators resulted from the conventional VAR using asynchronous data. Section 4.4 proposes a spatial proxy of the measurement error between true returns and asynchronous returns. Section 4.5 conducts an empirical analysis on seven international stock markets at different time regions. Section 4.6 concludes the study.

4.2 The conventional VAR model

Denote the latent variable as a $(J \times 1)$ vector of synchronous return for J markets at the point of synchronising time t by

$$R_t = \Delta P_t^s = \begin{pmatrix} \Delta P_t^{s,1} \\ \vdots \\ \Delta P_t^{s,J} \end{pmatrix}, \quad t = 1, 2, ..., T \tag{4.1}$$

where $P_t^{s,j} = \ln(p_t^{s,j})$ is the log of synchronous prices at time t for market j.

In the following, we define the usual VAR(1) model for the latent variable as

$$R_t = c + BR_{t-1} + \epsilon_t, \quad t = 1, ..., T \tag{4.2}$$

where T is the sample size, B is the $J \times J$ first order coefficient matrix and ϵ_t is a $j \times 1$ unobservable zero mean innovation vector process with covariance matrix Σ. To simplify the presentation of formulae along this research, the constant c is set to be zero. The empirical section allows for a non-zero constant term.

Assuming the synchronous return vector R_t follows a multivariate normal distribution with mean and autocovariance functions given, respectively, by

$$E[R_t] = \mu = (I_J - B)^{-1}c = 0,$$

$$\Gamma(0) = E[(R_t - \mu)(R_t - \mu)'] = \Sigma + B\Sigma B' + ... = \Sigma + B\Gamma(1)', \quad \text{for } h = 0$$

$$\Gamma(h) = E[(R_t - \mu)(R_{t-h} - \mu)'] = B\Gamma(h-1), \quad \text{for } h = 1, 2, 3, ...$$

where I_J denotes the $J \times J$ identity matrix.

Assume that the innovation process ϵ_t is multivariate normally distributed, the log-likelihood function in matrix form is given by

$$\begin{aligned} \ln\mathcal{L}(B, \Sigma) = & -\frac{TJ}{2}\ln(2\pi) - \frac{T}{2}\ln\,\det(\Sigma) \\ & -\frac{1}{2}\sum_{t=1}^{T}\big((R_t - BR_{t-1})'\Sigma^{-1}(R_t - BR_{t-1})\big) \end{aligned} \tag{4.3}$$

From the first order partial differentiation of the log-likelihood function with respect to B and Σ, we obtain the system of normal equations (see Appendix C for the subsequent proofs) for deriving the consistent maximum-likelihood estimators, which are given by

$$\hat{B}_{ML} = \{Q_{R_{t-1}}^{-1} Q_{R_{t-1}R_t}\}', \quad \text{and} \tag{4.4}$$

$$\hat{\Sigma}_{ML} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t', \tag{4.5}$$

where $Q_{R_{t-1}} = T^{-1}\sum_{t=1}^{T} R_{t-1}R_{t-1}'$, $Q_{R_{t-1}R_t} = T^{-1}\sum_{t=1}^{T} R_{t-1}R_t'$, and $\hat{\epsilon}_t = R_t - \hat{B}_{ML}R_{t-1}$.

Lemma 4.1. *Under the stationary conditions, the consistency of those maximum-likelihood estimators is assured as*

$$Q_{R_{t-1}} \xrightarrow{p} \Gamma(0), \quad Q_{R_{t-1}R_t} \xrightarrow{p} \Gamma(0)B',$$

and hence

$$\hat{B}_{ML} \xrightarrow{p} B.$$

where $\xrightarrow{p}$ *denotes convergence in probability asymptotically.*

Lemma 4.1 is a classical result from the literature on measurement error problem.

4.3 VAR with measurement error

Normally, we may not be able to observe daily stock returns synchronously at each international stock markets. In fact, people usually observe the asynchronous close-to-close stock returns r_t with error v_t such that

$$R_t = r_t + v_t, \tag{4.6}$$

where $r_t = \Delta \ln(p_t)$ is a $(J \times 1)$ log of close-to-close return vector at time t and $v_t = (v_{1t}, v_{2t}, ..., v_{kt})'$ is the $J \times 1$ vector of white noise measurement errors with mean zero and variance matrix Σ^v.

By substituting the measurement error equation (4.6) into the true VAR model at (4.2), we obtain

$$r_t = Br_{t-1} + e_t, \tag{4.7}$$

denote $e_t = Bv_{t-1} - v_t + \epsilon_t$.

People usually replace the synchronous return R_t with the observed asynchronous return r_t and estimate equation (4.7), without taking the measurement equation into account. The maximum likelihood estimators from this model are inconsistent since the error term has some elements which are correlated with the regressors. That is, $E[e_t e_s'] \neq 0$ $\forall\, t \neq s$. Denote the inconsistent estimator of B and Σ as b and Σ^e respectively.

Lemma 4.2. *Given (4.2), (4.6) and Lemma 4.1, the inconsistent estimators are as*

$$\hat{b} = \{Q_{r_{t-1}}^{-1} Q_{r_{t-1}r_t}\}' \qquad \text{and} \qquad \hat{\Sigma}^e = T^{-1} \sum_{t=1}^{T} \hat{e}_t \hat{e}_t',$$

where $Q_{r_{t-1}} = T^{-1} \sum_{t=1}^{T} r_{t-1} r_{t-1}'$, $Q_{r_{t-1}r_t} = T^{-1} \sum_{t=1}^{T} r_{t-1} r_t'$, $\hat{e}_t = r_t - \hat{b} r_{t-1}$, *and hence,*

$$\hat{b} \xrightarrow{p} \left(\Gamma(0) + \Sigma^v\right)^{-1} \Gamma(0) B'.$$

Again, lemma 4.2 is a classical result from measurement error problem.

Therefore, the measurement equation (4.6) should be included in the estimation procedure. The complete model (assumed a zero constant for simplicity) for estimation should be as follows

$$\begin{aligned} R_t &= BR_{t-1} + \epsilon_t \\ R_t &= r_t + v_t. \end{aligned} \tag{4.8}$$

Writing the system of equations in a single equation gives

$$r_t = Br_{t-1} - v_t + Bv_{t-1} + \epsilon_t. \tag{4.9}$$

People's interests on equation (4.9) could be the Granger causality of the (true) variable of interests which is measured by the coefficient matrix B; or could be the correlations and volatilities of true returns, which are measured by $Var[\epsilon_t]$. The coefficient B is a true parameter from the true model at (4.2), the parameter estimate of B is not identified if we do not observe R_t. The extreme case is that when the true parameter $B = 0$, the variance-covariance matrices of v_t and ϵ_t are confounded since $r_t = -v_t + \epsilon_t$. It is not possible to estimate Σ^v and Σ separately by observing only asynchronous returns r_t (see Patriota et al., 2010). Thus, the straightforward approach is to assume we have some knowledge to approximate the measurement error v_t, although v_t is not directly observed.

By correcting the unobserved measurement errors to the model, the maximum likelihood estimators from equation (4.9) are consistent with good asymptotic properties such as normality. The consistent estimators and their asymptotic distributions are proposed in the following.

Under the stationary condition of the true VAR model, the measurement errors v_t and v_{t-1} are assumed to be observable. Then the log-likelihood function is given by

$$\begin{aligned}\ln\mathcal{L}(B,\Sigma) = &-\frac{TJ}{2}\ln(2\pi) - \frac{T}{2}\ln\,\det(\Sigma)\\ &-\frac{1}{2}\sum_{t=1}^{T}\Big((r_t + v_t - Br_{t-1} - Bv_{t-1})'\Sigma^{-1}(r_t + v_t - Br_{t-1} - Bv_{t-1})\Big). \qquad (4.10)\end{aligned}$$

The likelihood function above has no change compared to equation (4.3) but only the measurement equation is substituted into the latent variable R_t. We must remark that the ML estimators derived below is under the assumption that measurement error is serially uncorrelated, there are more complicated forms presented at Appendix A if measurement error is serially correlated.

Lemma 4.3. *The consistent estimators are yielded by maximising the log-likelihood function (4.10), as*

$$\hat{B}_{ML} = \{(Q_{r_{t-1}} - \Sigma^v)^{-1}Q_{r_{t-1}r_t}\}',$$

$$\hat{\Sigma}_{ML} = T^{-1}\sum_{t=1}^{T}\hat{e}_t\hat{e}_t' - \Sigma^v - \hat{B}\Sigma^v\hat{B}',$$

where $Q_{r_{t-1}} = T^{-1}\sum_{t=1}^{T} r_{t-1}r_{t-1}'$, $Q_{r_{t-1}r_t} = T^{-1}\sum_{t=1}^{T} r_{t-1}r_t'$, *and* $\hat{e}_t = r_t - \hat{B}r_{t-1}$.

Theorem 4.1. *The ML estimators in Lemma 4.3 assured consistency as*

$$(Q_{r_{t-1}} - \Sigma^v)^{-1} \xrightarrow{p} \Gamma(0)^{-1}, \quad Q_{r_{t-1}r_t} \xrightarrow{p} \Gamma(0)B'$$

then

$$\hat{B}_{ML} \xrightarrow{p} B \qquad \text{and} \qquad \hat{\Sigma}_{ML} \xrightarrow{p} \Sigma$$

Theorem 4.2. *Subsequent to Theorem 4.1, the asymptotic distribution of the consistent ML estimator $\hat{B}$ is as*

$$\sqrt{T}(\hat{B} - B) \xrightarrow{D} \mathcal{N}(0, \Omega), \tag{4.11}$$

where $\Omega \equiv \Sigma^e\Gamma(0)^{-1} + \Gamma(0)^{-1}(\Sigma^e\Sigma^v - \Sigma^v B' B \Sigma^v)\Gamma(0)^{-1}$.

Comparing the estimators in Lemma 4.3 to the inconsistent estimators Lemma 4.2, the inconsistent estimators are biased by missing the measurement error variance. It can be seen that if $\Sigma^v = 0$, that is, when there is no measurement error, estimators in Lemma 4.2 reduced to the consistent estimators as in Lemma 4.3.

We can see the ML estimator of B by asynchronous VAR model $\hat{b}$ is not consistent, since it converges in probability to

$$\hat{b} \xrightarrow{p} B\big(I_J + \Sigma^v\Gamma(0)^{-1}\big)^{-1}. \tag{4.12}$$

The asymptotic variance of the inconsistent estimator $\hat{b}$ equals only the first term $\Sigma^e\Gamma(0)^{-1}$. If the variance bias correction is positive (given by $\Sigma^v \leq \Sigma^e$ and $|B| < 1$), The asymptotic variance of the inconsistent estimator is understated by the amount of the variance bias correction. Consequently, the statistical inference of the lagged return coefficients is misleading under the asynchronous VAR models. With a smaller asymptotic variance of the parameter estimate, the null hypothesis of parameter restrictions can be false rejected hence affecting the analysis of Granger causality.

From the above illustration, we see that if the measurement error (or at least its variance) is known or can be consistently estimated, i.e. $\hat{\Sigma^v} \xrightarrow{p} \Sigma^v$, then the asymptotic result derived in the above remains valid.

