2012

Pairs trading using cointegration approach

Heni Puspaningrum
University of Wollongong

Pairs Trading Using Cointegration Approach

A thesis submitted in fulfillment of the requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Heni Puspaningrum SSi (ITB), MCom (UNSW)

School of Mathematics and Applied Statistics
2012

Certification

I, Heni Puspaningrum, declare that this thesis, submitted in fulfillment of the requirement for the award of Doctor of Philosophy, in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Heni Puspaningrum
January 14, 2012

Abstract

Pairs trading strategy works by taking the arbitrage opportunity of temporary anomalies between prices of related assets which have long-run equilibrium. When such an event occurs, one asset will be overvalued relative to the other asset. We can then invest in a two-assets portfolio (a pair) where the overvalued asset is sold (short position) and the undervalued asset is bought (long position). The trade is closed out by taking the opposite positions of these assets after the asset prices have settled back into their long-run relationship. The profit is captured from this short-term discrepancies in the two asset prices. Since the profit does not depend on the movement of the market, pairs trading can be said as a market-neutral investment strategy.

There are four main approaches used to implement pairs trading: distance approach, combine forecasts approach, stochastic approach and cointegration approach. This thesis focuses on cointegration approach. It uses cointegration analysis in long/short investment strategy involving more than two assets. Cointegration incorporates mean reversion into pairs trading framework which is the single most important statistical relationship required for success. If the value of the portfolio is known to fluctuate around its equilibrium value then any deviations from this value can be traded against. Especially in this thesis, a pairs trading strategy is developed based on the cointegration coefficients weighted (CCW) rule. The CCW rule works by trading the number of unit in two assets based on their cointegration coefficients to achieve a guaranteed minimum profit per trade. The minimum profit per trade corresponds to the pre-set boundaries upper-bound U and lower-bound L chosen to open trades. The optimal pre-set boundary value is determined by maximising the minimum total profit (MTP) over a specified trading horizon. The MTP is a function of the minimum profit per trade and the number of trades during the trading horizon. This thesis provides the estimated number of trades. The number of trades is also influenced by the distance of the pre-set boundaries from the long-run cointegration equilibrium. The higher the pre-set boundaries for opening trades, the higher the minimum profit per trade but the trade numbers will be lower. The opposite applies for lowering the boundary values. The number of trades over a specified trading horizon is estimated jointly by the average trade duration and the average inter-trade interval. For any pre-set boundaries, both

of those values are estimated by making an analogy to the mean first-passage times for a stationary process.

Trading duration and inter-trade interval are derived using a Markov chain approach for a white noise and an AR(1) processes. However, to apply the approach to higher order AR(p) models, $p > 1$, is difficult. Thus, an integral equation approach is used to evaluate trading duration and inter-trade interval. A numerical algorithm is also developed to calculate the optimal pre-set upper-bound, denoted U_o, that would maximize the minimum total profit (MTP). The pairs trading strategy is applied to three empirical data examples. The pairs trading simulations show that the strategy works quite well for the data used.

Considering cointegration error ϵ_t may not follow a linear stationary AR(p) process but a nonlinear stationary process such as a nonlinear stationary exponential smooth transition autoregressive (ESTAR) model, this thesis extends Kapetanios's and Venetis's tests by considering a unit root test for a k-ESTAR(p) model with a different approach to Venetis. By using the approach in this thesis, the singularity problem can be avoided without adding the collinear regressors into the error term. For some cases, simulation results show that our approach is better than Venetis test, Kapetanios test and the Augmented Dickey-Fuller (ADF) test.

We also show that the mean first-passage time based on an integral equation approach can also be applied to the evaluation of trading duration and inter-trade interval for an 1-ESTAR(1) model and an 1-ESTAR(2) model. With some adjustments, a numerical algorithm for an AR(2) model can be used for an ESTAR model to calculate the optimal pre-set upper-bound. An application to pairs trading assuming ϵ_t is a nonlinear stationary ESTAR model is examined using simulated data.

This thesis also discusses cointegration and pairs trading between future prices and spot index prices of the S&P 500. There are three major issues discussed in Chapter 8. The first issue is to identify an appropriate model for the S&P 500 basis. The S&P 500 basis is defined as the difference between the log of future prices and the log of spot index prices. Following the same analysis procedures and using currently available data period, we reached different conclusion to Monoyios. Instead of concluding the basis follows an ESTAR model, we concluded that it follows a LSTAR model. Furthermore, even though we can conclude that there is possibility nonlinearity in the basis, there is no significant difference between a nonlinear LSTAR model and a linear autoregressive model in fitting the data. We also have a concern in the way the basis is constructed. By pairing up the spot price with the future contract with the nearest maturity, it may produce artificial jumps at the time of maturity. The longer the time to maturity, the higher the difference between the log future price and the log spot price as described by a cost-carry model. Therefore, the second issue of the chapter examined the cointegration of f_t and s_t with a time trend for each future contract. Cointegration analysis with a time trend based on the Engle-Granger approach and the Johansen approach are used.

Only 19 out of 44 future contracts conclude that they are cointegrated with a time trend using both the Engle-Granger approach and the Johansen approach. Perhaps, high volatility during financial crisis in the data period affects the cointegration test and also limited number of observations. Based on the Engle-Granger approach and then running nonlinearity tests for the regression residuals, we can conclude that there is no strong evidence of nonlinearity in the relationship between future and spot prices of S&P 500 index. It gives further evidence on the basis for a linearity rather than nonlinearity. The third issue is to examine the application of pairs trading between the future and spot index prices. The stationary characteristics of an AR(p) model are used to develop pairs trading strategy between future contract and corresponding spot index. Pairs trading simulations between future contract and CFD (Contract For Difference) the S&P 500 index data during 1998 show very good results resulting a total return more than 75% during the 3 months trading periods.

Acknowledgements

This thesis would not have been possible without the help and support of many people. I would firstly like to thank my supervisors, Associate Professor Yan-Xia Lin and Dr. Chandra Gulati during 7 years of my PhD study. Thank you Yan-Xia for your wisdom, knowledge, guidance and great patience. I appreciate that your comments have always been positive and for believing that I would finish this thesis one day, even when I had trouble believing it would happen. To Chandra, thank you for the time you have spent working through most of the detail of this thesis and also for your contribution ideas for the direction of this thesis. Both of you have taught me so much. I have appreciated all discussions we have had to get to this point. Thank you also for your support for my career by giving me opportunities to become a casual tutor and lecturer and also excellent referees for my job applications so now I already have a job offer from a government agency as a business analyst.

I would also like to thank the Student Research Centre University of Wollongong for giving me a University Postgraduate Award scholarship for the first 3.5 years of my study. Travel grants to attend conferences from the Centre for Statistical & Survey Methodology and the Centre for Mathematical Finance are also appreciated.

I would also like to thank the staff in the School of Mathematics and Applied Statistics for their support over years of my study. In particular, Michael McCrae for your contribution ideas for the direction of this thesis. I also thank to Carolyn for administration matters and Joe for computer matters.

The journey of my PhD is more enjoyable if it can be shared with others in the same situation. Thanks to all my fellow PhD students, particularly Carole, Sandy, Raed and Diane, for sharing the experiences. Thanks especially to Carole for your understanding in sharing our office room and also your help with the Latex.

To my friends at JPI, PPIA and IndoWollongong thanks for your supports and friendship during my study in Wollongong. With you all, I feel that living in Australia is like not far away from home. I would also like to thank mbak Anggraeni and mbak Sitta at ANU for helping me to get access for Datastream database used in Chapter 8.

To my parents in Jakarta, parent-in-law in Surabaya and other families in Indonesia, thanks for your support and prayers. Even though we have separated between two continents, I know that your prayers have always had for me and my family here.

Most importantly, to my husband, for your patience, support and encouragement you have always given to me. Thank you for your graciousness in helping me with house works and looking after our son. I know I wouldn't have achieved this without your support. To my son, thank you for your understanding during my busy days. May Allah always gives you guidance in your life.

I thank to Allah for sustaining me throughout this time and all He has given to me. I am looking forward to the next chapter in my life by starting my job in Canberra. I know that it is not going to be easy. But I know that whatever it is, He is always with me and gives me strength.

Contents

List of Figures

List of Tables

List of Abbreviations

ADF	Augmented Dickey-Fuller
ADRs	American Depository Receipts
AKSS	Augmented Kapetanios, Shin and Snell
APPT	Average Profit per Trade
APT	Arbitrage Pricing Theory
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
ART	Average Return per Trade
ASX	Australian Stock Exchange
BHP	Broken Hill Proprietary
BN	Beveridge-Nelson
BOQ	Bank of Queensland (BOQ)
CAPM	Capital Asset Pricing Model
CBA	Commonwealth Bank of Australia
CCW	Cointegration Coefficients Weighted
CFD	Contract For Difference
CRDW	Cointegrating Regression Durbin-Watson
CV	Cointegrating Vector
DF	Dickey-Fuller
DJIA	Dow-Jones Industrial Average
DLC	Dual-listed Company
DW	Durbin-Watson
ESTAR	Exponential Smooth Transition Auto Regressive
FTSE 100	Financial Times Stock Exchange 100
GLS	Generalised Least Squares
IID	Identic Identical Distribution
IT	Inter-Trades interval
LR	Likelihood Ratio

LSTAR	Logistic Smooth Transition Auto Regressive
Ltd	Limited
MA	Moving Average
MCDM	Multi Criteria Decision Methods
MDS	Martingale Difference Sequence
MMSE	Minimum Mean Squared Error
MSE	Minimum Squared Error
MTP	Minimum Total Profit
NASDAQ	National Association of Securities Dealers Automated Quotations
NT	Number of Trades
NV	Naamloze Vennootschap
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
PLC	Public Limited Company
PPP	Purchasing Power Parity
RESET	Regression Equation Specification Error Test
S&P 500	Standard & Poor's 500
STAR	Smooth Transition Auto Regressive
TAR	Threshold Auto Regressive
TD	Trading Duration
TP	Total Profit
USA	United States of America
VECM	Vector Error Correction Model
VPP	Venetis, Paya and Peel
WBC	Wespact Banking Corporation

Chapter 1

Introduction

Pairs trading was first discovered in the early 1980s by the quantitative analyst Tartaglia and a team of physicists, computer scientists and mathematicians, who did not have a background in finance. Their idea was to develop statistical rules to find ways to perform arbitrage trades, and take the 'skill' out of trading (Gatev *et al.*, 2006).

Pairs trading works by taking the arbitrage opportunity of temporary anomalies between prices of related assets which have long-run equilibrium. When such an event occurs, one asset will be overvalued relative to the other asset. We can then invest in a two-assets portfolio (a pair) where the overvalued asset is sold (short position) and the undervalued asset is bought (long position). The trade is closed out by taking the opposite positions of these assets after the asset prices have settled back into their long-run relationship. The profit is captured from this short-term discrepancies in the two asset prices. Since the profit does not depend on the movement of the market, pairs trading can be said as a market-neutral investment strategy.

Mostly people apply pairs trading in a stock market due to its liquidity. The first extensive pairs trading examination was done by Gatev *et al.* (1998). They tested pairs trading strategy on Wall Street with daily data over the period 1962 through 1997. According to Gatev *et al.* (1998), the growing popularity of the pairs trading strategy may also pose a problem because the opportunities to trade become much smaller, as many other arbitrageurs are aware of the strategy and may choose to enter at an earlier point of deviation from the equilibrium. Profit from the pairs trading strategy in recent years has been less than it was when pairs trading strategy was first adopted. However, Gillespie and Ulph (2001), Habak (2002), Hong and Susmel (2003) and Do and Faff (2008) showed that significant positive returns could still be made. Andrade *et al.* (2005) and Engelberg *et al.* (2008) followed the pairs trading strategy of Gatev *et al.* (1998) to understand and describe the profitability of pairs trading. An extensive discussion of pairs trading can be found in Gatev *et al.* (2006), Vidyamurthy (2004), Whistler (2004) and Ehrman (2006).

There are four main approaches used to implement pairs trading, namely distance approach, combine forecasts approach, stochastic approach and cointegration approach.

Papers categorised in distance approach include Gatev *et al.* (1998, 2006), Nath (2003), Andrade *et al.* (2005), Engelberg *et al.* (2008), Papandakis and Wisocky (2008) and Do and Faff (2008). Under this approach, the co-movement in a pair is measured by what is referred to as the distance, or the sum of squared differences between the two normalized price series. The distance approach purely exploits a statistical relationship between a pair of securities. As Do *et al.* (2006) noted, it is model-free and consequently, it has the advantage of not being exposed to model mis-specification and mis-estimation. However, this non-parametric approach lacks forecasting ability regarding the convergence time or expected holding period. What is a more fundamental issue is its underlying assumption that its price level distance is static through time, or equivalently, that the returns of the two stocks are in parity. Although such an assumption may be valid in short periods of time, it is only so for a certain group of pairs whose risk-return profiles are close to identical.

Combine forecasts approach is promoted by Huck (2009, 2010). This framework can be seen as a sort of combined forecast specially designed for pairs trading. In brief, the method is based on three phases: forecasting, ranking and trading. This framework differs from the others on very essential point that it is developed without reference to any equilibrium model. Huck (2009, 2010) argued that the method provides much more trading possibilities and could detect the "birth" of the divergence that the other approaches cannot consider. However, there is no other research follows this approach so far, may be due to complexity involving a matrix of size n, where n is the number of stocks in a stock market considered.

Papers categorised in stochastic approach include Elliott *et al.* (2005), Do *et al.* (2006), Rampertshammer (2007), Mudchanatongsuk *et al.* (2008) and Bertram (2009). This approach explicitly model the mean reverting behaviour of the spread (the difference between the two stock prices[1]) in a continuous time setting. These pairs trading based on stochastic approaches relied on an assumption that the spread follows an Ornstein-Uhlenbeck process which actually is an AR(1) process in a continuous term. However, they did not mention how to choose trading pairs so that the spread can follow an Ornstein-Uhlenbeck process. Furthermore, the mean reverting property of the spread can not be justified. Generally, the long term mean of spread should not be constant, but widens as they go up and narrows as they go down, unless the stocks have similar price points or those stock prices are cointegrated with cointegration coefficient 1.

Studies categorised in cointegration approach include Lin *et al.* (2006), Vidyamurthy (2004), Gillespie and Ulph (2001) and Hong and Susmel (2003). Alexander and Dimitriu (2002) and Galenko *et al.* (2007) used cointegration in long/short investment strategy

[1]Mudchanatongsuk *et al.* (2008) and Bertram (2009) defined "spread" as the difference between the log of two stock prices while Do *et al.* (2006) defined "spread" as a function of the log of two stock prices plus Γ which is a vector of exposure differentials

involving more than two stocks. Cointegration incorporates mean reversion into pairs trading framework which is the single most important statistical relationship required for success. If the value of the portfolio is known to fluctuate around its equilibrium value then any deviations from this value can be traded against.

The difference between Lin *et al.* (2006) and other pairs trading cointegration approach studies is that Lin *et al.* (2006) developed a pairs trading strategy based on the cointegration coefficients weighted (CCW) rule. The CCW rule works by trading the number of stocks based on their cointegration coefficients to achieve a guaranteed minimum profit per trade. The minimum profit per trade corresponds to the pre-set boundaries upper-bound U and lower-bound L chosen to open trades. Lin *et al.* (2006) is briefly introduced below:

Consider two stocks $S1$ and $S2$ whose prices are I(1), i.e., stock prices series are not stationary but the first difference of the stock prices series are stationary. If the stock prices $P_{S1,t}$ and $P_{S2,t}$ are cointegrated, there exists cointegration coefficients 1 and β corresponding to $P_{S1,t}$ and $P_{S2,t}$ respectively, such that a cointegration relationship can be constructed as follows:

$$P_{S1,t} - \beta P_{S2,t} = \mu + \epsilon_t, \tag{1.1}$$

where ϵ_t is a stationary time series called cointegration errors and μ is the mean of cointegration relationship[2].

Let N_{S1} and N_{S2} denote the number of stocks $S1$ and $S2$ respectively. Two type of trades, upper-trades (U-trades) and lower-trades (L-trades), are considered. For a U-trade, a trade is opened, when the cointegration error is higher than or equal to the pre-set upper-bound U, by selling N_{S1} of stocks in $S1$ and buying N_{S2} of stocks in $S2$ and then closing the trade when the cointegration error is less than or equal to zero. This is done by buying N_{S1} of stocks in $S1$ and selling N_{S2} of stocks in $S2$. The opposite happens for L-trades. The CCW rule choose $N_{S1} = 1$ and $N_{S2} = \beta$, respectively, so that the profit P for a U-trade will be:

$$\begin{aligned} P &= N_{S2}(P_{S2,t_c} - P_{S2,t_o}) + N_{S1}(P_{S1,t_o} - P_{S1,t_c}) \\ &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [P_{S1,t_o} - P_{S1,t_c}] \\ &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [(\epsilon_{t_o} + \mu) + \beta P_{S2,t_o}) - (\epsilon_{t_c} + \mu) + \beta P_{S2,t_c})] \\ &= (\epsilon_{t_o} - \epsilon_{t_c}) \geq U, \end{aligned} \tag{1.2}$$

where t_o and t_c are times for opening and closing the pair trade.

It is assumed that the cointegration error (ϵ_t) is a stationary process and has a symmetric distribution, so the lengths from the upper-bound U to the mean and from the lower-bound L to the mean are the same. Thus, the lower-bound can be set as $L = -U$. Similarly for lower-trades, choosing $N_{S1} = 1$ and $N_{S2} = \beta$, the profit will be

[2]Other papers such as Schmidt (2008) calls ϵ_t as equilibrium error.

$P \geq -L = U$. For details, see Lin *et al.* (2006) and Chapter 2. A question rises how do we choose the upper-bound U? Chapters 3 and 4 in this thesis address that question.

Vidyamurthy (2004) also suggested a pairs trading strategy with similar fashion to the CCW rule. However, he used log prices instead of level prices. To test for cointegration, he adopted the Engle and Granger's 2 step approach (Engle and Granger, 1987). He also addressed the need to estimate number of trades to choose the thresholds maximising profit function. Assuming the cointegration residuals as an ARMA process, Rice's formula (Rice, 1945) was used to calculate the rate of thresholds crossing as an estimate the number of trades. Using Rice's formula to estimate the number of trades is not quite correct because it does not give a restriction when the pair trade will be closed. Instead of using Rice's formula, the first passage-time for stationary series should be used. The first passage-time for stationary series calculates the time needed for the series when it crosses the threshold and then come back to the mean series for the first time.

The following terms are required in further discussion. *Trade duration* is the time between opening and closing a U-trade (an L-trade). *Inter-trade interval* is the time between two consecutive U-trades (L-trades) or the time between closing a U-trade(an L-trade) and then opening the next U-trade(L-trade). We assume that there is no open trade (neither U-trade nor L-trade) if the previous trade has not been closed yet. *Period* is the sum of the trade duration and the inter-trade interval for U-trades (L-trades).

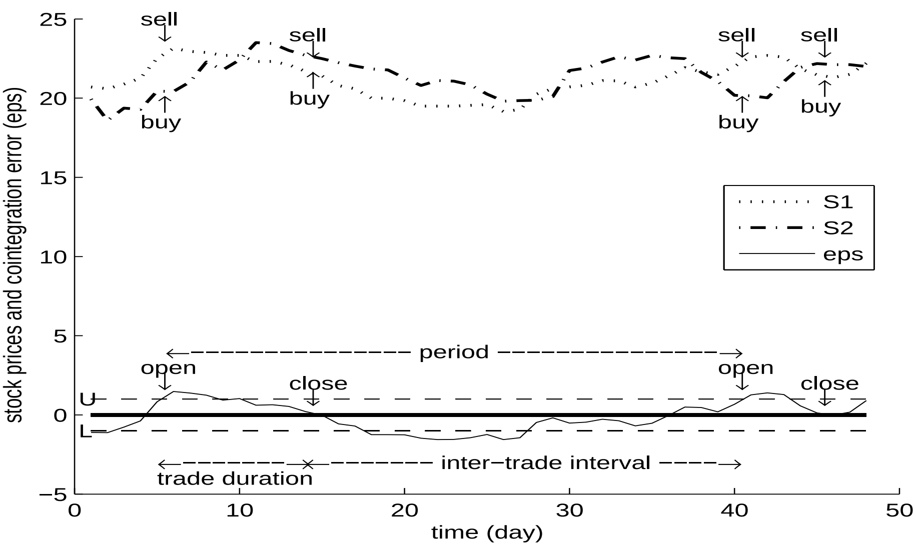


Figure 1.1: Example of two cointegrated shares ($S1$ and $S2$) with E(eps)=0.

Figure 1.1 shows two cointegrated shares called stock $S1$ and $S2$ and their cointegration error denoted by *eps* . At time t = 5, the cointegration error of the two stocks (*eps*) is higher than the upper-bound U, so a trade is opened by selling $S1$ and buying $S2$. At t = 14, $eps \ < 0$, so the trade is closed by taking the opposite position. Figure 1.1 also illustrates an example of trade duration, inter-trade interval and period for U-trades.

The optimal pre-set boundary values is determined by maximising the minimum total profit (MTP) over a specified trading horizon. The MTP is a function of the minimum profit per trade and the number of trades during the trading horizon. As the derivation of the pre-set minimum profit per trade is already provided in Lin *et al.* (2006), this thesis provides the estimated number of trades. The number of trades is also influenced by the distance of the pre-set boundaries from the long-run cointegration equilibrium. The higher the pre-set boundaries for opening trades, the higher the minimum profit per trade but the trade numbers will be lower. The opposite applies for lowering the boundary values. The number of trades over a specified trading horizon is estimated jointly by the average trade duration and the average inter-trade interval. For any pre-set boundaries, both of those values are estimated by making an analogy to the mean first-passage times for a stationary process.

In Chapter 3, trading duration and inter-trade interval are derived using a Markov chain approach for a white noise and an AR(1) processes. This approach has a disadvantage in determining the probability of the process from one state to another state. For a simple white noise and an AR(1) processes, it does not have any problems in determining the probability. However, to apply the approach to higher order AR(p) models, $p > 1$, is difficult. Thus, in Chapter 4, an integral equation approach in Basak and Ho (2004) is used to evaluate trading duration and inter-trade interval. In Chapter 4, a numerical algorithm is also developed to calculate the optimal pre-set upper-bound, denoted U_o, that would maximize the minimum total profit (MTP). Then, Chapter 5 provides empirical examples using pairs trading strategy outlined in Chapter 4.

Chapters 3–5 explain pairs trading strategies assuming that the cointegration error ϵ_t in (1.1) is a linear stationary AR(p) process. However, ϵ_t in (1.1) may not follow a linear stationary AR(p) process but a nonlinear stationary process. For example, Monoyios and Sarno (2002) concluded that there is a cointegration relationship between the log of future contract prices and the log of spot index S&P 500 and FTSE 100 and the basis [3] of the indexes follow a nonlinear stationary ESTAR (Exponential Smooth Transition Autoregressive) model [4].

Current developments in nonlinear unit root tests have raised as the standard linear unit root tests such as augmented Dickey-Fuller (ADF) (see Fuller, 1976) test cannot meet requirement for nonlinear processes. For example, Pippenger and Goering (1993) showed how the power of the standard DF tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. An STAR (Smooth Transition

[3]The difference between the log of future contract prices and the log of spot index prices.

[4]This paper has became the idea to develop a unit root test for an ESTAR model in this chapter, even though in Chapter 8 we show that their conclusion may not be true.

Autoregressive) model has become a popular model to analyse economic and finance data due to its generality and flexibility. Nonlinear adjustment in an STAR model allows for smooth rather than discrete adjustment in a TAR model. In a STAR model, adjustment takes place in every period but the speed adjustment varies with the extent of the deviation from equilibrium. A STAR(p) model can be expressed as follow:

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y_{t-j} \right] G(\theta, e, y_{t-d}) + \epsilon_t \tag{1.3}$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d \geq 1$ is a delay parameter; $(\theta, e) \in \{\mathbb{R}^+ \times \mathbb{R}\}$ where $\mathbb{R}$ denotes the real space $(-\infty, \infty)$ and $\mathbb{R}^+$ denotes the nonnegative real space $(0, \infty)$. The transition function $G(\theta, e, y_{t-d})$ determines the speed of adjustment to an equilibrium e. Two simple transition functions suggested by Granger and Terasvirta (1993) and Terasvirta (1994) are the logistic and exponential functions:

$$G(\theta, e, y_{t-d}) = \frac{1}{1 + \exp\{-\theta(y_{t-d} - e)\}} - \frac{1}{2}, \tag{1.4}$$

$$G(\theta, e, y_{t-d}) = 1 - \exp\{-\theta^2 (y_{t-d} - e)^2\}. \tag{1.5}$$

If the transition function $G(\theta, e, y_{t-d})$ is given by (1.4), (1.3) is called a logistic smooth transition autoregressive (LSTAR) model. If the transition function $G(\theta, e, y_{t-d})$ is given by (1.5), (1.3) is called an exponential smooth transition autoregressive (ESTAR) model.

An ESTAR model is more popular than a LSTAR model as it has a symmetric adjustment of the process for deviation above and below the equilibrium. Michael *et al.* (1997), Taylor *et al.* (2001) and Paya *et al.* (2003) used an ESTAR model to analyse real exchange rate and purchasing power parity (PPP) deviations. Terasvirta and Elliasson (2001), and Sarno *et al.* (2002) used an ESTAR model to analyse deviations from optimal money holding.

Kapetanios *et al.* (2003) considered a unit root test for an ESTAR(1) model and applied their test to real interest rates and rejected the null hypothesis for several interest rates considered, whereas Augmented Dickey-Fuller (ADF) tests failed to do so.

Venetis *et al.* (2009) developed a unit root test for k-ESTAR(p) model where k is the number of equilibrium levels and p denotes the order of the autoregressive lags. Thus, instead of (1.5) as a transition function, in their model, the transition function involves k equilibriums :

$$G(\theta, \mathbf{e}, y_{t-d}) = 1 - \exp\left[-\theta^2 \left(\prod_{i=1}^{k}(y_{t-d} - e_i)\right)^2\right] \tag{1.6}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_k)'$.

As noted by Venetis *et al.* (2009), many economic theories support the existence of multiple equilibria. However, even though Venetis *et al.* (2009) developed a more general form of ESTAR(p) model but their approach might cause singularity problem because some of the regressors might be collinear. To overcome the problem, they added the collinear regressors into the error term. Even though the test under alternative is consistent, but it may make a significant difference for some cases. Therefore, Chapter 6 in this thesis extends the work of Kapetanios *et al.* (2003) by considering a unit root test for a k-ESTAR(p) model with a different approach to Venetis *et al.* (2009). By using the approach in this thesis, the singularity problem can be avoided without adding the collinear regressors into the error term. For some cases, simulation results show that our approach is better than that of Venetis *et al.* (2009), Kapetanios *et al.* (2003) and the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979, 1981). The result from Chapter 6 can be used to analyse a possibility of nonlinear stationary ESTAR model against a linear unit root AR(p) model for other economics or finance data.

In Chapter 7, we show that the mean first passage-time based on an integral equation approach in Basak and Ho (2004) can also be applied to the evaluation of trading duration and inter-trade interval for an 1-ESTAR(1) model and an 1-ESTAR(2) model. With some adjustments, a numerical algorithm in Chapter 4 can also be used to calculate the optimal pre-set upper-bound, denoted U_o, that would maximize the minimum total profit (MTP) for an ESTAR model. An application to pairs trading assuming ϵ_t in (1.1) is a nonlinear stationary 1-ESTAR(2) model is examined in Chapter 7 using simulated data. As far as we know, no one has examined the mean first passage-time for an ESTAR model.

Chapter 8 focuses on cointegration and pairs trading between future prices and spot index prices of the S&P 500. There are three major issues discussed in this chapter.

The first issue in Chapter 8 is to identify an appropriate model for the S&P 500 basis. The S&P 500 basis is defined as the difference between the log of future prices and the log of spot index prices. Monoyios and Sarno (2002) concluded that the basis of S&P 500 follows an ESTAR model. Following the same analysis procedures, however, we have found different conclusions to Monoyios and Sarno (2002) using currently available data period. Instead of concluding the basis follows an ESTAR model, we conclude that it follows a LSTAR model. Furthermore, even though we can conclude that there is possibility nonlinearity in the basis, there is no significant difference between a nonlinear LSTAR model and a linear autoregressive model in fitting the data.

We also have a concern in the way the basis is constructed. By pairing up the spot price with the future contract with the nearest maturity, it may produce artificial jumps at the time of maturity. The longer the time to maturity, the higher the difference between the log future price and the log spot price as described by a cost-carry model. For example for S&P 500, it has 4 maturity times during a year which are the third Friday in March, June, September and December. We find that at those times, there are jumps in the basis. Monoyios and Sarno (2002) did not discuss this issue. Other papers such as Chowdhury (1991), Koutmos and Tucker (1996), Sarno and Valente (2000) and Kenourgios and Samitas (2008) which analysed the relationship between future and spot index prices also did not address this issue. McMillan and Speight (2005) gave an argument that they did not take into account a time-to-maturity variable because the effects of the cost-carry on the basis can be nullified as stocks go ex-dividend and interest is paid overnight. As such, the mean of the basis becomes zero. However, Ma *et al.* (1992) argued that the jumps may create volatility and bias in the parameter estimates. Kinlay and Ramaswamy (1988) using 15-minutes data of S&P 500 concluded that a time-to-maturity variable is significant when it is regressed with the misspricing future price series. They defined that misspricing is the difference between future price and its theoretical price. Other papers such as Milonas (1986) and Barnhill and Jordan (1987) also supported a maturity effect in the prices of futures financial assets contracts. Therefore, the second issue in Chapter 8 is about the examination of the cointegration between the future and spot index prices with a time trend (a time-to-maturity) for each future contract during 1998-2008. Cointegration analysis with a time trend based on the Engle-Granger approach and the Johansen approach are used. Only 19 out of 44 future contracts conclude that they are cointegrated with a time trend using both the Engle-Granger approach and the Johansen approach. Perhaps, high volatility during financial crisis in the data period affects the cointegration test and also limited number of observations, i.e. we only have maximum 2 years data for a future contract. Based on the Engle-Granger approach and then running nonlinearity tests for the regression residuals, we can say that there is no strong evidence of nonlinearity in the relationship between future and spot prices of S&P 500 index. It gives further evidence on the basis for a linearity rather than nonlinearity.

The third issue in Chapter 8 is to examine the application of pairs trading between the future and spot index prices. The stationary characteristics of an AR(p) model are used to develop pairs trading strategy between future contract and corresponding spot index. Pairs trading simulations between future contract and CFD (Contract For Difference) the S&P 500 index data during 1998 show very good results resulting a total return more than 75% during the 3 months trading periods. The pairs trading strategy involving a financial instrument called CFD is relatively a new idea. The use of CFD index as a substitute of the spot index price will avoid the need to trade hundreds of stocks in the index and large amount of money. Thus, it can be used for individual investors wanting to speculate in index markets. Papers such as Butterworth and Holmes

(2001), Holmes (1995) and Figlewski (1984b, 1985) analysed hedging performances of future index for a broadly diversified portfolio of stocks in the index. Essentially, those hedging strategies are used by management funds to lock-up the profit from the portfolios composing stocks in the index. Other papers, such as Cornell and French (1983), Figlewski (1984a), Modest and Sundaresan (1983) and Kinlay and Ramaswamy (1988) attempted to identify and measure the arbitrage trading boundaries between future and spot index prices. Results indicated that the basis (log future prices - log spot index prices) normally falls within boundaries determined by financing costs adjusted for dividend uncertainty, transactions costs and taxes. Stock index arbitragers compare the basis with net financing and transaction costs. While the basis normally falls within the no-arbitrage boundaries, it occasionally moves outside that range. When it exceeds the upper-bound, traders buy stocks and sell futures. They reverse the transactions later when the basis approaches zero at expiration. The opposite occurs when the basis falls below the lower-bound. As Kawaller *et al.* (1987) noted, those strategies generally involve at least three hundred stocks in the index and a minimum $5 million investment. Thus, those are not suitable for individual investors. Furthermore, none of those papers gave empirical trading simulation results.

Chapter 9 gives conclusion summary of this thesis and also suggestions for further research. The results from this thesis will extend literature in pairs trading strategy and can be beneficial for practitioners in financial markets who want to apply pairs trading strategy.

Results from this thesis have been presented and published or will be submitted to refereed journals, i.e.:

[1] *Trade Duration and Inter-trade Interval Using Markov Chain Approach for Cointegration Pairs Trading Strategy,* December 2004, Seminars in Mathematics Finance, University of Wollongong, Wollongong, Australia.

[2] *Optimal Pairs Trading Strategy Based on Cointegration Technique,* December 2005, The 15th Australasian Finance and Banking Conference, University of New South Wales, Sydney, Australia.

[3] *Finding the Optimal Pre-set Boundaries for Pairs Trading Strategy Based on Cointegration Technique,* 2009, Statistics Working Paper Series, No. 21-09, Centre for Statistical and Survey Methodology, University of Wollongong, Australia.

[4] *Finding the Optimal Pre-set Boundaries for Pairs Trading Strategy Based on Cointegration Technique,* 2010, Journal of Statistical Theory and Practice, Vol. **4**, No. 3, Page 391–419.

[5] *A Unit Root Test for Nonlinear ESTAR(2) Process,* July 2009, Society for Computational Economics, 15th International Conference on Computing in Economics and Finance, University Technology of Sydney, Sydney, Australia.

[6] *A Unit Root Test for a k-ESTAR(p) Model* will be submitted to a journal.

[7] *The Mean First Passage-time for an ESTAR Model Using an Integral Equation Approach and Application in Pairs Trading Strategy* will be submitted to a journal.