4.4 The spatio-temporal VAR model

To proximate the measurement error in asynchronous returns, we may think of the missing part from asynchronous returns compared to synchronous returns is larger for the markets with higher degree of time asynchronicity. Hence a spatial lag term is introduced to the observed asynchronous return vector, which is weighted by the closing time difference between market k to the synchronous market. In most related literature, the spatial weighted matrix is commonly formulated as a symmetric weighted matrix captures the symmetric spatial differences between the individuals, it is intuitive because the spatial differences refer to “physical” distances. Unlike the formulation of spatial matrix in most literature, the spatial weighted matrix stores the trading time differences $\Delta t_{i,j}$ within a day between earlier markets i and later markets j in which the weights refer to positive vector direction of trading time differences $\vec{\Delta t}$. Suppose the elements in the return vector are in an order of closing time differences from earliest to latest, the spatial weighted matrix has a form that 1) the upper diagonal elements are positive with the closing time differences between i and j for $i < j$ fractioned by 1-day; 2) the diagonal elements are zero since there is no trading time difference among the market itself; 3) the lower diagonal elements are also set to be zero because only the later close markets provide additional information for the earlier close markets to estimate their stock market values. The lower diagonal matrix for the spatial weights is possible when observed asynchronous returns are synchronised to the earliest market; although synchronising backward is not a practical interest for financial applications. For an illustrative example, the Japanese stock market closes fifteen hours ahead of the US market and nine hours ahead of the UK market, the spatial weights for JPN-US and JPN-UK is 15/24 and 9/24 respectively. For instance, a three-market stock returns $r_t = \{r_t^{jpn}, r_t^{uk}, r_t^{us}\}$ has a spatial weighted matrix as

$$A = \begin{bmatrix} 0 & \frac{9}{24} & \frac{15}{24} \\ 0 & 0 & \frac{4.5}{24} \\ 0 & 0 & 0 \end{bmatrix}$$

The motivation to impose the following assumption is that the asynchronicity has an obvious pattern of higher degree of time zone difference associated with larger measurement error. The spatial correlation matrix is set to assign larger weight for the market with higher time zone difference corresponds to larger correction of measurement error.

Assumption 4.1. *Assume that the $(J \times 1)$ vector of measurement error v_t^j for market i at time t is made up by a spatial lag on the vector of $(J \times 1)$ observed returns r_t^j, for $i, j = 1, 2, ..., J$ and $i < j$, where the matrix A is a $(J \times J)$ deterministic spatial weighted matrix, written as*

$$v_t = \rho A r_t. \tag{A4.1}$$

The spatial correlation parameter ρ is a scalar. Unlike other spatial models in the literature, the spatial weighted matrix in the above assumption is not row-standardised. It is because the lower-diagonal elements of the weighted matrix are set to be zero, row standardisation artificially inflates the values of non-zero elements at the same row.

By substituting (A4.1) into the measurement error equation, we obtain

$$R_t = r_t + \rho A r_t. \tag{4.13}$$

Synchronised returns at (4.13) are expressed as asynchronous returns with a spatial adjustment. Substituting (4.13) into the true VAR model, suppose the true model is a

VAR(1) process, we obtain a first-order spatio-temporal VAR as

$$\begin{aligned}(I+\rho A)r_t &= B(I+\rho A)r_{t-1}+\epsilon_t \\ \Leftrightarrow r_t &= Br_{t-1}-\rho Ar_t+B\rho Ar_{t-1}+\epsilon_t \end{aligned} \tag{4.14}$$

where the first term of (4.14) at the right-hand side is the temporal lag as in the conventional VAR, the second and the third terms are the spatial lags introduced to adjust the time zone asynchronicity.

An alternative proxy is presented in Appendix C. It is a temporal proxy instead of a spatial one, in which the unrecorded return is in a function of the current returns from other later markets. The use of such proxy leads to a transformed VAR(2) to be estimated; yet the estimation is complicated by using real data. Some simulation based methods can be conducted to obtain simulation results given the proposed maximum likelihood estimators. However, this study puts the main focus on the spatial-temporal VAR model.

4.4.1 The estimation

The spatio-temporal VAR(1) at (4.14) is the model to estimate. It can be seen as a temporal VAR(1) if pre-multiplying $(I-\rho A)=W$ on both sides, assuming $(I-\rho A)=W$ is invertible, we obtain

$$r_t=(I+\rho A)^{-1}B(I+\rho A)r_{t-1}+(I+\rho A)^{-1}\epsilon_t. \tag{4.15}$$

Recall that the asynchronous VAR(1) is given by $r_t=Br_{t-1}+e_t$ as in (4.7). Comparing the terms with (4.15), the disturbance terms ϵ_t in the true model can be expressed as the spatially weighted error terms from asynchronous VAR model, as

$$\epsilon_t=(I+\rho A)e_t \tag{4.16}$$

However, such a temporal VAR(1) is suffering from parameters under-identification. Two unknown parameters ρ and B are not identified by knowing only the estimate of the matrix product $(I+\rho A)^{-1}B(I+\rho A)$. Therefore, the conventional inference of vector autoregressive models should not be conducted.

Instead, this chapter uses an alternative estimation procedure which is inspired by the spatial cross-sectional models.

Let $(I + \rho A) = W$, then (4.13) becomes

$$\Rightarrow W r_t = BW r_{t-1} + \epsilon_t \tag{4.17}$$

with $\epsilon_t \sim N(0, \Sigma)$.

The log-likelihood function is as follows

$$\begin{aligned} \ln\mathcal{L} &= -\frac{T}{2}\ln\,\det(\Sigma) + \ln\,\det(W) - \frac{T}{2}\ln(2\pi) \\ &- \frac{1}{2}\sum_{t=1}^{T}\Big((Wr_t - BWr_{t-1})'\Sigma^{-1}(Wr_t - BWr_{t-1})\Big). \end{aligned} \tag{4.18}$$

The ML estimator of B is given by

$$\begin{aligned} \frac{\partial \ln\mathcal{L}}{\partial B} = 0 &\Rightarrow \sum_{t=1}^{T} Wr_{t-1}(Wr_t)' = \sum_{t=1}^{T} Wr_{t-1}(Wr_{t-1})'\hat{B}' \\ &\Rightarrow \sum_{t=1}^{T}(Wr_{t-1}r_t'W') = \sum_{t=1}^{T} Wr_{t-1}r_{t-1}'W'\hat{B}' \\ &\Rightarrow \hat{B}' = \Big(\sum_{t=1}^{T} Wr_{t-1}r_{t-1}'W'\Big)^{-1}\Big(\sum_{t=1}^{T} Wr_{t-1}r_t'W'\Big). \end{aligned}$$

Since $(I + \rho A) = W$, the ML estimate of B can be computed if ρ is known.

Next, the ML estimator of Σ can be written as

$$\begin{aligned} \hat{\Sigma} &= T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t' \\ &= T^{-1}\sum_{t=1}^{T} W\hat{e}_t\hat{e}_t'W', \end{aligned}$$

where $\hat{e}_t = (I+\rho A)^{-1}\hat{\epsilon}_t$. The estimated residuals $\hat{e}_t$ is from the VAR(1) of asynchronous returns. Similarly, the estimate of Σ can be computed if ρ is known.

Following Anselin and Bera (1998), the spatial correlation parameter ρ can be estimated by maximum-likelihood provided by the concentrated log-likelihood function in terms of ρ only, as

$$\begin{aligned} \ln\mathcal{L}^*(\rho) &= C - \frac{T}{2}\ln\,\det(\Sigma) + \ln\,\det(W) \\ &= C - \frac{T}{2}\ln\,\det(\hat{e}_t + \rho A\hat{e}_t)(\hat{e}_t + \rho A\hat{e}_t)' + \ln\,\det(I + \rho A) \end{aligned} \tag{4.19}$$

where C is irrelevant to ρ, ln det(W) is the log-determinant of $(I + \rho A)$, and $\ln\mathcal{L}^*(\rho)$ is a non-linear function in terms of the parameter ρ that must be maximised. Once the

ML estimate of ρ is yielded, the estimates of Σ, W and B can be obtained by working backwards.

Recall the restriction imposed in the spatial weighted matrix A such that matrix A is a upper diagonal matrix. The determinant of $I + \rho A$ is the product of the diagonal entries equals unity, and hence $log(1) = 0$.

The concentrated log-likelihood in terms of the unknown parameter ρ is reduced to

$$\ln\mathcal{L}^*(\rho) = -\frac{T}{2}\ln(\hat{e}_t + \rho A\hat{e}_t)'(\hat{e}_t + \rho A\hat{e}_t). \tag{4.20}$$

The ML estimator of ρ is given by

$$\begin{aligned}\frac{\partial \ln\mathcal{L}^*}{\partial \rho} = 0 &\Rightarrow \sum_{t=1}^{T}(\hat{e}_t' A' \hat{e}_t) = \sum_{t=1}^{T}(\hat{e}_t' A A' \hat{e}_t)\hat{\rho} \\ &\Rightarrow \hat{\rho} = \Big(\sum_{t=1}^{T}(\hat{e}_t' A' A \hat{e}_t)\Big)^{-1}\Big(\sum_{t=1}^{T}(\hat{e}_t' A' \hat{e}_t)\Big).\end{aligned}$$

4.5 Empirical application: a spatio-temporal VAR

In order to compare the estimation results and forecasting performance from two different synchronisation algorithms proposed, this empirical analysis uses the same dataset as in the previous chapter. The proposed spatio-temporal VAR model is applied on seven international stock markets from the eastern to western time zones including Japan, Australia, Hong Kong, Germany, the United Kingdom, Canada and the United States. The use of their daily close-to-close returns leaves time zone differences non-captured. This application assumed a dynamic conditional correlation (DCC) structure on the asynchronous variance-covariance matrix to allow time-varying correlations.

4.5.1 Data description

The data consists of seven stock market closing price series collected at the local closing time of each market, including Nikkei Stock Average (NIKKEI225) of Japan, Australian Stock Exchange (ASX) of Australia, Heng Seng Index (HSI) of Hong Kong, German stock index (DAX) of Germany, Financial Times Stock Exchange 100 index (FTSE100) of the United Kingdom, Toronto Stock Exchange (TSE) of Canada and Standard & Poor 500 index (S&P500) of the United States. The data is extracted from Bloomberg Database for the period 1 January 2005 - 30 September 2015. After dropping the observations with holidays/non-trading days in the time series data panel (i.e. the whole observation for each market j at day t is dropped if day t is the holiday for at least one of the investigated stock markets), there are 2360 common trading days for each stock price series. The data set is divided into two sub-samples, the first sample is ranging from 1 January 2005 to 31 December 2014, with 2190 trading days in total, for the purpose of model estimation. The second sample is ranging 1 January - 30 September 2015, with 170 trading days, for the purpose of out-of-sample forecast.

4.5.2 The spatial weighted matrix

The vector of observed close-to-close returns r_t for the seven market indices ($J = 7$) is calculated by taking the natural logarithmic difference of the closing prices p_t for each market j at the local closing time t_j in a time order of the earliest close to the latest close as

$$r_t = \{r_{t_1}^{ASX}, r_{t_2}^{NIK}, r_{t_3}^{HSI}, r_{t_4}^{DAX}, r_{t_5}^{FTSE}, r_{t_6}^{TSE}, r_{t_7}^{S\&P}\}' = \begin{pmatrix} \Delta\ln(p_{t_1}^{ASX}) \\ \vdots \\ \Delta\ln(p_{t_7}^{S\&P}) \end{pmatrix} = \Delta\ln(p_t)$$

These markets are located at different time zones from eastern to the western hemisphere, the local closing times and the time zone differences correspond to the US time (EDT) for the seven market stock indices are presented in Table 4.1. Expect the observed returns of the latest stock markets TSE and S&P500 are evaluated at the latest time of a day t, the observed returns of other earlier markets are recorded at $t_j < t$, denote the time differences between market i and j for all $i < j$ as $\Delta t_{i,j}$; in other words, the time direction for the earlier markets are out of phase and shifted backward.

Table 4.1 about here (see P.96).

Consider a $(J \times J)$ vector of synchronous returns R_t for the seven markets ($J = 7$) as a log difference of the synchronous prices p_t^s for each market j evaluated at a synchronised time t as

$$R_t = \{R_t^{ASX}, R_t^{NIK}, R_t^{HSI}, R_t^{DAX}, R_t^{FTSE}, R_t^{TSE}, R_t^{S\&P}\}' = \begin{pmatrix} \Delta\ln(p_t^{s,ASX}) \\ \vdots \\ \Delta\ln(p_t^{s,S\&P}) \end{pmatrix} = \Delta\ln(p_t^s)$$

where the closing time of the TSE and S&P500 is considered as the synchronised time t in this application.

Since synchronous returns R_t are not observed for each international market at synchronised time t, instead, asynchronous returns r_t are observed locally at local closing time

t_j. The differences in between are defined as measurement errors v_t(as in 4.6), and a (7×1) vector of measurement errors is as

$$v_t = R_t - r_t = \begin{pmatrix} \Delta \ln(p_t^{s,ASX}) - \Delta \ln(p_t^{ASX}) \\ \vdots \\ \Delta \ln(p_t^{s,S\&P}) - \Delta \ln(p_t^{S\&P}) \end{pmatrix} = \begin{pmatrix} v_t^{ASX} \\ \vdots \\ v_t^{S\&P} \end{pmatrix}$$

The assumption 1 in (A4.1) approximates the measurement errors v_t^i for earlier markets i as the spatial weights $A_{i,j}$ on the current observed returns $r_{t_j}^j$ $\forall\ i < j$ and $i, j \in J$, where t he spatial weights capture the positive directions of time differences from the earlier to the latest closing time. In matrix form, the spatial proxy is as

$$\begin{pmatrix} v_t^{ASX} \\ v_t^{JPN} \\ \vdots \\ v_t^{S\&P} \end{pmatrix} = \rho \begin{pmatrix} 0 & A_{ASX,UK} & \dots & A_{ASX,US} \\ 0 & 0 & \dots & A_{JPN,US} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} r_{t_1}^{ASX} \\ r_{t_2}^{JPN} \\ \vdots \\ r_{t_7}^{S\&P} \end{pmatrix}$$

where the spatial correlation parameter ρ is a scalar.

In particular to this empirical analysis, the spatial weighted matrix $A_{(7\times7)}$ is numerically defined as

$$A = \begin{pmatrix} A_{ASX,j} \\ A_{NIK,j} \\ A_{HSI,j} \\ A_{DAX,j} \\ A_{FTSE,j} \\ A_{TSE,j} \\ A_{S\&P,j} \end{pmatrix} = \begin{pmatrix} 0 & 1/24 & 3/24 & 11.5/24 & 11.5/24 & 16/24 & 16/24 \\ 0 & 0 & 2/24 & 10.5/24 & 10.5/24 & 15/24 & 15/24 \\ 0 & 0 & 0 & 8.5/24 & 8.5/24 & 13/24 & 13/24 \\ 0 & 0 & 0 & 0 & 0 & 4.5/24 & 4.5/24 \\ 0 & 0 & 0 & 0 & 0 & 4.5/24 & 4.5/24 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

4.5.3 Estimating the spatio-temporal VAR

The (7×1) vector of observed returns of the seven stock markets is modelled by the proposed "first-order spatio-temporal VAR" in (4.14). Additionally, the time-varying dynamics of the conditional variance-covariance is modelled through a Multivariate Dynamic Conditional Correlation (DCC) model (Engle, 2002). The model specification is as follows.

$$\text{True model: } R_t = \bar{c} + BR_{t-1} + \epsilon_t$$

$$\text{Spatial Proxy: } R_t = (I + \rho A) r_t$$

The spatio-temporal VAR is given by substituting the spatial proxy equation into the true VAR model, as

$$r_t = \bar{c} + Br_{t-1} - \rho Ar_t + B\rho Ar_{t-1} + \epsilon_t \tag{4.21}$$

where

B is the (7×7) coefficient matrix of the temporal lag;

$-\rho A$ is the (7×7) spatial coefficient matrix of the spatial lag;

$B\rho A$ is the (7×7) spatio-temporal coefficient matrix of the lagged spatial lag.

The hypothesis tests of Granger Causality can be constructed on parameter matrix B from the true model as long as the parameter B can be consistently estimated from the spatio-temporal VAR model in (4.21). However, the parameters are under-identified through the spatio-temporal VAR in (4.21). Thus the parameter estimates are produced by an alternative estimation procedure as follows.

$$Wr_t = \bar{c} + BWr_{t-1} + \epsilon_t \qquad \text{with } \epsilon_t|\mathcal{F}_{t-1} \sim N(0, H_t) \tag{4.22}$$

where

$$\begin{aligned}
&W \equiv (I + \rho A) \\
&H_t = D_t \tilde{R}_t D_t \\
&D_t = diag(H_t^{\frac{1}{2}}) \\
&H_{i,i,t} = \omega_i + \alpha_i' \epsilon_{t-1}\epsilon_{t-1}' + \beta_i H_{i,i,t-1} \qquad \text{for } \ i = 1, 2, ..., 7 \\
&\tilde{R}_t = diag(Q_t^{-\frac{1}{2}}) Q_t diag(Q_t^{-\frac{1}{2}}) \\
&Q_t = \Omega + \alpha\epsilon_{t-1}\epsilon_{t-1}' + \beta Q_{t-1}
\end{aligned}$$

This research considers the DCC(1,1) model. The conditional covariance matrix in the DCC model is decomposed into a relation between the estimated univariate GARCH variances D_t and the conditional correlation matrix $\tilde{R}_t$. H_t and $\tilde{R}_t$ are positive definite, D_t is a diagonal matrix with the elements of the estimated univariate GARCH variances, Q_t is the quasi-correlation matrix which is rescaled by $\rho_{i,j,t} = Q_{i,i,t}^{-\frac{1}{2}} Q_{i,j,t} Q_{j,j,t}^{-\frac{1}{2}}$ to ensure the correlation estimate is $[0, 1]$ bounded. The parameters (α_i, β_i) in the GARCH process are positive $\forall i$ and has a sum less than unity to ensure the stationary condition.

The maximum likelihood estimation procedure is given step-by-step as follows.

Step 1: Estimate the (7×1) vector of e_t from asynchronous VAR: $r_t = c_0 + br_{t-1} + e_t$ by maximum-likelihood, assume the error terms are multivariate normally distributed.

Step 2: Estimate the spatial correlation parameter ρ through the concentrated log-likelihood function as given by (4.20).

Step 3: Given the maximum likelihood estimate $\hat{\rho}$ in step 2, $\hat{W} = I + \rho A$ is obtained immediately. Then the maximum likelihood estimators B and H_t can be obtained by maximising the likelihood function of model (4.2).

The estimation results are reported in Table 4.2. Any insignificant elements are set to be zero for clearer presentation. The parameter estimate $\hat{b}$ is the coefficient matrix estimated from the asynchronous VAR(1) which is inconsistent and cannot be used for conducting tests of Granger causality. After performing the spatial adjustment to the asynchronous data, most of the parameter estimates of B from the true model are zero, except the first five elements of the last column are statistically significant and maternal in values. The last column is corresponding to the dependences between the US S&P500 and other earlier markets.

Table 4.2 about here (see P.97).

4.5.4 Spatially adjusted correlations

The (7×1) vector of estimated synchronised returns is given by

$$\hat{R}_t = \{\hat{R}_t^{ASX}, \hat{R}_t^{NIK}, \hat{R}_t^{HSI}, \hat{R}_t^{DAX}, \hat{R}_t^{FTSE}, \hat{R}_t^{TSE}, \hat{R}_t^{S\&P}\}' = (I + \hat{\rho}A)r_t$$

The spatially adjusted correlations of synchronised returns can be directly calculated, they are presented in Table 4.3. As expected, the contemporaneous correlations among the seven stock indices after the synchronisation adjustment are higher than the correlations from asynchronous data. The largest increase are the correlations between the Pacific stock markets and the North American markets from 0.2 to 0.6. The correlations between European markets and the North American markets also has slightly increase from 0.6 to 0.7. The correlations among the stock indices in the same time zone regions, such as DAX and FTSE100, TSE and S&P500 are almost unchanged because they are relatively synchronised.

Table 4.3 about here (see P.98).

4.5.5 Specification test

The specification test is performed to see whether the spatial autocorrelation exhibit is important or not. The most common test for the existence of spatial dependence is Moran's I by Moran (1948; 1950). The Moran's I statistic is defined for data vector x

with R number of spatial regions and the spatial weights $w_{i,j}$ between individual i and j by

$$I = \frac{R}{\sum_i \sum_j \omega_{ij}} \frac{\sum_i \sum_j \omega_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

and the variance of Moran's I is given by

$$Var(I) = \frac{1}{S_0^2(R^2 - 1)}(R^2 S_1 - RS_2 + 3S_0^2) - [E(I)]^2$$

where

$$E(I) = \frac{-1}{R-1} \text{ is the expectation of Moran's I}$$
$$S_0 = \sum_i \sum_j \omega_{ij}$$
$$S_1 = \frac{\sum_i \sum_j (\omega_{ij} + \omega_{ji})^2}{2}$$
$$S_2 = \sum_i (\omega_{i\cdot} + \omega_{\cdot})^2 \text{ is the sum of square of the ith column plus the ith row}$$

Moran's $\mathbb{I}$ statistic is asymptotically normal given by

$$\mathbb{I} = \frac{I - E(I)}{\sqrt{Var(I)}}$$

The Moran's $\mathbb{I}$ test statistic from this empirical analysis is $\mathbb{I} = 2.979$ which rejects the null hypothesis of no spatial autocorrelation at 1% level of significance. Therefore, the spatial weights with time zone differences are found to be statistically significant to capture the spatial dependence among the asynchronous stock returns.

4.6 Conclusion

Different stock markets are trading at their only local time, the correlations among international stock markets are underestimated when the data asynchronicity is not adjusted. The use of asynchronous data on the multivariate time series models not only underestimates the correlation dynamics among the data, but also results in measurement error problem which leads to inconsistent parameter estimates. This research proved that as long as a good proxy is available to replace the measurement error, the maximum likelihood estimators are still asymptotically consistent.