[8] *Is the Basis of a Stock Index Futures Market Nonlinear?* 17-18 February 2011, proceeding the Fourth ASEARC Conference, University of Western Sydney, Sydney, Australia.

[9] *Is the Basis of a Stock Index Futures Market Nonlinear?* will be submitted to a journal.

[10] *Pairs Trading between Future and Spot Index Based on Cointegration Approach,* 14-15 July 2011, the Young Statisticians Conference, University of Queensland, Brisbane, Australia.

[11] *Pairs Trading between Future and Spot Index Based on Cointegration Approach* will be submitted to a journal.

In this thesis, [1] is part of Chapter 3; [2]–[4] are parts of Chapters 4&5; [5]–[6] are parts of Chapter 6; [7] is part of Chapter 7 and [8]–[11] are parts of Chapter 8.

Chapter 2

Preliminaries and Literature Review

In this chapter, we will briefly introduce basic concepts used in this thesis and review literature concerning pairs trading strategy.

2.1 Preliminary Concepts

2.1.1 An Introduction to Cointegration

Cointegration was first attributed to the work of Engle and Granger (1987), which earned them the Nobel Prize 2003 for economics. Cointegration has since found many applications in macroeconomic analysis and more recently it has played an increasingly prominent role in funds management and portfolio construction. The statistical properties of cointegration make it such an attractive possibility across a range of applications for academics and practitioners alike.

Consider a set of economic variables $\{y_{i,t}\}$, $i = 1, \ldots, p$, in long-run equilibrium when

$$\beta_1 \, y_{1,t} + \beta_2 \, y_{2,t} + \ldots + \beta_p \, y_{p,t} = \mu + \epsilon_t \tag{2.1}$$

where p is number of variables in the cointegration equation, μ is the long-run equilibrium and ϵ_t is the cointegration error.

For notational simplicity, (2.1) can be represented in matrix form as

$$\boldsymbol{\beta}' \mathbf{y}_t = \mu + \epsilon_t \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p \ )'$ and $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{p,t})'$.

The cointegration error is the deviation from the long-run equilibrium and can be represented by

$$\epsilon_t = \boldsymbol{\beta}' \mathbf{y}_t - \mu. \tag{2.3}$$

The equilibrium is only meaningful if the residual series or cointegration error ϵ_t is stationary.

Stationarity and Linear Unit Root Tests

A time series $\{\epsilon_t\}$ is a weakly stationary series if its mean, variance and autocovariances depend only on the time lag $|t-s|$ and not on time points t and s. This can be summarised by the following conditions:

(i) $E(\epsilon_t) = E(\epsilon_{t-s}) = \mu$

(ii) $E(\epsilon_t - \mu)^2 = E(\epsilon_{t-s} - \mu)^2 = \sigma_\epsilon^2$

(iii) $E(\epsilon_t - \mu)(\epsilon_{t-s} - \mu) = E(\epsilon_{t-j} - \mu)(\epsilon_{t-j-s} - \mu) = \delta_s$

where μ, σ_ϵ^2, s, and δ_s are all constants.

Financial time series such as stock prices sometimes can be described as a random walk process which is a non-stationary process with a unit root. There are several ways to test whether the series is stationary or non-stationary with a unit root. The famous one is the Dickey-Fuller (DF) test (Dickey and Fuller, 1979, Fuller, 1976). It tests the null hypothesis that a series does contain a unit root (i.e., it is non-stationary) against the alternative of stationary. There are other tests, such as CRDW test (Sargan and Bhargava, 1983) based on the usual Durbin-Watson statistic; and the non-parametric tests developed by Phillips and Perron based on the Z-test (Phillips and Perron, 1988a), which involves transforming test statistic to eliminate autocorrelation in the model. DF test tends to be more popular due to its simplicity and its more general nature. The augmented Dickey-Fuller (ADF) test is used if the series ϵ_t follows an AR(p) process, $p > 1$, to eliminate autocorrelation in the residuals. Thus, assuming that ϵ_t follows a pth order autoregressive process, the ADF test is:

$$\begin{aligned} \epsilon_t &= \theta_1\, \epsilon_{t-1} + \theta_2\, \epsilon_{t-2} + \ldots + \theta_p\, \epsilon_{t-p} + \mu + \gamma\, t + \eta_t \\ \Delta\epsilon_t &= \theta\, \epsilon_{t-1} + \sum_{j=1}^{(p-1)} \theta_j^*\, \Delta\epsilon_{t-j} + \mu + \gamma\, t + \eta_t \end{aligned} \tag{2.4}$$

where μ is a constant, t is a time trend, $\theta = (\theta_1 + \ldots + \theta_p) - 1$, $\theta_j^* = -\sum_{i=j+1}^{p} \theta_i$, for $j = 1, \ldots, (p-1)$ and η_t is the error term. If $\theta = 0$ against alternative $\theta < 0$, then ϵ_t contains a unit root. To test the null hypothesis, a t-statistic is calculated. However, under non-stationary, the statistic does not have a standard t-distribution. So that critical values for the statistic are computed using Monte Carlo techniques. The critical values also depend on whether the deterministic components (i.e., a constant μ and/or time trend t) are included in (2.4).

A Definition of Cointegration

We say that components of vector $\mathbf{y}_t$ in (2.2) are cointegrated of order (d, b) which is denoted by $\mathbf{y}_t \sim CI(d, b)$ if:

(1) All components of $\mathbf{y}_t$ are integrated of order d. [1]

(2) There exists a vector $\boldsymbol{\beta}$ such that the linear combination $\boldsymbol{\beta}'\mathbf{y}_t$ is integrated of order $(d - b)$, where $b > 0$ and $\boldsymbol{\beta}$ is the cointegrating vector (CV) for $\mathbf{y}_t$.

Important aspects of cointegration:

(1) Cointegration refers to a linear combination of non-stationary variables. The CV is not unique. For example, if $(\beta_1, \beta_2, \ldots, \beta_p)'$ is a CV, then for a non-zero λ, $(\lambda\beta_1, \lambda\beta_2, \ldots, \lambda\beta_p)'$ is also a CV. Typically the CV is normalised with respect to $y_{1,t}$ by selecting $\lambda = 1/\beta_1$.

(2) All variables must be integrated of the same order. This is a prior condition for the presence of a cointegrating relationship. The inverse is not true which means that all similarly integrated variables do not imply that they are cointegrated.

(3) If the vector $\mathbf{y}_t$ has p components, there may be as many as $(p - 1)$ linearly cointegrating vectors. For example, if $p = 3$ then there can be at most two independent CVs.

Testing for Cointegration: The Engle-Granger Approach

Consider a simple model

$$P_{S1,t} = \beta P_{S2,t} + \mu + \epsilon_t \tag{2.5}$$

where $P_{S1,t}$ and $P_{S2,t}$ denote the prices of asset $S1$ and $S2$, respectively.

Harris (1995) noted that estimating (2.5) using OLS achieves a consistent estimate of the long-run steady-state relationship between the variables in the model, and all dynamics and endogeneity issues can be ignored asymptotically. This arises because of what is termed the "superconsistency" property of the OLS estimator when the series are cointegrated. According to the "superconsistency" property, if $P_{S1,t}$ and $P_{S2,t}$ are both non-stationary $I(1)$ variables and $\epsilon_t \sim I(0)$, then as sample size, T, becomes larger the OLS parameter estimates of (2.5) converges to their true values at a much faster rate than the usual OLS estimator with stationary $I(0)$ variables. However, in finite samples, Phillips and Durlauf (1986) conclude that the asymptotic distribution of the OLS estimators and associated t-statistics are highly complicated and non-normal and thus the standard tests are invalid. The next step of Engle-Granger approach is to test whether the residuals $\epsilon_t \sim I(1)$ against $\epsilon_t \sim I(0)$. Engle and Granger (1987) suggested the augmented Dickey-Fuller (ADF) test on the estimated residuals as follow:

$$\Delta\hat{\epsilon}_t = \rho\,\hat{\epsilon}_{t-1} + \sum_{i=1}^{p-1} \omega_i\,\Delta\hat{\epsilon}_{t-i} + \upsilon_t \tag{2.6}$$

[1] A series is integrated of order d if it must be differenced d times in order to become stationary. A stationary series by definition is an I(0) process.

where $\upsilon_t \sim$ IID $N(0, \sigma^2)$ and $\{\hat{\epsilon}_t\}$ are obtained from estimating (2.5).

The null hypothesis of nonstationary (i.e., the series has a unit root) and thus no cointegration, $\rho = 0$, can be tested using a t-statistic with a non-normal distribution. The lag length of the augmentation terms, p, is chosen as the minimum necessary to reduce the residuals to white noise. Critical values for this test have been calculated using Monte Carlo methods available in Fuller (1976). However, unless the true parameter values in (2.5) are known, it is not possible to use the standard Dickey-Fuller tables of critical values. There are two reasons for this. First, the ADF tests will tend to over-reject the null because the OLS estimator "chooses" the residuals in (2.5) to have the smallest sample variance. Even if the variables are not cointegrated, the OLS makes the residuals appear as stationary as possible. Second, the distribution of the test statistic under the null is affected by the number of regressors included in (2.5). Thus, different critical values are needed as changes and also whether constant and/or trend are included along with the sample size. Taking into account all of these, MacKinnon (1992) linked the critical values for particular tests to a set of parameters of a equation of the response surfaces. As the number of regressor in (2.5) is one (excluding the constant and trend), the critical values by MacKinnon (1992) will be the same as the ADF tests.

Other alternatives to test a unit root are the Z-tests suggested by Phillips (1987), Perron (1988) and Phillips and Perron (1988b). Monte Carlo results in Schwert (1989) suggest that these tests have poor size properties (i.e., the tendency to over-reject the null when it is true) when MA terms present in underlying time series. Ng and Perron (2001) developed a unit root test claimed robust to ARIMA representations with large negative MA term and also improves the power of standard ADF tests by using a GLS detrending/demeaning procedure. However, Crowder and Phengpis (2005) concluded that there are no significant difference of results from the ADF tests and Ng and Perron (2001) when they apply the tests to the S&P 500 future and spot index prices. Therefore, in this thesis only the ADF test is used.

Testing for Cointegration: The Johansen Approach

Johansen (1988, 1991) developed cointegration analysis based on a multivariate system. Compared to Engle-Granger approach, it has an advantage that it analyses the cointegration relationship between variables simultaneously in one system. Especially if we have more than 2 variables, the Engle-Granger approach which involves one equation will be a problem. Consider a vector autoregressive (VAR) process for $\mathbf{y}_t$ as follow:

$$\mathbf{y}_t = \sum_{j=1}^{k} \Phi_j \ \mathbf{y}_{t-j} + \boldsymbol{\epsilon}_t \tag{2.7}$$

where $\mathbf{y}_t$ is p-dimentional vector of variables integrated with the same order; Φ_j are

$(p \times p)$ coefficient matrices; a positive integer k is chosen as the minimum necessary to reduce the residuals $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{p,t})'$ to be a white noise error vector with a covariance matrix Ω.

The Johansen tests are calculated from the transformed version of (2.7) into a vector error-correction model (VECM) as follow:

$$\Delta \mathbf{y}_t = \Pi\, \mathbf{y}_{t-1} + \sum_{j=1}^{k-1} \Gamma_j\, \Delta \mathbf{y}_{t-j} + \boldsymbol{\epsilon}_t \tag{2.8}$$

where $\Gamma_j = \sum_{i=1}^{j} \Phi_i - I$ and $\Pi = \sum_{j=1}^{k} \Phi_j - I$.

The long-run multiplier matrix Π can be decomposed into two $(p \times r)$ matrices such that $\Pi = \boldsymbol{\alpha}\, \boldsymbol{\beta}'$. The $(p \times r)$ matrix $\boldsymbol{\beta}$ represents the cointegration vectors or the long-run equilibria of the system equations. The $(p \times r)$ matrix $\boldsymbol{\alpha}$ is the matrix of error-correction coefficients which measure the rate each variable adjusts to the long-run equilibrium.

Maximum likelihood estimation of (2.8) can be carried out using a reduced rank regression. Johansen (1988, 1991) suggested first to concentrate on the short-run dynamics by regressing $\Delta \mathbf{y}_t$ and $\mathbf{y}_{t-1}$ on $\Delta \mathbf{y}_{t-1}, \Delta \mathbf{y}_{t-2}, \cdots, \Delta \mathbf{y}_{t-k+1}$, and save the residuals as R_{0t} and R_{1t}, respectively. Next, calculate the residuals (product moment) matrix $S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}$ and find eigenvalues and eigenvectors for $S_{10} S'_{00} S_{01}$. Then order the eigenvalues from the largest to the smallest so that $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_p)'$ and the corresponding eigenvectors $\hat{\boldsymbol{V}} = (\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_p)'$. The maximum likelihood estimate of β is obtained as the eigenvectors corresponding to the r largest eigenvalues $(0 \leq r \leq p)$. To test the null hypothesis that there are at most r cointegration vectors, Johansen (1988, 1991) suggested two measurements. The first is called as the trace statistic. It imposes a restriction on different values of r and then the log of the maximised likelihood function for the restricted model is compared to the log of the maximised likelihood function of the unrestricted model and a standard likelihood ratio test computed.

$$\lambda_{trace} = -2 \log(Q) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i), \qquad r = 0, 1, \cdots, (p-1) \tag{2.9}$$

where Q is restricted maximised likelihood divided by unrestricted maximised likelihood.

The second measurement is to test the significance of the largest λ_r. Thus it is called as maximal eigenvalue or λ_{max} statistic. This statistic is testing that there are r cointegrating vectors against the alternative that $r + 1$ exist.

$$\lambda_{max} = -T \log(1 - \hat{\lambda}_{r+1}), \qquad r = 0, 1, \cdots, (p-1). \tag{2.10}$$

Reimers (1992) showed that in small sample situations, the Johansen procedure over-rejects when the null is true. Thus, he suggested taking into account of the number of parameters to be estimated in the model and making an adjustment for the degrees of freedom by replacing T in (2.9) and (2.10) by $T - pk$, where T is the sample size, p is the number variables in the model and k is the lag-length set when estimating (2.8).

2.1.2 Nonlinear Models and Non-linear Unit Root

Threshold Autoregressive (TAR) Model

The TAR model was introduced by Tong (1978) and has since become quite popular in nonlinear time series. Inference theories about the model have been developed. For examples, Chan (1991) and Hansen (1996) described the asymptotic distribution of the likelihood ratio test for the model, Chan (1993) showed that the least squares estimate of the threshold is super-consistent and found its asymptotic distribution, Chen and Lee (1995) used Bayesian approach to developed the asymptotic distribution while Hansen (1997, 2000) developed an alternative approximation to the asymptotic distribution, and Chan and Tsay (1998) analysed the related continuous TAR model and found the asymptotic distribution of the parameter estimates in this model.

Since the introduction of Dickey-Fuller (DF) unit root tests in Fuller (1976) and then Dickey and Fuller (1979, 1981), many new types of unit root tests have been developed. Developments in nonlinear unit root tests raised as the standard linear unit root tests cannot meet requirement for nonlinear processes. For example, Pippenger and Goering (1993) showed how the power of the standard DF tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. Other researchers have attempted to address similar issues in the context of a TAR model, for examples, Balke and Fomby (1997), Enders and Granger (1998), Berben and van Dijk (1999), Caner and Hansen (2001) and Lo and Zivot (2001).

The TAR model can be described as follow:

$$\Delta y_t = \boldsymbol{\theta}_1' \; \mathbf{x}_{t-1} \; 1_{Z_{t-1}<\lambda} + \boldsymbol{\theta}_2' \; \mathbf{x}_{t-1} \; 1_{Z_{t-1}\geq\lambda} + \epsilon_t, \tag{2.11}$$

$t = 1, \ldots, T$ where $\mathbf{x}_{t-1} = (y_{t-1}, \mathbf{r}_t', \Delta y_{t-1}, \ldots, \Delta y_{t-k})'$, $1_{(.)}$ is the indicator function, ϵ_t is an IID error, $Z_{t-1} = y_t - y_{t-1}$ and $\mathbf{r}_t$ is a vector of deterministic components including an intercept and possibly a linear time trend.

The threshold λ is unknown. It takes on values in the interval $\bigwedge = [\lambda_1, \lambda_2]$ where λ_1 and λ_2 are picked so that $P(Z_t \leq \lambda_1) = \phi_1 > 0$ and $P(Z_t \leq \lambda_2) = \phi_2 < 1$. It is typical to treat ϕ_1 and ϕ_2 symmetrically so that $\phi_2 = 1 - \phi_1$, which imposes the restriction that no "regime" has less than $\phi_1 \times 100\%$ of the total sample. The particular choice for ϕ_1 is somewhat arbitrary and in practice must be guided by the consideration that each "regime" needs to have sufficient observations to adequately identify the regression parameters.

The threshold function Z_{t-1} should be predetermined, strictly stationary and ergodic with a continuous distribution function. The choice of $Z_{t-1} = y_t - y_{t-1}$ is convenient because it is ensured to be stationary under alternative assumptions that y_t is $I(1)$.

For some analysis such as a unit root test, it will be convenient to separately discuss the component of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, i.e.:

$$\boldsymbol{\theta}_1 = \begin{pmatrix} \rho_1 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_1 \end{pmatrix}, \qquad \boldsymbol{\theta}_2 = \begin{pmatrix} \rho_2 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\alpha}_2 \end{pmatrix},$$

where ρ_1 and ρ_2 are scalars, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have the same dimension as $\mathbf{r}_t$, and $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are k-vectors. Equation (2.12) can be arranged as follow:

$$\begin{aligned} \Delta y_t &= (\rho_1 \, 1_{Z_{t-1}<\lambda} + \rho_2 \, 1_{Z_{t-1}\geq\lambda}) \, y_{t-1} + (\boldsymbol{\beta}_1^{'} \, 1_{Z_{t-1}<\lambda} + \boldsymbol{\beta}_2^{'} \, 1_{Z_{t-1}\geq\lambda}) \, \mathbf{r}_t \\ &\quad + (\boldsymbol{\alpha}_1^{'} \, 1_{Z_{t-1}<\lambda} + \boldsymbol{\alpha}_2^{'} \, 1_{Z_{t-1}\geq\lambda}) \, \boldsymbol{\Delta y}_{t-1} + \epsilon_t, \end{aligned} \tag{2.12}$$

where (ρ_1, ρ_2) are slope coefficients on y_{t-1}, (β_1, β_2) are the slopes on the deterministic components and (α_1, α_2) are the slope coefficients on $\boldsymbol{\Delta y}_{t-1} = (\Delta y_{t-1}, \ldots, \Delta y_{t-k})'$ in the two regimes. The two regimes are where $Z_{t-1} \geq \lambda$ and $Z_{t-1} \leq \lambda$.

Under the null hypothesis of a unit root of TAR model, $\rho_1 = \rho_2 = 0$ while $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$ fulfill stationary conditions for Δy_t. Caner and Hansen (2001) used the Wald test to obtain the asymptotic null distribution.

Smooth Transition Autoregressive (STAR) Model

The smooth transition autoregressive (STAR) process developed by Granger and Terasvirta (1993) has been a popular process for modelling economic and finance data due to its generality and flexibility. Nonlinear adjustment in a STAR model allows for smooth rather than discrete adjustment in a TAR model. In a STAR model, adjustment takes place in every period but the speed adjustment varies with the extent of the deviation from equilibrium. A STAR(p) model can be expressed as follow:

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j} \, y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j} \, y_{t-j} \right] G(\theta, e, y_{t-d}) + \epsilon_t \tag{2.13}$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d \geq 1$ is a delay parameter; $(\theta, e) \in \{\mathbb{R}^+ \times \mathbb{R}\}$ where $\mathbb{R}$ denotes the real space $(-\infty, \infty)$ and $\mathbb{R}^+$ denotes the nonnegative real space $(0, \infty)$. The transition function $G(\theta, e, y_{t-d})$ determines the speed of adjustment to the equilibrium e. Two simple transition functions suggested by Granger and Terasvirta (1993) and Terasvirta (1994) are the logistic and exponential functions:

$$G(\theta, e, y_{t-d}) = \frac{1}{1+\exp\{-\theta(y_{t-d}-e)\}} - \frac{1}{2}, \tag{2.14}$$

$$G(\theta, e, y_{t-d}) = 1 - \exp\{-\theta^2(y_{t-d}-e)^2\}. \tag{2.15}$$

If the transition function $G(\theta, e, y_{t-d})$ is given by (2.14), (2.13) is called a logistic smooth transition autoregressive (LSTAR) model. If the transition function $G(\theta, e, y_{t-d})$ is given by (2.15), (2.13) is called an exponential smooth transition autoregressive (ESTAR) model.

Granger and Terasvirta (1993), Terasvirta (1994) and Luukkonen *et al.* (1988) suggested to apply a sequence of linearity tests to an artificial regression of Taylor series expansions of (2.14) or (2.15). Specification, estimation and evaluation of STAR models based on Terasvirta (1994) are explained as follow:

For LSTAR model, replace (2.14) by

$$G(z) = \frac{1}{1+\exp(-z)} - \frac{1}{2}, \tag{2.16}$$

where $z = \theta(y_{t-d} - e)$. Thus, (2.16) can be approximated using third-order Taylor approximation[2] by

$$\begin{aligned} G(z) &\approx \left.\frac{\partial G}{\partial z}\right|_{z=0} z + \frac{1}{2}\left.\frac{\partial^2 G}{\partial z^2}\right|_{z=0} z^2 + \frac{1}{6}\left.\frac{\partial^3 G}{\partial z^3}\right|_{z=0} z^3, \\ &= g_1\, \theta(y_{t-d}-e) + g_3\, \theta^3\, (y_{t-d}-c)^3, \end{aligned} \tag{2.17}$$

where $g_1 = \left.\frac{\partial G}{\partial z}\right|_{z=0}$ and $g_3 = \frac{1}{6}\left.\frac{\partial^3 G}{\partial z^3}\right|_{z=0}$. Note that $\left.\frac{\partial^2 G}{\partial z^2}\right|_{z=0} = 0$. Inserting (2.17) into (2.13) will result in an artificial model as follow:

$$\begin{aligned} y_t &= \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-j} + \theta_{2,0}\left(g_1\theta\, (y_{t-d}-e) + g_3\, \theta^3\, (y_{t-d}-e)^3\right) \\ &\quad + \sum_{j=1}^{p} \theta_{2,j}\, y_{t-j}\left(g_1\, \theta\, (y_{t-d}-e) + g_3\, \theta^3\, (y_{t-d}-e)^3\right) + \epsilon_t^*, \end{aligned}$$

$$\begin{aligned} y_t &= \phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j}\, y_{t-j} + \sum_{j=1}^{p} \phi_{1,j}\, y_{t-j}\, y_{t-d} + \sum_{j=1}^{p} \phi_{2,j}\, y_{t-j}\, y_{t-d}^2 \\ &\quad + \sum_{j=1}^{p} \phi_{3,j}\, y_{t-j}\, y_{t-d}^3 + \epsilon_t^* \end{aligned} \tag{2.18}$$

[2]Saikonnen and Luukkonen (1988) used first-order Tailor approximation, but Terasvirta (1994) argued that the test has low power against the alternative. Therefore, he suggested to use the third-order Taylor approximation.

where

$$
\begin{aligned}
\phi_{0,0} &= \theta_{1,0} - \theta_{2,0}\left(g_1\ \theta\ e + g_3\ \theta^3\ e^3\right), \\
\phi_{0,j} &= \theta_{1,j} + \theta_{2,0}\ g_1\ \theta 1_d + \theta_{2,j}\ g_1\ \theta\ e + \theta_{2j}\ g_3\ \theta^3\ e^3, \\
\phi_{1,j} &= \theta_{2,j}\ (g_1\ \theta\ e + 3g_3\ \theta^3\ e^2) - 3\theta_{2,0}\ g_3\ \theta^3 e\ 1_d, \\
\phi_{2,j} &= -3\theta_{2,j}\ g_3\ \theta^3\ e - \theta_{2,0}\ g_3\ \theta^3\ 1_d, \\
\phi_{3,j} &= \theta_{2,j}\ g_3\ \theta^3,
\end{aligned}
$$

for $j = 1, \ldots, p$ and $1_d = 1$ for $j = d$ and $1_d = 0$ for $j \neq d$.

For ESTAR model, replace (2.15) by

$$G(z) = 1 - \exp(-z^2), \tag{2.19}$$

where $z = \theta(y_{t-d} - e)$. Thus, (2.19) can be approximated using third-order Taylor approximation by ,

$$
\begin{aligned}
G(z) &\approx \left.\frac{\partial G}{\partial z}\right|_{z=0} z + \frac{1}{2}\left.\frac{\partial^2 G}{\partial z^2}\right|_{z=0} z^2 + \frac{1}{6}\left.\frac{\partial^3 G}{\partial z^3}\right|_{z=0} z^3, \\
&= g_2\ \theta^2\ (y_{t-d} - e)^2, \qquad (2.20)
\end{aligned}
$$

where $g_2 = \frac{1}{2}\left.\frac{\partial^2 G}{\partial z^2}\right|_{z=0}$. Note that $\left.\frac{\partial G}{\partial z}\right|_{z=0} = 0$ and $\left.\frac{\partial^3 G}{\partial z^3}\right|_{z=0} = 0$. Inserting (2.20) into (2.13) will result in an artificial model as follow:

$$
\begin{aligned}
y_t &= \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\ y_{t-j} + \theta_{2,0}\left(g_2\ \theta^2\ (y_{t-d} - e)^2\right) \\
&\quad + \sum_{j=1}^{p} \theta_{2,j}\ y_{t-j}\left(g_2\ \theta^2\ (y_{t-d} - e)^2\right) + \epsilon_t^*, \\
&= \phi_{0,0} + \sum_{j=1}^{p} \phi_{0,j}\ y_{t-j} + \sum_{j=1}^{p} \phi_{1,j}\ y_{t-j}\ y_{t-d} + \sum_{j=1}^{p} \phi_{2,j}\ y_{t-j}\ y_{t-d}^2 + \epsilon_t^*, \qquad (2.21)
\end{aligned}
$$

where

$\phi_{0,0} = \theta_{1,0} + \theta_{2,0}\ g_2\ \theta^2\ e^2$, $\phi_{0,j} = \theta_{1,j} - 2\theta_{2,0}\ g_2\ \theta^2\ e\ 1_d$

$\phi_{1,j} = -\theta_{2,j}\ g_2\ \theta^2\ e + \theta_{2,0}\ g_2\ \theta^2\ 1_d$, $\phi_{2,j} = \theta_{2,j}\ g_2\ \theta^2$,

for $j = 1, \ldots, p$ and $1_d = 1$ for $j = d$ and $1_d = 0$ for $j \neq d$.

Comparing (2.18) for LSTAR model and (2.21) for ESTAR model, the artificial model for ESTAR model does not include the cubic term of y_{t-d}. Thus, this difference can be used to differentiate LSTAR model and ESTAR model. Firstly, using a regression model in (2.18) and keeping the delay parameter d fixed, testing the null hypothesis

$$H_{01} : \phi_{1,j} = \phi_{2,j} = \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \ldots, p\}$$

provides a test of linear AR(p) model against a general nonlinear STAR model. Denote this general test as LM^G. If the LM^G test can be rejected, testing the null hypothesis

$$H_{02} : \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}$$

provides a test of ESTAR nonlinearity against LSTAR-type nonlinearity. Denote this test as LM^3. Finally, if the restrictions $\phi_{3,j} = 0$ cannot be rejected at a chosen significance level, then a more powerful test for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_{03} : \phi_{1,j} = \phi_{2,j} = 0 | \ \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}.$$

Denote this test as LM^E.

2.2 Pairs Trading

The traditional focus of equity investing has been on finding stocks to buy long that offer opportunity for appreciation. Institutional investors have given little if any thought to incorporate short-selling into their equity strategies to capitalize on over-valued stocks. More recently however, a growing number of investors have begun holding both long and short positions in their equity portfolios.

Short-selling is the practice of selling stock at current prices, but delaying the delivery of the stocks to its new owner. The idea is that the seller can then purchase the stock at a later date for delivery at a cheaper price than they collected for the stock. The difference between the sale price and the purchase price is the profit made by the short-seller. Obviously an investor is only willing to open a short position in a stock when they expect the price of that stock to fall.

Jacobs and Levy (1993) categorised long/short equity strategies as market neutral, equitised, and hedge strategies. The market neutral strategy holds both long and short positions with equal market risk exposures at all times. This is done by equating the weighted betas of both the long position and the short position within portfolio. This approach eliminates net equity market exposure so that the return realised should not be correlated with those of the market portfolio. This is equivalent to a zero-beta portfolio. Returns on these portfolios are generated by the isolation of alpha, which is a proxy for excess return to active management, adjusted for risk (Jensen, 1969). The funds received from the short sale are traditionally used to fund the long side, or invested at the cash rate.

The equitised strategy, in addition to hold stocks long and short in equal dollar balance, adds a permanent stock index futures overlay in an amount equal to the invested capital. Thus, the equitised portfolio has a full equity market exposure at all times. Profits are made from long/short spread, which is contingent upon the investors' risk.

The hedge strategy also holds stocks long and short in equal dollar balance but also has a variable equity market exposure based on a market outlook. The variable market exposure is achieved using stocks index futures. Once again, profits are made from the long/short spread as well as the exposure to the changing stock index future position. This approach is similar to typical hedge fund management but is more structured. Hedge funds sell stocks short to partially hedge their long exposures and to benefit from depreciating stocks. This differs from investing the entire capital both long and short to benefit from full long/short spread and then obtaining the desired market exposure through the use of stock index futures.

Do *et al.* (2006) structured slightly different approaches to long/short equity investing with Jacobs and Levy (1993). They tried to clarify the position of a pair trading strategy amongst other seemingly related hedge fund strategies. Do *et al.* (2006) noted that due to strategies fundamentals, which involve the simultaneous purchase of under-valued stocks and the shorting of over-valued stocks, pairs trading is essentially a form of long/short equity investing. After consulting academic sources and informal, internet-based sources, they declared that long/short equity strategies can be classified as either market neutral strategies or pairs trading strategies. This interpretation can be reconciled to that proposed by Jacobs and Levy (1993) since all three of their strategies include elements of market neutrality, even if the resulting portfolio may exhibit some market risk.

Ultimately, the difference between the strategies originates from their definition of "mispricing". The long/short strategies described by Jacobs and Levy (1993) refer to an "absolute mispricing". Those strategies require the identification of stocks that are either over-valued or under-valued relative to some risk-return equilibrium relationship such as the Arbitrage Pricing Theory (APT) model or the Capital Asset Pricing Model (CAPM). Pairs trading strategies, as defined by Do *et al.* (2006) require the simultaneous opening of long and short positions in different stocks and thus, fall under the "umbrella of long/short equity investments" (Do *et al.*, 2006). Debate continues amongst academics and practitioners alike regarding the role, if any, of market neutrality for a successful pairs trading strategy.

2.2.1 Why Pairs Trading Strategy?

Jacobs and Levy (1993) and Jacobs *et al.* (1999) suggested that investors who are able to overcome short-selling restrictions and have the flexibility to invest in both long and short positions can benefit from both winning and losing stocks. Traditional fund management does not allow investment managers to utilise short positions in their portfolio construction and so the investment decision-making process focuses on identifying undervalued stocks which can be expected to generate positive alpha. In effect, managers can only profit from those over-performing stocks, and any firm-specific information which suggests future under-performance is essentially worthless, investors can only benefit from half the market.

Another benefit of long/short investing is that, potentially, short positions provide greater opportunities than long positions. The search for undervalued stock takes place in a crowded field because most traditional investors look only for undervalued stocks. Because of various short-selling impediments, relatively few investors search for overvalued stocks. Furthermore, security analysts issue far more buy recommendations than sell recommendations. Buy recommendations have much more commission-generating power than sells, because all customers are potential buyers, but only those customers having current holdings are potential sellers, and short-sellers are few in number. Analysts may also be reluctant to express negative opinions. They need open lines of communication with company management, and in some cases management has cut them off and even threatened libel suits over negative opinions. Analysts have also been silenced by their own employers to protect their corporate finance business, especially their underwriting relationships (Jacobs and Levy, 1993, p. 3). Shorting opportunities may also arise from management fraud, "window-dressing" negative information, for which no parallel opportunity exists on the long side.

2.2.2 A Role of Market Neutrality in a Pairs Trading Strategy

Do *et al.* (2006) categorised the set of long/short equity strategies as either belonging to a "market neutral" or a "pair trading" sub-category. Consequently, does this apparent mutual-exclusivity render the constraint of market neutrality irrelevant to a successful pairs trading strategy?

To answer this question what is required is classification of what is meant by "market neutrality". According to Fund and Hsieh (1999), a strategy is said to be market neutral if it generates returns which are independent of relevant market returns. Market neutral funds actively seek to avoid major risk factors, and instead take bets on relative price movements. A market neutral portfolio exhibits zero systematic risk and is practically interpreted to possess a market beta equal to zero.

Lin *et al.* (2006) and Nath (2003) implicitly described pair trading as an implementation of market neutral investing. Both sets of authors repeatedly described pairs trading as "riskless", suggesting that the riskless nature of pairs trading stems from the simultaneous long/short opening market positions and that the opposing positions ideally immunise trading outcomes against systematic market-wide movements in prices that work against uncovered positions.