This study provides analytical evidence that the maximum likelihood estimators are inconsistent if using asynchronous data on the multivariate time series model. The measurement error bias is closely associated with the degree of asynchronicity among two markets. Therefore, this study proposed a spatial proxy to the measurement error which is inspired by the fact that the amount of underestimation depends on the degree of trading asynchronicity. The spatial weighted matrix captures the time zone differences between the early and later close markets, in which the latest close market is taken as the base. By using the spatial proxy, the temporal vector autoregressive process of the unobserved true returns can be expressed as a spatio-temporal vector autoregressive process of the observed asynchronous returns. In this model, the test of Granger causality can be conducted on the true coefficient matrix B. The Moran's I specification test also confirmed that the spatial autocorrelation is significant on the asynchronous returns, thus the proposed model can be useful to capture the spatial time zone differences of data.

Tables of Chapter 4

Index	Closing at Local Time	Closing at US Time	Trading Overlap with US?	Closing time diff. with US (in hours)
ASX	16:00	00:00	non-overlap	16
NIKKEI225	15:00	01:00	non-overlap	15
HSI	16:00	03:00	non-overlap	13
DAX	17:30	11:30	partially	4.5
FTSE100	16:30	11:30	partially	4.5
TSE	16:00	16:00	overlap	0
S&P500	16:00	16:00	overlap	0

Table 4.1: Closing times, overlapping and closing time differneces corresponds to the US markets for seven international stock indices.

Step 1 & 2: Spatial Correlation Parameter ρ

Parameter	ASX	NIK	HSI	DAX	FTSE	TSE	S&P
$\rho = 0.2752^{***}$							
$b_{ASX,i}$	-0.097	-0.047	-0.067	0.061	0.059	0.094	0.348
$b_{NIK,i}$	-0.095	-0.082	0	0.187	0	0	0.374
$b_{HSI,i}$	0	-0.079	-0.101	0	0	0	0.406
$b_{DAX,i}$	0	0	0	-0.064	-0.189	-0.066	0.355
$b_{FTSE,i}$	0	0	0	-0.094	-0.172	0	0.321
$b_{TSE,i}$	0	0	0	0	0	-0.106	0.077
$b_{S\&P,i}$	0	0	0	0	0	-0.084	-0.071

Step 3: Spatio-Temporal VAR $Wr_t = \bar{c} + BWr_{t-1} + \epsilon_t$

Parameter	ASX	NIK	HSI	DAX	FTSE	TSE	S&P
$B_{ASX,i}$	0	0	0	0	0	0	0.480
$B_{NIK,i}$	0	-0.089	0	0.197	0	0	0.531
$B_{HSI,i}$	0	-0.080	-0.102	0	0	0	0.510
$B_{DAX,i}$	0	0	0	0	-0.196	0	0.366
$B_{FTSE,i}$	0	0	0	-0.099	-0.176	0	0.333
$B_{TSE,i}$	0	0	0	0	0	-0.121	0
$B_{S\&P,i}$	0	0	0	0	0	-0.099	0
$\bar{c}_i$	0.002	0.002	0.002	0.002	0.001	0.001	0.001
ω_i	0.000	0.000	0.000	0.000	0.000	0.000	0.000
α_i	0.069	0.097	0.070	0.073	0.073	0.062	0.087
β_i	0.903	0.868	0.907	0.900	0.899	0.919	0.882

Note: All the parameter estimates are statistically significant at 1% level of significance.

Table 4.2: The estimation results of spatio-temporal VAR(1)-DCC(1,1) by the maximum likelihood.

Panel A: Asynchronous Correlations

Index	ASX	NIKKEI	HSI	DAX	FTSE	TSE	S&P
ASX	1.000	0.691	0.676	0.414	0.462	0.315	0.218
NIKKEI225	0.691	1.000	0.650	0.397	0.419	0.298	0.214
HSI	0.676	0.650	1.000	0.441	0.472	0.382	0.300
DAX	0.414	0.397	0.441	1.000	0.868	0.599	0.664
FTSE100	0.462	0.419	0.472	0.868	1.000	0.638	0.636
TSE	0.315	0.298	0.382	0.599	0.638	1.000	0.785
S&P500	0.218	0.214	0.300	0.664	0.636	0.785	1.000

Panel B: Spatially Adjusted Correlations

Index	ASX	NIKKEI	HSI	DAX	FTSE	TSE	S&P
ASX	1.000	0.817	0.810	0.707	0.741	0.624	0.561
NIKKEI225	0.817	1.000	0.768	0.634	0.651	0.546	0.483
HSI	0.810	0.768	1.000	0.629	0.655	0.574	0.511
DAX	0.707	0.634	0.629	1.000	0.884	0.642	0.703
FTSE100	0.741	0.651	0.655	0.884	1.000	0.684	0.683
TSE	0.624	0.546	0.574	0.642	0.684	1.000	0.785
S&P500	0.561	0.483	0.511	0.703	0.683	0.785	1.000

Table 4.3: The asynchronous daily correlations (Panel A) and the spatially adjusted correlations (Panel B) for seven markets for the period 3 January 2005 - 31 December 2014.

Appendix A

Appendix of Chapter 2

A.1 Proof of Theorem 2.1

According to the standard expression of LM statistic (2.16), the pointwise LM statistic evaluated under the null in (2.14) is given as

$$LM(\beta,\gamma) = S_T^*(\beta,\gamma)'\Xi_T^*(\beta,\gamma)^{-1}S_T^*(\beta,\gamma)$$

where the product matrices are given by

$$\begin{aligned}
\Xi_T^*(\beta,\gamma) &= \Xi_T(\beta,\gamma) - M_T(\beta,\gamma)M_T(\beta)^{-1}\Xi_T(\beta,\gamma) - \Xi_T(\beta,\gamma)M_T(\beta)^{-1}M_T(\beta,\gamma)\\
&\quad + M_T(\beta,\gamma)M_T(\beta)^{-1}\Xi_T(\beta)M_T(\beta)^{-1}M_T(\beta,\gamma),\\
S_T^*(\beta,\gamma) &= S_T(\beta,\gamma) - M_(\beta,\gamma)M_T(\beta)^{-1}S_T(\beta),\\
M_T(\beta,\gamma) &= I_p \otimes \frac{1}{T}\sum_{t=1}^{T} D_{1t}(\beta,\gamma)z_t(\beta)z_t(\beta)' \quad \text{and} \quad M_T(\beta) = I_p \otimes \frac{1}{T}\sum_{t=1}^{T} z_t(\beta)z_t(\beta)'\\
\Xi_T(\beta,\gamma) &= \frac{1}{T}\sum_{t=1}^{T} D_{1t}(\beta,\gamma)(\tilde{\epsilon}_t\tilde{\epsilon}_t' \otimes z_t(\beta)z_t(\beta)') \quad \text{and} \quad \Xi(\beta) = \frac{1}{T}\sum_{t=1}^{T}(\tilde{\epsilon}_t\tilde{\epsilon}_t' \otimes z_t(\beta)z_t(\beta)'),\\
S_T(\beta,\gamma) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} D_{1t}(\beta,\gamma)(z_{0,t} \otimes z_t(\beta)) \quad \text{and} \quad S_T(\beta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{z}_{0,t} \otimes z_t(\beta))\\
&\text{where} \quad \tilde{z}_{0,t} = \Delta^{\tilde{d}}(X_t - \tilde{\mu}).
\end{aligned}$$

Given by assumption 1(5), the estimate of β does not enter into the optimality of supLM statistic. Then

$$\sup LM = \sup_{\gamma\in\Gamma^*} LM(\beta_0,\gamma) = \sup_{r\in\Lambda} LM(\beta_0, \mathcal{F}^{-1}(r))$$

where $r = \mathcal{F}(\gamma)$ given in Definition 1. Since $LM(\beta_0,\gamma)$ is a function of γ only through $1(e_{t-1} \leq \gamma) = 1(\omega_{t-1} \leq r)$, it follows that

$$\sup LM = \sup_{r\in\Lambda} LM(\beta_0, r)$$

Under $\mathcal{H}_0$ and replacing $1(e_{t-1} \leq \gamma)$ with $1(\omega_{t-1} \leq r)$ for the above product matrices, the LM statistic simplifies to

$$LM(r) = S_T^*(r)'\Xi_T^*(r)^{-1}S_T^*(r)$$

where

$$\begin{aligned}
\Xi_T^*(r) &= \Xi_T(r) - M_T(r)M_T^{-1}\Xi_T(r) - \Xi_T(r)M_T^{-1}M_T(r) + M_T(r)M_T^{-1}\Xi_T M_T^{-1}M_T(\beta,\gamma),\\
S_T^*(r) &= S_T(r) - M_(r)M_T^{-1}S_T,\\
M_T(r) &= I_p \otimes \frac{1}{T}\sum_{t=1}^{T} 1(\omega_{t-1} \leq r)z_t z_t' \quad \text{and} \quad M_T = I_p \otimes \frac{1}{T}\sum_{t=1}^{T} z_t z_t'\\
\Xi_T(r) &= \frac{1}{T}\sum_{t=1}^{T} 1(\omega_{t-1} \leq r)(\tilde{\epsilon}_t\tilde{\epsilon}_t' \otimes z_t z_t') \quad \text{and} \quad \Xi = \frac{1}{T}\sum_{t=1}^{T}(\tilde{\epsilon}_t\tilde{\epsilon}_t' \otimes z_t z_t'),\\
S_T(r) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} 1(\omega_{t-1} \leq r)(z_{0,t} \otimes z_t) \quad \text{and} \quad S_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{z}_{0,t} \otimes z_t).
\end{aligned}$$

The stated rsult then follows from the joint convergence

$$\begin{aligned}
M_T(r) &\Rightarrow M(r),\\
\Omega_t(r) &\Rightarrow \Omega(r),\\
S_T(r) &\Rightarrow S(r).
\end{aligned}$$

The sketch proof above is reproduced from Theorem 1 of Hansen and Seo (2002) and Theorem 3 of Hansen (1996).

A.2 Matlab codes

```
%% TWO REGIME THRESHOLD FCVAR MODEL %%
%-----------------------------------------------------------------------%
%-----------------------------------------------------------------------%
% This ThresholdGrid.m file is written by Chi Wan Cheang for the paper
% "Threshold Fractionally Cointegrated VAR Model and Application to VIX Index".
%
% The Matlab code provided in this file is based on the Matlab programme
% for the (Linear) Fractionally Cointegrated VAR Model written by Nielsen
% and Popiel (2016), and the Matlab code for the Threshold Cointegration by
% Hansen and Seo (2002).
%
% The proposed 2-regime threshold model is compatable with Johansen and
% Nielsen (2012)'s FCVAR model and Hansen and Seo (2002)'s threshold
% cointegration. Level parameter is assumed in the FCVAR and is assumed to
% be fixed across regimes. Option for a regime-switching unrestricted
% constant (Xi) is also allowed for programme users. When restriction
% d=b=1 is imposed, the threshold FCVAR model reduced to the standard
% threshold cointegration model.

% -------- Import Data ----------%
clc;
clear all;

[data, date, raw] = xlsread('vix', 'Sheet1', 'A2:D3216');
vix = [data(:,1), data(:,2)];

% Add path containing Auxillary files required for estimation.
addpath Auxiliary/

% A bivariate time series.
x = vix(:, [1 2]);
sp = data(:,3);

%% -------- INITIALIZATION ----------%
p                = size(x, 2); % system dimension.
T                = size(x, 1); % number of observations in the sample.
kmax             = 3;    % maximum number of lags for VECM.
order            = 12;   % number of lags for white noise test in lag selection.
printWNtest      = 1;    % to print results of white noise tests post-estimation.

% -------- Choosing estimation options ----------%
opt = EstOptions; % Define variable to store Estimation Options (object).
opt.dbMin        = [0.01 0.01]; % lower bound for d,b.
opt.dbMax        = [1.00 1.00]; % upper bound for d,b.
opt.unrConstant  = 0; % include an unrestricted constant? 1 = yes, 0 = no.
opt.rConstant    = 0; % include a restricted constant? 1 = yes, 0 = no.
opt.levelParam   = 1; % include level parameter? 1 = yes, 0 = no.
opt.constrained  = 0; % impose restriction dbMax >= d >= b >= dbMin ? 1 = yes, 0 = no.
opt.restrictDB   = 0; % impose restriction d=b ? 1 = yes, 0 = no.
opt.db0          = [0.75 0.75]; % set starting values for optimization algorithm.
opt.N            = 0; % number of initial values to condition upon.
opt.print2screen = 1; % print output.
opt.printRoots   = 1; % do not print roots of characteristic polynomial.
opt.plotRoots    = 0; % plot roots of characteristic polynomial.
opt.gridSearch   = 0; % For more accurate estimation, perform the grid search.
opt.plotLike     = 1; % Plot the likelihood (if gridSearch = 1).
opt.progress     = 1; % Show grid search progress indicator waitbar.
opt.updateTime   = .5; % How often progress is updated (seconds).
```

```
% Linux example:
opt.progLoc = '"/usr/bin/fdpval"';  % location path with program name
                                    % of fracdist program, if installed
                                    % Note: use both single (outside) and double
                                    % quotes (inside). This is especially important
                                    % if path name has spaces.

DefaultOpt = opt; % store the options for restoring them in between hypothesis tests.

startProg = tic(); % start timer


%% --------- LINEAR FCVAR MODEL ESTIMATION ---------- %

k = 0;             % number of lags (short run dynamic).
r = 1;             % number of cointegrating rank.
opt1 = DefaultOpt; % define an estimation option.

% Estimate the linear FCVAR model.
m = FCVARestn(x, k, r, opt1);

% Call the coefficient estimates from the estimated linear model.
db = m.coeffs.db;
beta0 = m.coeffs.betaHat;
muHat = m.coeffs.muHat;

%% --------- THRESHOLD FCVAR MODEL ----------- %

% --------- Initialization --------- %
  tn = 100;          % number of grid points for threshold parameter gamma.
  bn = 100;          % number of grid points for beta.
  TotIters = tn*bn; % total iterations in the grid search.
  trim = 0.10;      % trimming (%) from the probability distribution of e0.
  trim_test = 0.15; % trimming for threshold test.
  beta_gs = 1;      % set to 1 to perform grid search around beta estimate.
  plotlike = 1;     % set to 1 to plot the likelihood functions of beta
                    % and gamma.
  n = size(x,1);   % number of observations.
  nlag = n - 1;
  x1 = x(2:n, :);

  % ------ Construct the grid --------- %
  beta = beta0(2);

  % Extract the standard error from long-run relation.
  seb = 0.01;

  % Calculate the threshold variable.
  if (opt.levelParam)
        dx = x - ones(size(x,1),1)*muHat; % Demean by mu.
        dx1 = x1 - ones(size(x1,1),1)*muHat; % Demean by mu.
        dxlag = dx(1:nlag,:);
        e0 = dxlag*beta0;
        q = unique(e0);
  else
        xlag = x(1:nlag,:);
        e0 = xlag*beta0;
        q = unique(e0);
  end
```

```
  % Define a set of threshold parameters given the number of grids.
  gamma1=q(round((1/(tn+1):1/(tn+1):(tn/(tn+1)))*size(q,1)));

  % Define the threshold parameters for hypothesis testing.
  gamma2=q(round((trim_test:((1-2*trim_test)/tn):((tn-1)*((1-2*trim_test)/tn)↙
+trim_test))*size(q,1)));

  % Construct evenly spaced grid on beta.
  Lbeta = beta - 6*seb;
  Ubeta = beta + 6*seb;
  inc_step = (2*6)*seb/(bn-1);

  if beta_gs == 1
        if bn == 1
            beta == beta0(2);
        else
        beta =(Lbeta:inc_step:Ubeta);
        end;
   else
        beta = beta0(2);
        bn = 1;
   end;

   % Create a matrix of NaN's to store likelihoods of grid search, we use
   % NaN's here because NaN entries are not plotted and do not affect the
   % finding of the maximum.
   like = ones(tn,bn)*NaN;

   % -------- Calculate likelihood over each grid. -------- %
   j = 1;
while j <= tn
    gamma=gamma1(j);
    bj = 1;
    while bj <= bn
        beta_g = [1; beta(bj)];
      if (opt.levelParam)
        e = dxlag*beta_g;
      else
        e = xlag*beta_g;
      end
        d1 = (e<=gamma); % logical result (Yes = 1, No = 0)
                         % error correction variable less than/equal to gamma
        d2 = 1 - d1;     % error correction variable larger than gamma
        n1 = sum(d1)';   % number of "Yes"

        if min([n1;nlag-n1])/n>trim    % split the sample into two
            if (opt.levelParam)
                col = size(dx1,2);
                r1=[dx1.*(d1*ones(1,col))];  % sample at regime 1
                r2=[dx1.*(d2*ones(1,col))];  % sample at regime 2
            else
                col = size(x1,2);
                r1=[x1.*(d1*ones(1,col))];  % sample at regime 1
                r2=[x1.*(d2*ones(1,col))];  % sample at regime 2
            end

    % Stack the observations into the corresponding regime.
        % Regime 1
         index = 0;
```

```
          for i=1:size(d1,1)
            if d1(i)==1
              if index ==0
                y1 = r1(i,:);
                index = 1;
              else
                y1 = [y1;r1(i,:)];
              end;
            end;
          end;

        % Regime 2
         index = 0;
          for i=1:size(d2,1)
            if d2(i)==1
              if index ==0
                y2 = r2(i,:);
                index = 1;
              else
                y2 = [y2; r2(i,:)];
              end;
            end;
          end;

     % Impose restriction to beta in the estimation option.
     opt1.R_Beta = [1 1];
     opt1.r_Beta = beta_g(1,1) + beta_g(2,1);

     t1 = size(y1,1);
     T1 = t1 - opt1.N;      % opt1.N is the number of initial values.
     t2 = size(y2,1);
     T2 = t2 - opt1.N;
     p = size(y1,2);
     opt1 = updateRestrictions(opt1,p,r);

     % Calculate the parameter estimates subject to the constraint on beta.
     estimates1 = GetParamsT(y1, k, r, db, beta_g, opt1);
     estimates2 = GetParamsT(y2, k, r, db, beta_g, opt1);

     % Calculate the new likelihood value given the new coefficient matrix.
     like1 = - T1*p/2*( log(2*pi) + 1)  - T1/2*log(det(estimates1.OmegaHat));
     like2 = - T2*p/2*( log(2*pi) + 1)  - T2/2*log(det(estimates2.OmegaHat));

     % Store the likelihood for each iteration.
     like(j,bj)= -(like1+like2);
     end;
       bj=bj+1;
    end;
     j=j+1;
end;

% -------------- Identify the MLE of beta and gamma --------------%
[temp, mlike] = min(like);
[temp, c] = min(diag(like(mlike,:)));
v = mlike(c);
gammaHat=gamma1(v);
b1=beta(c);
beta_H = [1 ;beta(c)];
minlike=like(v,c);
```

```
if (opt.levelParam)
    e_star = dxlag*beta_H;
else
    e_star = xlag*beta_H;
end
d1=(e_star<=gammaHat);
d2=1-d1;
R1=[x1.*(d1*ones(1,size(x1,2)))];
R2=[x1.*(d2*ones(1,size(x1,2)))];

index = 0;
for i=1:length(d1)
   if d1(i)==1
      if index ==0
         regime1 = R1(i,:);
         index = 1;
        else
         regime1 = [regime1; R1(i,:)];
      end;
   end;
end;

index = 0;
for i=1:length(d2)
    if d2(i)==1
        if index ==0
         regime2 = R2(i,:);
         index = 1;
        else
         regime2 = [regime2; R2(i,:)];
        end;
    end;
end;

% Piecewise linear estimation of threshold model.
Sample1 = FCVARestnT(regime1, k, r, db, muHat, beta_H, opt1);
Sample2 = FCVARestnT(regime2, k, r, db, muHat, beta_H, opt1);
```

Appendix B

Appendix of Chapter 3

The synchronised return model at equation (3.14) involves the fraction parameter matrix A and the autoregressive parameter matrix B. The matrix B can be estimated from the first-order autoregressive process of asynchronous returns (A2), and the matrix A can then be identified by an addition location-scale model at (A3).

Alternatively, instead of imposing Assumption 3.3 a location-scale model to synchronised returns, an auxiliary regression is proposed to identify the fraction parameter A.

B.1 Auxiliary regression

Suppose a vector of proxy X_t is observable has a linear relationship with the synchronised return R_t, we have a vector auxiliary regression such that

$$X_t = MR_t + v_t \tag{B.1}$$

with parameter M a $J \times J$ matrix.

The linear relation of the proxy variable X_t on synchronised returns R_t may be econometrically problematic if that auxiliary regression in which not only R_t affects the proxy X_t but also X_t affects R_t. If this is the case X_t and R_t have feedback effects instead. The unidirectional regression will lead to bias and inconsistent estimators, given by a non-zero covariance between R_t and the error v_t. To avoid this model misspecification, the choice of the proxy X_t is the key; alternatively it will be convenient to set-up a simultaneous regression model.

Recall that synchronised returns R_t are not observable but it can be written in terms of asynchronous returns as (3.14). Therefore, we are setting up a simultaneous regression model for the proxy X_t on asynchronous returns r_t.

Firstly we substitute R_t on (B1) by (3.14) and yield

$$\begin{aligned} X_t &= M[(I_J + AB)r_t - ABr_{t-1}] + v_t \\ &= \beta_1 r_t + \beta_2 r_{t-1} + v_t, \end{aligned} \tag{B.2}$$

let $\beta_1 = M(I_J + AB)$ and $\beta_1 = -MAB$.

Note that the fraction parameter A can be algebraically identified by

$$A = -(\beta_1 + \beta_2)^{-1}\beta_2 B^{-1}. \tag{B.3}$$

Referring to figure 1 for illustration, there is possibility that asynchronous returns r_t affect on both X_t and X_{t-1}. Thus we suppose the reverse causality of r_t on X_t is given by

$$r_t = \Gamma_1 X_t + \Gamma_2 X_{t-1} + u_t. \tag{B.4}$$

Considering equation (B2) and (B4) as a simultaneous equation model:

$$\begin{aligned} X_t &= \beta_1 r_t + \beta_2 r_{t-1} + v_t \\ r_t &= \Gamma_1 X_t + \Gamma_2 X_{t-1} + u_t \end{aligned}$$

Possibly, the asynchronous return r_t and the proxy X_t are *endogenous*, while the lag-1 proxy X_{t-1} and the lag-1 asynchornous return r_{t-1} are *exogenous.*

As the standard OLS method is not appropriate for simultaneous equation models, we should perform the parameter estimation by obtaining the reduced form model.