To some extent both authors are correct, however as Alexander and Dimitriu (2002) explained the reasoning proposed is not substantial enough to guarantee a market neutral portfolio. Although long/short equity strategies are often seen as being market neutral by construction, unless they are specifically designed to have zero-beta, long/short strategies are not necessarily market neutral. For example, in a recent paper Brooks and Kat (2002) found evidence of significant correlation of classic long/short equity hedge funds indexes with equity market indexes such as S & P500, DJIA, Rusell 2000 and NASDAQ, correlation which may still be under-estimated due to the auto-correlation of returns. Alexander and Dimitriu (2002) provided an alternative explanation which suggests that market neutrality in long/short equity strategies is derived from proven interdependencies within the chosen stocks. Such interdependencies, which can take the form of convergence, ensure that over a given time horizon the equities will reach as assumed equilibrium pricing relationship. In this case, the portfolio does not require a beta of zero to immunise it against systematic risk. This is handled by the assumed equilibrium pricing relationship, for example, a cointegrating relationship.

Next section will provide a brief review of different pairs trading approaches in the literature.

2.3 Pairs Trading Approaches

In this section, we introduce four main approaches used to implement pairs trading: distance approach, stochastic approach, combine forecasts approach and cointegration approach.

2.3.1 Distance Approach

Papers categorised in this approach include Gatev *et al.* (1998, 2006), Nath (2003), Andrade *et al.* (2005), Engelberg *et al.* (2008), Papandakis and Wisocky (2008) and Do and Faff (2008). Under this approach, the co-movement in a pair is measured by what is referred to as the distance, or the sum of squared differences between the two normalized price series.

Gatev *et al.* (1998, 2006) are the most cited papers in pairs trading. They constructed a cumulative total returns index for each stock over the formation period and then choose a matching partner for each stock by finding the security that minimizes the sum of squared deviations between the two normalized price series. Stock pairs are formed by exhaustive matching in normalized daily “price” space, where price includes reinvested dividends. In addition to “unrestricted” pairs, the study also provides results by sector, where they restrict stocks to the same broad industry categories defined by S&P. Those “restricted” pairs act as a test for robustness of any net profits identified using the unrestricted sample of pair trades. Their trading rules for opening and closing positions are based on a standard deviation metric. An opening long/short trade occurs when prices diverge by more than two standard deviations, estimated during the pair formation period. Opened positions are closed-out at the next crossing of the prices.

Gatev *et al.* (1998, 2006) showed that their pairs trading strategy after costs can be profitable. However, replicating Gatev *et al.* (1998, 2006) methodology with more recent data, Do and Faff (2008) reported the profit results of the strategy were declining. Engelberg *et al.* (2008) indicated the profitability from this strategy decreases exponentially over time.

Nath (2003) also used measure of distance to identify potential pair trades, although his approach does not identify mutually exclusive pairs. He kept a record of distances for each pair in the universe of securities, in a empirical distribution format so that each time an observed distance crosses over the 15 percentile, a trade is opened for that pair. A discrepancy between the two approaches is that Gatev *et al.* (1998, 2006) simplistically made no attempt to incorporate any risk management measure into their trading approach while Nath (2003) incorporated a stop-loss trigger to close the position whenever the distance moves against him to hit the 5 percentile. Additionally in Nath (2003), a maximum trading period is incorporated, in which all open positions are closed if distances have not reverted to their equilibrium state inside a given time-frame, as well as a rule which states that if any trades are closed early prior to mean reversion, then new trades on that particular pair are prohibited until such time as the distance or price series has reverted.

The distance approach purely exploits a statistical relationship between a pair of securities. As Do *et al.* (2006) noted, it is model-free and consequently, it has the advantage of not being exposed to model mis-specification and mis-estimation. However, this non-parametric approach lacks forecasting ability regarding the convergence time or expected holding period. A more fundamental issue is its underlying assumption that its price level distance is static through time, or equivalently, that the returns of the two stocks are in parity. Although such an assumption may be valid in short periods of time, it is only so for a certain group of pairs whose risk-return profiles are close to identical.

2.3.2 Stochastic Approach

Papers categorised in this approach include Elliott *et al.* (2005), Do *et al.* (2006), Rampertshammer (2007), Mudchanatongsuk *et al.* (2008) and Herlemont (2008).

Elliott *et al.* (2005) outlined an approach to pairs trading which explicitly models the mean reverting behaviour of the spread in a continuous time setting. The observed spread y_t is defined as the difference between the two stock prices. It is assumed that the observed spread is driven mainly by a state process x_k plus some measurement error, ω_k, i.e.:

$$y_k = x_k + D\omega_k, \tag{2.22}$$

where x_k denotes the value of variable at time $t_k = k\ \tau$, for $k = 0, 1, 2, \ldots$ and $\omega_k \sim$ IID $N(0, 1)$ with $D > 0$ is a constant measure of errors. The state variable x_k is assumed to follow a mean reverting process:

$$x_{k+1} - x_k = (a - bx_k)\ \tau + \sigma\sqrt{\tau}\ \epsilon_{k+1}, \tag{2.23}$$

where $\sigma > 0$, $b > 0$, $a > 0$, and $\{\epsilon_k\}$ is IID $N(0,1)$ and independent of $\{\omega_k\}$ in (2.22). The process mean reverts to $\mu = a/b$ with "strength" b and $x_k \sim N(\mu_k, \sigma_k)$, where

$$\mu_k \to a/b \qquad \text{as } k \to \infty,$$

and

$$\sigma_k^2 \to \frac{\sigma^2\tau}{1-(1-b\tau)^2} \qquad \text{as } k \to \infty.$$

Equation (2.23) can be written as

$$x_k = A + Bx_{k-1} + C\epsilon_k, \tag{2.24}$$

with $A = a\tau \geq 0$, $0 < B = 1 - b\tau < 1$, and $C = \sigma\sqrt{\tau}$. The discrete process in (2.24) can be approximated by a continuous process, i.e. $x_k \cong X_t$ where $\{X_t | t \geq 0\}$ satisfies the stochastic differential equation :

$$dX_t = \rho(\mu - X_t)\ dt + \sigma\ dB_t \tag{2.25}$$

where $\rho = b$, $\mu = a/b$ and $\{B_t |\ t \geq 0\}$ is a standard Brownian motion. By assuming x_k follows the process in (2.24), actually Elliott *et al.* (2005) considered the observed process y_k which is mainly driven by an AR(1) process in a discrete context.

Using the Ornstein-Uhlenbeck process as an approximation to (2.25), the first passage time result for X_t will be[3],

$$T = \inf\{t \geq 0, X_t = \mu | X_0 = \mu + c\ \sigma/\sqrt{2\rho}\} = \hat{t}\ \rho \tag{2.26}$$

where

$$\hat{t} = 0.5 \ln\left[1 + 0.5\left(\sqrt{(c^2-3)^2 + 4C^2} + C^2 - 3\right)\right]$$

and T in (2.26) means the time needed for the process X_t to reach μ for the first time given at $t = 0$, $X_0 = \mu + c\ \sigma/\sqrt{2\rho}$, c is a constant. Coefficients A, B, C and D in (2.22) and (2.24) are estimated using the state space model and the Kalman filter. For an introduction to the state space model and Kalman filter, see Durbin and Koopman (2001) and Elliott *et al.* (1995).

Elliott *et al.* (2005) suggested a pairs trading strategy by firstly choose a value of $c > 0$. Enter a pair trade when $y_k \geq \mu + c\ (\sigma/\sqrt{2\rho})$ and unwind the trade at time T later where T is defined in (2.26). A corresponding pair trade would be performed when $y_k \leq \mu - c\ (\sigma/\sqrt{2\rho})$ and unwind the trade at time T later. However, they did not give clear explanation how to choose the optimal value of c which can be regarded

[3]See Finch (2004) for the first passage time of the the Ornstein-Uhlenbeck process.

as a threshold to open a pair trade.

Do *et al.* (2006) noted that Elliott *et al.* (2005) offers three major advantages from the empirical prospective. Firstly, it captures mean reversion which underpins pairs trading. However, according to Do *et al.* (2006), the observed spread should be defined as the difference in logarithms of the prices. Do *et al.* (2006) suggested that generally, the long term mean of the level difference in two stocks should not be constant. The exception is when the stocks trade at similar price points. Therefore, defining the spread series as log differences, this is no longer a problem.

However, Schmidt (2008) noted that there is an issue with Do *et al.* (2006) and Elliott *et al.* (2005) remarks. If the spread series does not exhibit mean reversion then simply taking the logarithms should not result in a mean reverting series. This transformation simply forces the spread series to appear to converge with large deviations appear less pronounced. In effect, it gives the spread series the appearance of mean reverting property without providing any solid justification for its occurrence. Generally speaking, the spread of an arbitrary pair of stocks is not expected to has a long-run relationship or mean reversion unless those stocks are cointegrated.

The second advantage offered by Elliott *et al.* (2005) according to Do *et al.* (2006) is that it is a continuous time model, and, as such, it is a convenient vehicle for forecasting purposes. Importantly, traders can compute the expected time that the spread converges back to its long term mean, so that questions critical to pairs trading such as the expected holding period and expected return can be answered explicitly. In fact, there are explicit first passage time results available for the Orntein-Uhlenbeck dynamics.

The third advantage is that the model is completely tractable, with its parameters easily estimated by the Kalman filter in a state space setting. The estimator is a maximum likelihood estimator and optimal in the sense of minimum mean squared error (MMSE).

Despite several advantages, Do *et al.* (2006) criticised that this approach does have fundamental limitation in that it restricts the long-run relationship between the two stocks to one of return parity (Do *et al.*, 2006). That is, in the long-run, the stock pairs chosen must provide the same return such that any departure from it will be expected to be corrected in the future which Do *et al.* (2006) gave the proof of this. This severely limits this models generality as in practice it is rare to find two stocks with identical return series. While the risk-return models such as Arbitrage Pricing Theory (APT) and Capital Asset Pricing Model (CAPM) could suggest that two stocks with similar risk factors should exhibit identical expected returns, in reality it is not necessarily the case because each stock is subject to firm-specific risks which differentiate the return series of the two firms. It is also important to note that the Markovian concept of diversification does not apply here since a pairs trading portfolio is not sufficiently diversified.

Given this fundamental limitation, in what circumstances can this approach be applicable? One possibility is the case where companies adopt a dual-listed company (DLC) structure; essentially a merger between two companies domiciled in two different countries with separate shareholder registries and identities. Globally, there are only a small number of dual listed companies, which are notable examples including Uniliver NV/PLC, Royal Dutch Petroleum/Shell, BHP Billiton Ltd/PLC and Rio Tinto Ltd/PLC. In a DLC structure, both groups of shareholders are entitled to the same cash flows, although shares are traded on two separate exchanges and often attract different valuations. The fact that the shares cannot be exchanged each other preclude riskless arbitrage although there is a clear opportunity for pairs trading. Another candidate for pairs trading assuming returns parity is companies that follow cross listing. A cross listing occurs when an individual company is listed in multiple exchanges, the most prominent from being via American Depository Receipts (ADRs). Companies may also cross list within different exchanges within a country, such as the NASDAQ and NYSE in the USA. See Bedi and Tennant (2002) for more information on DLC's and cross listing.

Do *et al.* (2006) proposed a pairs trading strategy which differentiates itself from existing approaches by modelling mispricing at the return level, as opposed to the more traditional price level. The model also incorporates a theoretical foundation for a stock pairs pricing relationship in an attempt to remove ad hoc trading rules which are prevalent in Elliott *et al.* (2005).

The proposed model adopted more general modelling than Elliott *et al.* (2005), i.e:

$$x_k = A + Bx_{k-1} + C\epsilon_k, \tag{2.27}$$

$$y_k = x_k + \Gamma U_k + D\omega_k. \tag{2.28}$$

where x_k, for $k = 0, 1, \ldots$, is a state process defined the same as x_k in (2.24). Equation (2.28) is almost the same as (2.22), except y_k is the observed spread defined as the difference of the asset returns, i.e.: $y_k = R_{S1,k} - R_{S2,k}$, where $R_{S1,k}$ and $R_{S2,k}$ are the returns of asset $S1$ and $S2$, respectively at time k. In continuous context, $R_{S1,k}$ and $R_{S2,k}$ can be regarded as the first difference of log asset prices $S1$ and $S2$, denoted as $P_{S1,k}$ and $P_{S2,k}$ respectively[4]. Another difference is the existence of variable U_k where U_k is an exogenous input. If $\Gamma = 0$, the model here will be the same as the model in (2.22). Do *et al.* (2006) tried to explain the derivation of (2.28). However, the explanations are not clear. For example, they began with the asset pricing theory (APT) (Ross, 1976) and wrote in Do *et al.* (2006, p.12):

"The APT asserts that the return on a risky asset, over and above a risk free rate, should be the sum of risk premiums times the exposure, where specification of the risk

[4] $R_{S1,k} = (P_{S1,k} - P_{S1,(k-1)})/P_{S1,(k-1)} \approx \log(P_{S1,k}) - \log(P_{S1,(k-1)})$ and $R_{S2,k} = (P_{S2,k} - P_{S2,(k-1)})/P_{S2,(k-1)} \approx \log(P_{S2,k}) - \log(P_{S2,(k-1)})$.

factors is flexible, and may, for instance, take the form of Fama-French 3-factor model:

$$R = R_f + \beta r^m + \eta^i \tag{2.29}$$

where $\beta = (\beta_1^i \beta_2^i \ldots \beta_n^i)$ and $r^m = [(R^1 - r_f)(R^2 - r_f)\ldots(R^n - r_f)]^T$, with R^i denotes the raw return on the ith factor. The residual η has expected value of zero, reflecting the fact that the APT works on a diversified portfolio such that firm-specific risks are unrewarded, although its actual value may be non-zero."

They did not give clear explanation what they mean by n, R_f, r_f and T in (2.29). Reading literature about APT, CAPM and Fama-French 3-factor model, [5] we try to summarise Do *et al.* (2006) using our understanding.

Equation (2.29) would be better if it is written as follow:

$$R = r_f + \boldsymbol{\beta}' \, \mathbf{r} + \eta \tag{2.30}$$

where R is the return of an asset ; r_f is the risk free rate; $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)'$ is a vector $n \times 1$; $\mathbf{r} = (r_1, r_2, \ldots, r_n)' = ((R_1 - r_f), (R_2 - r_f), \ldots, (R_n - r_f))'$ is a vector $n \times 1$ with R_j denotes the return on the jth factor affecting the asset's return; η is the risky asset's idiosyncratic random shock with mean zero for asset i and the apostrophe $'$ denotes a transpose. Equation (2.30) states that the return of an asset is a linear function of the risk free rate plus the asset's sensitivities to the n factors affecting the asset's return.

CAPM uses 1 factor in model, i.e. it assumes that return of an asset depends only on market return:

$$R = r_f + \beta_1 \, r_1 + \eta \tag{2.31}$$

where $r_1 = (R_1 - r_f)$ is the excess return of the whole stock market R_1 over the risk free rate and β_1 is a constant describing the asset's sensitivities to the market. The return of stock market index is usually used as a proxy of R_1.

By using only 1 factor, CAPM oversimplifies the complex market. Therefore Fama and French added 2 other factors to CAPM. Fama-French 3-factor model uses 3 factors in the model, i.e.:

$$R = r_f + \beta_1 \, r_1 + \beta_2 \, r_2 + \beta_3 \, r_3 + \eta \tag{2.32}$$

where r_1 is the same as in (2.31); $r_2 = SMB$ stands for " small (market capitalization) minus big" and $r_3 = HML$ stands for "high (book-to-price ratio) minus low"; they measure the historic excess returns of small caps and value stocks over growth stocks; β_1, β_2 and β_3 are constants.

In fact, Do *et al.* (2006) should be aware that they adopted the CAPM instead of Fama-French 3-factor model in their simulation and empirical testing as they only used 1 factor, i.e. market return denoted as R_m. They defined a "relative" prices on two

[5] See for example Bodie *et al.* (2009) and references therein.

stocks $S1$ and $S2$ can be written as:

$$R_{S1,t} = R_{S2,t} + \Gamma\, r_{m,t} + e_t \tag{2.33}$$

where R_{S1} and R_{S2} are returns of assets $S1$ and $S2$; $r_m = R_m - r_f$ denotes the excess of market return over risk free rate; $\Gamma = \beta_{S1} - \beta_{S2}$; β^{S1} and β^{S2} describe the sensitivity of assets $S1$ and $S2$ to the market; and e_t is residuals noise term.

Equation (2.33) can be obtained using the CAPM model in (2.31). From (2.31),

$$R_{S1,t} = r_f + \beta_{S1}\, r_{m,t} + \eta_{S1,t}, \tag{2.34}$$

$$R_{S2,t} = r_f + \beta_{S2}\, r_{m,t} + \eta_{S2,t}. \tag{2.35}$$

Combining (2.34) and (2.35),

$$\begin{aligned} R_{S1,t} &= (R_{S2,t} - \beta_{S2}\, r_{m,t} - \eta_{S2,t}) + \beta_{S1}\, r_{m,t} + \eta_{S1,t}, \\ &= R_{S2,t} + (\beta_{S1} - \beta_{S2})\ r_{m,t} + (\eta_{S1,t} - \eta_{S2,t})\,, \end{aligned}$$

which is the same as (2.33) if $\Gamma = \beta_{S1} - \beta_{S2}$ and $e_t = \eta_{S1,t} - \eta_{S2,t}$.

Embracing the above equilibrium model allows the specification of the residual spread function G:

$$G_t = G(P_{S1,t}, P_{S2,t}, U_t) = R_{S1,t} - R_{S2,t} - \Gamma\, r_{m,t} \tag{2.36}$$

Thus, from (2.36), Do *et al.* (2006) defined U_t in (2.28) as $r_{m,t}$. If the values of Γ is known and $r_{m,t}$ is specified, G_t is observable. However, Γ is unknown, so that it needs to be estimated. One solution they suggested is to redefine the observation in state space y_t instead of using G_t such that

$$y_t = x_t + \Gamma\, r_{m,t} + D\, \omega_t$$

which is similar to (2.28) where

$$y_t = \left[\log(P_{S1,t}) - \log(P_{S1,(t-1)})\right] - \left[\log(P_{S2,t}) - \log(P_{S2,(t-1)})\right] \approx R_{S1,t} - R_{S2,t}.$$

In Do *et al.* (2006, p.15), they proposed a trading rule strategy:

“ A trading rule for this strategy is to take a long-short position whenever the accumulated spread $\delta_k = \sum_{i=k-l}^{k} E[x_i|\mathcal{F}_i]$, [6] with l less than or equal to the current time k, exceed θ by a certain threshold [7]. The trader will have to fix a base from which to determine point l where $\delta_l = 0$. One may wish to compute the expected convergence time, that is the expectation of $T > k$ such that δ_T first crosses 0, given $\delta_k = c$.”

There are still some questions about their trading rule, e.g.: how do we choose the

[6] $\mathcal{F}_i = \sigma\{y_1, y_2, \ldots y_i\}$ denotes all information up dated to y_i.

[7] $\theta = \mu$ in (2.25) describing the mean of state process x_t

threshold c?, how do we determine the expected holding period, hence the expected return? In fact, in their empirical examples, they did not provide trading simulation results. Furthermore, basing the strategy on mispricing at the return level, would make misjudgement in the strategy. Figure 2.1(a) shows the plots of simulated prices of stock A and B. They are clearly not cointegrated. However, in Figure 2.1(d), the residual spreads denoted as $R_t = R_{S1,t} - R_{S2,t}$ seems stationary, apart from large deviation in the last observations. This happens because $R_{S1,t}$ and $R_{S2,t}$ are usually stationary series, so that a linear combination of these series would make a stationary series as well.

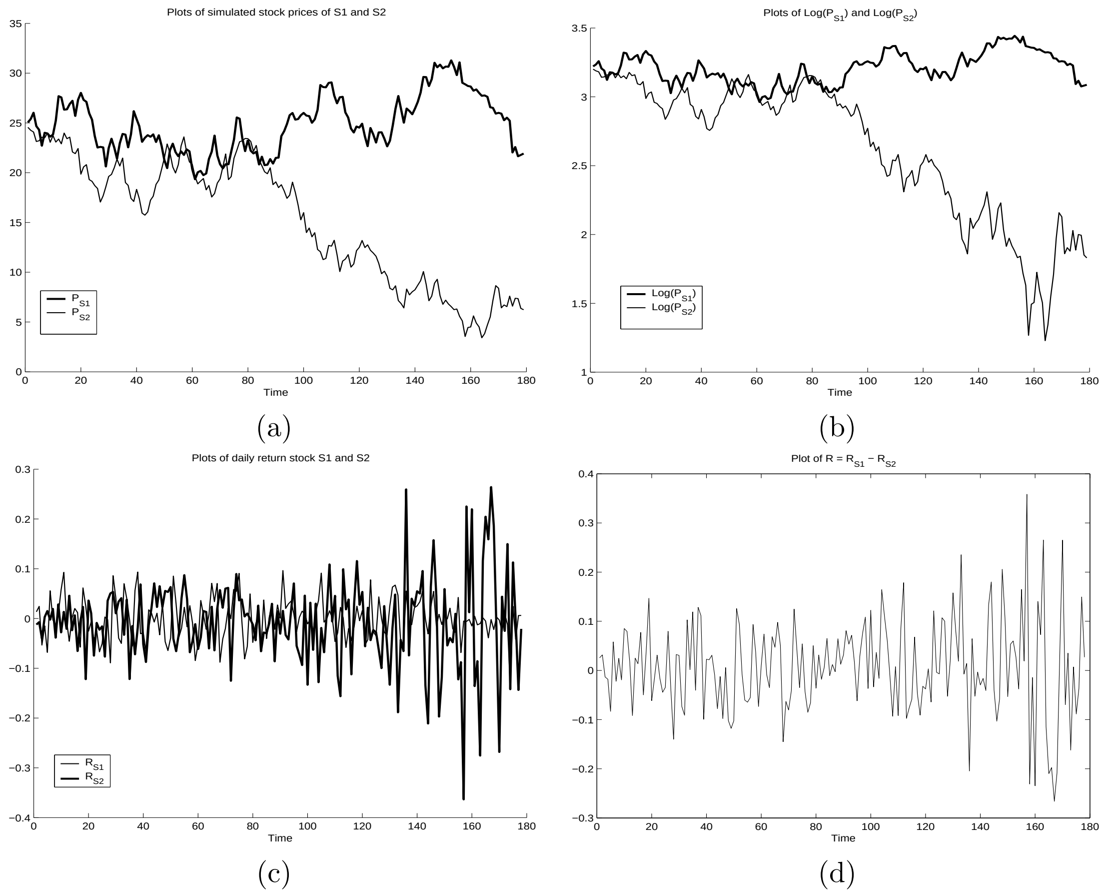


Figure 2.1: Residual spread simulation data

2.3.3 Combine Forecasts Approach

This approach is promoted by Huck (2009, 2010). The framework of the approach can be seen as a sort of combined forecast specially designed for pairs trading. The combination of forecasts can be done using Multi Criteria Decision Methods (MCDM). The approach can be described as follows:

1. Consider n stocks, $n(n-1)/2$ different pairs can thus be formed.

2. For each pair, forecast the cumulative spread return the stocks up to time d.
3. Define a symmetric matrix of size $n \times n$. Each element of the matrix will be the anticipated spread computed before. This matrix will be the input of a MCDM in order to rank stocks in terms of anticipated returns.
4. The strategic behavior is thus quite simple: Firstly, buy the first stocks of the ranking and sell the last ones. Buying and selling positions have the same weight. The strategy is, by construction, dollar neutral. Secondly, at time d, close all positions. If the ranking was relevant, a profit is made.

In brief, the method is based on three phases: forecasting, ranking and trading. This framework differs from the others on very essential point that it is developed without reference to any equilibrium model. Huck (2009, 2010) argued that the method provides much more trading possibilities and could detect the "birth" of the divergence that the other approaches cannot achieve. However, there is no other research follows this approach so far, may be due to complexity involving a matrix of size $n \times n$. Furthermore, Huck (2009, 2010) did not give detail on how to choose time d.

2.3.4 Cointegration Approach

Cointegration is a statistical relationship where two time series that are both integrated of same order d, I(d), can be linearly combined to produce a single time series which is integrated of order $d - b$, where $b > 0$. In its application to pairs trading, we refer to the case where two stocks price series are I(1) combined to produce a stationary, or $I(0)$, portfolio time series. Consider again,

$$P_{S1,t} - \beta\ P_{S2,t} = \mu + \epsilon_t, \tag{2.37}$$

where $P_{S1,t}$ and $P_{S2,t}$ are the prices of assets $S1$ and $S2$ respectively; ϵ_t is a stationary time series called cointegration errors and μ is the mean of cointegration relationship.

Cointegration incorporates mean reversion into pairs trading framework which is the single most important statistical relationship required for success. If the value of the portfolio is known to fluctuate around its equilibrium value then any deviations from this value can be traded against. Studies categorized in this approach include Lin *et al.* (2006), Vidyamurthy (2004) and Galenko *et al.* (2007), Gillespie and Ulph (2001) and Hong and Susmel (2003). Alexander and Dimitriu (2002) and Galenko *et al.* (2007) used cointegration in long/short investment strategy involving more than two stocks.

Lin *et al.* (2006) developed a pairs trading strategy based on a cointegration technique called the cointegration coefficients weighted (CCW) rule. They derived the minimum profit per trade using pairs trading based on cointegration. Consider the following assumptions:

1. The two share price series are cointegrated over the relevant time period.

2. Long (buy) and short (sell) positions always apply to the same shares in the share-pair.

3. We can not open a trade if the previous open trade is not closed yet.

4. Short sales are permitted or possible through a broker and there is no interest charged for the short sales and no cost for trading.

5. $\beta > 0$ in (2.37).

Assumptions 1, 2 and 3 are fairly straight forward. Assumption 4 is applied to simplify the analysis. The fifth assumption is required to carry out pairs trading strategy and given that the prices of the two assets have the same pattern, β in (2.37) should be positive.

Upper-trades (U-trades)

Consider two cointegrated assets, $S1$ and $S2$ as in (2.37). By using Assumption 1, we can conclude that

- When $\epsilon_t \geq U$, the price of one unit share $S1$ is higher than or equal to the price of β unit shares $S2$, relative to their equilibrium relationship. In other words, $S1$ is overvalued while $S2$ is undervalued. A trade is opened at this time. Let t_o represent the time of opening a trade position.

- If $\epsilon_t \leq 0$, the price of one unit share $S1$ is less than or equal the price of β unit shares $S2$, relative to their equilibrium relationship. In other words, $S1$ is undervalued while $S2$ is over-valued. The trade is closed at this time. Let t_c represent the time of closing out a trade position.

When the cointegration error is higher than or equal to the pre-set upper-bound U at time t_o, a trade is opened by selling N_{S1} of $S1$ shares at time t_o for $N_{S1}P_{S1,t_0}$ dollars and buying N_{S2} of $S2$ at time t_o for $N_{S2}P_{S2,t_o}$ dollars.

When the cointegration error has settled back to its equilibrium at time t_c, the positions are closed out by simultaneously selling the long position shares for $N_{S2}P_{S2,t_c}$ dollars and buying back the N_{S1} of $S1$ shares for $N_{S1}P_{S1,t_c}$ dollars.

Profit per trade will be

$$P = N_{S2}(P_{S2,t_c} - P_{S2,t_o}) + N_{S1}(P_{S1,t_o} - P_{S1,t_c}). \tag{2.38}$$

According to the CCW rule as in Lin *et al.* (2006), if the weight of N_{S2} and N_{S1} are chosen as a proportion of the cointegration coefficients, i.e. $N_{S1} = 1$ and $N_{S2} = \beta$, the minimum profit per trade can be determined as follows: [8]

$$\begin{aligned} P &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [P_{S1,t_o} - P_{S1,t_c}] \\ &= \beta[P_{S2,t_c} - P_{S2,t_o}] + [(\epsilon_{t_o} + \mu) + \beta P_{S2,t_o}) - (\epsilon_{t_c} + \mu) + \beta P_{S2,t_c})] \\ &= (\epsilon_{t_o} - \epsilon_{t_c}) \geq U. \end{aligned} \tag{2.39}$$

Thus, by trading the shares with the weight as a proportion of the cointegration coefficients, the profit per trade is at least U dollars.

Lower-trades (L-trades)

For an L-trade, the pre-set lower-bound L can be set to be $-U$. So, a trade is opened when $\epsilon_t \leq -U$ by selling $S2$ and buying $S1$.

Profit per trade will be:

$$P = N_{S2}(P_{S2,t_o} - P_{S2,t_c}) + N_{S1}(P_{S1,t_c} - P_{S1,t_o}). \tag{2.40}$$

Analogous to the derivation of minimum profit per trade for an U-trade, let $N_{S2} = \beta$ and $N_{S1} = 1$. Thus,

$$P = \beta[P_{S2,t_o} - P_{S2,t_c}] + [P_{S1,t_c} - P_{S1,t_o}] = (\epsilon_{t_c} - \epsilon_{t_o}) \geq U. \tag{2.41}$$

So, trading 1 unit share $S1$ and β unit shares $S2$, either in U-trades or L-trades would make a minimum profit per trade at least U.

The profit does not consider the time value of money as they consider pairs trading is not a long-term investment so that the change in value over time is negligible. Also note that the pairs trading strategy here is not a dollar-neutral as at the beginning of the trade the stock value in long position is not the same as the stock value in short position. Gatev *et al.* (2006) used a dollar-neutral strategy with the distance method so they assumed that the proceed from the short selling can be used to buy stocks in long position. However, in practice, investors do not have available the proceed of the short-sale at the beginning of the trade. To establish a short position in the US market for example, investors need to open a margin account, which requires a 50% collateral deposit for a given open position. Therefore, a dollar-neutral pairs trading strategy is hard to accomplish in practice. Furthermore, even though Gatev *et al.* (2006) showed that they can produce significant profits, the profit made from the portfolios have exposures to the systematic risks. On the other hand, the strategy in Lin *et al.* (2006) can be categorized as a market neutral strategy[9]. This strategy holds both long and short positions with weights correspond to the cointegration coefficients so that the portfolio's outcomes is immunized against systematic market risk and the pre-set profit

[8] For simplicity, fractional share holdings are permitted.

[9] See Vidyamurthy (2004) and Schmidt (2008)

per trade can be achieved. However, Lin *et al.* (2006) did not give further explanation how to choose the upper-bound U.

Vidyamurthy (2004) also suggested pairs trading based on cointegration. To test for cointegration, he adopted the Engle and Granger's 2 step approach (Engle and Granger, 1987) in which the log price of stock $S1$ is first regressed against log price of stock $S2$ in what we refer to as the cointegrating equation:

$$\log(P_{S1,t}) - \gamma \ \log(P_{S2,t}) = \mu + \epsilon_t \tag{2.42}$$

where γ is the "cointegrating coefficient" ; the constant term μ captures some sense of "premium" in stock $S1$ versus stock $S2$; and ϵ_t is the residuals term with mean 0 and variance σ_ϵ^2. The estimated residual series $\hat{\epsilon}_t$ is then tested for stationary using the augmented Dickey-Fuller test (ADF).

Equation (2.42) says that a portfolio comprising long 1 unit of stock $S1$ and short γ units of stock $S2$ has a long-run equilibrium value of μ and any deviations from this value are merely temporary fluctuations (ϵ_t). The portfolio will always revert to its long-run equilibrium value since ϵ_t is known to be an I(0). Vidyamurthy (2004) developed trading strategies based on the assumed dynamics of portfolio. The basic trading idea is to open a long position in portfolio when it is sufficiently below it long-run equilibrium ($\mu - \Delta$) and similarly, short the portfolio when it is sufficiently above its long-run value($\mu + \Delta$). Once the portfolio mean reverts to its long-run equilibrium value then position is closed and profit is earned equal to Δ per trade. The key question when developing a trading strategy is what value of Δ is going to maximise the profit function. The profit function is the profit per trade multiplied by number of trades. The number of trades is estimated by calculating the rate of zero crossings and level crossing for different values of Δ. Assuming the residuals as an ARMA process, Rice' formula (Rice, 1945) is used to calculate the rate of zero crossings and level crossing. The value of Δ which maximises the profit function is chosen as the trading trigger.

Another alternative is non-parametric approach constructing an empirical distribution of zero and level crossings based on the estimation sample. The optimal Δ is chosen so as to maximise the profit function from the estimated sample. This value is then applied to real time portfolio construction. A fundamental assumption of this non-parametric approach to determining Δ is that the observed dynamics of ϵ continue into the future. This approach appears to be favoured by Vidyamurthy (2004) due to its simplicity and avoidance of model mis-specification.

Apart from being rather as hoc , Do *et al.* (2006) criticised that Vidyamurthy's approach may be exposed to errors arising from the econometric techniques employed. Firstly, the 2-step cointegration procedure renders results sensitive to the ordering of variables, therefore the residuals may have different sets of statistical properties. Secondly, if the bivariate series are not cointegrated, the "cointegrating equation" results in spurious estimators (Lim and Martin, 1995). This would have the effect of making any mean reversion analysis of the residuals unreliable. Do *et al.* (2006) also criticized the difficulty of cointegration approach in associating it with theories on asset pricing.

Furthermore, using Rice's formula to estimate the number of trades is not quite right because it calculate number of crossing without giving restriction when the trade start or open. Instead of using Rice's formula, the first passage time of stationary series should be used.

Choosing the number of assets traded as a proportion of cointegration coefficients, the approaches in Lin *et al.* (2006) and Vidyamurthy (2004) are not dollar neutral as the amount of money from short asset is not the same as the money needed to buy the long asset.

2.3.5 Cointegration and Correlation in Pairs Trading Strategies

Following the seminal work of Markowitz (1959), Sharpe (1964), Lintner (1965) and Black (1972), the fundamental statistical tool for traditional portfolio optimisation is correlation analysis of assets returns. Optimization models for portfolio construction focus on minimizing the variance of the combined portfolio, for a given return, with additional constraints concerning certain investment allowances, short-sale restrictions and associated costs of rebalancing the portfolio.