To do that, we substitute X_t given by (B2) into (B4) obtain

$$\begin{aligned} r_t &= \Gamma_1(\beta_1 r_t + \beta_2 r_{t-1} + v_t) + \Gamma_2 X_{t-1} + u_t \\ &= \Gamma_1\beta_1 r_t + \Gamma_1\beta_2 r_{t-1} + \Gamma_2 X_{t-1} + \Gamma_1 v_t + u_t \\ &= \Psi_1 r_{t-1} + \Psi_2 X_{t-1} + U_{1t}, \end{aligned} \tag{B.5}$$

where $\Psi_1 = (I - \Gamma_1\beta_1)^{-1}\Gamma_1\beta_2$, $\Psi_2 = (I - \Gamma_1\beta_1)^{-1}\Gamma_2$, $U_{1t} = (I - \Gamma_1\beta_1)^{-1}(\Gamma_1 v_t + u_t)$.

Then, we replace r_t on (B2) with its expression on (B5) to obtain

$$\begin{aligned} X_t &= \beta_1(\Psi_1 r_{t-1} + \Psi_2 X_{t-1} + U_{1t}) + \beta_2 r_{t-1} + v_t \\ &= (\beta_1\Psi_1 + \beta_2)r_{t-1} + \beta_1\Psi_2 X_{t-1} + \beta_1 U_{1t} + v_t \\ &= \Phi_1 r_{t-1} + \Phi_2 X_{t-1} + U_{2t}, \end{aligned} \tag{B.6}$$

where $\Phi_1 = \beta_1\Psi_1 + \beta_2$, $\Phi_2 = \beta_1\Psi_2$ and $U_{2t} = \beta_1 U_{1t} + v_t$.

B.2 Parameters identification

Now we can correctly estimate (B5) and (B6) by OLS and obtain their parameters Ψ_1, Ψ_2 and Φ_1, Φ_2 respectively. Since we have four unknown structural parameters β_1, β_2, Γ_1 and Γ_2 with four equations given by Ψ_1, Ψ_2, Φ_1 and Φ_2, we are able to identify the structural parameters and hence calculate the fraction parameter A as follows.

There are four equations for coefficients

i) $\Psi_1 = (I - \Gamma_1\beta_1)^{-1}\Gamma_1\beta_2$,

ii) $\Psi_2 = (I - \Gamma_1\beta_1)^{-1}\Gamma_2$,

iii) $\Phi_1 = \beta_1\Psi_1 + \beta_2$, and

iv) $\Phi_2 = \beta_1\Psi_2$.

From iv), $\beta_1 = \Phi_2\Psi_2^{-1}$. Substituting β_1 into iii) we obtain

$$\begin{aligned} \Phi_1 &= \Phi_2\Psi_2^{-1}\Psi_1 + \beta_2 \\ \rightarrow \beta_2 &= \Phi_1 - \Phi_2\Psi_2^{-1}\Psi_1. \end{aligned}$$

Recall that the fraction parameter A can be expressed in terms of β_1 and β_2 as (B3); therefore, the matrix A in terms of those four structural parameters is given by

$$\begin{aligned} A &= -(\beta_1 + \beta_2)^{-1}\beta_2 B^{-1} \\ &= -(\Phi_2\Psi_2^{-1} + \Phi_1 - \Phi_2\Psi_2^{-1}\Psi_1)^{-1}(\Phi_1 - \Phi_2\Psi_2^{-1}\Psi_1)B^{-1}. \end{aligned} \tag{B.7}$$

Appendix C

Appendix of Chapter 4

C.1 Proof of inconsistency of MLE from asynchronous VAR

I. Maximum likelihood estimators of B and Σ from true model (4.2), let $c = 0$ for simplification.

- *MLE of B*:

$$\begin{aligned}\frac{\partial ln\mathcal{L}}{\partial B} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial B}\big((R_t - BR_{t-1})'\Sigma^{-1}(R_t - BR_{t-1})\big)\\ &= \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial B}\big(R_t'\Sigma^{-1}BR_{t-1} + R_{t-1}'B'\Sigma^{-1}R_t - R_{t-1}'B'\Sigma^{-1}BR_{t-1}\big)\\ &= \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial B}\big(tr(R_{t-1}R_t'\Sigma^{-1}B) + tr(\Sigma^{-1}R_tR_{t-1}'B') - tr(R_{t-1}R_{t-1}'B'\Sigma^{-1}B)\big)\\ &= \frac{1}{2}\sum_{t=1}^{T}\big(R_{t-1}R_t'\Sigma^{-1} + R_{t-1}R_t'\Sigma^{-1} - 2R_{t-1}R_{t-1}'B'\Sigma^{-1}\big)\\ &= \sum_{t=1}^{T}\big(R_{t-1}R_t'\Sigma^{-1} - R_{t-1}R_{t-1}'B'\Sigma^{-1}\big)\end{aligned}$$

$\frac{\partial ln\mathcal{L}}{\partial B} = 0$ gives the maximum likelihood estimator of B as follows:

$$\begin{aligned}\frac{\partial ln\mathcal{L}}{\partial B} = 0 &\rightarrow \sum_{t=1}^{T}R_{t-1}R_t' = \sum_{t=1}^{T}R_{t-1}R_{t-1}'\hat{B}'\\ &\rightarrow T^{-1}\sum_{t=1}^{T}\big(R_{t-1}R_{t-1}'\big)\hat{B}' = T^{-1}\sum_{t=1}^{T}R_{t-1}R_t'\\ &\rightarrow \hat{B}'_{ML} = \big(\sum_{t=1}^{T}R_{t-1}R_{t-1}'\big)^{-1}\big(\sum_{t=1}^{T}R_{t-1}R_t'\big).\end{aligned}$$

- *MLE of* Σ:

$$
\begin{aligned}
\frac{\partial ln\mathcal{L}}{\partial\Sigma} &= -\frac{T}{2}\frac{\partial}{\partial\Sigma}ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial\Sigma}\big(\epsilon_t'\Sigma^{-1}\epsilon_t\big) \\
&= -\frac{T}{2}\frac{1}{\Sigma} - \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial\Sigma}\big(tr(\epsilon_t\epsilon_t'\Sigma^{-1})\big) \\
&= -\frac{T}{2\Sigma} + \frac{1}{2}\sum_{t=1}^{T}\epsilon_t\epsilon_t'\Sigma^{-2}
\end{aligned}
$$

$\frac{\partial ln\mathcal{L}}{\partial\Sigma} = 0$ gives the maximum likelihood estimator of Σ as follows:

$$
\begin{aligned}
\frac{\partial ln\mathcal{L}}{\partial\Sigma} = 0 \rightarrow \frac{T}{\hat{\Sigma}} &= \sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t{}'\hat{\Sigma}^{-2} \\
\rightarrow \hat{\Sigma} &= T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t{}'.
\end{aligned}
$$

II. Maximum likelihood estimators of B and Σ from the model of observed returns r_t with the consideration of measurement equations, assuming serially uncorrelated measurement errors.

- *MLE of* B:

$$
\begin{aligned}
\frac{\partial ln\mathcal{L}}{\partial B} &= -\frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial B}\big((r_t+v_t-B(r_{t-1}+v_{t-1}))'\Sigma^{-1}(r_t+v_t-B(r_{t-1}+v_{t-1}))\big) \\
&= \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial B}\big(2(r_t+v_t)'\Sigma^{-1}B(r_{t-1}+v_{t-1}) - (r_{t-1}+v_{t-1})'B'\Sigma^{-1}B(r_{t-1}+v_{t-1})\big) \\
&= \sum_{t=1}^{T}\big((r_{t-1}+v_{t-1})(r_t+v_t)'\Sigma^{-1} - (r_{t-1}+v_{t-1})(r_{t-1}+v_{t-1})'B'\Sigma^{-1}\big)
\end{aligned}
$$

$\frac{\partial ln\mathcal{L}}{\partial B} = 0$ gives the maximum likelihood estimator of B as follows:

$$
\begin{aligned}
\frac{\partial ln\mathcal{L}}{\partial B} = 0 &\rightarrow \sum_{t=1}^{T}(r_{t-1}+v_{t-1})(r_t+v_t)' = \sum_{t=1}^{T}(r_{t-1}+v_{t-1})(r_{t-1}+v_{t-1})'\hat{B}' \\
&\rightarrow T^{-1}\sum_{t=1}^{T}(r_{t-1}r_t') = T^{-1}\sum_{t=1}^{T}(r_{t-1}r_{t-1}' - v_{t-1}v_{t-1}')\hat{B}' \\
&\rightarrow \hat{B}'_{ML} = \big(\sum_{t=1}^{T}(r_{t-1}r_{t-1}' - v_{t-1}v_{t-1}')\big)^{-1}\big(\sum_{t=1}^{T}r_{t-1}r_t'\big),
\end{aligned}
$$

denote $Q_{r_{t-1}} = T^{-1}\sum_{t=1}^{T}r_{t-1}r_{t-1}'$, $Q_{r_{t-1}r_t} = T^{-1}\sum_{t=1}^{T}r_{t-1}r_t'$, $T^{-1}\sum_{t=1}^{T}v_{t-1}v_{t-1}' = \Sigma^v$,

$$
\rightarrow \hat{B}'_{ML} = \big(Q_{r_{t-1}} - \Sigma^v\big)^{-1}Q_{r_{t-1}r_t}.
$$

- *MLE of* Σ:

$$\frac{\partial ln\mathcal{L}}{\partial\Sigma} = -\frac{T}{2}\frac{\partial}{\partial\Sigma}ln\Sigma - \frac{1}{2}\sum_{t=1}^{T}\frac{\partial}{\partial\Sigma}\left(\epsilon_t'\Sigma^{-1}\epsilon_t\right)$$
$$= -\frac{T}{2\Sigma} + \frac{1}{2}\sum_{t=1}^{T}\epsilon_t\epsilon_t'\Sigma^{-2}$$

$\frac{\partial ln\mathcal{L}}{\partial\Sigma} = 0$ gives the maximum likelihood estimator of Σ as follows:

$$\frac{\partial ln\mathcal{L}}{\partial\Sigma} = 0 \rightarrow \hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t{}'$$
$$\rightarrow \hat{\Sigma} = T^{-1}\sum_{t=1}^{T}(r_t + v_t - \hat{B}r_{t-1} - \hat{B}v_{t-1})(r_t + v_t - \hat{B}r_{t-1} - \hat{B}v_{t-1})'$$
$$\rightarrow \hat{\Sigma} = T^{-1}\sum_{t=1}^{T}(r_t - \hat{B}r_{t-1})(r_t - \hat{B}r_{t-1})' - T^{-1}\sum_{t=1}^{T}v_tv_t' - T^{-1}\sum_{t=1}^{T}\hat{B}v_tv_t'\hat{B}'$$
$$\rightarrow \hat{\Sigma} = \Sigma^e - \Sigma^v - \hat{B}\Sigma^v\hat{B}'.$$

III. Repeat II, assuming serially correlated measurement errors.

- *MLE of* B:

$$\frac{\partial ln\mathcal{L}}{\partial B} = 0 \rightarrow \sum_{t=1}^{T}(r_{t-1}r_t' - v_{t-1}v_t') = \sum_{t=1}^{T}(r_{t-1}r_{t-1}' - v_{t-1}v_{t-1}')\hat{B}'$$
$$\rightarrow \hat{B}_{ML}' = \left(\sum_{t=1}^{T}(r_{t-1}r_{t-1}' - v_{t-1}v_{t-1}')\right)^{-1}\left(\sum_{t=1}^{T}(r_{t-1}r_t' - v_{t-1}v_t')\right),$$

denote $Q_{v_{t-1}v_t} = T^{-1}\sum_{t=1}^{T}v_{t-1}v_t'$,

$$\rightarrow \hat{B}_{ML}' = \left(Q_{r_{t-1}} - \Sigma^v\right)^{-1}\left(Q_{r_{t-1}r_t} - Q_{v_{t-1}v_t}\right).$$

- *MLE of* Σ:

$$\frac{\partial ln\mathcal{L}}{\partial\Sigma} = 0 \rightarrow \hat{\Sigma} = T^{-1}\sum_{t=1}^{T}(r_t + v_t - \hat{B}r_{t-1} - \hat{B}v_{t-1})(r_t + v_t - \hat{B}r_{t-1} - \hat{B}v_{t-1})'$$
$$\rightarrow \hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\left((r_t - \hat{B}r_{t-1})(r_t - \hat{B}r_{t-1})' - v_tv_t' - \hat{B}v_tv_t'\hat{B}' + \hat{B}v_{t-1}v_t' + v_tv_{t-1}'\hat{B}'\right)$$
$$\rightarrow \hat{\Sigma} = \Sigma^e - \Sigma^v - \hat{B}\Sigma^v\hat{B}' + \hat{B}Q_{v_{t-1}v_t} + Q_{v_{t-1}v_t}'\hat{B}'.$$

IV. Proof of Theorem 4.1.

Proof.