In the last decade the concept of cointegration has been widely applied in financial econometrics in connection with time series analysis and macroeconomics. It has evolved as an extremely powerful statistical technique because it allows the application of simple estimation methods (such as least squares regression and maximum likelihood) to non-stationary variables. Still, its relevance to investment analysis has been rather limited thus far, mainly due to the fact that the standard in portfolio management and risk management is the correlation of asset returns.

However as Alexander and Dimitriu (2002) noted, correlation analysis is only valid for stationary variables. This requires prior de-trending of prices and other levels of financial variables, which are usually found to be integrated of order one or higher. Taking the first difference in log prices is the standard procedure for ensuring stationary and leads all further inference to be based on returns. However, this procedure has the disadvantage of loosing valuable information. In particular, de-trending the variables before analysis removes any possibility to detect common trends in prices. Furthermore, when variables in a system are integrated of different orders, and therefore require different orders of differences to become stationary, the interpretation of the results becomes difficult. By contrast, the aim of the cointegration analysis is to detect any stochastic trend in the price data and use these common trends for a dynamic analysis of correlation in returns (Alexander *et al.*, 2001).

The fundamental remark justifying the application of cointegration concept to stock price analysis is that a system of non-stationary stock prices in level form can share common stochastic trends (Stock and Watson, 1991). According to Beveridge and Nelson (1981), a variable has a stochastic trend if it has a stationary invertible ARMA(p,q) representation plus a deterministic component. Since ARIMA(p,1,q) models seem to characterize many financial variables, it follows that the growth in these variables can be described by stochastic trends.

The main advantage of cointegration analysis, as compared to the classical but rather limited concept of correlation, is that it enables the use the entire information set comprised in the levels of financial variables. Furthermore, a cointegrating relationship is able to explain the long-run behaviour of cointegrated series, whereas correlation, as a measure of co-dependency, usually lacks stability, being only a short-run measure. While the amount of history required to support the cointegrating relationship may be large, the attempt to use the same sample to estimate correlation coefficients may face many obstacles such as outliers in the data sample and volatility clustering (Alexander and Dimitriu, 2005). The enhanced stability of cointegrated relationship generates a number of significant advantages for a trading strategy. These include the reduction of the amount of rebalancing of trades in a hedging strategy and consequently, the associated transaction costs.

When applied to stock prices and stock market indexes, usually found to be integrated of order one, cointegration requires the existence of at least one stationary linear combination between them. A stationary linear combination of stock prices/market indexes can be interpreted as mean reversion in price spreads. The finding that the spread in a system of prices is mean reverting does not provide any information for forecasting the individual prices in the system, or the position of the system at some point in the future, but it does provide the valuable information that, irrespective to its position, the prices in the system will evolve together over the long term.

If two stocks prices are cointegrated, then a combination of these may be formed such that their spread is stationary, or mean-reverting. Pairs trading seeks to identify stocks whereby some form of relative pricing measure can be approximated by a long-run equilibrium relationship. Thus, pairs trading approaches based on cointegration can guarantee mean reversion, which is the single most important feature of successful pairs trading strategy. No other approach can guarantee this property.

As explained in the previous sections, there are some pairs trading strategies in literature. Each of them has own advantages and disadvantage. Despite some criticisms, this thesis develops pairs trading strategy based on cointegration. Especially, this thesis extends the results in Lin *et al.* (2006) and Vidyamurthy (2004). The main reason for doing this is the fact that any long/short trading strategy always uses the assumption of mean reverting relationship between the assets traded. Cointegration test is a statistical tool to determine that there exists mean reverting relationship between the assets. As suggested by Do *et al.* (2006), this thesis also uses Johansen cointegration test which is more rigorous test compared to Engle-Granger cointegration test. We are aware that the pairs strategy developed in this thesis still far away from perfect to be applied in practical situation. However, we hope that the results will give some ideas for further developments.

Chapter 3

Trade Duration and Inter-trades Interval for Pairs Trading Using Markov Chain Approach

3.1 Introduction

Consider again a cointegration relationship between asset prices P_{S1} and P_{S2} in (1.1):

$$P_{S1,t} - \beta\, P_{S2,t} = \mu + \epsilon_t. \tag{3.1}$$

If P_{S1} and P_{S2} are cointegrated, then the cointegration error series $\{\epsilon_t\}$ is a stationary process. As a stationary process, $\{\epsilon_t\}$ might follow a linear stationary process (e.g.: white noise, autoregressive, moving average, and autoregressive-moving average processes) or a non-linear stationary process. In this chapter, we concentrate with $\{\epsilon_t\}$ satisfying the Markov property (e.g.: white noise and autoregressive processes). Assuming $\{\epsilon_t\}$ is a white noise or an AR(1) model, trade duration and inter-trade interval for pairs trading using Marcov chain approach are discussed.

3.2 Basic Concepts of Markov Chain

The Markov chain concepts used here are mainly taken from Kemeny and Snell (1960). Let $\{\epsilon_t\}$ be a real-value discrete-time Markov process with the state space, S = {1,2,3}. We also assume that it is a regular Markov chain[1].

The transition matrix will be :

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \tag{3.2}$$

[1]If $p_{ij}^{(n)} > 0$ for all n where n is the number of steps from state i to state j, then the chain with the transition matrix P in (3.2) is a regular Markov chain.

where p_{ij} denotes the transition probability from state i to state j.

If state 3 is an absorbing state, the transition matrix in (3.2) will become:

$$P^* = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.3}$$

For a regular Markov chain, the first passage-time from state 1 to state 3 denoted by f_{13} is the number of steps to enter state 3 for the first time from state 1. The mean time to go from state 1 to state 3 for the first time with the transition matrix P in (3.2) is the same as the mean time before absorption from state 1 in the transition matrix P^* in (3.3) (see Kemeny and Snell, 1960, p. 78).

The canonical form of matrix P^* is

$$P^* = \begin{bmatrix} p_{11} & p_{12} & \vdots & p_{13} \\ p_{21} & p_{22} & \vdots & p_{23} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 1 \end{bmatrix} = \begin{bmatrix} Q & \vdots & R \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 1 \end{bmatrix} \tag{3.4}$$

where

$$Q = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \text{ and } R = \begin{bmatrix} p_{13} \\ p_{23} \end{bmatrix}. \tag{3.5}$$

For an absorbing Markov chain, we define the fundamental matrix N as follow:

$$\begin{aligned} N &= 1 + Q + Q^2 + Q^3 + \ldots = (I - Q)^{-1} = \begin{bmatrix} 1 - p_{11} & -p_{12} \\ -p_{21} & 1 - p_{22} \end{bmatrix}^{-1} \\ &= \frac{1}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \begin{bmatrix} 1 - p_{22} & p_{12} \\ p_{21} & 1 - p_{11} \end{bmatrix}. \end{aligned} \tag{3.6}$$

The mean of time to the absorption state from state 1 will be equal to the sum of the first row of the matrix N (see Kemeny and Snell, 1960, p. 51). Hence, the mean of first passage-time from state 1 to state 3 denoted as $E(f_{13})$ is:

$$E(f_{13}) = N_{11} + N_{12} = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} \tag{3.7}$$

where N_{ij} denotes the entry of matrix N at ith row and jth column.

3.3 Transition Matrix Construction for Trade Duration in Pairs Trading Strategy

Consider again the pairs trading strategy based on the cointegration coefficients weighted (CCW) rule from Lin *et al.* (2006) explained in Subsection 2.3.4. We assume that $\{\epsilon_t\}$

has following properties:

1. $\{\epsilon_t\}$ is a real-value discrete-time process.

2. $\{\epsilon_t\}$ satisfies the Markov property.

Since share prices data are usually taken in a certain interval of time (e.g.: hourly, daily or weekly) and the share prices are real number , justify the first assumption. The second assumption is made as we are only concern with the cointegration error satisfying the Markov property. Because the cointegration error is a stationary process, it would not be a controversial if we assume that it is also a stationary Markon chain as in the second assumption.

According to the pairs trading strategy for upper-trades, when $\epsilon_t \geq U$ at time t_o, a pair trade is opened by selling N_{S1} of $S1$ shares and buying N_{S2} of $S2$ shares . When $\epsilon_t \leq 0$ at time t_c, the positions are closed out by simultaneously selling the long position the N_{S2} of $S2$ shares and buying back the N_{S1} of $S1$ shares. Trade duration is $t_c - t_o$, which is the time needed to unwind a pair trade. Using the Marcov chain approach, we divide the value of ϵ_t into 3 states. Therefore, the state space of the Markov chain becomes as follows:

State 1: $\epsilon_t \geq U$

State 2: $0 < \epsilon_t < U$

State 3: $\epsilon_t \leq 0$

where U is the upper-bound defined as $U = k\sigma_\epsilon$; σ_ϵ is the sample standard deviation of ϵ_t and k is a constant.

Since at any time and at any state, ϵ_t can move from one state to other states, $p_{ij} > 0$ for every i and j in matrix P. Hence, $p_{ij}^{(n)} > 0$ for all n. So we can call the chain in matrix P as a regular Markov chain. However, when we want to calculate f_{13}, we make state 3 as an absorbing state as in matrix P^*.

The trading duration is analogous to the first passage-time from state 1 to state 3. Following the derivation of the first passage-time in Section 3.2, we will derive the trade duration for white noise and AR(1) processes. Without loss generality, only upper-trades case are derived as white noise and AR(1) processes are symmetric, so that we can consider that the trade duration for lower trades is the same as the trade duration for upper trades.

3.3.1 Case 1: White Noise Process

Consider that ϵ_t is a white noise process with distribution as follows:

$$\epsilon_t \sim IID\ N(0, \sigma_\epsilon^2), \qquad \forall\ t. \tag{3.8}$$

Since ϵ_t is a white noise process, there is no correlation between ϵ_t and ϵ_{t+1} or $E[\epsilon_t \epsilon_{t+1}] = 0$, i.e. ϵ_t and ϵ_{t+1} are independent. So, for this case, it is easy to calculate p_{11}, p_{22}, p_{12} and p_{21}.

$$\begin{aligned} p_{11} &= P[\epsilon_{t+1} \geq U | \epsilon_t \geq U] = P[\epsilon_{t+1} \geq U], \\ p_{22} &= P[0 < \epsilon_{t+1} < U | 0 < \epsilon_t < U] = P[0 < \epsilon_{t+1} < U], \\ p_{21} &= P[\epsilon_{t+1} \geq U | 0 < \epsilon_t < U] = P[\epsilon_{t+1} \geq U] = p_{11}, \\ p_{12} &= P[0 < \epsilon_{t+1} < U | \epsilon_t \geq U] = P[0 < \epsilon_{t+1} < U] = p_{22}. \end{aligned}$$

By using (3.7), the expected of trading duration $E(TD_U)$ for this case is:

$$E(TD_U) = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} = \frac{1}{1 - (p_{11} + p_{22})}.$$

Since $p_{11} + p_{22} = \int_U^\infty f(\epsilon_t)\, d\epsilon_t + \int_0^U f(\epsilon_t)\, d\epsilon_t = 1/2$, with $f(\epsilon_t)$ is a probability density function of ϵ_t,

$$E(TD_U) = 2. \tag{3.9}$$

Therefore, for this case, the cointegration error variance σ_ϵ^2 and the cointegration coefficient β do not have any effect to the trading duration. If the unit time is one day, (3.9) describes that the average of trade duration is 2 days.

3.3.2 Case 2: AR(1) Process

Consider that $\{\epsilon_t\}$ is a simple AR(1) process[2].

$$\epsilon_t = \phi\, \epsilon_{t-1} + a_t, \tag{3.10}$$

where a_t IID $N(0, \sigma_a^2)$ and $|\phi| < 1$. It is well known that the simple AR(1) process also follows the Markov chain process (Lai, 1978).

The AR(1) process has properties as follows:

$$E[\epsilon_t] = E[a_t] = 0, \tag{3.11}$$

$$Var[\epsilon_t] = \frac{Var[a_t]}{1 - \phi^2} = \frac{\sigma_a^2}{1 - \phi^2} = \sigma_\epsilon^2, \tag{3.12}$$

$$Cov[\epsilon_t, \epsilon_{t+1}] = \frac{\phi\, \sigma_a^2}{1 - \phi^2} = \phi\, \sigma_\epsilon^2. \tag{3.13}$$

For a stationary AR(1) process, the expectation, variance and covariance in (3.11)-(3.13) are free from t. Assume that ϵ_0 has normal distribution $N(0, \sigma_\epsilon^2)$. Therefore, the ϵ_t will be distributed as $N(0, \sigma_a^2/(1 - \phi^2))$. However, the ϵ_t are correlated. To derive the joint probability density function $f(\epsilon_t, \epsilon_{t+1})$, we consider $a_{t+1} = \epsilon_{t+1} - \phi\, \epsilon_t$. a_{t+1} and ϵ_t are independent each other. Hence, the joint probability density function of

[2]Without loss generality, it is assumed that $\{\epsilon_t\}$ has a zero mean.

(ϵ_t, a_{t+1}) is

$$\begin{aligned} f(\epsilon_t, a_{t+1}) &= f(\epsilon_t) f(a_{t+1}) \\ &= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{1}{2}\left[\frac{\epsilon_t - 0}{\sigma_\epsilon}\right]^2\right\} \times \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left\{-\frac{1}{2}\left[\frac{a_{t+1} - 0}{\sigma_a}\right]^2\right\}. \end{aligned}$$

To get $f(\epsilon_t, \epsilon_{t+1})$, we need to obtain Jacobian transformation. As $\epsilon_t = \epsilon_t$ and $a_{t+1} = \epsilon_{t+1} - \phi\, \epsilon_t$, the Jacobian transformation is

$$J = \begin{vmatrix} \partial\epsilon_t/\partial\epsilon_t & \partial\epsilon_t/\partial\epsilon_{t+1} \\ \partial a_{t+1}/\partial\epsilon_t & \partial a_{t+1}/\partial\epsilon_{t+1} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -\phi & 1 \end{vmatrix} = 1.$$

It follows that

$$\begin{aligned} f(\epsilon_t, \epsilon_{t+1}) &= f(\epsilon_t, a_{t+1})|J| \\ &= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left\{-\frac{1}{2}\left[\frac{\epsilon_t - 0}{\sigma_\epsilon}\right]^2\right\} \times \\ &\quad \frac{1}{\sqrt{2\pi\sigma_\epsilon^2(1-\phi^2)}} \exp\left\{-\frac{1}{2}\left[\frac{(\epsilon_{t+1} - \phi\,\epsilon_t) - 0}{\sigma_\epsilon\sqrt{(1-\phi^2)}}\right]^2\right\} \\ &= \frac{1}{2\pi\sigma_\epsilon^2\sqrt{(1-\phi^2)}} \exp\left\{-\frac{1}{2}\left[\frac{\epsilon_t^2}{\sigma_\epsilon^2} + \frac{\epsilon_{t+1}^2 - 2\phi\,\epsilon_t\,\epsilon_{t+1} + \phi^2\epsilon_t^2}{\sigma_\epsilon^2(1-\phi^2)}\right]\right\} \\ &= \frac{1}{2\pi\sigma_\epsilon^2\sqrt{(1-\phi^2)}} \exp\left\{-\frac{1}{2(1-\phi^2)}\left[\frac{\epsilon_t^2 - 2\phi\,\epsilon_t\,\epsilon_{t+1} + \epsilon_{t+1}^2}{\sigma_\epsilon^2}\right]\right\}. \end{aligned} \tag{3.14}$$

Thus, $f(\epsilon_t, \epsilon_{t+1})$ in (3.14) is a joint probability density function for bivariate Normal distribution with mean zero and the variance-covariance matrix can be determined by using (3.12)-(3.13).

Now, we can calculate p_{11}, p_{22}, p_{12} and p_{21} as follow

$$p_{11} = P[\epsilon_{t+1} \geq U | \epsilon_t \geq U] = \frac{\int_U^\infty \int_U^\infty f(\epsilon_t, \epsilon_{t+1})\, d\epsilon_t\, d\epsilon_{t+1}}{\int_U^\infty f(\epsilon_t)\, d\epsilon_t}, \tag{3.15}$$

$$p_{22} = P[0 < \epsilon_{t+1} < U | 0 < \epsilon_t < U] = \frac{\int_0^U \int_0^U f(\epsilon_t, \epsilon_{t+1})\, d\epsilon_t\, d\epsilon_{t+1}}{\int_0^U f(\epsilon_t)\, d\epsilon_t}, \tag{3.16}$$

$$p_{12} = P[0 < \epsilon_{t+1} < U | \epsilon_t \geq U] = \frac{\int_0^U \int_U^\infty f(\epsilon_t, \epsilon_{t+1})\, d\epsilon_t\, d\epsilon_{t+1}}{\int_U^\infty f(\epsilon_t)\, d\epsilon_t}, \tag{3.17}$$

$$p_{21} = P[\epsilon_{t+1} \geq U | 0 < \epsilon_t < U] = \frac{\int_U^\infty \int_0^U f(\epsilon_t, \epsilon_{t+1})\, d\epsilon_t\, d\epsilon_{t+1}}{\int_0^U f(\epsilon_t)\, d\epsilon_t}. \tag{3.18}$$

By using (3.7), the expected of trading duration ($E(TD_U)$) for this case is:

$$E(TD_U) = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}}. \tag{3.19}$$

For this case, we can see that the trading duration is influenced by σ_ϵ because the bivariate Normal distribution, $f(\epsilon_t, \epsilon_{t+1})$, is a function of σ_ϵ. However, β still does not have any impact to the trading duration, because the trading duration only depend on the distribution of $\{\epsilon_t\}$.

In Proposition 3.1, we show that the minimum expected trading duration is 1. It happens when $p_{11} = p_{22} = p_{12} = p_{21} = 0$ in (3.19) and it will be achieved if ϕ is close to -1.

Proposition 3.1 *If $U \geq 0$ and $\phi \downarrow -1$ then $E(TD_U) \to 1$ which is the minimum trading duration.*

Proof:
Consider

$$\int_0^\infty \int_0^\infty f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1} \tag{3.20}$$

where

$$f(\epsilon_t, \epsilon_{t+1}) = \frac{1}{2\pi\sigma_\epsilon^2\sqrt{1 - \phi^2}} \exp\left\{-\frac{\epsilon_t^2 - 2\phi\ \epsilon_t\ \epsilon_{t+1} + \epsilon_{t+1}^2}{2\sigma_\epsilon^2(1 - \phi^2)}\right\}.$$

As $\phi \downarrow -1$, without loss generality, assume that $-1 < \phi < 0$. Since $\epsilon_t, \epsilon_{t+1} \geq 0$ in (3.20), then

$$\exp\left\{-\frac{\epsilon_t^2 - 2\phi\ \epsilon_t\ \epsilon_{t+1} + \epsilon_{t+1}^2}{2\sigma_\epsilon^2(1 - \phi^2)}\right\} \leq \exp\left\{-\frac{\epsilon_t^2 + \epsilon_{t+1}^2}{2\sigma_\epsilon^2(1 - \phi^2)}\right\}.$$

So,

$$\begin{aligned} 0 \ &\leq\ A = \int_0^\infty \int_0^\infty f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1} \\ &\leq\ \frac{1}{2\pi\sigma_\epsilon^2\sqrt{1 - \phi^2}} \int_0^\infty \int_0^\infty \exp\left\{-\frac{\epsilon_t^2 + \epsilon_{t+1}^2}{2\sigma_\epsilon^2(1 - \phi^2)}\right\}\ d\epsilon_t\ d\epsilon_{t+1}. \end{aligned}$$

Let $\epsilon_t = r\sin(\theta)$, and $\epsilon_{t+1} = r\cos(\theta)$. Then,

$$0 \leq A \leq \frac{1}{2\pi\sigma_\epsilon^2\sqrt{1 - \phi^2}} \int_0^{\pi/2} \int_0^\infty \exp\left\{-\frac{r^2}{2\sigma_\epsilon^2(1 - \phi^2)}\right\}\ r\ dr\ d\theta$$

Let $x = \frac{r^2}{2\sigma_\epsilon^2(1-\phi^2)}$, so $dx = \frac{rdr}{\sigma_\epsilon^2(1-\phi^2)}$. Hence,

$$0 \leq A \leq \frac{\sqrt{1 - \phi^2}}{2\pi} \int_0^{\pi/2} \int_0^\infty \exp(-x)\ dx\ d\theta = \frac{\sqrt{1 - \phi^2}}{4}$$

Thus, $0 \leq A \leq 0$ as $\phi \downarrow -1$,

$$\lim_{\phi \downarrow -1} A = \lim_{\phi \downarrow -1} \int_0^\infty \int_0^\infty f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1} = 0.$$

As the numerators in (3.15), (3.17), (3.18) and (3.16) are less than $\int_0^\infty \int_0^\infty f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1}$, hence,

$$\lim_{\phi \downarrow -1} p_{11}(\phi) = 0, \qquad \lim_{\phi \downarrow -1} p_{22}(\phi) = 0,$$

$$\lim_{\phi \downarrow -1} p_{21}(\phi) = 0, \qquad \lim_{\phi \downarrow -1} p_{12}(\phi) = 0.$$

As a result, the trading duration expectation in (3.19) tend to 1 as ϕ tend to -1.

To show that $\inf_{-1<\theta<1} E(TD_U) = 1$, with $U \geq 0$, assume that there is a value of U such that

$$E(TD_U) = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} < 1.$$

This condition can be happened if

$$\begin{aligned}
1 - p_{22} + p_{12} &< (1 - p_{11})(1 - p_{22}) - p_{12}p_{21} \\
1 &< 1 - p_{11} - p_{22} + p_{11}p_{22} - p_{12}p_{21} + p_{22} - p_{12} \\
0 &< p_{11}p_{22} - p_{12}p_{21} - p_{11} - p_{12} \\
0 &< p_{11}(p_{22} - 1) - p_{12}(1 + p_{21}).
\end{aligned}$$

This contradicts with the facts that $0 \leq p_{11} \leq 1$, $0 \leq p_{22} \leq 1$, $0 \leq p_{21} \leq 1$, and $0 \leq p_{12} \leq 1$.

3.4 Transition Matrix Construction for Inter-trades Interval in Pairs Trading Strategy

Consider again the pairs trading strategy based on the cointegration coefficients weighted (CCW) rule from Lin *et al.* (2006) explained in Subsection 2.3.4. For upper-trades case, inter-trade interval is the time needed to open another upper-trade after the previous upper-trade is unwound. Similar to transition matrix construction for trade duration in 3.3, the state space of the Markov chain for inter-trade interval becomes as follows:

State 1: $\epsilon_t \leq 0$

State 2: $0 < \epsilon_t < U$

State 3: $\epsilon_t \geq U$

The inter-trade interval is analogous to the first passage-time from state 1 to state 3. Following the derivation of the first passage-time in Subsection 3.2, we will derive the inter-trade interval for white noise and AR(1) processes. Without loss generality, only upper-trades case are derived as white noise and AR(1) processes are symmetric, so that we can consider that the trade interval for lower trades is the same as the inter-trade interval for upper trades.

3.4.1 Case 1: White Noise Process

Similar to trade duration derivation for a white noise process in Section 3.3.1, p_{11}, p_{22}, p_{12} and p_{21} for inter-trade interval estimation can be calculated as follow:

$$\begin{aligned}
p_{11} &= P[\epsilon_{t+1} \leq 0 | \epsilon_t \leq 0] = P[\epsilon_{t+1} \leq 0], \\
p_{22} &= P[0 < \epsilon_{t+1} < U | 0 < \epsilon_t < U] = P[0 < \epsilon_{t+1} < U], \\
p_{21} &= P[\epsilon_{t+1} \geq U | 0 < \epsilon_t < U] = P[\epsilon_{t+1} \geq U] = p_{11}, \\
p_{12} &= P[0 < \epsilon_{t+1} < U | \epsilon_t \leq 0] = P[0 < \epsilon_{t+1} < U] = p_{22}.
\end{aligned}$$

By using (3.7), the expected of inter-trade interval $E(IT_U)$ for this case is:

$$E(IT_U) = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}} = \frac{1}{1 - (p_{11} + p_{22})}.$$

Note that

$$p_{11} + p_{22} = \int_{-\infty}^{0} f(\epsilon_t)\ d\epsilon_t + \int_{0}^{U} f(\epsilon_t)\ d\epsilon_t = \Phi\left(\frac{U}{\sigma_\epsilon}\right),$$

where $\Phi(.)$ denotes the cumulative distribution of a standard normal distribution. Thus,

$$E(IT_U) = \left(1 - \Phi\left(\frac{U}{\sigma_\epsilon}\right)\right)^{-1}. \tag{3.21}$$

Therefore, for this case, the inter-trade interval depends on the values of the cointegration error variance σ_ϵ^2 and the pre-set upper-bound U but not on the value of the cointegration coefficient β.

3.4.2 Case 2: AR(1) Process

Similar to trade duration derivation for an AR(1) process in Subsection 3.3.2, p_{11}, p_{22}, p_{12} and p_{21} for inter-trade interval estimation can be calculated as follow:

$$p_{11} = P[\epsilon_{t+1} \leq 0 | \epsilon_t \leq 0] = \frac{\int_{-\infty}^{0} \int_{-\infty}^{0} f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1}}{\int_{-\infty}^{0} f(\epsilon_t)\ d\epsilon_t}, \tag{3.22}$$

$$p_{22} = P[0 < \epsilon_{t+1} < U | 0 < \epsilon_t < U] = \frac{\int_{0}^{U} \int_{0}^{U} f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1}}{\int_{0}^{U} f(\epsilon_t)\ d\epsilon_t}, \tag{3.23}$$

$$p_{12} = P[0 < \epsilon_{t+1} < U | \epsilon_t \leq 0] = \frac{\int_0^U \int_{-\infty}^0 f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1}}{\int_{-\infty}^0 f(\epsilon_t)\ d\epsilon_t}, \quad (3.24)$$

$$p_{21} = P[\epsilon_{t+1} \leq 0 | 0 < \epsilon_t < U] = \frac{\int_{-\infty}^0 \int_0^U f(\epsilon_t, \epsilon_{t+1})\ d\epsilon_t\ d\epsilon_{t+1}}{\int_0^U f(\epsilon_t)\ d\epsilon_t}. \quad (3.25)$$

By using (3.7), the expected of inter-trade interval ($E(IT_U)$) for this case is:

$$E(IT_U) = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}\ p_{21}}. \quad (3.26)$$

For this case, we can see that the inter-trade interval is influenced by σ_ϵ and the pre-set upper-bound U. However, β still does not have any impact to the inter-trade interval.

3.5 Simulation on the Evaluation of the Trade Duration and Inter-trades Interval

Cointegration error series $\{\epsilon_t\}$ data are generated from simulations. Equation (3.10) is used to generate $\{\epsilon_t\}$ for simulations where a_t in (3.10) are generated from a standard normal distribution with zero mean and standard deviation σ_a. Then, $\{\epsilon_t\}$ is generated with a_t and θ. Totally 50 independent simulations are carried out and each simulation has 1000 observations. To compare the trading duration and inter-trade interval with different variances at a fixed value of U, for simplicity we choose $U = 1.5$. Following the pairs trading strategy mentioned in Subsection 2.3.4, if $\epsilon_t \geq U$, we will open a pair trade, and then close the trade if $\epsilon_t \leq 0$. If there is an up-crossing of upper-bound U, but the previous open trade has not been closed, we can not open another pair trade. We calculate the average of trade duration and inter-trade interval for each simulation and then we take again the average from the 50 simulations. The simulation results are compared to the theoretical values.

From Table 3.1, except for inter-trade interval with $\sigma_a = 0.75$ for all $\theta = (0.5, 0, -0.5)$ producing slightly different results to the theoretical values, for other values of σ_a, the values from simulations are very close to the theoretical values. It follows that the Markov chain approach can be used as an approximation to determine trading duration and inter-trade interval for the cointegration error following white noise or AR(1) processes. Computer program for Table 3.1 is provided in Appendix A.1.

Table 3.1: Trade duration and inter-trade interval in theory and simulation with pre-set upper-bound $U = 1.5$

σ_a	Trade Duration		Inter-trade Interval	
	Theoretical	Simulation	Theoretical	Simulation
$\phi = 0.5$				
0.75	3.9570	4.0499 (0.5439)	38.9959	31.0623 (4.9090)
1	3.8132	3.8806 (0.3978)	15.3101	15.5109 (1.5794)
2	3.4471	3.4172 (0.2761)	6.2209	6.1380 (0.5571)
$\phi = 0$				
0.75	2	2.0197 (0.3485)	43.9558	42.6117 (7.3821)
1	2	2.0319 (0.1823)	14.9684	14.9040 (2.0706)
2	2	2.0066 (0.1346)	4.4125	4.4659 (0.2885)
$\phi = -0.5$				
0.75	1.1807	1.1959 (0.1063)	27.2184	24.0312 (4.5256)
1	1.2575	1.2550 (0.0601)	9.4585	9.1442 (0.9594)
2	1.3905	1.3748 (0.0519)	3.0293	3.0287 (0.1697)

Note: values in the parentheses are the standard deviations.

3.6 Conclusion

In this chapter, the problem to determine trade duration and inter-trade interval for cointegration error following white noise or AR(1) processes in pairs trading strategy is solved using Markov chain approach. The results from simulations are actually encouraging. However, it is difficult to extend the results for general AR(p) processes, $p > 1$. Even for a simple AR(1) process, it involves a double integral to calculate $p_{i,j}$. Therefore, in the next chapter another approach, namely integral equation approach, is used.

Chapter 4

Optimal Threshold for Pairs Trading Using Integral Equation Approach

4.1 Introduction

In Chapter 3, we derived the expected values of trade duration and inter-trade interval for pairs trading using Markov chain approach. We applied the derivation for cointegration error $\{\epsilon_t\}$ following a white noise or an AR(1) process. However, we find difficulty to extend the results for general AR(p) processes, $p > 1$. Therefore, in this chapter, we derive the expected of trade duration and inter-trade interval for pairs trading using integral equation approach. Using those trade duration and inter-trade interval estimates and for a given upper-bound U, a number of trade during a trading period is derived. Then, we propose a numerical algorithm to calculate the optimal pre-set upper-bound U_o. Firstly, we apply for an AR(1) process and then for an AR(2) process.

4.2 Mean First-Passage Time for AR(p) Processes Using Integral Equation Approach

Background material in this section is mainly taken from Basak and Ho (2004). We define a discrete-time real-valued Markov process $\{y_t\}$ on a probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$. Consider $\{y_t\}$ is an AR(p) process as follow [1]:

$$y_t = \phi_1\ y_{t-1} + \phi_2\ y_{t-2} + \ldots + \phi_p\ y_{t-p} + \xi_t \tag{4.1}$$

where $\{\xi_t\}$ are IID $N(0, \sigma_\xi^2)$ random variables. Now, if $\mathbf{y}_t = (y_{t-p+1}, \ldots, y_{t-1}, y_t)'$, the state vector $\mathbf{y}_t$ consists of exactly the past p states of the original process. Suppose that $\mathbf{y}_0 = (y_{-p+1}, \ldots, y_{-1}, y_0)'$ and that we are given level $b > a$, where $\mathbf{y}_0 \in [a, b]^p$.

[1] Without loss of generality, it is assumed that $E(y_t) = 0$.

Define

$$P_n(\mathbf{y}_0) := P\left(a \le y_1 \le b, \ldots, a \le y_{n-1} \le b, y_n > b \text{ or } y_n < a|\mathbf{y}_0\right).$$

By looking at the first step and using the Markov property, we have, for $n \ge 2$,

$$P_n(\mathbf{y}_0) = \int_a^b P_{n-1}(\mathbf{y}_1)\ f(\mathbf{y}_0, y)\ dy \tag{4.2}$$

where the term $f(\mathbf{y}_0, y)$ denotes the transition density of reaching y at the next step given that the present state is $\mathbf{y}_0$, $\mathbf{y}_0 = (y_{-p+1}, \ldots, y_{-1}, y_0)'$ and $\mathbf{y}_1 = (y_{-p+2}, \ldots, y_0, y)'$.

Define $P_n(\mathbf{y}_j, z) = P_n(\mathbf{y}_j)z^n$, $j = 0, 1 \ldots$ and $0 < z \le 1$ is a radius of convergence. Thus,

$$\begin{aligned} P_n(\mathbf{y}_0)z^n &= z\int_a^b P_{n-1}(\mathbf{y}_1)z^{n-1}\ f(\mathbf{y}_0, y)\ dy \\ P_n(\mathbf{y}_0, z) &= z\int_a^b P_{n-1}(\mathbf{y}_1, z)\ f(\mathbf{y}_0, y)\ dy. \end{aligned} \tag{4.3}$$

Summing the term $P_n(\mathbf{y}_0, z)$ in (4.3) for $n \ge 2$ gives

$$\begin{aligned} \sum_{n=2}^{\infty} P_n(\mathbf{y}_0, z) &= z\int_a^b \sum_{n=1}^{\infty} P_n(\mathbf{y}_1, z) f(\mathbf{y}_0, y) dy, \\ P(\mathbf{y}_0, z) &= z\int_a^b P(\mathbf{y}_1, z) f(\mathbf{y}_0, y) dy + zP_1(\mathbf{y}_0), \end{aligned} \tag{4.4}$$

where $P(\mathbf{y}_j, z) := \sum_{n=1}^{\infty} P_n(\mathbf{y}_j, z) = \sum_{n=1}^{\infty} P_n(\mathbf{y}_j)z^n$, $0 < z \le 1$, $j = 0, 1$. $P(\mathbf{y}_0, z)$ in (4.4) is a probability generating function for the first-passage time of $\{y_t\}$ crossing an interval [a,b] for the first time given $\mathbf{y}_0$. Equation (4.4) is a Fredholm integral equation of the second kind. Basak and Ho (2004) provided theorems below to guarantee the existence and uniqueness of a solution (4.4).

Theorem 4.1 *(Theorem 4.1 in Basak and Ho (2004)) The equation (4.4) exhibits a unique solution in [a,b] given that*

$$\int_a^b f(y_0, y) dy < 1 \textit{ for all } y_0 \in [a, b]. \tag{4.5}$$

Theorem 4.2 *(Theorem 4.2 in Basak and Ho (2004)) Under the condition (4.5), $P(\mathcal{T}_{a,b}(y_0) < \infty) = 1$. Here $\mathcal{T}_{a,b}(y_0)$ denotes the exit time of interval [a,b] provided that the Markov process starts at $y_0 \in [a, b]$.*

The first-passage time $\mathcal{T}_{a,b}(\mathbf{y}_0)$ is defined as

$$\mathcal{T}_{a,b}(\mathbf{y}_0) = \min\{t : y_t > b \text{ or } y_t < a|\mathbf{y}_0 \in [a, b]^p\} \tag{4.6}$$

where $\mathbf{y}_0 = (y_{-p+1}, \ldots, y_{-1}, y_0)'$ and $y_t \in [a, b]$ for $t = (-p + 1), \ldots, -1, 0$.

Particularly,

$$\mathcal{T}_a(\mathbf{y}_0) = \mathcal{T}_{a,\infty}(\mathbf{y}_0) = \min\{t : y_t < a|\mathbf{y}_0\} \tag{4.7}$$

where $y_t > a$ for $t = (-p+1), \ldots, -1, 0$ and

$$\mathcal{T}_b(\mathbf{y}_0) = \mathcal{T}_{-\infty,b}(\mathbf{y}_0) = \min\{t : y_t > b | \mathbf{y}_0\} \tag{4.8}$$

where $y_t < b$ for $t = (-p+1), \ldots, -1, 0$.

Let $E(\mathcal{T}_{a,b}(\mathbf{y}_0))$, $E(\mathcal{T}_a(\mathbf{y}_0))$ and $E(\mathcal{T}_b(\mathbf{y}_0))$ denote the mean first-passage time of $\mathcal{T}_{a,b}(\mathbf{y}_0)$, $\mathcal{T}_a(\mathbf{y}_0)$ and $\mathcal{T}_b(\mathbf{y}_0)$ respectively.

For an AR(p) process, (4.4) becomes

$$P(\mathbf{y}_0, z) = \frac{z}{\sqrt{2\pi}\sigma_\xi} \int_a^b P(\mathbf{y}_1, z) \exp\left[-\frac{(y - \boldsymbol{\phi}'\ \mathbf{y}_0)^2}{2\sigma_\xi^2}\right] dy + zP_1(\mathbf{y}_0) \tag{4.9}$$

where $\boldsymbol{\phi} = (\phi_p, \ldots, \phi_1)'$.

As Equation (4.9) is the probability generating function of first-passage time for an AR(p) process, differentiating this equation with respect to z and evaluating at $z = 1$ will give the mean first-passage time. Thus,

$$\begin{aligned}
E(\mathcal{T}_{a,b}(\mathbf{y}_0)) &= \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b E(\mathcal{T}_{a,b}(\mathbf{y}_1)) \exp\left(-\frac{(y - \boldsymbol{\phi}'\ \mathbf{y}_0)^2}{2\sigma_\xi^2}\right) dy \\
&\quad + \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b P(\mathbf{y}_1, 1) \exp\left(-\frac{(y - \boldsymbol{\phi}'\ \mathbf{y}_0)^2}{2\sigma_\xi^2}\right) dy + P_1(\mathbf{y}_0) \qquad (4.10) \\
&= \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b E(\mathcal{T}_{a,b}(\mathbf{y}_1)) \exp\left(-\frac{(y - \boldsymbol{\phi}'\ \mathbf{y}_0)^2}{2\sigma_\xi^2}\right) dy + P(\mathbf{y}_0, 1) \\
&= \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b E(\mathcal{T}_{a,b}(\mathbf{y}_1)) \exp\left(-\frac{(y - \boldsymbol{\phi}'\mathbf{y}_0)^2}{2\sigma_\xi^2}\right) dy + 1 \qquad (4.11)
\end{aligned}$$

where the last term is equal to 1 by Theorem 4.2.

4.3 Pairs Trading with Cointegration Error Following an AR(1) Process

4.3.1 Mean First-passage Time Numerical Scheme for an AR(1) Process

Consider again an AR(1) process:

$$y_t = \phi\ y_{t-1} + \xi_t, \tag{4.12}$$

where $-1 < \phi < 1$ and $\xi_t \sim$ IID $N(0, \sigma_\xi^2)$.

Following (4.11) with $\mathbf{y}_0 = y_0 \in [a, b]$ and $\boldsymbol{\phi} = \phi$, the mean first-passage time for an AR(1) process in (4.12) will become:

$$E(\mathcal{T}_{a,b}(y_0)) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b E(\mathcal{T}_{a,b}(u)) \exp\left(-\frac{(u-\phi y_0)^2}{2\sigma_\xi^2}\right) du + 1. \tag{4.13}$$

The integral equation in (4.13) is a Fredholm type of the second kind and it can be solved numerically by using the Nystrom method (Atkinson, 1997). Basak and Ho (2004) did not provide the numerical scheme for mean first-passage time of an AR(1) process. Therefore, in this subsection, we provide a numerical scheme based on the Nystrom method (Atkinson, 1997).

Since $E(\mathcal{T}_{a,b}(y_0))$ converges monotonically to $E(\mathcal{T}_b(y_0))$ as $a \to -\infty$, the approximation of $E(\mathcal{T}_b(y_0))$ can be obtained by evaluating $E(\mathcal{T}_{a,b}(y_0))$, then, letting $a \to -\infty$.

Define $h = (b-a)/n$, where n is the number of partitions in $[a, b]$ and h is the length of each partition. Recall the trapezoid integration rule (Atkinson, 1997):

$$\int_a^b f(u)du \approx \frac{h}{2}\left[w_0 f(u_0) + w_1 f(u_1) + \cdots + w_{n-1} f(u_{n-1}) + w_n f(u_n)\right], \tag{4.14}$$

where $u_0 = a, u_j = a+jh, u_n = b, j = 1, \ldots, n$ and the weights w_j for the corresponding nodes are

$$w_j = \begin{cases} 1, & \text{for } j = 0 \text{ and } j = n, \\ 2, & \text{for } j = 1, \ldots, (n-1). \end{cases}$$

The integral term in (4.13) can be approximated by

$$\int_a^b E(\mathcal{T}_{a,b}(u)) \exp\left(-\frac{(u-\phi y_0)^2}{2\sigma_\xi^2}\right) du \approx \frac{h}{2} \sum_{j=0}^{n} w_j E(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi y_0)^2}{2\sigma_\xi^2}\right). \tag{4.15}$$

Let $E_n(\mathcal{T}_{a,b}(y_0))$ denote the approximation of $E(\mathcal{T}_{a,b}(y_0))$ using n partitions. Thus, the expectation in (4.13) using n partitions can be estimated by

$$E_n(\mathcal{T}_{a,b}(y_0)) \approx \frac{h}{2\sqrt{2\pi}\sigma_\xi} \sum_{j=0}^{n} w_j E_n(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi y_0)^2}{2\sigma_\xi^2}\right) + 1. \tag{4.16}$$

Set y_0 as u_i for $i = 0, 1, \ldots, n$ and reformulate (4.16) as follows

$$E_n(\mathcal{T}_{a,b}(u_i)) - \sum_{j=0}^{n} \frac{h}{2\sqrt{2\pi}\sigma_\xi} w_j E_n(\mathcal{T}_{a,b}(u_j)) \exp\left(-\frac{(u_j-\phi u_i)^2}{2\sigma_\xi^2}\right) = 1. \tag{4.17}$$

In matrices terms, (4.17) becomes

$$\begin{pmatrix} 1-K(u_0,u_0) & -K(u_0,u_1) & \dots & -K(u_0,u_n) \\ -K(u_1,u_0) & 1-K(u_1,u_1) & \dots & -K(u_1,u_n) \\ \vdots & \vdots & \vdots & \vdots \\ -K(u_n,u_0) & -K(u_n,u_1) & \dots & 1-K(u_n,u_n) \end{pmatrix} \begin{pmatrix} E_n(\mathcal{T}_{a,b}(u_0)) \\ E_n(\mathcal{T}_{a,b}(u_1)) \\ \vdots \\ E_n(\mathcal{T}_{a,b}(u_n)) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{4.18}$$

where

$$K(u_i,u_j) = \frac{h}{2\sqrt{2\pi}\sigma_\xi} w_j \exp\left(-\frac{(u_j-\phi u_i)^2}{2\sigma_\xi^2}\right), \quad i,j=0,1,\dots,n.$$

Solving the linear equations in (4.18), one will obtain approximation values of $E_n(\mathcal{T}_{a,b}(u_j))$. The mean first-passage time of crossing [a,b], with an initial point $y_0 \in [a,b]$, can be approximated by substituting the values of $E_n(\mathcal{T}_{a,b}(u_j))$ into (4.16).

Examples of the mean first-passage time using an integral equation approach and simulation for some AR(1) processes are provided in Table 4.1. For given σ_ξ^2, ϕ, and $y_0 = 1.5$, the mean first-passage time of level zero, using an integral equation approach, is calculated. We use different b and n in order to make the length of partition h the same for each case. Results show that $h = 0.1$ is enough to get results similar to the simulation. For simulation, we generate an AR(1) process as in (4.12) for given σ_ξ^2, and ϕ. Using the initial state $y_0 = 1.5$, the time needed for the process to cross zero for the first time is calculated. The simulation is repeated 10000 times and then we calculate the average. Table 4.1 shows that simulation results and results from the integral equation approach are very close. Computer program for Table 4.1 is provided in Appendix A.2.1.

Table 4.1: Mean first-passage time of level 0, given $y_0 = 1.5$ for y_t in (4.12).

σ_ξ^2	ϕ	Integral equation		Simulation
0.49	0.5	3.9181	(b = 5, n= 50)	3.9419 (2.8605)
	0.0	2.0000	(b =5, n = 50)	1.9918 (1.4662)
	-0.5	1.2329	(b=5, n= 50)	1.2341 (0.6405)
1.00	0.5	3.5401	(b = 7, n= 70)	3.5571 (2.7667)
	0.0	2.0000	(b = 7, n= 70)	2.0055 (1.4058)
	-0.5	1.3666	(b=7, n= 70)	1.3636 (0.8186)
4.00	0.5	3.0467	(b = 14, n= 140)	3.0512 (2.6549)
	0.0	2.0000	(b = 14, n = 140)	1.9959 (1.4168)
	-0.5	1.5626	(b=14, n= 140)	1.5725 (0.9496)

Note: values in the parentheses are the standard deviations.

4.3.2 Trade Duration and Inter-trade Intervals for an AR(1) Process

Consider the cointegration error (ϵ_t) in (1.1) follows an AR(1) process, i.e.:

$$\epsilon_t = \phi\ \epsilon_{t-1} + a_t,\ \text{where } a_t \sim \text{ IID } N(0, \sigma_a^2). \tag{4.19}$$

As explained in Subsection 2.3.4, the trade duration is the time between opening and closing a trade. For a U-trade, a trade is opened when ϵ_t is higher than or equal to the pre-set upper-bound U and it is closed when ϵ_t is less than or equal to 0 which is the mean of the cointegration error. Suppose ϵ_o is at U , so a U-trade is opened. To calculate the expected trade duration, we would like to know the time needed on average for ϵ_t to pass 0 for the first time. Thus, calculating the expected trade duration is the same as calculating the mean first-passage time for ϵ_t to pass 0 for the first time, given the initial value is U. Let TD_U denote the expected trade duration corresponding to the pre-set upper-bound U. Using (4.13), TD_U is defined as follows:

$$TD_U := E(\mathcal{T}_{0,\infty}(U)) = \lim_{b\to\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_0^b E(\mathcal{T}_{0,b}(s)) \exp\left(-\frac{(s-\phi U)^2}{2\sigma_a^2}\right)\ ds + 1. \tag{4.20}$$

As for trade duration, the inter-trade interval is the waiting time needed to open a trade after the previous trade is closed. For a upper-trade, if there is an open upper-trade while ϵ_t is at 0 during trading, the trade has to be closed. To calculate the expected inter-trade interval, we would like to know the time needed on average for ϵ_t to pass the pre-set upper-bound U for the first time, so we can open an upper-trade again. Thus, calculating the expected inter-trade interval is the same as calculating the mean first-passage time for ϵ_t to pass U given the initial value is 0. Let IT_U denote the expected inter-trade interval for the pre-set upper-bound U.

$$IT_U := E(\mathcal{T}_{-\infty,U}(0)) = \lim_{-b\to-\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_{-b}^U E(\mathcal{T}_{-b,U}(s)) \exp\left(-\frac{s^2}{2\sigma_a^2}\right)\ ds + 1. \tag{4.21}$$

4.3.3 Number of Trades Over a Trading Horizon for an AR(1) Process

The expected number of upper-trades $E(N_{UT})$ and the expected number of periods corresponding to upper-trades $E(N_{UP})$ over a time horizon [0,T] are defined as follow:

$$E(N_{UT}) = \sum_{k=1}^{\infty} kP(N_{UT} = k) \quad \text{ and } \quad E(N_{UP}) = \sum_{k=1}^{\infty} kP(N_{UP} = k).$$

In this subsection, we want to derive the expected number of upper-trades $E(N_{UT})$ over a specified trading horizon. It is difficult to evaluate the exact value of $E(N_{UT})$. Thus, a possible range of values of $E(N_{UT})$ is provided.

We define $Period_U$ as the sum of the trade duration and the inter-trade interval for upper-trades given an upper-bound U(see Figure 1.1 in Chapter 1). Thus, the expected $Period_U$ is given by,

$$E(Period_U) = TD_U + IT_U.$$

First, we will evaluate the expected number of $Period_U$'s $E(N_{UP})$ in the time horizon [0,T] as it has a direct connection to the trade duration and the inter-trade interval. Then, the relationship of N_{UT} and N_{UP} will be used to obtain a possible range of values of $E(N_{UT})$.

Let $Period_{Ui}$ denote the length of the period corresponding to the ith upper-trade. Thus,

$$T \geq E\left(\sum_{i=1}^{N_{UP}} (Period_{Ui})\right) = \sum_{k=1}^{\infty} \left[\sum_{i=1}^{k} E(Period_{Ui})\right] P(N_{UP} = k). \tag{4.22}$$

Since the period depends on the distribution of ϵ_t, which is a stationary time series, $E(Period_{Ui})$ will be the same for all i. Thus, $E(Period_{Ui}) = E(Period_U)$ and

$$T \geq E(Period_U) \sum_{k=1}^{\infty} kP(N_{UP} = k) = E(Period_U)E(N_{UP}). \tag{4.23}$$

Thus,

$$E(N_{UP}) \leq \frac{T}{E(Period_U)} = \frac{T}{TD_U + IT_U}. \tag{4.24}$$

As for the derivation that leads to (4.24),

$$T < E\left(\sum_{i=0}^{N_{UP}+1} (Period_{Ui})\right) = E(Period_U)E(N_{UP} + 1), \tag{4.25}$$

giving

$$E(N_{UP}) > \frac{T}{E(Period_U)} - 1 = \frac{T}{TD_U + IT_U} - 1. \tag{4.26}$$

Thus,

$$\frac{T}{TD_U + IT_U} \geq E(N_{UP}) > \frac{T}{TD_U + IT_U} - 1. \tag{4.27}$$

However, the relationship between number of upper-trades (N_{UT}), and number of $Period_U$'s (N_{UP}) is $N_{UT} = N_{UP}$ in the case of the last upper-trade is not closed before T or $N_{UT} = N_{UP} + 1$ in the case of the last upper trade is closed before T.

Thus,

$$\frac{T}{TD_U + IT_U} + 1 \geq E(N_{UP}) + 1 \geq E(N_{UT}) \geq E(N_{UP}) > \frac{T}{TD_U + IT_U} - 1. \tag{4.28}$$

Comparing the number of upper-trades results in Tables 4.2 from the integral equation approach and simulation, we see that for $\phi = 0.5$, the estimates of the number of upper-trades using the integral equation are higher than those given by the simulation results. The differences are getting larger when σ_a^2 is larger. For example, the number of upper-trades estimate is around 21 by integral equation approach which is similar to the result from the simulation for $\sigma_a^2 = 0.49$ but the number of upper-trades estimate is around 115 by integral equation approach while the result from simulation is 103 for $\sigma_a^2 = 4$. The opposite happens if $\phi = -0.5$ where the estimates of the number of U-trades using the integral equation are smaller than those given by the simulation results. The differences are due to a slight difference in the framework underpinning the theory of integral equations and that for simulation from the simulation data as by using the real data we usually open a trade at slightly higher than U, not exactly at U. Computer program for Table 4.2 is provided in Appendix A.2.1.

Table 4.2: Estimation of the number of upper-trades with $U = 1.5$.

		Integral Equation			Simulation			
ϕ	σ_a^2	TD_U	I_U	$\hat{N}_{UT} = \frac{1000}{TD_U+I_U} - 1$	TD_U	IT_U	N_{UT} simulation	$\hat{N}_{UT} = \frac{1000}{TD_U+IT_U} - 1$
0.5	0.49	3.9181	40.6074	21.459	4.054(0.585)	42.801(12.340)	21.725(5.277)	20.342
	1	3.5401	14.6006	54.125	3.780(0.308)	15.153(1.838)	52.650(5.226)	51.817
	4	3.0469	5.5679	115.079	3.407(0.255)	6.254(0.466)	103.000(6.421)	102.508
-0.5	0.49	1.2329	32.6253	28.535	1.170(0.084)	28.958(4.760)	32.025(5.091)	32.191
	1	1.3666	10.523	83.1071	1.242(0.072)	9.206(0.952)	95.175(8.311)	94.706
	4	1.5626	3.6220	191.879	1.385(0.051)	3.030(0.151)	225.500(8.741)	225.500

Note: For the integral equation approach, we use $\hat{N}_{UT} = \frac{1000}{TD_U+IT_U} - 1$ to estimate the expected number of upper-trades within [0,T], T=1000. TD_U and IT_U are calculated by using (4.20) and (4.21) respectively. For simulation, 1000 observations are generated from the model described in (4.19) for each simulation with $\epsilon_0 = 0$. Simulation is independently repeated 50 times . The values in parentheses are the standard deviations. For each generated simulation data, we start to open a trade when ϵ_t exceeds $U = 1.5$ and close the trade when ϵ_t goes below zero. We record the average trade duration TD_U, the average inter-trade interval IT_U and the number of upper-trades N_{UT} for each simulation. At the end of all 50 repeated simulations, we calculate the mean of TD_U, IT_U and N_{UT} from all simulations as well as the standard deviations. Furthermore, the last column shows the number of trades using $\hat{N}_{UT} = \frac{1000}{TD_U+IT_U} - 1$.

4.3.4 Minimum Total Profit and the Optimal Pre-set Upper-bound for an AR(1) Process

This section will combine the pre-set minimum profit per trade in Lin *et al.* (2006) and the number of U-trades Subsection 4.3.3 to define minimum total profit (MTP) over the time horizon [0,T]. The optimal pre-set upper-bound, denoted by U_o, is determined by maximising the MTP.

Let TP_U denote the total profit from U-trades within the time horizon [0,T] for a pre-set upper-bound U. Thus,

$$TP_U = \sum_{i=1}^{N_{UT}} (\text{ Profit from the } i\text{th U-trade}) \tag{4.29}$$

assumed that the last U-trade is completed[2] before time T.

Using (2.39) and (4.28), i.e.:

$$\text{Profit per trade} \geq U \qquad \text{and} \qquad E(N_{UT}) \geq \frac{T}{TD_U + IT_U} - 1,$$

we define the minimum total profit within the time horizon [0,T] by

$$MTP(U) := \left(\frac{T}{TD_U + IT_U} - 1 \right) U.^3 \tag{4.30}$$

Then, considering all $U \in [0, b]$, the optimal pre-set upper-bound U_o is chosen such that $MTP(U_o)$ takes the maximum at that U_o. In practice, the value of b is set up as $5\sigma_\epsilon$ because ϵ_t is a stationary process, and the probability that $|\epsilon_t|$ is greater than $5\sigma_\epsilon$ is close to zero.[4]

The numerical algorithm to calculate the optimal pre-set upper-bound U_o is as follows:

1. Set up the value of b as $5\sigma_\epsilon$.
2. Decide a sequence of pre-set upper-bounds U_i, where $U_i = i \times 0.01$, and $i = 0, \ldots, b/0.01$.
3. For each U_i,
   (a) calculate $E(\mathcal{T}_{0,b}(U_i))$ as the trade duration (TD_{U_i}) using (4.20).
   (b) calculate $E(\mathcal{T}_{-b,U_i}(0))$ as the inter-trade interval (IT_{U_i}) using (4.21).

[2] We use this assumption to simplify the derivation. In the empirical examples section, we use the strategy that the last trade is forced to be closed at the end of the period if it has not been closed yet, possibly at a loss. We call it as an incomplete trade. If T is long enough to have significant number of trades, the loss from incomplete trade could be off set by the actual total profit from the previous complete trades, so that the average profit per trade is still higher than U.

[3] We adopt the notation $MTP(U)$ since the Minimum Total Profit is a function of U.

[4] σ_ϵ is the standard deviation of ϵ_t.

Table 4.3: Numerical results in determining the optimal pre-set upper-bound U_o

	$\phi = -0.8$		$\phi = -0.5$		$\phi = -0.2$	
σ_a^2	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o
0.25	91.7097	0.59	77.0414	0.50	66.4935	0.47
0.49	128.3609	0.83	107.8254	0.70	93.0673	0.65
1	183.3448	1.19	154.0117	1.00	132.9287	0.93
2.25	274.9967	1.78	230.9996	1.49	199.3710	1.40
4	366.6515	2.37	307.9922	1.99	265.8216	1.86
		$\approx 1.2\sigma_a$		$\approx \sigma_a$		$\approx 0.93\sigma_a$
		$\approx 0.72\sigma_\epsilon$		$\approx 0.87\sigma_\epsilon$		$\approx 0.91\sigma_\epsilon$

	$\phi = 0.2$		$\phi = 0.5$		$\phi = 0.8$	
σ_a^2	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o	$MTP(U_o)$	U_o
0.25	55.1798	0.47	46.7138	0.53	34.7004	0.70
0.49	77.2190	0.66	65.3655	0.74	48.5545	0.97
1	110.2877	0.95	93.3549	1.05	69.3438	1.39
2.25	165.4104	1.42	140.0095	1.58	103.9991	2.09
4	220.5361	1.89	186.6704	2.10	138.6582	2.78
		$\approx 0.95\sigma_a$		$\approx 1.05\sigma_a$		$\approx 1.4\sigma_a$
		$\approx 0.93\sigma_\epsilon$		$\approx 0.91\sigma_\epsilon$		$\approx 0.84\sigma_\epsilon$

(c) calculate $MTP(U_i) = \left(\frac{T}{TD_{U_i} + IT_{U_i}} - 1\right) U_i$.

4. Find $U_o \in \{U_i\}$ such that $MTP(U_o)$ is the maximum .

Examples of numerical results from some AR(1) processes described in (4.19) using $T = 1000$ are shown in Table 4.3. It shows that for a given ϕ, U_o increases as σ_a increases. The last two rows of each ϕ show the approximation of U_o as a proportion of σ_a and σ_ϵ. Those approximations can be used as a general rule in choosing U_o. For example if we have the adjusted cointegration error ϵ_t with an AR(1) process and the ϕ is -0.5 or 0.5, quickly we can choose $U_o = \sigma_a$. Computer program for Table 4.3 is provided in Appendix A.2.1.

4.4 Pairs Trading with Cointegration Error Following an AR(2) Processes

4.4.1 Mean First-passage Time for AR(2) Process

On probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$, define an AR(2) process as follows:

$$y_t = \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \xi_t, \tag{4.31}$$

where ξ_t are IID $N(0, \sigma_\xi^2)$.

Following (4.11), the mean first-passage time for an AR(2) process is

$$E\left(\mathcal{T}_{a,b}(\mathbf{y}_1)\right) = \frac{1}{\sqrt{2\pi}\sigma_\xi}\int_a^b E\left(\mathcal{T}_{a,b}(\mathbf{y}_1)\right)\exp\left(-\frac{\left(u_1-\boldsymbol{\phi}'\ \mathbf{y}_0\right)^2}{2\sigma_\xi^2}\right)\ du_1+1, \tag{4.32}$$

where $\mathbf{y}_0 = (y_{-1}, y_0)'$, $\mathbf{y}_1 = (y_0, u_1)'$ and $\boldsymbol{\phi} = (\phi_2, \phi_1)'$. Similarly,

$$E\left(\mathcal{T}_{a,b}(\mathbf{y}_1)\right) = \frac{1}{\sqrt{2\pi}\sigma_\xi}\int_a^b E\left(\mathcal{T}_{a,b}(\mathbf{y}_2)\right)\exp\left(-\frac{\left(u_2-\boldsymbol{\phi}'\ \mathbf{y}_1\right)^2}{2\sigma_\xi^2}\right)\ du_2+1 \tag{4.33}$$

where $\mathbf{y}_2 = (u_1, u_2)'$.

After substituting $E\left(\mathcal{T}_{a,b}(\mathbf{y}_1)\right)$ in (4.32) with (4.33), we obtain:

$$\begin{aligned}E\left(\mathcal{T}_{a,b}(\mathbf{y}_0)\right) &= \frac{1}{2\pi\sigma_\xi^2}\int_a^b\int_a^b E\left(\mathcal{T}_{a,b}(\mathbf{y}_2)\right)\exp\left(-\frac{\sum_{i=1}^2\left(\mathbf{e}_1'\ \mathbf{y}_i-\boldsymbol{\phi}'\ \mathbf{y}_{i-1}\right)^2}{2\sigma_\xi^2}\right)\ du_2\ du_1\\ &+\frac{1}{\sqrt{2\pi}\sigma_\xi}\int_a^b \exp\left(-\frac{\left(u_1-\boldsymbol{\phi}'\ \mathbf{y}_0\right)^2}{2\sigma_\xi^2}\right)\ du_1+1\end{aligned} \tag{4.34}$$

where $\mathbf{e}_1 = (0, 1)'$. Equation (4.34) is the same as Equation (4.15) in Basak and Ho (2004).

4.4.2 Numerical Procedures

Basak and Ho (2004) provided numerical scheme of the mean first-passage time for an AR(2) process. The first step of the numerical procedures is to divide the domain of integration into small triangles and apply a quadrature rule to perform numerical integration. Since our domain here simply the square $[a, b] \times [a, b]$, we naturally divide it into $2n^2$ triangles where n is a number of partition in $[a, b]$.

Suppose that $\triangle$ is one such triangle, with vertices $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$ and $v_3 = (x_3, y_3)$. Through the change of variables, $(x, y) = T(s, t)$, where

$$T(s,t) = (1-s-t)v_1 + sv_2 + tv_3, \qquad s,t \in [0,1].$$

For examples, $v_1 = T(0, 0)$, $v_2 = T(1, 0)$ and $v_3 = T(0, 1)$.

Based on the quadrature rule of a seven-nodes formula provided in Atkinson (1997),

$$\begin{aligned}\int_\triangle g(x,y)\ dx\ dy &\approx 2\ \text{area}(\triangle)\left(\frac{1}{40}\left[g(T(0,0))+g(T(0,1))+g(T(1,0))\right]\right.\\ &+\left.\frac{9}{40}g(T(1/3,1/3))+\frac{1}{15}\left[g(T(0,1/2)),g(T(1/2,1/2)),g(T(1/2,0))\right]\right)\end{aligned} \tag{4.35}$$

where a function $g : [a, b]^2 \to \mathbb{R}$ is continuous and integrable and $\mathbb{R}$ a real number space. This rule has degree of precision equal to 3. It means that there is no error if g is a polynomial of degree no greater than 3, and it can be derived using the method of undetermined coefficients. When it is used in the composite formula, the nodes

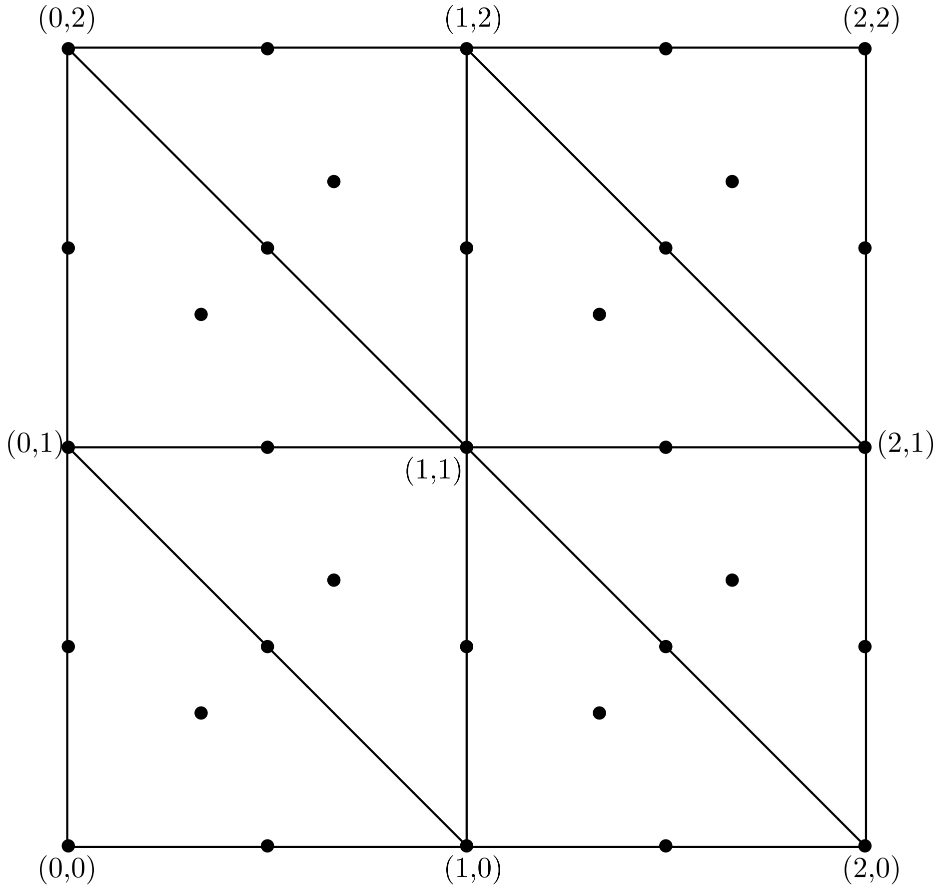


Figure 4.1: Example of determining the nodes based on the quadrature rule of a seven-node formula.

from adjacent triangles overlap. So, the total number of integration nodes will be $6n^2 + 4n + 1 = 2n^2 + (2n + 1)^2$ instead of $14n^2 = 7 \times 2n^2$.

For an example, see Figure 4.1. With $n = 2$, there are $2n^2 = 8$ triangles. Looking at the first triangle with 3 corner vertices $v_1 = (0, 0)$, $v_2 = (1, 0)$ and $v_3 = (0, 1)$, it has 7 nodes used to approximate the double integral, i.e.: {(0,0),(1,0),(0,1),(1/3,1/3),(1/2,0), (1/2,1/2),(0,1/2)} corresponding to {T(0,0),T(1,0),T(0,1),T(1/3,1/3),T(1/2,0), T(1/2,1/2),T(0,1/2)}. For the second triangle with 3 corner vertices $v_1 = (1, 1)$, $v_2 = (1, 0)$ and $v_3 = (0, 1)$, the 7 nodes are {(1,1),(1,0),(0,1),(2/3,2/3),(1,1/2),(1/2,1/2),(1/2,1)} corresponding to {T(0,0),T(1,0),T(0,1),T(1/3,1/3),T(1/2,0),T(1/2,1/2),T(0,1/2)}. Similarly for the other triangles. However, if we use all the 8 triangles, there are some overlap nodes. Thus, instead of having $7 \times 2n^2 = 56$ nodes, we only have $6n^2 + 4n + 1 = 32$ nodes.