- The probability limit of $Q_{r_{t-1}}$ when $T \to \infty$.

$$\begin{aligned} Q_{r_{t-1}} &= T^{-1}\sum_{t=1}^{T} r_{t-1}r'_{t-1} \\ &= T^{-1}\sum_{t=1}^{T}(R_{t-1} - v_{t-1})(R_{t-1} - v_{t-1})' \\ &= T^{-1}\sum_{t=1}^{T}(R_{t-1}R'_{t-1} - R_{t-1}v'_{t-1} - v_{t-1}R'_{t-1} + v_{t-1}v'_{t-1}) \\ &= Q_{R_{t-1}} + Q_{v_{t-1}} + O_p(n^{-1/2}), \end{aligned}$$

  then $Q_{r_{t-1}} \xrightarrow{p} \Gamma(0) + \Sigma^v$; hence, $Q_{r_{t-1}} - \Sigma^v \xrightarrow{p} \Gamma(0)$.

- The probability limit of $Q_{r_{t-1}r_t}$ when $T \to \infty$.

$$\begin{aligned} Q_{r_{t-1}r_t} &= T^{-1}\sum_{t=1}^{T} r_{t-1}r'_{t} \\ &= T^{-1}\sum_{t=1}^{T}(R_{t-1}R'_{t} - R_{t-1}v'_{t} - v_{t-1}R'_{t} + v_{t-1}v'_{t}) \\ &= Q_{R_{t-1}R_t} + O_p(n^{-1/2}), \end{aligned}$$

  then

$$Q_{r_{t-1}r_t} \xrightarrow{p} \Gamma(0)B'.$$

□

V. Proof of Theorem 4.2.

Proof.

- Asymptotic mean of $\hat{B}_{ML}$:

$$\begin{aligned}
\text{plim}\hat{B}'_{ML} &= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} Q_{r_{t-1}r_t}\big) \\
&= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} T^{-1} \sum_{t=1}^{T} r_{t-1} r'_t\big) \\
&= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} T^{-1} \sum_{t=1}^{T} r_{t-1} (B r_{t-1} + e_t)'\big) \\
&= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} (T^{-1} \sum_{t=1}^{T} r_{t-1} r'_{t-1} B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t)\big) \\
&= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} (Q_{r_{t-1}} B' - T^{-1} \sum_{t=1}^{T} r_{t-1} (v_t - B v_{t-1} + \epsilon_t)')\big) \\
&= \text{plim}\big((Q_{r_{t-1}} - \Sigma^v)^{-1} (Q_{r_{t-1}} - \Sigma^v) B'\big) \\
&= B'
\end{aligned}$$

- Asymptotic variance of $\hat{B}_{ML}$: since

$$\begin{aligned}
\hat{B}' &= (Q_{r_{t-1}} - \Sigma^v)^{-1} Q_{r_{t-1}r_t} \\
&= (Q_{r_{t-1}} - \Sigma^v)^{-1} T^{-1} \sum_{t=1}^{T} r_{t-1} (B r_{t-1} + e_t)' \\
&= (Q_{r_{t-1}} - \Sigma^v)^{-1} T^{-1} \sum_{t=1}^{T} r_{t-1} (B r_{t-1} - \Sigma^v B' + \Sigma^v B' + e_t)' \\
&= B' + (Q_{r_{t-1}} - \Sigma^v)^{-1} (\Sigma^v B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t) \\
\rightarrow \hat{B}' - B' &= (Q_{r_{t-1}} - \Sigma^v)^{-1} (\Sigma^v B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t)
\end{aligned}$$

where $e_t = -v_t + B v_{t-1} + \epsilon_t$, then

$$\begin{aligned}
Var\big(\sqrt{T}(\hat{B}' - B')\big) &= T\Gamma(0)^{-1} Var(\Sigma^v B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t) \Gamma(0)^{-1} \\
&= T\Gamma(0)^{-1} E\big((\Sigma^v B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t)(\Sigma^v B' + T^{-1} \sum_{t=1}^{T} r_{t-1} e'_t)'\big) \Gamma(0)^{-1} \\
&= \Gamma(0)^{-1} E\big(-\Sigma^v B' B \Sigma^v + Q_{r_{t-1}} (\Sigma^v + B \Sigma^v B' + \Sigma)\big) \Gamma(0)^{-1}
\end{aligned}$$

since $Q_{r_{t-1}} \xrightarrow{p} \Gamma(0) + \Sigma^v$, and $\Sigma^e = \Sigma^v + B\Sigma^v B' + \Sigma$, then

$$Var\big(\sqrt{T}(\hat{B}' - B')\big) \xrightarrow{p} \Sigma^e \Gamma(0)^{-1} + \Gamma(0)^{-1} (\Sigma^v \Sigma^e - \Sigma^v B' B \Sigma^v) \Gamma(0)^{-1} \equiv \Omega. \qquad \square$$

C.2 An alternative proxy

This alternative proxy is derived from the definition of asynchronous returns in relation to the true return variable and the measurement error. Recall that the $(J \times 1)$ vector of asynchronous returns is defined as the log difference of two consecutive closing prices at the local closing time of each market, written as $r_t = \Delta P_t$; similarly, the $(J \times 1)$ vector of true returns is defined as the log difference of two consecutive synchronous prices $R_t = \Delta P_t^s$.

By adding and subtracting the true returns from asynchronous returns, we obtain an expression of asynchronous returns as true returns plus some non-captured returns, given by

$$\begin{aligned} r_t &= r_t - R_t + R_t \\ &= \Delta P_t^s - (P_t^s - P_t) + (P_{t-1}^s - P_{t-1}) \\ &= R_t - (P_t^s - P_t) + (P_{t-1}^s - P_{t-1}) \end{aligned}$$

denote $P_t^s - P_t = R_t^*$ as the $(J \times 1)$ vector of missing returns, then

$$r_t = R_t - R_t^* + R_{t-1}^*.$$

Recall that the measurement error equation is as $r_t = R_t + v_t$, thus the measurement error should be the uncaptured returns, that is

$$v_t = -R_t^* + R_{t-1}^*.$$

The good thing is that we now have a meaningful definition to the measurement error, the question needed to solve is that such missing return R_t^* is unobserved since it is made up by the unobserved synchronous price. Therefore, the following assumption can be an option to proxy the missing return R_t^*.

Assumption C.1. *Assume that the $(J \times 1)$ vector of missing returns $R_t^{*,j}$ for market j at time t is made up by a $(J \times J)$ fraction matrix A of the observed return r_t^j for $j = 1, 2, ..., J$, written as*

$$R_t^* = Ar_t \tag{C1}$$

where the $(J \times J)$ fraction matrix A is closed and bounded on $[0, 1]$.

By substituting (C1) into the measurement error equation, we obtain

$$\begin{aligned} r_t &= R_t - Ar_t + Ar_{t-1} \\ \Leftrightarrow R_t &= (I + A)r_t - Ar_{t-1} \end{aligned}$$

denote $I + A = \alpha$, then

$$R_t = \alpha r_t + (I - \alpha)r_{t-1}$$

the $(J \times J)$ fraction parameter α is also $[0,1]$ bounded given A is closed and bounded on $[0,1]$.

Next the expression of true returns in terms of observed returns is substituted into the true VAR model, suppose the true model is a VAR(1) such that $R_t = BR_{t-1} + \epsilon_t$, obtain

$$\begin{aligned} \alpha r_t + (I-\alpha) r_{t-1} &= B(\alpha r_t + (I-\alpha) r_{t-1}) + \epsilon_t \\ \Leftrightarrow \alpha r_t &= (B\alpha - I + \alpha) r_{t-1} + B(I-\alpha) r_{t-2} + \epsilon_t \\ \Leftrightarrow r_t &= \alpha^{-1}(B\alpha - I + \alpha) r_{t-1} + \alpha^{-1} B(I-\alpha) r_{t-2} + \alpha^{-1}\epsilon_t \end{aligned}$$

denote $\beta_1 = \alpha^{-1}(B\alpha - I + \alpha)$, $\beta_2 = \alpha^{-1}B(I-\alpha)$, and $e_t = \alpha^{-1}\epsilon_t$, the equation above is a VAR(2) process of observed returns, that is

$$r_t = \beta_1 r_{t-1} + \beta_2 r_{t-2} + e_t.$$

This transformed VAR(2) model can be estimated by the usual maximum likelihood method, but our aims are identifying the true parameter B, the fraction parameter α and the estimate of the variance matrix of the true disturbance term, $Var[\epsilon_t] = \Sigma$. Therefore, instead of estimating β_1 and β_2, the maximum likelihood estimators of α, B and Σ are derived from the transformed VAR(2).

The maximum likelihood estimators assuming normally distributed error are derived as follows.

$$\hat{B}'_{ML} = [\sum_{t=1}^{T} \hat{R}_{t-1}\hat{R}'_{t-1}]^{-1}[\sum_{t=1}^{T} \hat{R}_{t-1}\hat{R}'_{t}],$$

$$\begin{aligned} \hat{\alpha}' = &\Big(\sum_{t=1}^{T}(r_t - (I+\hat{B})r_{t-1} + \hat{B}r_{t-2})(r_t - (I+\hat{B})r_{t-1} + \hat{B}r_{t-2})'\Big)^{-1} \\ &\times \Big(\sum_{t=1}^{T}(r_t - (I+\hat{B})r_{t-1} + \hat{B}r_{t-2})(\hat{B}r_{t-2} - r_{t-1})'\Big), \end{aligned}$$

$$\hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}'_t.$$

where $\hat{R}_t = \hat{\alpha} r_t + (I - \hat{\alpha}) r_{t-1}$, and $\hat{\epsilon}_t = \hat{\alpha} r_t - (\hat{B}\hat{\alpha} - I + \hat{\alpha}) r_{t-1} + \hat{B}(I - \hat{\alpha}) r_{t-2}$.