Define

$$K(y_{-1}, y_0, u_1, u_2) = \frac{1}{2\pi\sigma_\xi^2} \exp\left(-\frac{\sum_{i=1}^{2} \left(\mathbf{e}_1' \ \mathbf{y}_i - \boldsymbol{\phi}' \ \mathbf{y}_{i-1}\right)^2}{2\sigma_\xi^2} \right) \tag{4.36}$$

where $\mathbf{y}_0 = (y_{-1}, y_0)'$,$\mathbf{y}_1 = (y_0, u_1)'$, $\mathbf{y}_2 = (u_1, u_2)'$, $\mathbf{e}_1 = (0, 1)'$, $\boldsymbol{\phi} = (\phi_2, \phi_1)'$ and σ_ξ^2 is the variance of ξ_t. We can now approximate the double integral in (4.34) by

$$\begin{aligned} &\int_a^b \int_a^b K(y_{-1}, y_0, u_1, u_2) E(\mathcal{T}_{a,b}(u_1, u_2)) \ du_1 \ du_2 \\ &\qquad \approx 2 \sum_{k=1}^{2n^2} \operatorname{area}(\triangle_k) \sum_{i=1}^{7} w_i K(y_{-1}, y_0, T_k(\mu_i)) E\left(\mathcal{T}_{a,b}(T_k(\mu_i))\right) \end{aligned} \tag{4.37}$$

where $\mu_i \in \{(1/3, 1/3), (0, 0), (0, 1), (1, 0), (0, 1/2), (1/2, 1/2), (1/2, 0)\}$ and $w_i \in \{9/40, 1/40, 1/15\}$ is the coefficient corresponding to μ_i. Instead of using $14n^2$ nodes, (4.37) can be evaluated by using $6n^2 + 4n + 1$ nodes as follow:

$$\begin{aligned} &\int_a^b \int_a^b K(y_{-1}, y_0, u_1, u_2) E(\mathcal{T}_{a,b}(u_1, u_2)) \ du_1 \ u_2 \\ &\qquad = \sum_{j=1}^{m} \omega_j K(y_{-1}, y_0, u_{1j}, u_{2j}) E(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) \end{aligned} \tag{4.38}$$

where $m = 6n^2 + 4n + 1$ and ω_j is the weight corresponding to the jth node. Thus, using n partitions in $[a, b]$, (4.34) can be approximated by

$$E_n(\mathcal{T}_{a,b}(y_{-1}, y_0)) \approx \sum_{j=1}^{m} \omega_j K(y_{-1}, y_0, u_{1j}, u_{2j}) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) + \psi(y_{-1}, y_0) \tag{4.39}$$

where

$$\psi(y_{-1}, y_0) = 1 + \frac{1}{\sqrt{2\pi}\sigma_\xi} \int_a^b \exp\left(-\frac{\left(u_1 - \boldsymbol{\phi}' \ \mathbf{y}_0\right)^2}{2\sigma_\xi^2} \right) du_1. \tag{4.40}$$

Given the values of a, b, $\mathbf{y}_0 = (y_{-1}, y_0)'$, $\boldsymbol{\phi}$ and σ_ξ, (4.40) can be obtained.

We can then evaluate the values of $E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j}))$ on the node points by solving the following system of linear equations:

$$E_n(\mathcal{T}_{a,b}(u_{1i}, u_{2i})) - \sum_{j=1}^{m} \omega_j K(u_{1i}, u_{2i}, u_{1j}, u_{2j}) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) = \psi(u_{1i}, u_{2i}) \tag{4.41}$$

where $i, j = 1, \ldots, m$.

Finally, using the values of $E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j}))$ from solving (4.41), we can use the Nystrom interpolation formula to obtain the value of $E_n(\mathcal{T}_{a,b}(y_{-1}, y_0))$:

$$E_n(\mathcal{T}_{a,b}(y_{-1}, y_0)) = \sum_{j=1}^{m} \omega_j K(y_{-1}, y_0, u_{1j}, u_{2j}) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) + \psi(y_{-1}, y_0) \tag{4.42}$$

Table 4.4: Examples of the mean first-passage time numerical results for an AR(2) process crossing 0 given initial values (y_{-1}, y_0).

(y_{-1}, y_0)	Simulation		Integral equation		
(ϕ_2, ϕ_1)	(0.5,1.5)	(2,1.5)	(0.5,1.5)	(2,1.5)	Note
(0.3,0.5)	6.3408 (6.4180)	7.7185 (6.9802)	6.4235	7.5454	b = 9, n = 18
(0.3,-0.5)	1.5255 (1.0601)	1.8490 (1.2398)	1.5381	1.8450	b = 9, n = 18
(-0.3,0.5)	2.5567 (1.6894)	2.1686 (1.5708)	2.5528	2.1662	b = 6, n = 12
(-0.3,-0.5)	1.2558 (0.6214)	1.1243 (0.4607)	1.2557	1.1268	b = 6, n = 12
(0.5,0.3)	6.0747 (6.4493)	7.9658 (7.1644)	6.0336	7.9330	b = 9, n = 18
(0.5,-0.3)	2.0553 (1.6506)	2.7987 (1.8705)	2.0545	2.7715	b = 9, n = 18
(-0.5,0.3)	1.8868 (1.1026)	1.4454 (0.8605)	1.8875	1.4456	b = 8, n = 16
(-0.5,-0.3)	1.3103 (0.6300)	1.0960 (0.3871)	1.3124	1.0993	b = 8, n = 16

Note: For simulation results, given initial values $\mathbf{y}_0 = (y_{-1}, y_0)'$, y_t for $t \geq 1$ are generated based on an AR(2) process in (4.31) with $\sigma_\xi^2 = 1$. The first-passage time is recorded when $y_t \leq 0$ for the first time. The simulation is repeated 10000 times and the mean of the first-passage time is reported in this table. The values in the parentheses are the standard deviations. Equation (4.34) is used for integral equation with $a = 0$ while b and n are defined in the last column.

where $j = 1, \ldots, m$.

The example of numerical results are provided in Table 4.4. We can see that the simulation results and integral equation results are very close. Computer program for Table 4.4 is provided in Appendix A.2.2.

4.4.3 Application to Pairs Trading Strategy

Consider the cointegration error, ϵ_t in (1.1) follows an AR(2) process, i.e:

$$\epsilon_t = \phi_1\ \epsilon_{t-1} + \phi_2\ \epsilon_{t-2} + a_t, \text{ where } a_t \sim \text{ IID } N(0, \sigma_a^2). \tag{4.43}$$

As for an AR(1) process, given a pre-set upper-bound U, we need to estimate trade duration TD_U and inter-trade interval IT_U using integral equation approach. Not like AR(1) process where the estimations only depend on ϵ_0 which is set up as U for calculating trading duration and the mean of ϵ_t for inter-trade interval, for an AR(2) process, the estimations also depend on ϵ_{-1}. We solve the problem by choosing ϵ_{-1} as the mean of ϵ_t. To calculate trade duration TD_U and inter-trade interval IT_U for an AR(2) process, we use (4.34) by modifying the initial vector state $\boldsymbol{\epsilon}_0 = (\epsilon_{-1}, \epsilon_0)'$ and the double integral region. Usually we choose $b = 5 * std(\epsilon_t)$ and since we assume $a_t \sim$ IID $N(0, \sigma_a^2)$, so $E(\epsilon_t) = 0$, thus the lower integrand, $a = 0$. Furthermore, as the mean of ϵ_t is 0, we set $\boldsymbol{\epsilon}_0 = (0, U)'$ for trade duration and $\boldsymbol{\epsilon}_0 = (0, 0)'$ for inter-trade interval. Thus, TD_U and IT_U are defined as follow:

$$TD_U := E\left(\mathcal{T}_{0,\infty}(0,U)\right) =$$
$$\lim_{b\to\infty} \frac{1}{2\pi\sigma_a^2} \int_0^b \int_0^b E\left(\mathcal{T}_{a,b}(u_1,u_2)\right) \exp\left(-\frac{\sum_{i=1}^2 \left(\mathbf{e}_1'(\boldsymbol{\epsilon}_i - \boldsymbol{\phi}'\ \boldsymbol{\epsilon}_{i-1}\right)^2}{2\sigma_a^2}\right)\, du_2\, du_1$$
$$+1 + \lim_{b\to\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_0^b \exp\left(-\frac{(u_1 - \boldsymbol{\phi}'\ \boldsymbol{\epsilon}_0)^2}{2\sigma_a^2}\right)\, du_1, \tag{4.44}$$

where $\mathbf{e}_1 = (0,1)'$, $\boldsymbol{\epsilon}_0 = (0,U)'$, $\boldsymbol{\epsilon}_1 = (U,u_1)'$, $\boldsymbol{\epsilon}_2 = (u_1,u_2)'$, and $\boldsymbol{\phi} = (\phi_2,\phi_1)'$.

$$IT_U := E\left(\mathcal{T}_{-\infty,U}(0,0)\right) =$$
$$\lim_{b\to-\infty} \frac{1}{2\pi\sigma_a^2} \int_b^U \int_b^U E\left(\mathcal{T}_{a,b}(u_1,u_2)\right) \exp\left(-\frac{\sum_{i=1}^2 \left(\mathbf{e}_1'\ \boldsymbol{\epsilon}_i - \boldsymbol{\phi}'\ \boldsymbol{\epsilon}_{i-1}\right)^2}{2\sigma_a^2}\right)\, du_2\, du_1$$
$$+1 + \lim_{b\to-\infty} \frac{1}{\sqrt{2\pi}\sigma_a} \int_b^U \exp\left(-\frac{(u_1 - \boldsymbol{\phi}'\ \boldsymbol{\epsilon}_0)^2}{2\sigma_a^2}\right)\, du_1, \tag{4.45}$$

where $\mathbf{e}_1 = (0,1)'$, $\boldsymbol{\epsilon}_0 = (0,0)'$, $\boldsymbol{\epsilon}_1 = (0,u_1)'$, $\boldsymbol{\epsilon}_2 = (u_1,u_2)'$, and $\boldsymbol{\phi} = (\phi_2,\phi_1)'$.

Table 4.5 shows the comparison of number of trades using integral equation approach and simulation with $U = 1.5$. We see that if the both of coefficients are positive, the results from integral equation approach are quite close to the simulation results. The opposite thing happens if ϕ_1 is negative which result in very different results compared to the simulation. The differences are due to a slight difference in the framework underpinning the theory of integral equations and that for simulation. For simulation we usually open an upper-trade at slightly higher than U, not exactly at U. However, since most of the results from integral equation are less or almost equal to the simulation results, we still can derive the minimum total profit (MTP) to generate the optimal upper-bound U_o that maximise the MTP. Computer program for Table 4.5 is provided in Appendix A.2.2.

Table 4.5: Number of upper-trades ($\hat{N}_{UT}$) for AR(2) processes with $U = 1.5$.

(ϕ_2, ϕ_1)	Simulation				Integral Equation		
	TD_U	IT_U	$\hat{N}_{UT}$	$\hat{N}_{UT}$ simulation	TD_U	IT_U	$\hat{N}_{UT}$
(0.3, 0.5)	7.2875 (1.0469)	17.9776 (2.9194)	38.5802	39.1400 (4.5085)	6.0199	17.7877	42.003
(0.3, -0.5)	1.1604 (0.0550)	5.3334 (0.6010)	152.9940	153.3000 (14.5788)	1.4478	9.1548	94.316
(-0.3, 0.5)	2.7113 (0.1867)	12.4512 (1.2950)	63.8895	65.5600 (5.3952)	2.6716	12.3538	66.554
(-0.3, -0.5)	1.3169 (0.0567)	9.5667(0.9586)	90.8753	91.5400 (8.0412)	1.3117	10.3655	85.637
(0.5, 0.3)	6.3234 (1.0123)	16.8054 (3.2551)	42.2362	42.9200 (5.9790)	5.3249	18.4235	42.108
(0.5, -0.3)	1.3139 (0.0979)	5.7954 (0.5756)	139.6608	139.4000 (12.0915)	1.8191	10.3912	81.898
(-0.5, 0.3)	2.0280 (0.0908)	10.3282 (1.0721)	79.9314	80.5600 (7.3183)	2.0359	10.7962	77.929
(-0.5, -0.3)	1.4413 (0.0655)	8.9418 (0.9405)	95.3106	95.9200 (8.7081)	1.4157	10.1947	86.129

Note: For simulation results, 1000 observations are generated from the model described in (4.43) for each simulation with $(\epsilon_{-1}, \epsilon_0) = (0, 0)$ are randomly drawn from $N(0, 1)$ and $a_t \sim$ IID $N(0, 1)$. Simulation is independently repeated 50 times. The values in parentheses are the standard deviations. For each generated simulation data, we start to open a trade when ϵ_t exceeds $U = 1.5$ and close the trade when ϵ_t goes below zero. We calculate the average trade duration TD_U, the average inter-trade interval IT_U and the number of upper-trades N_{UT} for each simulation. At the end of all 50 repeated simulations, we calculate the mean of TD_U, IT_U and N_{UT} from all simulations as well as the standard deviations. Furthermore, we also calculate the number of trades using $\hat{N}_{UT} = \frac{1000}{TD_U+IT_U} - 1$. For integral equation results, trade duration is calculated by using (4.44) with $(\epsilon_{-1}, \epsilon_0) = (0, 1.5)$ while inter-trade interval is calculated by using (4.45) with $(\epsilon_{-1}, \epsilon_0) = (0, 0)$. Number of trades is calculated using $\hat{N}_{UT} = \frac{1000}{TD_U+IT_U} - 1$.

Analogous to the derivation of MTP for AR(1) process in (4.30), let $MTP(U)$ denotes the minimum total profit corresponding to an upper-bound U:

$$\begin{aligned} MTP(U) &= \hat{N}_{UT} \times \text{Profit per trade} \\ &\geq \left(\frac{T}{TD_U + IT_U} - 1\right) U. \end{aligned} \tag{4.46}$$

Using the function of the minimum total profit (MTP) in (4.46), we want to find the value of U, so that the MTP is the maximum where $U \in [0, b]$. That optimal value of U is denoted as U_o.

The numerical algorithm to calculate the optimal pre-set upper-bound U_o is as follows:

1. Set up the value of b as $5\sigma_\epsilon$.
2. Decide a sequence of pre-set upper-bounds U_i, where $U_i = i \times 0.1$, and $i = 0, \ldots, b/0.1$.
3. For each U_i,
   (a) calculate $E(\mathcal{T}_{0,b}(0, U_i))$ as the trade duration (TD_{U_i}) in (4.44).
   (b) calculate $E(\mathcal{T}_{-b,U_i}(0, 0))$ as the inter-trade interval (IT_{U_i}) in (4.45).
   (c) calculate $MTP(U_i) = \left(\frac{T}{TD_{U_i}+IT_{U_i}} - 1\right) U_i$.
4. Find $U_o \in \{U_i\}$ such that $MTP(U_o)$ is the maximum .

Example of numerical results are shown in Table 4.6. Computer program for Table 4.6 is provided in Appendix A.2.2.

Table 4.6: The optimal pre-set upper-bound U_o for an AR(2) process.

σ_a^2	(ϕ_2, ϕ_1)	max MTP	U_o	σ_ϵ^2	relationship U_o and σ_ϵ^2
1	(0.3, 0.5)	65.4268	1.2	2.2436	$U_o \approx 0.80\sigma_\epsilon$
	(0.3, -0.5)	154.7184	1.1	2.2436	$U_o \approx 0.73\sigma_\epsilon$
	(-0.3, 0.5)	114.4106	1.1	1.2897	$U_o \approx 0.97\sigma_\epsilon$
	(-0.3, -0.5)	159.3847	1.0	1.2897	$U_o \approx 0.88\sigma_\epsilon$
0.49	(0.3, 0.5)	45.8919	0.8	1.0994	$U_o \approx 0.76\sigma_\epsilon$
	(0.3, -0.5)	108.2194	0.8	1.0994	$U_o \approx 0.76\sigma_\epsilon$
	(-0.3, 0.5)	80.16096	0.7	0.6319	$U_o \approx 0.88\sigma_\epsilon$
	(-0.3, -0.5)	111.6703	0.7	0.6319	$U_o \approx 0.88\sigma_\epsilon$

4.5 Conclusion

In this chapter, trade duration and inter-trade interval for cointegration error following AR(1) and AR(2) processes in pairs trading strategy are evaluated by using integral equation approach. The results from simulations are encouraging. The methodology can be extended for AR(p) processes, $p > 2$. However, there is a challenge in numerical scheme for AR(p) processes, $p > 2$, in calculating trade duration and inter-trade interval as it involves more than 2 integrals. For example, to calculate trade duration and inter-trade interval for an AR(2) process, it involves 2 integrals (see (4.44) and (4.45)) and by using $n = 12$, it needs to calculate an inverse matrix of $(m \times m)$ where $m = 6n^2 + 4n + 1 = 913$. Thus, for $p > 2$ the numerical scheme will become more complex.

Chapter 5

Empirical Examples for Cointegration Error Following an AR(1) Process

In this chapter, results of pair trading strategy using three cointegrated pair of finance stocks are given. Selection of trading pairs, determination of training and trading periods, and measurement of profits and returns are discussed. For the three pairs selected, trading strategy and resulting profits and trades conducted are given.

5.1 Trading Pairs Selection

To short-list candidates of trading pairs from a universe of stocks in a equity market, Do *et al.* (2006) and Schmidt (2008) suggested to choose pairs on the basis of industry similarity while Bertram (2009) and Hong and Susmel (2003) suggested to choose dual-listed securities.

For pairs trading strategy based on cointegration, cointegration tests are the key to choose the trading pairs. In order to determine whether cointegration exists between two time series, there are two techniques that are generally used: the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the technique developed by Johansen (1988). The Engle-Granger approach uses OLS (Ordinary Least Squares) to estimate the long-run steady-state relationship between the variables in the model, and then tests whether residuals series is stationary or not by using the augmented Dickey-Fuller test (ADF). The Engle-Granger approach is quite easy to use and residuals from the OLS can be regarded as cointegration error in (1.1), because it already has a zero mean. Alexander and Dimitriu (2002) and Vidyamurthy (2004) used the Engle-Granger approach. However, there are some criticisms of this approach, e.g., (i) the 2-step cointegration procedure is sensitive to the ordering variables, therefore the residuals may have different sets of properties; (ii) If the bivariate series are not cointegrated, it will result in spurious regression (Lim and Martin, 1995). Spurious regression arises from the static regression of non-stationary processes. There are two types of spurious regression: (i) falsely rejecting an existing cointegration relationship; (ii) falsely accept a non-existing cointegration relationship [1]. To overcome the problems found in the Engle-Granger approach, the Johansen's approach uses a vector error-correction model (VECM) so that all variables can be endogenous. More discussion about the two cointegration test methods can be found in Harris (1995). To apply cointegration model in (1.1), we only obtain the cointegration coefficients if we use the Johansen method. Thus, the cointegration error from the Johansen method will be:

$$\epsilon_t^* = P_{S1,t} - \beta P_{S2,t} \qquad \text{and} \qquad \epsilon_t = \epsilon_t^* - \mu$$

where ϵ_t^* is the cointegration error from the Johansen's approach, ϵ_t is the adjusted cointegration error so that it has a zero mean, and μ is the mean of ϵ_t^*.

To combine the Johansen method and Engle-Granger method, firstly, Johansen method is used to make sure that the pairs of stocks have a long-run equilibrium. Secondly, Engle-Granger approach is applied to confirm the cointegration and obtain the cointegration error. By doing the two cointegration approaches, the type 2 spurious regression problem in the Engle-Granger method will be avoided and we are really sure that the pairs are cointegrated. In this thesis, we use the residuals series from the OLS of Engle-Granger method as the cointegration error instead of using the cointegration error from the Johansen method due to a practical reason that we do not need to make adjustment for the residuals.

The aim of this chapter is to demonstrate the method discussed in Chapter 4 to determine the optimal pre-set boundaries for pairs trading based on cointegration in practice. We use 3 pairs suggested by Schmidt (2008) , i.e.,WBC-BOQ,WBC-FKP, and CBA-ASX [2].

5.2 Determining In-sample and Out-sample Period

There is no standard rule for deciding the lengths of the training period (in-sample data) and the trading period (out-sample data). However, the training period needs to be long enough so that we can determine whether a cointegration relationship actually exists, but not so long that there is not enough information left for the study for the trading period. For trading period, it needs to be long enough to have opportunities to open and close trades and test the strategy, but it can not too long because it is possible that the cointegration relationship between two tested stocks may change.

[1] See Chiarella *et al.* (2008) for details

[2] Wespact Banking Corporation (WBC), Bank of Queensland (BOQ), Commonwealth Bank of Australia (CBA), FKP Property Group (FKP), Australian Stock Exchange (ASX). Schmidt (2008) actually suggested 5 trading pairs but we only use 3 as data for one pair (WBC-SGB) are not available due to merger of WBC and SGB and another pair (ASX-LLC) is not cointegrated based on 3 years data 2003-2005.

Johansen tests for the 3 pairs with 1 year data for 2003 show there were no indication of cointegration for all pairs. Using 2 years data, only WBC-FKP was found to have a significant cointegration relationship. Using 3 years data of 2003, 2004 and 2005 Johansen test results presented in Table 5.1 shows that all 3 pairs are significantly cointegrated [3].

Table 5.1: Cointegration analysis using the Johansen method with data set 2^{nd} January 2003 to 30^{th} December 2005.

Pairs	Ho:rank=p	$-T\log(1-\lambda_{p+1})$	$-T\sum_{i=p+1}^{2}\log(1-\lambda_i)$
WBC-BOQ	p == 0	17.64*	17.86*
	p <= 1	0.2162	0.2162
WBC-FKP	p == 0	14.68*	14.69
	p <= 1	0.007119	0.007119
CBA-ASX	p == 0	17.86*	20.59**
	p <= 1	2.734	2.734

Note: (1) λ_i is the ith eigen value; (2) * and ** mark significance at 95% and 99%; (3) We use 1 lag for the analysis.

For the trading period (out-sample), we use 6-months trading period in line with previous works [4]. We also try 3-months trading period for comparison. Table 5.2 shows the 3-years training period and the 6-months trading period as well as the 3-years training period and the 3-months trading period.

Table 5.2: Training periods and trading periods:

The 3-years training period and the 6-months trading period	
Training period	**Trading period**
January 2003 - December 2005	January 2006 - June 2006
July 2003 - June 2006	July 2006 - December 2006
The 3-years training period and the 3-months trading period	
Training period	**Trading period**
January 2003 - December 2005	January 2006 - March 2006
April 2003 - March 2006	April 2006 - June 2006
July 2003 - June 2006	July 2006 - September 2006
October 2003 - September 2006	October 2006 - December 2006

[3] WBC-BOQ and CBA-ASX are significant by both λ_{max} and λ_{trace} tests and WBC-FKP is significant by λ_{max} test.

[4] Gatev *et al.* (2006), Gillespie and Ulph (2001), Habak (2002) and Do and Faff (2008).

5.3 Measuring Profits and Returns

In deriving the optimal upper-bound U_o by using (4.29), we assume that all trades are completed before the end of period T. However, in the study of empirical examples, we use the strategy that the last trade is forced to be closed at the end of period if it has not been closed yet, possibly at a loss. We call this trade as an incomplete trade. Based on this strategy, for an empirical example, we define:

$$\begin{aligned} \text{Total Profit (TP)} &= \sum_{i=1}^{NT1} (\text{Profit or Loss})_i, && (5.1) \\ \text{Average Profit per Trade (APPT)} &= TP/NT1, && (5.2) \end{aligned}$$

where $(\text{Profit or Loss})_i$ is the profit or loss from the ith trade and $NT1$ is the number of complete and incomplete trades.

Amount of profit is not the only measure to determine whether our investment is good or not. The returns of our investments also need to be considered. As Hong and Susmel (2003) noted, measuring returns for pairs trading might be complicated. They considered three measures of return. In this thesis, the third measure in Hong and Susmel (2003) is used as it is the most realistic measure of return. However, in Australia only 20% of the stock values deposit is needed[5]. Therefore, for our study, we use the third return measure and use 20% margin deposit. We define "Average Return per Trade (ART)":

$$ART = \sum_{i=1}^{NT1} \left(\frac{(\text{Profit or Loss})_i}{Vl_{0i} + 0.2 * Vs_{0i}} \times 100 \right) \frac{1}{NT1}, \tag{5.3}$$

where Vl_{0i} is the stock value in a long position, Vs_{0i} is the stock value in a short position at the beginning of the ith trade and $NT1$ is the number of complete and incomplete trades.

5.4 Pairs Trading Strategy with Empirical Data

We employ the following 6 steps to apply our pairs trading strategy to the empirical data:

Step 1: Run cointegration Johansen and Engle-Granger tests for the 3 trading pairs with the training period data. Remove the pairs if they are not significantly cointegrated using both tests. We use PcGive (Hendry and Doornik, 1996) and PcFiml 9.0 (Doornik and Hendry, 1997) to analyse the cointegration based on Johansen method.

[5]See the ASX media release about short selling on 6 March 2008.

Step 2: Obtain residual series (cointegration errors) ϵ_t of the pairs from the Engle-Granger cointegration equation:

$$P_{S1,t} = \mu + \beta P_{S2,t} + \epsilon_t, \tag{5.4}$$

where $P_{S1,t}$ and $P_{S2,t}$ are the prices of stock $S1$ and $S2$ respectively and μ is a constant.

Step 3: For the cointegration errors from Step 2, analyse whether an AR(1) model is appropriate. If it can not be fitted with an AR(1) model, remove the pair.

Step 4: From the parameter estimations of the AR(1) model from Step 3, apply the integral equation approach to estimate the average of trading duration and the average of inter-trade interval. Then use these estimations to estimate the number of trades and decide the optimal pre-set boundaries (thresholds) for the pairs trading strategy.

Step 5: Forecast the residual series for the trading period by using parameter estimations from the model in the training period. Using the optimal pre-set boundaries from Step 4, apply the pairs trading strategy in Lin *et al.* (2006) which also explained in Subsection 2.3.4 for residual series in the training period and the trading period data. If the last open trade is not closed yet, we have to close it at the end of period (possibly at loss). As we use daily closing prices for this empirical examples, we assume that the trades occur at the end of the trading day even if the thresholds may be reached during the day.

From trading simulation in in the training period and the trading period data, calculate:

(a) Total profit (TP) as in (5.1).

(b) $NT1$ = number of complete and incomplete trades.

(c) $NT2$ = number of complete trades.

(d) Average trading duration = $\sum_{i=1}^{NT2} TD_i/NT2$, where TD_i is the trading duration from the ith trade.

(e) Average profit per trade (APT) as in (5.2).

(f) Average return per trade (ART) as in (5.3).

(g) Annualized return[6] =

(i) $ART * (NT2/3)$ for the 3-years training period.

(ii) $ART * (NT2 * 2)$ for the 6-months trading period.

(iii) $ART * (NT2 * 4)$ for the 3-months trading period.

[6] In calculating the annualized return, we prefer to use $NT2$ instead of $NT1$ as $NT2 \leq NT1$, so that if we use $NT1$, the annualized return will tend to over estimate

Step 6: Repeat the above procedure by updating the cointegration model for the next training period and trading period.

5.5 Trading Simulation Results with the 6-months Trading Period

In this part, we report the trading simulation using the steps described in Section 5.4.

5.5.1 Training Period : January 2003 - December 2005 and Trading Period: January 2006 - June 2006

Step 1: From Table 5.1 , the 3 trading pairs, i.e. WBA-BOQ, WBA-FKP and CBA-ASX, are significantly cointegrated using Johansen method. Using the Engle-Granger methods, the ADF unit root test of the residual series (cointegration errors) in Table 5.3 confirm that the three pairs are cointegrated during the training period.

Step 2: The integration equations based on Engle-Granger method for the WBC-BOQ, WBC-FKP and CBA-ASX:

$$WBC_t = 6.34113 + 1.1233 \times BOQ_t + rWbcBoq_t$$
$$WBC_t = 11.328 + 2.1156 \times FKP_t + rWbcFkp_t$$
$$CBA_t = 18.724 + 0.73815 \times ASX_t + rCbaAsx_t$$

where $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$ are the residual series corresponding to their trading pairs. The plots of these 3 series are presented in Figure 5.1.

Step 3: AR(1) models analysis for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$. Consider an AR(1) model in (4.19):

From Table 5.4, all of the AR(1) coefficients are significant and the Pr > Chisq to lag 48 show that there are no autocorrelation for the residuals from the AR(1) models with 1% significant level. Thus, we conclude that an AR(1) model is appropriate for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$.

Step 4: Table 5.5 shows the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 6-months trading period, the estimation of the minimum total profit and the number of trades will be one sixth of the estimation for the 3-years training period.

Step 5: Trading simulation results are presented in Tables 5.6 and 5.7:

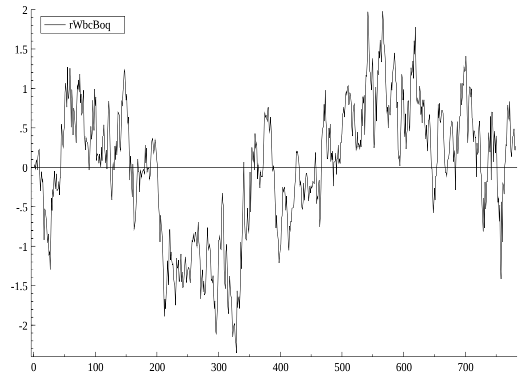
Table 5.3: The ADF unit root test for cointegration error from the OLS with data set 2^{nd} January 2003 to 30^{th} December 2005.

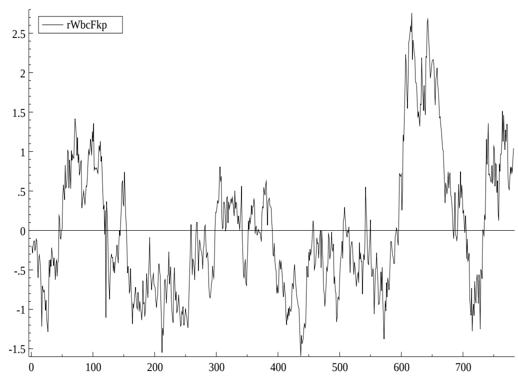
Critical values: 5%=-1.94 1%=-2.568

	t-adf	beta Y_1	\sigma	lag	t-DY_lag	t-prob	F-prob
Rwbc_boq	-2.9516**	0.96691	0.23692	10	-2.0941	0.0366	
Rwbc_boq	-3.2140**	0.96414	0.23745	9	-3.1193	0.0019	0.0366
Rwbc_boq	-3.6349**	0.95957	0.23881	8	0.68215	0.4954	0.0009
Rwbc_boq	-3.5781**	0.96055	0.23872	7	-0.00048280	0.9996	0.0023
Rwbc_boq	-3.6111**	0.96054	0.23857	6	1.2625	0.2072	0.0058
Rwbc_boq	-3.4789**	0.96228	0.23866	5	0.11460	0.9088	0.0065
Rwbc_boq	-3.4945**	0.96243	0.23851	4	-0.078367	0.9376	0.0131
Rwbc_boq	-3.5356**	0.96232	0.23835	3	-0.73919	0.4600	0.0238
Rwbc_boq	-3.6665**	0.96128	0.23828	2	-0.32254	0.7471	0.0334
Rwbc_boq	-3.7459**	0.96083	0.23814	1	-2.0482	0.0409	0.0517
Rwbc_boq	-4.0793**	0.95771	0.23864	0			0.0212
Rwbc_fkp	-3.3511**	0.96585	0.22623	10	1.9682	0.0494	
Rwbc_fkp	-3.1283**	0.96830	0.22666	9	1.1971	0.2317	0.0494
Rwbc_fkp	-3.0114**	0.96967	0.22672	8	1.0408	0.2983	0.0709
Rwbc_fkp	-2.9150**	0.97082	0.22674	7	-0.052648	0.9580	0.0946
Rwbc_fkp	-2.9406**	0.97076	0.22659	6	1.8711	0.0617	0.1722
Rwbc_fkp	-2.7555**	0.97271	0.22696	5	-0.42213	0.6730	0.0790
Rwbc_fkp	-2.8185**	0.97226	0.22684	4	-1.3050	0.1923	0.1224
Rwbc_fkp	-2.9826**	0.97081	0.22694	3	-0.34581	0.7296	0.1089
Rwbc_fkp	-3.0442**	0.97043	0.22681	2	-1.5496	0.1217	0.1566
Rwbc_fkp	-3.2538**	0.96860	0.22702	1	-3.1725	0.0016	0.1130
Rwbc_fkp	-3.7015**	0.96440	0.22835	0			0.0070
Rcba_asx	-2.4055*	0.98203	0.32310	10	-1.9261	0.0545	
Rcba_asx	-2.5981**	0.98065	0.32368	9	-0.077918	0.9379	0.0545
Rcba_asx	-2.6197**	0.98059	0.32347	8	-0.11116	0.9115	0.1567
Rcba_asx	-2.6449**	0.98051	0.32326	7	0.045204	0.9640	0.2931
Rcba_asx	-2.6548**	0.98054	0.32305	6	0.99024	0.3224	0.4443
Rcba_asx	-2.5725**	0.98123	0.32304	5	-1.0236	0.3063	0.4529
Rcba_asx	-2.6860**	0.98050	0.32305	4	1.0067	0.3144	0.4516
Rcba_asx	-2.6010**	0.98120	0.32306	3	0.24461	0.8068	0.4539
Rcba_asx	-2.5910**	0.98137	0.32286	2	-1.4670	0.1428	0.5554
Rcba_asx	-2.7512**	0.98031	0.32310	1	0.44810	0.6542	0.4402
Rcba_asx	-2.7214**	0.98063	0.32293	0			0.5157

Table 5.4: AR(1) models analysis.

ϵ	$\hat{\phi}$	t−prob	$\hat{\sigma}_a$	Pr > Chisq to lag 50
rWbcBoq	0.95762	0.0000	0.238204	0.0936
rWbcFkp	0.96460	0.0000	0.227874	0.0167
rCbaAsx	0.98057	0.0000	0.32196	0.2679





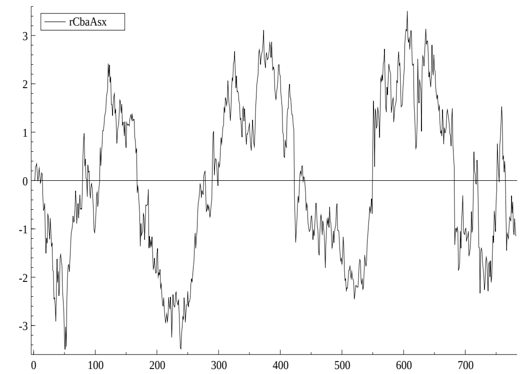


Figure 5.1: Cointegration errors with data set 2003 to 2005.

Table 5.5: Estimation from integral equations:

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	6 months	Threshold	Duration	3 years	6 months
$rWbcBoq_t$	\$ 13.9206	\$2.3201	0.51	19.8766 days	27.2950	4.5492
$rWbcFkp_t$	\$ 12.3106	\$2.0517	0.51	22.6302 days	24.1384	4.0231
$rCbaAsx_t$	\$ 13.1158	\$ 2.1859	0.79	33.7642 days	16.6024	2.7671

Table 5.6: Trading simulation results using training period data (in-sample) in January 2003 - December 2005 by using the optimal thresholds from Table 5.5:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	\$ 18.4964	\$ 12.9940	\$20.5269
Number of complete and incomplete trades	24	20	13
Number of complete trades	23	19	12
Average trading duration	42.7 days	31.3 days	56.8 days
Average profit per trade	\$ 0.7707	\$ 0.6497	\$ 1.5799
Average return per trade	4.56%	4.85%	6.01%
Annualized Return	36.50% pa	32.37% pa	26.05% pa

Table 5.7: Trading simulation results using trading period data (out-sample) in January 2006 - June 2006 by using the optimal thresholds from Table 5.5:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	\$ 6.2328	\$ 4.0548	\$ 9.2094
Number of complete and incomplete trades	7	4	6
Number of complete trades	7	4	6
Average trading duration	10.8571 days	29.75 days	15.83 days
Average profit per trade	\$ 0.8904	\$ 1.013712	\$ 1.5349
Average return per trade	3.33%	5.97%	3.85%
Annualized Return	46.59 % pa	47.81 % pa	46.24 % pa

5.5.2 Training Period : July 2003 - June 2006 and Trading Period: July 2006 - December 2006

Step 1: The 3 trading pairs, i.e. WBA-BOQ, WBC-FKP and CBA-ASX, are significantly cointegrated by using both the Johansen and the Engle-Granger methods The Johansen test results are presented in Table 5.8 and the Engle-Granger test results for the 3 pairs are presented in Table 5.9.

Table 5.8: Cointegration analysis using the Johansen method with data set 1^{st} July 2003 to 30^{th} June 2005

Pairs	Ho:rank=p	$-T\log(1-\lambda_{p+1})$	$-T\sum_{i=p+1}^{2}\log(1-\lambda_i)$
WBC-BOQ	p == 0	20.67**	21.32**
	p <= 1	0.6572	0.6572
WBC-FKP	p == 0	16.52*	16.95*
	p <= 1	0.4323	0.4323
CBA-ASX	p == 0	20.83**	20.83**
	p <= 1	0.006521	0.006521

Note: (1) λ_i is the ith eigen value; (2) * and ** mark significance at 95% and 99%; (3) We use 1 lag for the analysis.

Step 2: The integration equations for the WBC-BOQ, WBC-FKP and CBA-ASX:

$$WBC_t = 5.6739 + 1.1831 \times BOQ_t + rWbcBoq_t$$

$$WBC_t = 10.336 + 2.4211 \times FKP_t + rWbcFkp_t$$

$$CBA_t = 18.344 + 0.77127 \times ASX_t + rCbaAsx_t$$

Step 3: AR(1) models analysis for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$. Consider an AR(1) model in (4.19):

Table 5.9: The ADF unit root test for cointegration error from the OLS with data set 1^{st} July 2003 to 30^{th} June 2005.

Critical values: 5%=-1.94 1%=-2.568

	t-adf	beta Y_1	\sigma	lag	t-DY_lag	t-prob	F-prob
Rwbc_boq	-3.4830**	0.95735	0.25470	10	-1.0123	0.3117	
Rwbc_boq	-3.6447**	0.95574	0.25471	9	-3.6810	0.0002	0.1674
Rwbc_boq	-4.2166**	0.94897	0.25681	8	1.8113	0.0705	0.0021
Rwbc_boq	-3.9932**	0.95210	0.25719	7	0.49573	0.6202	0.0012
Rwbc_boq	-3.9647**	0.95292	0.25706	6	1.0784	0.2812	0.0020
Rwbc_boq	-3.8547**	0.95463	0.25709	5	-0.16266	0.8708	0.0024
Rwbc_boq	-3.9168**	0.95437	0.25693	4	0.14012	0.8886	0.0042
Rwbc_boq	-3.9382**	0.95460	0.25676	3	-0.87681	0.3809	0.0071
Rwbc_boq	-4.1081**	0.95314	0.25672	2	0.12314	0.9020	0.0089
Rwbc_boq	-4.1358**	0.95334	0.25656	1	-1.7585	0.0791	0.0139
Rwbc_boq	-4.4518**	0.95030	0.25691	0			0.0084
Rwbc_fkp	-3.6864**	0.95907	0.25666	10	1.6368	0.1021	
Rwbc_fkp	-3.5012**	0.96141	0.25694	9	3.1368	0.0018	0.1127
Rwbc_fkp	-3.1455**	0.96536	0.25843	8	1.3299	0.1840	0.0055
Rwbc_fkp	-3.0160**	0.96697	0.25856	7	-1.5611	0.1189	0.0053
Rwbc_fkp	-3.2180**	0.96496	0.25881	6	0.19014	0.8493	0.0039
Rwbc_fkp	-3.2199**	0.96521	0.25864	5	0.25126	0.8017	0.0069
Rwbc_fkp	-3.2145**	0.96552	0.25849	4	-0.56952	0.5692	0.0113
Rwbc_fkp	-3.3083**	0.96479	0.25837	3	-0.26204	0.7934	0.0163
Rwbc_fkp	-3.3686**	0.96445	0.25821	2	-3.3413	0.0009	0.0246
Rwbc_fkp	-3.8554**	0.95945	0.25992	1	-0.55004	0.5824	0.0011
Rwbc_fkp	-3.9810**	0.95860	0.25980	0			0.0017
Rcba_asx	-2.5006*	0.97969	0.37867	10	-2.8959	0.0039	
Rcba_asx	-2.7881**	0.97736	0.38051	9	1.1484	0.2512	0.0433
Rcba_asx	-2.6887**	0.97827	0.38059	8	0.77094	0.4410	0.0461
Rcba_asx	-2.6277**	0.97886	0.38049	7	-1.4025	0.1612	0.0611
Rcba_asx	-2.7799**	0.97774	0.38073	6	-0.75911	0.4480	0.0509
Rcba_asx	-2.8723**	0.97712	0.38063	5	0.083800	0.9332	0.0650
Rcba_asx	-2.8805**	0.97719	0.38038	4	0.40382	0.6865	0.0948
Rcba_asx	-2.8560**	0.97751	0.38017	3	1.5496	0.1217	0.1271
Rcba_asx	-2.7186**	0.97867	0.38052	2	-2.5138	0.0121	0.0945
Rcba_asx	-2.9926**	0.97657	0.38184	1	-0.25637	0.7977	0.0239
Rcba_asx	-3.0396**	0.97635	0.38161	0			0.0344

Table 5.10: AR(1) models analysis

ϵ	$\hat{\phi}$	t−prob	$\hat{\sigma}_a$	Pr > Chisq to lag 50
rWbcBoq	0.95087	0.0000	0.256753	0.1341
rWbcFkp	0.95846	0.0000	0.257833	0.0332
rCbaAsx	0.97620	0.0000	0.379376	0.1860

From Table 5.10, all of the AR(1) coefficients are significant and the Pr > Chisq to lag 48 show that there are no autocorrelation for the residuals from the AR(1) models with 3% significant level. Thus, we conclude that an AR(1) model is appropriate for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$.

Step 4: Table 5.11 presents the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 6-months trading period, the estimation of the minimum total profit and the number of trades will be one sixth of the estimation for the 3-years training period.

Table 5.11: Estimation from integral equations :

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	6 months	Threshold	Duration	3 years	6 months
$rWbcBoq_t$	$15.9734	$2.6622	0.53	17.87 days	30.1386	5.0231
$rWbcFkp_t$	$ 14.9298	$ 2.4883	0.55	20.07 days	27.145	4.5241
$rCbaAsx_t$	$ 17.0356	$ 2.8393	0.91	29.69 days	18.7204	3.1201

Step 5: Trading simulation results are presented in Tables 5.12 and 5.13:

Table 5.12: Trading simulation results using training period data (in-sample) in July 2003 - June 2006 with the thresholds from Table 5.11:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	$ 19.7015	$ 16.4417	$22.2689
Number of complete and incomplete trades	25	21	16
Number of complete trades	25	20	15
Average trading duration	21.8 days	29.15 days	45.73 days
Average profit per trade	$ 0.7881	$ 0.7829	$ 1.3918
Average return per trade	4.09%	5.09%	4.73%
Annualized return	34.090% pa	33.98% pa	23.65% pa

Generally, the trading simulations above show good results. The trading simulations results from both training period in Tables 5.6 and 5.12 produce total profits

Table 5.13: Trading simulation results using trading period data (out-sample) in July 2006 - December 2006 with the thresholds from Table 5.11:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total Profit	\$ 4.5721	\$ 2.2726	\$ 3.3738
Number of complete and incomplete trades	5	4	4
Number of complete trades	4	3	3
Average trading duration	25.5 days	27.33 days	32 days
Average profit per trade	\$ 0.9144	\$ 0.5681	\$ 0.8434
Average return per trade	3.81 %	2.75%	2.10 %
Annualized return	30.47 % pa	16.530 % pa	12.61 % pa

and average profit per trade higher than the minimum total profits and the minimum average profit per trade predicted from the integral equations in Tables 5.5 and 5.11. However, as we note in Section 3.4, the number of trades from the integral equations are over estimated while the average trading durations are under estimated compared to the trading simulations. Even though the differences between the average trading durations from the integral equation and the trading simulations are quite large, the differences between the number of complete trades from the integral equation and the trading simulations are not large. For example in Table 5.5, the number of trades estimations are 27, 24 and 16 for WBC-BOQ, WBC-FKP and CBA-ASX respectively compared to the actual number of trades in Table 5.6 which are 23, 19 and 12. Furthermore, the three trading pairs (i.e. WBC-BOQ, WBC-FKP and CBA-ASX) produce the average return per trade of above 4% and the annualised return of above 23% pa.

The trading simulation results from the trading period (out-sample) in January 2006-June 2006 in Table 5.7 show very good results. All of the measures, i.e. total profits, number of trades, average profit per trade, are higher than the expectations in Tables 5.5 and 5.6 and even the annualized returns are more than 46% pa. However, the trading simulation results from the trading period (out-sample) in July 2006-December 2006 presented in Table 5.13 are not as good as those from Table 5.7 but generally all measures are still quite close to the estimation from the integral equation in Tables 5.11 and 5.12. Generally, the pair WBC-BOQ is the best pair in this period producing total profit and average profit per trade higher than the expectations. The pair also produces the highest annualized return of 30.47% pa compared to WBC-FKP and CBA-ASX which only produce the annualized returns of 16.53% pa and 12.61% pa respectively.

5.6 Trading Simulation Results with the 3-months Trading Period

5.6.1 Training Period : January 2003 - December 2005 and Trading Period: January 2006 - March 2006

As the training period is the same as the training period in 5.5.1, we can use the results of Step 1-3 in that section.

Step 4: Table 5.14 shows the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 3-months trading period, the estimation of the minimum total profit and the number of trades will be one 12th of the estimation for the 3-years training period.

Table 5.14: Estimation from integral equations:

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	3 months	Threshold	Duration	3 years	3 months
$rWbcBoq_t$	\$ 13.9206	\$1.16005	0.51	19.8766 days	27.2950	2.2746
$rWbcFkp_t$	\$ 12.3106	\$1.0259	0.51	22.6302 days	24.1384	2.0115
$rCbaAsx_t$	\$ 13.1158	\$1.0929	0.79	33.7642 days	16.6024	1.3835

Step 5: Trading simulation results in the training period is the same as Table 5.6 while for the trading period the results are presented in 5.15:

Table 5.15: Trading simulation results using trading period data (out-sample) in January 2006 - March 2006 by using the optimal thresholds from Table 5.14:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	\$ 2.9312	\$-1.1973	\$ 4.0765
Number of complete and incomplete trades	3	1	4
Number of complete trades	3	0	3
Average trading duration	10 days	NA days	11 days
Average profit per trade	\$ 0.977	\$ -1.1973	\$1.01904
Average return per trade	4.1668%	% -7.6808	%2.2215
Annualized Return	50.00 % pa	NA pa	% 26.6579 pa

5.6.2 Training Period : April 2003 - March 2006 and Trading Period: April 2006 - June 2006

Step 1: The 3 trading pairs, i.e. WBA-BOQ, WBC-FKP and CBA-ASX, are significantly cointegrated by using both the Johansen and the Engle-Granger methods The

Johansen test results are presented in Table 5.16 and the Engle-Granger test results for the 3 pairs are presented in Table 5.17.

Table 5.16: Cointegration analysis using the Johansen method with data set 1st April 2003 to 31st March 2006

Pairs	Ho:rank=p	$-T\log(1-\lambda_{p+1})$	$-T\sum_{i=p+1}^{2}\log(1-\lambda_i)$
WBC-BOQ	p == 0	19.67**	20.37**
	p <= 1	0.7075	0.7075
WBC-FKP	p == 0	15.86*	15.89*
	p <= 1	0.02927	0.02927
CBA-ASX	p == 0	14.99*	15.03
	p <= 1	0.03961	0.03961

Note: (1) λ_i is the ith eigen value; (2) * and ** mark significance at 95% and 99%; (3) We use 1 lag for the analysis.

Step 2: The integration equations for the WBC-BOQ, WBC-FKP and CBA-ASX:

$$WBC_t = 6.4176 + 1.1159 \times BOQ_t + rWbcBoq_t$$

$$WBC_t = 10.936 + 2.2671 \times FKP_t + rWbcFkp_t$$

$$CBA_t = 19.071 + 0.72861 \times ASX_t + rCbaAsx_t$$

Step 3: AR(1) models analysis for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$. Consider an AR(1) model in (4.19):

From Table 5.18, all of the AR(1) coefficients are significant and the Pr > Chisq to lag 48 show that there are no autocorrelation for the residuals from the AR(1) models with 5% significant level. Thus, we conclude that an AR(1) model is appropriate for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$.

Step 4: Table 5.19 presents the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 3-months trading period, the estimation of the minimum total profit and the number of trades will be one 12th of the estimation for the 3-years training period.

Step 5: Trading simulation results are presented in Tables 5.20 and 5.21:

Table 5.17: The ADF unit root test for cointegration error from the OLS with data set 1^{st} April 2003 to 31^{th} March 2006.

Critical values: 5%=-1.94 1%=-2.568

	t-adf	beta Y_1	\sigma	lag	t-DY_lag	t-prob	F-prob
Rwbc_boq	-3.0209**	0.96494	0.24038	10	-2.0569	0.0400	0.0909
Rwbc_boq	-3.2718**	0.96220	0.24089	9	-3.5139	0.0005	0.0331
Rwbc_boq	-3.7303**	0.95694	0.24269	8	1.2547	0.2100	0.0005
Rwbc_boq	-3.6016**	0.95873	0.24278	7	0.018986	0.9849	0.0006
Rwbc_boq	-3.6299**	0.95876	0.24262	6	1.1726	0.2413	0.0012
Rwbc_boq	-3.5137**	0.96036	0.24268	5	-0.46920	0.6391	0.0013
Rwbc_boq	-3.6014**	0.95970	0.24255	4	-0.17885	0.8581	0.0022
Rwbc_boq	-3.6545**	0.95945	0.24240	3	-1.3100	0.1906	0.0037
Rwbc_boq	-3.8635**	0.95750	0.24251	2	0.011303	0.9910	0.0035
Rwbc_boq	-3.9003**	0.95752	0.24235	1	-2.3034	0.0215	0.0057
Rwbc_boq	-4.2823**	0.95375	0.24303	0			0.0017
Rwbc_fkp	-3.2594**	0.96571	0.24130	10	1.2420	0.2146	0.0646
Rwbc_fkp	-3.1337**	0.96725	0.24139	9	1.9778	0.0483	0.0632
Rwbc_fkp	-2.9315**	0.96949	0.24185	8	1.2841	0.1995	0.0266
Rwbc_fkp	-2.8118**	0.97088	0.24195	7	-0.33004	0.7415	0.0254
Rwbc_fkp	-2.8652**	0.97051	0.24181	6	1.3500	0.1774	0.0400
Rwbc_fkp	-2.7404**	0.97193	0.24194	5	-0.95984	0.3374	0.0352
Rwbc_fkp	-2.8570**	0.97090	0.24193	4	-1.5625	0.1186	0.0407
Rwbc_fkp	-3.0417**	0.96917	0.24216	3	-0.14464	0.8850	0.0296
Rwbc_fkp	-3.0794**	0.96901	0.24200	2	-1.7310	0.0838	0.0439
Rwbc_fkp	-3.3199**	0.96681	0.24232	1	-2.8846	0.0040	0.0275
Rwbc_fkp	-3.7531**	0.96269	0.24347	0			0.0034
Rcba_asx	-2.2995*	0.98129	0.34052	10	-2.9429	0.0034	0.9238
Rcba_asx	-2.5901**	0.97892	0.34224	9	-0.13424	0.8932	0.1254
Rcba_asx	-2.6188**	0.97881	0.34202	8	0.85379	0.3935	0.1881
Rcba_asx	-2.5478*	0.97949	0.34195	7	-0.60749	0.5437	0.2169
Rcba_asx	-2.6265**	0.97898	0.34181	6	-0.13976	0.8889	0.2680
Rcba_asx	-2.6587**	0.97886	0.34159	5	-1.0511	0.2935	0.3467
Rcba_asx	-2.7999**	0.97788	0.34162	4	0.78961	0.4300	0.3457
Rcba_asx	-2.7275**	0.97861	0.34153	3	1.1102	0.2673	0.3786
Rcba_asx	-2.6263**	0.97951	0.34158	2	-1.6915	0.0912	0.3669
Rcba_asx	-2.8299**	0.97803	0.34200	1	0.095505	0.9239	0.2595
Rcba_asx	-2.8399**	0.97812	0.34178	0			0.3214

Table 5.18: AR(1) models analysis

ϵ	$\hat{\phi}$	t−prob	$\hat{\sigma}_a$	Pr > Chisq to lag 50
rWbcBoq	0.95414	0.0000	0.242918	0.0658
rWbcFkp	0.96299	0.0000	0.24294	0.0558
rCbaAsx	0.97798	0.0000	0.340851	0.0632

Table 5.19: Estimation from integral equations :

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	3 months	Threshold	Duration	3 years	3 months
$rWbcBoq_t$	$14.683	$1.22358	0.51	18.7763 days	28.7902	2.39918
$rWbcFkp_t$	$ 13.3796	$ 1.11496	0.54	21.9768 days	24.7772	2.0647
$rCbaAsx_t$	$ 14.7524	$ 1.22936	0.83	31.3139 days	17.774	1.48116

Table 5.20: Trading simulation results using training period data (in-sample) in April 2003 - March 2006 with the thresholds from Table 5.11:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	$ 20.3924	$ 12.9575	$20.2703
Number of complete and incomplete trades	26	19	14
Number of complete trades	26	18	13
Average trading duration	22.57 days	31.55 days	50.92 days
Average profit per trade	$ 0.7843	$ 0.6819	$
Average return per trade	4.5625%	7.0945%	5.2190%
Annualized return	39.5414% pa	42.5672% pa	22.6158% pa

Table 5.21: Trading simulation results using trading period data (out-sample) in April 2006 - June 2006 with the thresholds from Table 5.19:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total Profit	$ 3.2998	$ 2.5443	$ 5.0459
Number of complete and incomplete trades	4	3	3
Number of complete trades	4	2	3
Average trading duration	11.5 days	23.5 days	17.33 days
Average profit per trade	$ 0.8249	$ 0.8481	$ 1.6820
Average return per trade	4.0568 %	4.4059%	4.7032%
Annualized return	64.9087 % pa	35.2472 % pa	56.4387 % pa

5.6.3 Training Period : July 2003 - June 2006 and Trading Period: July 2006 - September 2006

As the training period is the same as the training period in 5.5.2, we can use the results of Step 1-3 in that section.

Step 4: Table 5.22 shows the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 3-months trading period, the estimation of the minimum total profit and the number of trades will be one 12th of the estimation for the 3-years training period.

Table 5.22: Estimation from integral equations :

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	3 months	Threshold	Duration	3 years	3 months
$rWbcBoq_t$	$15.9734	$1.3311	0.53	17.87 days	30.1386	2.5115
$rWbcFkp_t$	$ 14.9298	$ 1.2441	0.55	20.07 days	27.145	2.2621
$rCbaAsx_t$	$ 17.0356	$ 1.4196	0.91	29.69 days	18.7204	1.5600

Step 5: Trading simulation results in the training period is the same as Table 5.12 while for the trading period the results are presented in 5.23:

Table 5.23: Trading simulation results using trading period data (out-sample) in July 2006 - September 2006 with the thresholds from Table 5.22:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total Profit	$ 1.9048	$ 1.5037	$ 2.7398
Number of complete and incomplete trades	3	3	2
Number of complete trades	2	2	1
Average trading duration	20 days	17 days	4 days
Average profit per trade	$ 0.6349	$ 0.5012	$ 1.3699
Average return per trade	% 2.83686	% 2.6362	%3.0469
Annualized return	% 22.6949 pa	% 21.0900 pa	%12.1878 pa

5.6.4 Training Period : October 2003 - September 2006 and Trading Period: October 2006 - December 2006

Step 1: The 3 trading pairs, i.e. WBA-BOQ, WBC-FKP and CBA-ASX, are significantly cointegrated by using both the Johansen and the Engle-Granger methods The Johansen test results are presented in Table 5.24 and the Engle-Granger test results for the 3 pairs are presented in Table 5.25.

Table 5.24: Cointegration analysis using the Johansen method with data set 1^{st} October 2003 to 29^{th} September 2006

Pairs	Ho:rank=p	$-T\log(1-\lambda_{p+1})$	$-T\sum_{i=p+1}^{2}\log(1-\lambda_i)$
WBC-BOQ	p == 0	21.05**	21.39**
	p <= 1	0.3394	0.3394
WBC-FKP	p == 0	16.39*	16.57*
	p <= 1	0.1743	0.1743
CBA-ASX	p == 0	17.75*	17.83*
	p <= 1	0.08457	0.08457

Note: (1) λ_i is the ith eigen value; (2) * and ** mark significance at 95% and 99%; (3) We use 1 lag for the analysis.

Step 2: The integration equations for the WBC-BOQ, WBC-FKP and CBA-ASX:

$$WBC_t = 5.5408 + 1.1908 \times BOQ_t + rWbcBoq_t$$
$$WBC_t = 10.720 + 2.3183 \times FKP_t + rWbcFkp_t$$
$$CBA_t = 17.756 + 0.80133 \times ASX_t + rCbaAsx_t$$

Step 3: AR(1) models analysis for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$. Consider an AR(1) model in (4.19):

From Table 5.26, all of the AR(1) coefficients are significant and the Pr > Chisq to lag 48 show that there are no autocorrelation for the residuals from the AR(1) models with 2% significant level. Thus, we conclude that an AR(1) model is appropriate for $rWbcBoq_t$, $rWbcFkp_t$ and $rCbaAsx_t$.

Step 4: Table 5.27 presents the estimations from integral equations for the 3-years training period. By assuming that the cointegration relationship in training period still continues in the 6-months trading period, the estimation of the minimum total profit and the number of trades will be one sixth of the estimation for the 3-years training period.

Step 5: Trading simulation results are presented in Tables 5.28 and 5.29:

Table 5.25: The ADF unit root test for cointegration error from the OLS with data set 1^{st} October 2003 to 31^{th} September 2006.

Critical values: 5%=-1.94 1%=-2.568

	t-adf	beta Y_1	\sigma	lag	t-DY_lag	t-prob	F-prob
Rwbc_boq	-3.5944**	0.95433	0.26515	10	-0.45880	0.6465	0.0940
Rwbc_boq	-3.6948**	0.95353	0.26501	9	-4.0790	0.0001	0.1416
Rwbc_boq	-4.3598**	0.94532	0.26773	8	2.0394	0.0418	0.0005
Rwbc_boq	-4.0935**	0.94912	0.26829	7	0.20033	0.8413	0.0002
Rwbc_boq	-4.1120**	0.94949	0.26812	6	1.4105	0.1588	0.0004
Rwbc_boq	-3.9496**	0.95194	0.26829	5	-0.024441	0.9805	0.0004
Rwbc_boq	-3.9948**	0.95190	0.26811	4	0.11008	0.9124	0.0007
Rwbc_boq	-4.0231**	0.95209	0.26794	3	-0.57612	0.5647	0.0013
Rwbc_boq	-4.1567**	0.95107	0.26782	2	0.78657	0.4318	0.0019
Rwbc_boq	-4.0864**	0.95243	0.26776	1	-2.5450	0.0111	0.0027
Rwbc_boq	-4.5439**	0.94763	0.26871	0			0.0005
Rwbc_fkp	-3.7134**	0.95913	0.26232	10	2.6868	0.0074	0.6344
Rwbc_fkp	-3.3811**	0.96295	0.26340	9	3.2447	0.0012	0.1021
Rwbc_fkp	-3.0062**	0.96707	0.26505	8	0.96601	0.3343	0.0038
Rwbc_fkp	-2.9168**	0.96826	0.26503	7	-1.8711	0.0617	0.0050
Rwbc_fkp	-3.1562**	0.96584	0.26547	6	-0.73671	0.4615	0.0025
Rwbc_fkp	-3.2749**	0.96484	0.26539	5	-0.70529	0.4808	0.0038
Rwbc_fkp	-3.3930**	0.96388	0.26530	4	-1.3651	0.1726	0.0055
Rwbc_fkp	-3.6106**	0.96190	0.26545	3	1.1800	0.2384	0.0049
Rwbc_fkp	-3.4832**	0.96356	0.26552	2	-2.4893	0.0130	0.0051
Rwbc_fkp	-3.8769**	0.95975	0.26642	1	0.073181	0.9417	0.0011
Rwbc_fkp	-3.9106**	0.95986	0.26625	0			0.0018
Rcba_asx	-2.8954**	0.97517	0.40887	10	-1.7550	0.0797	0.9957
Rcba_asx	-3.0982**	0.97355	0.40943	9	0.92469	0.3554	0.7510
Rcba_asx	-3.0176**	0.97439	0.40940	8	2.0194	0.0438	0.7446
Rcba_asx	-2.8330**	0.97601	0.41022	7	-2.2966	0.0219	0.4002
Rcba_asx	-3.0755**	0.97402	0.41138	6	-1.5897	0.1123	0.1381
Rcba_asx	-3.2632**	0.97257	0.41179	5	-0.27721	0.7817	0.0970
Rcba_asx	-3.3161**	0.97232	0.41154	4	-0.54296	0.5873	0.1345
Rcba_asx	-3.3998**	0.97181	0.41135	3	2.4929	0.0129	0.1700
Rcba_asx	-3.1505**	0.97392	0.41275	2	-1.6389	0.1017	0.0463
Rcba_asx	-3.3456**	0.97244	0.41321	1	0.20405	0.8384	0.0317
Rcba_asx	-3.3453**	0.97263	0.41295	0			0.0453

Table 5.26: AR(1) models analysis

ϵ	$\hat{\phi}$	t−prob	$\hat{\sigma}_a$	Pr > Chisq to lag 50
rWbcBoq	0.94834	0.0000	0.268199	0.0295
rWbcFkp	0.96010	0.0000	0.265	0.0885
rCbaAsx	0.97360	0.0000	0.411464	0.0853

Table 5.27: Estimation from integral equations :

Residuals	Min Total Profit		Optimal	Ave Trading	Trades Number	
	3 years	3 months	Threshold	Duration	3 years	3months
$rWbcBoq_t$	$17.0364	$ 1.4197	0.55	17.2973 days	30.9752	2.58126
$rWbcFkp_t$	$ 15.0756	$ 1.2563	0.58	20.8619 days	25.9922	2.1660
$rCbaAsx_t$	$ 19.3962	$ 1.61635	0.97	27.6925 days	19.996	1.6663

Table 5.28: Trading simulation results using training period data (in-sample) in October 2003 - September 2006 with the thresholds from Table 5.27:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	$ 20.9082	$ 14.8731	$25.6794
Number of complete and incomplete trades	27	19	16
Number of complete trades	26	18	16
Average trading duration	21.81 days	32.94 days	43.56 days
Average profit per trade	$ 0.7744	$ 0.7828	$ 1.6049
Average return per trade	4.7972%	4.8454%	4.6128%
Annualized return	41.5755% pa	29.0723% pa	24.6020% pa

Table 5.29: Trading simulation results using trading period data (out-sample) in October 2006 - December 2006 with the thresholds from Table 5.27:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total Profit	$ 2.2840	$ 0.4040	$ 3.0146
Number of complete and incomplete trades	3	2	4
Number of complete trades	2	1	3
Average trading duration	21 days	21 days	15 days
Average profit per trade	$ 0.7613	$ 0.2020	$ 0.7536
Average return per trade	2.9697%	0.8472%	1.7950 %
Annualized return	23.7577 % pa	3.3888 % pa	21.5404 % pa

5.7 Summary Results and Conclusion

Summary results from the 6-months trading periods and the 3-months trading periods are provided in Tables 5.30 and 5.31, respectively. From the summary trading simulation results in Tables 5.30 and 5.31, we can conclude that generally the results from the 6-months trading periods are better than the 3-months trading periods in terms of the average profit per trade and the average return per trade. It happens because when we use 3-months trading periods, the frequencies of making losses due to incomplete trades are higher than using 6-months trading periods. The significant impact can be seen from the pair WBC-FKP which has significant smaller profit per trade and the average return per trade when we use 3-months trading periods compared to these indicators when we use 6-months trading periods. Results given in this chapter are a revised version of Puspaningrum *et al.* (2009).

Table 5.30: Summary trading simulation result which is a combination of Tables 5.7 and 5.13 of 6-months trading periods:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	\$ 10.8049	\$ 6.3274	\$ 12.5832
Total number of complete and incomplete trades	12	8	10
Total number of complete trades	11	7	9
Average trading duration	18.2 days	28.54 days	23.91 days
Average profit per trade	\$ 0.9024	\$ 0.7909	\$ 1.1891
Average return per trade	3.57 %	4.36%	2.97 %
Average annualized return	38.53 % pa	32.17 % pa	29.42 % pa

Table 5.31: Summary trading simulation results of 3-months trading periods which is a combination of Tables 5.15, 5.21, 5.23 and 5.29:

Pairs:	WBC-BOQ	WBC-FKP	CBA-ASX
Total profit	\$ 10.4198	\$ 3.2547	\$ 14.8764
Total number of complete and incomplete trades	13	9	13
Total number of complete trades	11	5	10
Average trading duration	15.62 days	15.375 days	11.83 days
Average profit per trade	\$ 0.7995	\$ 0.0885	\$ 1.2061
Average return per trade	3.505%	0.05%	2.94%
Average annualized return	40.34 % pa	14.93% pa	29.20% pa

Chapter 6

Unit Root Tests for ESTAR Models

6.1 Introduction

Since the introduction of unit root tests in Fuller (1976) and then Dickey and Fuller (1979, 1981), many new types of unit root tests have been developed. Developments in nonlinear unit root tests occurred as the standard linear unit root tests performed poorly for nonlinear processes. For example, Pippenger and Goering (1993) showed that the power of the standard DF tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. Other researchers have attempted to address similar issues in the context of a TAR model, for example, Balke and Fomby (1997), Enders and Granger (1998), Berben and van Dijk (1999), Caner and Hansen (2001), and Lo and Zivot (2001) (see Subsection 2.1.2 for a brief description of TAR model).

The smooth transition autoregressive (STAR) process developed by Granger and Terasvirta (1993) has been a popular process for modelling economic and finance data due to its generality and flexibility. Nonlinear adjustment in a STAR model allows for smooth rather than discrete adjustment in a TAR model. In a STAR model, adjustment takes place in every period but the speed adjustment varies with the extent of the deviation from equilibrium. A STAR(p) model can be expressed as follow:

$$y_t = \theta_{1,0}+\sum_{j=1}^{p}\theta_{1,j}\ y_{t-j}+\left[\theta_{2,0}+\sum_{j=1}^{p}\theta_{2,j}\ y_{t-j}\right]G(\theta, e, y_{t-d})+\epsilon_t, \qquad t=1,2,\ldots,T, \quad (6.1)$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d \geq 1$ is a delay parameter; $(\theta, e) \in \{\mathbb{R}^+ \times \mathbb{R}\}$ where $\mathbb{R}$ denotes the real space $(-\infty, \infty)$ and $\mathbb{R}^+$ denotes the nonnegative real space $(0, \infty)$. The transition function $G(\theta, e, y_{t-d})$ determines the speed of adjustment to the equilibrium e. Two simple transition functions suggested by Granger and Terasvirta (1993) and Terasvirta (1994) are the logistic and exponential functions:

$$G(\theta, e, y_{t-d}) = \frac{1}{1 + \exp\{-\theta(y_{t-d} - e)\}} - \frac{1}{2}, \tag{6.2}$$

$$G(\theta, e, y_{t-d}) = 1 - \exp\{-\theta^2(y_{t-d} - e)^2\}. \tag{6.3}$$

If the transition function $G(\theta, e, y_{t-d})$ is given by (6.2), (6.1) is called a logistic smooth transition autoregressive (LSTAR) model. If the transition function $G(\theta, e, y_{t-d})$ is given by (6.3), (6.1) is called an exponential smooth transition autoregressive (ESTAR) model.