The maximum likelihood estimators derived above invloved complicated non-linear optimisations that are difficult to applied in the real data. Further research could conduct some Monte Carlo simulation studies to estimate the parameters and to evaluate the adequacy of the asymptotic distributions of the estimators.

Bibliography

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856.

Andrews, D. W. and Guggenberger, P. (2003). A bias–reduced log–periodogram regression estimator for the long–memory parameter. *Econometrica*, 71(2):675–712.

Andrews, D. W. and Sun, Y. (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614.

Anselin, L. and Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, 155:237–290.

Asgharian, H., Hess, W., and Liu, L. (2013). A spatial analysis of international stock market linkages. *Journal of Banking & Finance*, 37(12):4738–4754.

Audrino, F. and Bühlmann, P. (2002). Synchronizing multivariate financial time series. Technical report, Università della Svizzera italiana.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, 73(1):5–59.

Baillie, R. T. and Bollerslev, T. (1994). Cointegration, fractional cointegration, and exchange rate dynamics. *The Journal of Finance*, 49(2):737–745.

Baillie, R. T. and Kapetanios, G. (2007). Testing for neglected nonlinearity in long-memory models. *Journal of Business & Economic Statistics*, 25(4):447–461.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International economic review*, pages 627–645.

Bandi, F. M. and Perron, B. (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, 4(4):636–670.

Bell, S. W., Schmieta, S. H., and Siu, F. P.-H. (2013). Returns-timing for multiple market factor risk models. US Patent 8,533,107.

Burns, P., Engle, R. F., and Mezrich, J. J. (1998). Correlations and volatilities of asynchronous data. *the Journal of Derivatives*, 5(4):7–18.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596.

Caporin, M., Ranaldo, A., and De Magistris, P. S. (2013). On the predictability of stock prices: A case for high and low prices. *Journal of Banking & Finance*, 37(12):5132–5146.

Chan, K.-S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The annals of statistics*, pages 520–533.

Cheung, Y.-W. and Lai, K. S. (1993). A fractional cointegration analysis of purchasing power parity. *Journal of Business & Economic Statistics*, 11(1):103–112.

Choi, I. and Saikkonen, P. (2010). Tests for nonlinear cointegration. *Econometric Theory*, 26(3):682–709.

Chow, G. C., Lawler, C. C., et al. (2003). A time series analysis of the shanghai and new york stock price indices. *Annals of Economics and Finance*, 4:17–36.

Christensen, B. J. and Nielsen, M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133(1):343–371.

Christensen, B. J. and Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150.

Cohen, K. J., Hawawini, G. A., Maier, S. F., Schwartz, R. A., and Whitcomb, D. K. (1980). Implications of microstructure theory for empirical research on stock price behavior. *The Journal of Finance*, 35(2):249–257.

Cohen, K. J., Hawawini, G. A., Maier, S. F., Schwartz, R. A., and Whitcomb, D. K. (1983). Friction in the trading process and the estimation of systematic risk. *Journal of Financial Economics*, 12(2):263–278.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of econometrics*, 105(1):131–159.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7(2):197–226.

Dittmann, I. and Granger, C. W. (2002). Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics*, 110(2):113–133.

Dolatabadi, S., Nielsen, M. Ø., and Xu, K. (2015). A fractionally cointegrated VAR analysis of price discovery in commodity futures markets. *Journal of Futures Markets*, 35(4):339–356.

Dolatabadi, S., Nielsen, M. Ø., and Xu, K. (2016). A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets. *Journal of Empirical Finance*, 38:623–639.

Dueker, M. and Startz, R. (1998). Maximum-likelihood estimation of fractional cointegration with an application to us and canadian bond rates. *Review of Economics and Statistics*, 80(3):420–426.

Durante, F. and Foscolo, E. (2013). An analysis of the dependence among financial markets by spatial contagion. *International Journal of Intelligent Systems*, 28(4):319–331.

Enders, W. and Siklos, P. L. (2001). Cointegration and threshold adjustment. *Journal of Business & Economic Statistics*, 19(2):166–176.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

Fernandez, V. (2011). Spatial linkages in international financial markets. *Quantitative Finance*, 11(2):237–245.

Figuerola-Ferretti, I. and Gonzalo, J. (2010). Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*, 158(1):95–107.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4):221–238.

Giraitis, L. and Surgailis, D. (1985). Clt and other limit theorems for functionals of gaussian processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 70(2):191.

Gonzalo, J. and Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110(2):319–352.

Gonzalo, J. and Pitarakis, J.-Y. (2006a). Threshold effects in cointegrating relationships. *Oxford Bulletin of Economics and Statistics*, 68(s1):813–833.

Gonzalo, J. and Pitarakis, J.-Y. (2006b). *Threshold effects in multivariate error correction models*. Palgrave Macmillan.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1):121–130.

Granger, C. W. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of economics and statistics*, 48(3):213–228.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of empirical finance*, 11(3):399–421.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29.

Granger, C. W. and Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics letters*, 62(2):161–165.

Greene, W. H. (2003). *Econometric analysis.* Pearson Education India.

Haldrup, N. and Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of econometrics*, 135(1):349–376.

Haldrup, N., Vera-Valdés, J., et al. (2017). Long memory, fractional integration, and cross-sectional aggregation. Technical report.

Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys*, 13(5):551–576.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430.

Hansen, B. E. (1997). Inference in tar models. *Studies in nonlinear dynamics & econometrics*, 2(1).

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface*, 4(2):123–127.

Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of econometrics*, 110(2):293–318.

Hansen, L. P., Sargent, T. J., et al. (1981). Exact linear rational expectations models: Specification and estimation. Technical report, Federal Reserve Bank of Minneapolis.

Heckman, J. J. (1977). Sample selection bias as a specification error (with an application to the estimation of labor supply functions).

Hosking, J. R. (1981). Fractional differencing. *Biometrika*, pages 165–176.

Jensen, A. N. and Nielsen, M. Ø. (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis*, 35(5):428–436.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: Journal of the Econometric Society*, pages 1551–1580.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press.

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24(03):651–676.

Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, 158(1):51–66.

Johansen, S. and Nielsen, M. Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, 80(6):2667–2732.

Johansen, S. and Nielsen, M. Ø. (2016). The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory*, 32(5):1095–1139.

Jones, M. E., Nielsen, M. Ø., and Popiel, M. K. (2014). A fractionally cointegrated var analysis of economic voting and political support. *Canadian Journal of Economics/Revue canadienne d'économique*, 47(4):1078–1130.

Kanzler, L. et al. (1998). Gph: Matlab module to calculate geweke-porter-hudak long memory statistic. *Statistical Software Components*.

Kapetanios, G., Shin, Y., and Snell, A. (2006). Testing for cointegration in nonlinear smooth transition error correction models. *Econometric Theory*, 22(2):279–303.

Kristensen, D. and Rahbek, A. (2013). Testing and inference in nonlinear cointegrating vector error correction models. *Econometric Theory*, 29(6):1238–1288.

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society*, volume 1, pages 67–74. VNU Science Press Utrecht.

Łasak, K. (2010). Likelihood based testing for no fractional cointegration. *Journal of Econometrics*, 158(1):67–77.

Lee, S., Seo, M. H., and Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association*, 106(493):220–231.

Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1-2):181–211.

Lobato, I. N. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of time series analysis*, 18(2):137–155.

Marinucci, D. and Robinson, P. M. (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics*, 105(1):225–247.

Martens, M. and Poon, S.-H. (2001). Returns synchronization and daily correlation dynamics between international stock markets. *Journal of Banking & Finance*, 25(10):1805–1827.

Mohanty, S., Peterson, E. W. F., and Smith, D. B. (1998). Fractional cointegration and the false rejection of the law of one price in international commodity markets. *Journal of Agricultural and Applied Economics*, 30(02):267–276.

Nielsen, M. and Popiel, M. (2016). A matlab program and user's guide for the fractionally cointegrated var model. Technical report, Queen's University, Department of Economics.

Nielsen, M. Ø. (2007). Local whittle analysis of stationary fractional cointegration and the implied–realized volatility relation. *Journal of Business & Economic Statistics*, 25(4):427–446.

Patriota, A. G., Sato, J. R., and Achic, B. G. B. (2010). Vector autoregressive models with measurement errors for testing granger causality. *Statistical Methodology*, 7(4):478–497.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, pages 1361–1401.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Ramchand, L. and Susmel, R. (1998). Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, 5(4):397–416.

Robinson, P. (1991). Time series with strong dependence. In *Advances in econometrics, sixth world congress*, volume 1, pages 47–95.

Robinson, P. and Marinucci, D. (2003). Fractional cointegration. *Time Series with Long Memory*, page 334.

Robinson, P. M. (1994). Semiparametric analysis of long-memory time series. *The Annals of Statistics*, pages 515–539.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, pages 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072.

Robinson, P. M. (2008). Multiple local whittle estimation in stationary systems. *The Annals of Statistics*, 36(5):2508–2530.

Robinson, P. M. and Hidalgo, F. (1997). Time series regression with long-range dependence. *The Annals of Statistics*, 25(1):77–104.

Rossi, E. and Santucci de Magistris, P. (2013). A no-arbitrage fractional cointegration model for futures and spot daily ranges. *Journal of Futures Markets*, 33(1):77–102.

Saikkonen, P. (2008). Stability of regime switching error correction models under linear cointegration. *Econometric Theory*, 24(1):294–318.

Saikkonen, P. and Choi, I. (2004). Cointegrating smooth transition regressions. *Econometric Theory*, 20(2):301–340.

Scherer, B. (2013). Synchronize your data or get out of step with your risks. *The Journal of Derivatives*, 20(3):75–84.

Scholes, M. and Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of financial economics*, 5(3):309–327.

Seo, B. (2003). Nonlinear mean reversion in the term structure of interest rates. *Journal of Economic Dynamics and Control*, 27(11):2243–2265.

Seo, M. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics*, 134(1):129–150.

Seo, M. H. (2011). Estimation of nonlinear error correction models. *Econometric Theory*, 27(2):201–234.

Shi, X. and Phillips, P. C. (2012). Nonlinear cointegrating regression under weak identification. *Econometric Theory*, 28(3):509–547.

Shimotsu, K. (2012). Exact local whittle estimation of fractionally cointegrated systems. *Journal of Econometrics*, 169(2):266–278.

Shimotsu, K. and Phillips, P. C. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

Simon, D. P. and Campasano, J. (2014). The vix futures basis: Evidence and trading strategies. *The Journal of Derivatives*, 21(3):54–69.

Sun, Y. and Phillips, P. C. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, 115(2):355–389.

Taqqu, M. S. (1979). Convergence of integrated processes of arbitrary hermite rank. *Probability Theory and Related Fields*, 50(1):53–83.

Theodossiou, P., Kahya, E., Koutmos, G., and Christofi, A. (1997). Volatility reversion and correlation structure of returns in major international stock markets. *Financial Review*, 32(2):205–224.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 245–292.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *journal of the american statistical association*, 93(443):1188–1202.

Van Dijk, D., Franses, P. H., and Paap, R. (2002). A nonlinear long memory model, with an application to us unemployment. *Journal of Econometrics*, 110(2):135–165.

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management*, 26(3):12–17.