Granger and Terasvirta (1993), Terasvirta (1994) and Luukkonen *et al.* (1988) suggested applying a sequence of linearity tests to an artificial regression of Taylor series expansions of (6.2) or (6.3). This allows detection of general nonlinearity through the significance of the higher-order terms in the Taylor expansions, with the value of delay parameter d selected as that yielding the largest value of the test statistics. The tests can also be used to discriminate between ESTAR or LSTAR models, since some higher-order terms disappear in the Taylor expansion of the ESTAR transition function.

The logistic transition function in (6.2) is bounded between -1/2 and 1/2 and implies asymmetric behavior of y_t depending upon whether it is above or below the equilibrium level (see Figure 6.1(a)). On the other hand, the exponential transition function in (6.3) is bounded between zero and unity and symmetrically inverse-bell shaped around equilibrium level e (see Figure 6.1(b)). These properties of an ESTAR model are more attractive in the present modelling context than a LSTAR model because it allows a smooth transition between regimes and symmetric adjustment of y_t for deviation above and below the equilibrium level e. The transition parameter θ in ESTAR models determines the speed of transition between the two extreme regimes, with lower absolute value of θ implying slower transition. The inner regime in ESTAR models corresponds to $y_{t-d} = e$, so that $G(\theta, e, y_{t-d}) = 0$ and (6.1) becomes a linear AR(p) model:

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{6.4}$$

The outer regime of ESTAR models corresponds to $\lim_{(y_{t-d}-e)\to\pm\infty} G(\theta, e, y_{t-d}) = 1$, for a given θ, so that (6.1) becomes a different linear AR(p) model as follow:

$$y_t = \theta_{1,0} + \theta_{2,0} + \sum_{j=1}^{p} (\theta_{1,j} + \theta_{2,j})\, y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{6.5}$$

An ESTAR model has become a popular model to analyse some economic and finance data. Michael *et al.* (1997), Taylor *et al.* (2001) and Paya *et al.* (2003) used an ESTAR model to analyse real exchange rate and purchasing power parity (PPP) deviations. Terasvirta and Elliasson (2001), and Sarno *et al.* (2002) used an ESTAR

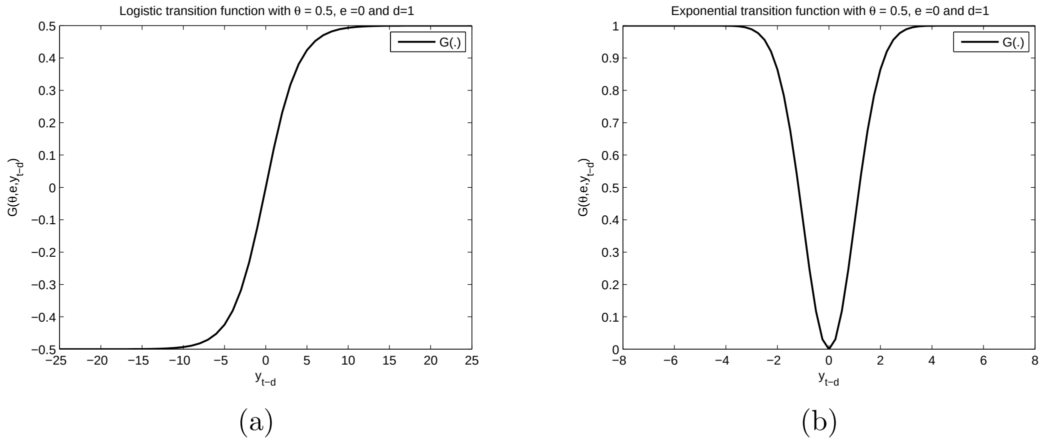


Figure 6.1: Plots of transition functions with $\theta = 0.5$, $e = 0$ and $d = 1$.

model to analyse deviations from optimal money holding. Monoyios and Sarno (2002) found that symmetric deviations from arbitrage processes such as stock index futures follow ESTAR models.

In economics and finance theories such as real exchange rate, PPP, and arbitrage processes, an ESTAR model can be characterised by a unit root behavior in the inner regime, but for large deviations, the process is mean reverting. Kapetanios *et al.* (2003) considered a unit root test for an ESTAR(1) model and applied their test to real interest rates and rejected the null hypothesis for several interest rates considered, whereas Augmented Dickey-Fuller (ADF) tests failed to do so.

Venetis *et al.* (2009) developed a unit root test for k-ESTAR(p) model where k is the number of equilibrium levels. In their model, the transition function involves k equilibriums:

$$G(\theta, \mathbf{e}, y_{t-d}) = 1 - \exp\left[-\theta^2 \left(\prod_{i=1}^{k}(y_{t-d} - e_i)\right)^2\right] \tag{6.6}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_k)'$.

As noted by Venetis *et al.* (2009), many economic theories support the existence of multiple equilibria. For example, in the case of inflation, attempts by governments to finance substantial proportion of expenditure by seigniorage can lead to multiple inflationary equilibria (see Cagan, 1956 and Sargent and Wallace, 1973). Theoretical models suggest that, in these circumstances, inflation follows a non-linear process and that the stability characteristics depend on expectations formation. In the case of unemployment, shocks from public produce not merely fiscal and monetary (demand policy) responses but also changes in supply-side policy (effecting the equilibrium values of real variables or "natural rate")(see Diamond, 1982 and Layard *et al.*, 1991). With regard to monetary policy rules, some models suggest that once you take into account the zero bound on nominal interest rate, real interest rates might follow a number equilibria (see Benhabib *et al.*, 1999).

Even though Venetis *et al.* (2009) developed a unit root test for a more general form of ESTAR(p) model but their approach might cause singularity problem because some of the regressors might be collinear. To overcome the problem, they added the collinear regressors into the error term. Even though the test under alternative is consistent, but it may make a significant difference for some cases.

This chapter extends the work of Kapetanios *et al.* (2003) by considering a unit root test for a k-ESTAR(p) model with a different approach to Venetis *et al.* (2009). By using a new approach given in this chapter, the singularity problem can be avoided without adding the collinear regressors into the error term. For some cases, simulation results show that our approach is better than Venetis *et al.* (2009) , Kapetanios *et al.* (2003) and the Augmented Dickey-Fuller (ADF) test Dickey and Fuller (1979, 1981).

In this chapter, Section 6.2 will give overview about unit root test for an 1-ESTAR(1) model from Kapetanios *et al.* (2003). Section 6.3 explains unit root test derivation for a k-ESTAR(p) model based on Venetis *et al.* (2009). Section 6.4 explains the new unit root test derivation for a k-ESTAR(p) model. As the asymptotic distribution of the test for a k-ESTAR(2) model does not contain a nuisance parameter while that for a k-ESTAR(p) model contains nuisance parameters, Section 6.5 will give further analysis of unit root test for a k-ESTAR(2) model and Section 6.6 will give further analysis of unit root test for a k-ESTAR(p) model, $p > 2$. Section 6.7 presents conclusions from this chapter.

Some symbols are used in the following derivation, i.e.: $\int W = \int_0^1 W(s)ds$ where $W(s)$ is the standard Brownian motion defined on $s \in [0,1]$; " $\Rightarrow$" means convergence in distribution; " $\rightarrow$" means convergence in probability; $X_t = o_p(1)$ means that $X_t \rightarrow 0$ in probability as $t \rightarrow \infty$; $X_t = O_p(1)$ means that X_t is bounded in probability, i.e. for every $\varepsilon > 0$ there is an $M < \infty$ such that $P(|X_t| > M) < \varepsilon$ for all t.

6.2 Unit Root Test for an 1-ESTAR(1) Model

Kapetanios *et al.* (2003) developed a simple unit root test for an ESTAR(1) model by assuming that y_t is a mean zero stochastic process. Thus, (6.1) can be simplified as follow:

$$y_t = \theta_{1,1}\, y_{t-1} + \theta_{2,1}\, y_{t-1} G(\theta, y_{t-d}) + \epsilon_t, \qquad t = 1, 2, \ldots, T \tag{6.7}$$

where $G(\theta, y_{t-d}) = 1 - \exp\{-\theta^2\ y_{t-d}^2\}$, $\theta > 0$ and $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2.

Equation (6.7) can be reparamaterised as

$$\Delta y_t = \phi\, y_{t-1} + \theta_{2,1}\, y_{t-1} \left[1 - \exp(-\theta^2\ y_{t-d}^2)\right] + \epsilon_t, \qquad t = 1, 2, \ldots, T \tag{6.8}$$

where $\phi = \theta_{1,1} - 1$.

They imposed $\phi = 0$ or $\theta_{1,1} = 1$ implying that y_t follows a unit root process in the inner regime. They also set $d = 1$, so that they had a specific ESTAR(1) as

$$\Delta y_t = \theta_{2,1}\, y_{t-1} \left(1 - \exp(-\theta^2\, y_{t-1}^2)\right) + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{6.9}$$

Their test directly focused on parameter θ, where they tested

$$H_0 : \theta = 0 \tag{6.10}$$

against the alternative

$$H_1 : \theta > 0. \tag{6.11}$$

Thus, under the null hypothesis, y_t has a linear unit root and under the alternative hypothesis, y_t is a nonlinear globally stationary ESTAR(1) process.

Testing the null hypothesis (6.10) for (6.9) directly is not feasible, since $\theta_{2,1}$ is not identified under the null. To overcome this problem, they followed Luukkonen *et al.* (1988) by using first-order Taylor approximation around $\theta^2 = 0$ to (6.9) and then derived a t-type test statistic. Thus, under the null, they obtained the auxiliary regression

$$\Delta y_t = \delta y_{t-1}^3 + \text{ error }, \qquad t = 1, 2, \ldots, T \tag{6.12}$$

where $\delta = \theta_{2,1}\, \theta^2$.

They obtained the t-statistics for $\delta = 0$ against $\delta < 0$,[1]

$$t_{KSS} = \frac{\hat{\delta}}{\text{s.e.}(\hat{\delta})}, \tag{6.13}$$

where $\hat{\delta}$ is the OLS estimate of δ and s.e.$(\hat{\delta})$ is the standard error of $\hat{\delta}$.

They found that the t_{KSS} statistic defined by (6.13) does not have an asymptotic standard normal distribution but has the following asymptotic distribution:

$$t_{KSS} \Rightarrow \frac{\frac{1}{4}W(1)^4 - \frac{3}{2}\int_0^1 W(s)^2 ds}{\sqrt{\int_0^1 W(s)^6 ds}}, \qquad T \to \infty, \tag{6.14}$$

where $W(s)$ is a standard Brownian motion on [0,1]. Under the alternative hypothesis (6.11) for ESTAR(1) model in (6.9), the t_{KSS} statistic is consistent.

[1] Under the alternative hypothesis, the requirement for y_t to be globally stationary is $|\theta_{1,1} + \theta_{2,1}G(\theta, y_{t-d})| < 1$. Since $\theta_{1,1} = 1$, then $-2 < \theta_{2,1}G(\theta, y_{t-d}) < 0$. Since $1 \geq G(\theta, y_{t-d}) \geq 0$, then $-2 < \theta_{2,1} < 0$.

6.3 Unit Root Test for a k-ESTAR(p) Model Based on Venetis *et al.* (2009)

In this section, a unit root test for k-ESTAR(p) based on Venetis *et al.* (2009) will be explained. Consider a k-ESTAR(p) model as follow,

$$y_t = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\, y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\, y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \qquad t = 1, 2, \ldots, T \tag{6.15}$$

with

$$G(\theta, \mathbf{e}, y_{t-d}) = 1 - \exp\left\{-\theta^2 \left[\prod_{i=1}^{k}(y_{t-d} - e_i)\right]^2\right\}, \tag{6.16}$$

where ϵ_t is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2, $\theta_{\mathbf{1}} = (\theta_{1,0}, \theta_{1,1}, \cdots, \theta_{1,p})'$, $\theta_{\mathbf{2}} = (\theta_{2,0}, \theta_{2,1}, \cdots, \theta_{2,p})'$, $\mathbf{e} = (e_1, \ldots, e_k)'$, d and θ are unknown parameters. The variable y_{t-d} in function $G(\theta, \mathbf{e}, y_{t-d})$ is called a transition variable. Define the polynomial $\theta_1(L) = 1 - \sum_{j=1}^{p} \theta_{1,j}$ and $\theta_2(L) = 1 - \sum_{j=1}^{p} \theta_{2,j}$ as the "linear and nonlinar autoregressive polynomials". Venetis *et al.* (2009) interested in the special case of a unit root $\theta_1(L) = 0$ in the linear polynomial, thus, the subsequent analysis is based on the restriction

$$\sum_{j=1}^{p} \theta_{1,j} = 1. \tag{6.17}$$

As Venetis *et al.* (2009) noted that the transition function $G(.)$ no longer follows the familiar U-shape of a 1-ESTAR(p) model although it is still bounded between 0 and 1. The smoothness or transition speed parameter θ is one of the factors determining the speed of transition between regimes $G(.) = 0$ and $G(.) = 1$ along with the distance of y_{t-d} from a specified location e_i as in the typical ESTAR model (see Figure 6.2 for two transition function with different speeds adjustment and same two "fixed" points). However, a k-ESTAR(p) model supports a much wider dynamic behavior as the adjustment speed need not be symmetric around any location point depending on the number of location points as well as their relative distance.

Not like k-ESTAR(1) models, it is not easy to determine the stationary conditions for k-ESTAR(p) models, $p > 1$. However, similar geometric ergodicity and associated globally stationary conditions in Kapetanios *et al.* (2003) hold for (6.15) and following Bhattacharya and Lee (1995, Theorem 1), we consider the condition $\left|\sum_{j=1}^{p}(\theta_{1,j} + \theta_{2,j})\right| < 1$ which is a necessary stationary condition for k-ESTAR(p).

For the k-ESTAR(p) model in (6.15), an equilibrium level e_i can be defined as any

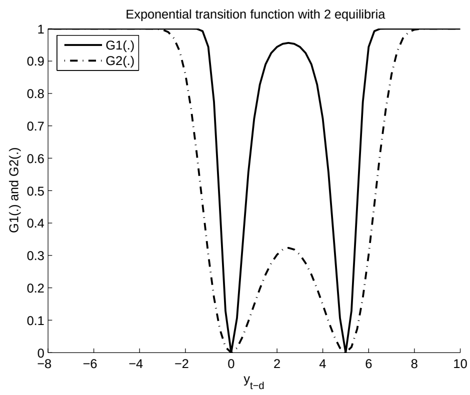


Figure 6.2: Plots of transition functions $G1 = 1 - exp[-0.08(y_{t-d} - 0)^2(y_{t-d} - 5)^2]$ and $G2 = 1 - exp[-0.01(y_{t-d} - 0)^2(y_{t-d} - 5)^2]$.

real number y^* that solves

$$y^* = \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\ y^* + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\ y^*\right] G(\theta, \mathbf{e}, y^*). \tag{6.18}$$

Venetis *et al.* (2009) rearranged (6.15) to become

$$\begin{aligned}\Delta y_t &= \theta_{1,0} + \sum_{j=1}^{p} \theta_{1,j}\ y_{t-1} + \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\ y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ &\quad t = 1, 2, \ldots, T \end{aligned} \tag{6.19}$$

where $\theta^*_{1,j} = -\sum_{k=j+1}^{p} \theta_{1,k}, \quad j = 1, \cdots, (p-1)$.

When $y_{t-d} = e_1$ or $y_{t-d} = e_2$ or $\cdots$ or $y_{t-d} = e_k$, and there is a unit root in the linear term, i.e. $\sum_{j=1}^{p} \theta_{1,j} = 1$, the k-ESTAR(p) model allows for multiple "inner" regimes with $G(.) = 0$ and (6.19) reduces to

$$\Delta y_t = \theta_{1,0} + \sum_{j=1}^{p-1} \theta^*_{1,j}\ \Delta y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{6.20}$$

Thus, y_t behaves as a random walk process (with a drift if $\theta_{1,0} \neq 0$). For 1-ESTAR(p) models, this case is consistent with the existence of an equilibrium around which the series behaves as a random walk.

In the "outer" regimes when function $G(.) \approx 1$, model (6.19) reduces to

$$\Delta y_t = (\theta_{1,0} + \theta_{2,0}) + \sum_{j=1}^{p} \theta_{2,j}\ y_{t-j} + \sum_{j=1}^{p-1} (\theta^*_{1,j} + \theta^*_{2,j})\ \Delta y_{t-j} + \epsilon_t, \qquad t = 1, 2, \ldots, T. \tag{6.21}$$

Assuming $|\sum_{j=1}^{p}(\theta_{1,j} + \theta_{2,j})| < 1$, y_t will be a linear stationary AR(p) model.

Venetis *et al.* (2009) developed an F-type test for the null hypothesis of unit root, $H_0 : \theta = 0$ in (6.15). However, testing H_0 in (6.15) cannot be performed directly due to a well known identification problem (see Luukkonen *et al.*, 1988, and Terasvirta, 1994, for details). Following Luukkonen *et al.* (1988), the identification problem is circumvented by using a Taylor approximation of the nonlinear function $G(\theta, \mathbf{e}, y_{t-d})$ around the null hypothesis.

The second order Taylor series approximation of $G(\theta) = G(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ is

$$\begin{aligned} G(\theta) &\approx G(0) + \left.\frac{\partial G}{\partial \theta}\right|_{\theta=0} \theta + \frac{1}{2} \left.\frac{\partial^2 G}{\partial \theta^2}\right|_{\theta=0} \theta^2 + R \\ &= \theta^2 \prod_{i=1}^{k} (y_{t-d} - e_i)^2 + R, \end{aligned} \tag{6.22}$$

where R is the remainder, $G(0) = 0$ and $\frac{\partial G}{\partial \theta}\big|_{\theta=0} = 0$.

Under the unit root assumption, $\sum_{j=1}^{p} \theta_{1,j} = 1$ and substituting (6.22) into (6.19), we obtain

$$\begin{aligned} \Delta y_t &= \theta_{1,0} + \sum_{j=1}^{p-1} \theta_{1,j}^* \, \Delta y_{t-j} + \theta^2 \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j} \, y_{t-j} \right] \prod_{i=1}^{k} (y_{t-d} - e_i)^2 + \epsilon_t^*, \\ &\quad t = 1, 2, \ldots, T \end{aligned} \tag{6.23}$$

where $\epsilon_t^* = \left(\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j} \, y_{t-j} \right) R + \epsilon_t$.

Under the null hypothesis of a unit root process against the globally stationary process generated by (6.15) is equivalent to testing

$$H_0 : \theta^2 = 0 \tag{6.24}$$

in (6.23). Under the null hypothesis $\epsilon_t^* = \epsilon_t$, so an F-type test can be constructed. However, it is clear that the approach results in overfitting even for moderate autoregressive polynomial orders p (assuming a reasonable value of k). The general auxiliary regression through which (6.23) will be tested can be written as

$$\begin{aligned} \Delta y_t &= \theta_{1,0} + \sum_{j=1}^{p-1} \theta_{1,j}^* \, \Delta y_{t-j} \; + \\ &\quad \theta^2 \left[\theta_{2,0} \, \delta_0 + \delta_0 \sum_{j=1}^{p} \theta_{2,j} \, y_{t-j} + \theta_{2,0} \sum_{s=1}^{2k} \delta_s \, y_{t-d}^s + \sum_{j=1}^{p} \sum_{s=1}^{2k} \theta_{2,j} \, \delta_s \, y_{t-j} \, y_{t-d}^s \right] + \epsilon_t^*, \\ &= \lambda_0 + \sum_{j=1}^{p-1} \theta_{1,j}^* \, \Delta y_{t-j} \; + \sum_{j=1}^{p} \lambda_{1,j} \, y_{t-j} + \sum_{s=1}^{2k} \lambda_{2,s} \, y_{t-d}^s + \sum_{j=1}^{p} \sum_{s=1}^{2k} \lambda_{3,js} \, y_{t-j} \, y_{t-d}^s + \epsilon_t^*, \end{aligned} \tag{6.25}$$

where $\prod_{i=1}^{k}(y_{t-d}-e_i)^2=\delta_0+\sum_{s=1}^{2k}\delta_s y_{t-d}^s$ with parameters δ_s being functions of the location parameter e_i; $\lambda_0=\theta_{1,0}+\theta^2\theta_{2,0}\,\delta_0$; $\lambda_{1,j}=\theta^2\delta_0\theta_{2,j}$; $\lambda_{2,s}=\theta^2\theta_{2,0}\delta_s$; $\lambda_{3,js}=\theta^2\theta_{2,j}\,\delta_s$; $j=1,\ldots,p$ and $s=1,\ldots,2k$. In particular, $\delta_0=\prod_{i=1}^{k}e_i^2$ and $\delta_{2k}=1$. In addition $\delta_s=0$ for $s=0,1,\ldots,(2k-1)$ if $\mathbf{e}=0$.

Finally, if some $e_i=0$, then $\delta_0=0$ and (6.25) will become

$$\Delta y_t \;=\; \theta_{1,0}+\sum_{j=1}^{p-1}\theta_{1,j}^*\,\Delta y_{t-j}\;+\sum_{s=2}^{2k}\lambda_{2,s}\,y_{t-d}^s+\sum_{j=1}^{p}\sum_{s=2}^{2k}\lambda_{3,js}\,y_{t-j}\,y_{t-d}^s+\epsilon_t^*. \tag{6.26}$$

If $e_i\neq 0$ for all i, then testing the null hypothesis of a unit root against the alternative of a globally stationary k-ESTAR(p) model is equivalent with testing

$$H_0:\lambda_{1,j}=\lambda_{2,s}=\lambda_{3,js}=0,\ \text{for all } s,j, \tag{6.27}$$

in (6.25). If $e_i=0$ for a certain i, then testing the null hypothesis of a unit root against the alternative of a globally stationary k-ESTAR(p) model is equivalent with testing

$$H_0:\lambda_{2,s}=\lambda_{3,js}=0,\ \text{for all } s,j, \tag{6.28}$$

in (6.26).

Setting $\theta_{1,0}=0$ assuming random walk without drift when $y_{t-d}=e_i$, (6.25) can be written as partitioned regression model,

$$Y=X_1\ \mathbf{b}_1+X_2\ \mathbf{b}_2+\epsilon_t^* \tag{6.29}$$

where Y is data matrix of dependent variables on the left hand side of (6.25); X_1 is data matrix including the $p-1$ regressors on the right hand side of (6.25), i.e. $X_1=(\Delta y_{t-1},\Delta y_{t-2},\ldots,\Delta y_{t-p+1})'$, that are stationary under the null hypothesis; X_2 is data matrix of the remaining regressors on the right hand side of (6.25) and $\epsilon_t^*=\epsilon_t+R$, R is the remainder.

The F-type statistic to test the null hypothesis in (6.27) is

$$F_{VPP}=\frac{1}{\hat{\sigma}^2_{\epsilon^*}}\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right)'(X_2'M_1X_2)\left(\hat{\mathbf{b}}_2-\mathbf{b}_2\right) \tag{6.30}$$

where $M_1=I-X_1(X_1'X_1)^{-1}X_1'$ is the orthogonal to X_1 projection matrix and $\hat{\sigma}^2_{\epsilon^*}$ is the maximum likelihood estimator of the error variance.

Venetis *et al.* (2009) argued that not all regressors in X_2 are necessary for testing procedure because under the null of non-stationarity and for finite orders p, the regressors $y_{t-1},y_{t-2},\ldots,y_{t-p}$ are collinear asymptotically. The same conclusion is reached for regressors involving powers of the transition variable where $y_{t-d}^s, y_{t-1}y_{t-d}^{s-1},\ldots,y_{t-p}y_{t-d}^{s-1}$, $s=2,\ldots,2k$, are also collinear asymptotically. These collinear regressors result in a singularity problem in calculating F_{VPP} in (6.30). In order to overcome this difficulty, they re-specify the auxiliary regression model in (6.25) into

$$\Delta y_t = \sum_{j=1}^{p-1} \theta_{1,j}^* \, \Delta y_{t-j} + \lambda_0^* + \lambda_1 \, y_{t-1} + \sum_{s=2}^{2k} \lambda_{2,s} \, y_{t-d}^s + \lambda_3 \, y_{t-1} \, y_{t-d}^{2k} + \upsilon_t, \tag{6.31}$$

where $\lambda_0^* = \theta^2 \theta_{2,0} \, \delta_0$; and

$$\upsilon_t = \epsilon_t^* + \sum_{j=2}^{p-1} \lambda_{1,j} \, y_{t-j} + \lambda_{2,1} y_{t-d} + \sum_{j=2}^{p} \sum_{s=1}^{2k-1} \lambda_{3,js} \, y_{t-j} \, y_{t-d}^s.$$

Thus, they added some regressors into the error term υ_t. If $e_i = 0$ for a certain i, then (6.31) becomes

$$\Delta y_t = \sum_{j=1}^{p-1} \theta_{1,j}^* \, \Delta y_{t-j} + \sum_{s=2}^{2k} \lambda_{2,s} \, y_{t-d}^s + \lambda_3 \, y_{t-1} \, y_{t-d}^{2k} + \upsilon_t. \tag{6.32}$$

Using the auxiliary regression model in (6.31), if $e_i \neq 0$ for all i, then testing the null hypothesis of a unit root against the alternative of a globally stationary k-ESTAR(p) model is equivalent with testing

$$H_0 : \lambda_1 = \lambda_{2,s} = \lambda_3 = 0, \text{ for all } s = 2, \ldots, 2k. \tag{6.33}$$

If $e_i = 0$ for a certain i, then testing the null hypothesis of a unit root against the alternative of a globally stationary k-ESTAR(p) model is equivalent with testing

$$H_0 : \lambda_{2,s} = \lambda_3 = 0, \text{ for all } s = 2, \ldots, 2k, \tag{6.34}$$

in (6.32).

The partitioned regression model,

$$Y = X_1 \, \mathbf{b}_1 + X_{2*} \, \mathbf{b}_{2*} + \epsilon_t^* \tag{6.35}$$

where Y is data matrix of dependent variables on the left hand side of (6.31); X_1 is data matrix including the $p-1$ regressors on the right hand side of (6.31), i.e. $(\Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-p+1})$, that are stationary under the null hypothesis; X_{2*} is data matrix of the remaining regressors on the right hand side of (6.31) and $\epsilon_t^* = \epsilon_t + R$, R is the remainder.

The F-type statistic to test the null hypothesis in (6.31) is

$$F_{VPP} = \frac{1}{\hat{\sigma}_\upsilon^2} \left(\hat{\mathbf{b}}_{2*} - \mathbf{b}_{2*} \right)' (X_{2*} M_1 X_{2*}) \left(\hat{\mathbf{b}}_{2*} - \mathbf{b}_{2*} \right) \tag{6.36}$$

where $M_1 = I - X_1 (X_1' X_1)^{-1} X_1'$ is the orthogonal to X_1 projection matrix and $\hat{\sigma}_\upsilon^2$ is the maximum likelihood estimator of the error variance.

If $e_i \neq 0$ for all i, under the null hypothesis $H_0 : \theta = 0$,

$$F_{VPP} \Rightarrow G'_{1*}(W)\ G^{-1}_{2*}(W)\ G_{1*}(W) \tag{6.37}$$

where

$$G_{1*}(W) = \left(W(1),\ \int W dW,\ \int W^2 dW,\ \ldots,\ \int W^{2k+1} dW\right)',$$

$$G_{2*}(W) = \begin{bmatrix} 1 & \int W & \int W^2 & \cdots & \int W^{2k+1} \\ \int W & \int W^2 & \int W^3 & \cdots & \int W^{2k+2} \\ \int W^2 & \int W^3 & \int W^4 & \cdots & \int W^{2k+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \int W^{2k+1} & \int W^{2k+2} & \int W^{2k+3} & \cdots & \int W^{(4k+2)} \end{bmatrix},$$

$\int W^x = \int_0^1 W(s)^x ds$ and $\int W^x dW = \int_0^1 W(s)^x dW(s)$ where $W(s)$ is the standard Brownian motion defined on $s \in [0,1]$ and x in an integer number.

If $e_i = 0$ for a certain i, then $G_{1*}(W)$ and $G_{2*}(W)$ in (6.37) will be:

$$G_{1*}(W) = \left(\int W^2 dW,\ \int W^3 dW,\ \ldots,\ \int W^{2k+1} dW\right)',$$

$$G_{2*}(W) = \begin{bmatrix} \int W^4 & \int W^5 & \cdots & \int W^{2k+3} \\ \int W^5 & \int W^6 & \cdots & \int W^{2k+4} \\ \vdots & \vdots & \ddots & \vdots \\ \int W^{2k+2} & \int W^{2k+3} & \cdots & \int W^{(4k+2)} \end{bmatrix}.$$

6.4 A New Approach of Unit Root Test for a k-ESTAR(p) Model

In this section we develop a unit root test for a k-ESTAR(p) model with a slightly different approach compared to Venetis *et al.* (2009). Especially, we are interested in the case where y_t is a zero mean series[2] and has a unit root in the linear term, i.e. $\theta_{1,0} = 0$ and $\sum_{j=1}^{p} \theta_{1,j} = 1$ [3]. Consider a k-ESTAR(p) model in (6.15) and (6.16). An equilibrium level e_i can be defined as any real number y^* that solves

$$0 = \left[\theta_{2,0} + \sum_{j=1}^{p} \theta_{2,j}\ y^*\right] G(\theta, \mathbf{e}, y^*). \tag{6.38}$$

[2]If y_t is not a zero mean series, we can de-mean the series so that the adjustment series will has a zero mean. This de-mean strategy was also applied in He and Sandberg (2005b). Empirical examples in Venetis *et al.* (2009) and Monoyios and Sarno (2002) support the assumption is satisfied in practice.

[3]Note that when $\sum_{j=1}^{p} \theta_{1,j} = 1$ hold, the restriction $-2 < \sum_{j=1}^{p} \theta_{2,j} < 0$ ensures ergodicity of the process.

One of solutions for (6.38) is $y_1^* = -\frac{\theta_{2,0}}{\sum_{j=1}^{p}\theta_{2,j}}$ where $\sum_{j=1}^{p}\theta_{2,j} \neq 0$. This solution is named as the first equilibrium e_1 while other solutions as i^{th} equilibriums[4], i.e. $y_i^* = e_i$, $i = 2, \cdots, k$. Note that since under the alternative $\sum_{j=1}^{p}\theta_{2,j} \neq 0$ and $e_1 = -\frac{\theta_{2,0}}{\sum_{j=1}^{p}\theta_{2,j}}$, then $e_1 = 0$ if only if $\theta_{2,0} = 0$. Therefore, if $\theta_{2,0} = 0$, one of the equilibrium should be zero.

Instead of rearrange (6.15) to become (6.19), we rearrange (6.15) to become

$$\begin{aligned} y_t &= \theta_{1,0} + \sum_{j=1}^{p}\theta_{1,j}\; y_{t-1} + \sum_{j=1}^{p-1}\theta_{1,j}^*\; \Delta y_{t-j} \\ &+ \left[\theta_{2,0} + \sum_{j=1}^{p}\theta_{2,j}\; y_{t-1} + \sum_{j=1}^{p-1}\theta_{2,j}^*\; \Delta y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ &\quad t = 1, 2, \ldots, T \end{aligned} \tag{6.39}$$

where $\theta_{i,j}^* = -\sum_{k=j+1}^{p}\theta_{i,k}$, $j = 1, \cdots, (p-1)$ and $i = 1, 2$.

Let $\theta_{1,0} = 0$ and $\sum_{j=1}^{p}\theta_{1,j} = 1$ meaning that $\{y_t\}$ has a unit root without a drift in the linear term[5]. Furthermore, without loss generality, assume that $\theta_{2,0} = 0$ so that $e_1 = 0$. Thus, (6.39) can be arranged to become

$$\begin{aligned} \Delta y_t &= \sum_{j=1}^{p-1}\theta_{1,j}^*\; \Delta y_{t-j} + \left[\sum_{j=1}^{p}\theta_{2,j}\; y_{t-1} + \sum_{j=1}^{p-1}\theta_{2,j}^*\; \Delta y_{t-j}\right] G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ &\quad t = 1, 2, \ldots, T, \end{aligned} \tag{6.40}$$

where

$$G^*(\theta, \mathbf{e}, y_{t-d})) = 1 - \exp\left\{-\theta^2 y_{t-d}^2 \left[\prod_{i=2}^{k}(y_{t-d} - e_i)\right]^2\right\}. \tag{6.41}$$

6.4.1 F-test Procedure

In this section, we develop a F-test for testing the null hypothesis of unit root, $H_0 : \theta = 0$ for k-ESTAR(p) model in (6.40). Like 1-ESTAR(1) model in Kapetanios *et al.* (2003) and k-ESTAR(p) in Venetis *et al.* (2009), testing H_0 can not be done directly due to a well known identification problem. We use the same strategy in Venetis *et al.* (2009) to solve the problem by using second order Taylor approximation of the nonlinear function around $\theta = 0$ in (6.22).

[4]Following Bair and Haesbroeck (1997) further differentiation reveals that e_i, $i = 2, 3, \cdots, k$, is monotonously semistable from below if $e_i > -\theta_{2,0}/\sum_{j=1}^{p}\theta_{2,j}$, and monotonously semistable from above if $e_i < -\theta_{2,0}/\sum_{j=1}^{p}\theta_{2,j}$.

[5]This thesis only considers test for a unit root without a drift because in we want to apply the series for pair trading which is not possible if the series has a drift or trend.

$$
\begin{aligned}
G^*(\theta, \mathbf{e}, y_{t-d})) &= 1 - \exp\left\{ -\theta^2 \ y_{t-d}^2 \left[\prod_{i=2}^{k} (y_{t-d} - e_i) \right]^2 \right\} \\
&\approx \theta^2 \ y_{t-d}^2 \left[\prod_{i=2}^{k} (y_{t-d} - e_i) \right]^2 + R \\
&= \theta^2 \ y_{t-d}^2 \left(\delta_0 + \sum_{s=1}^{2(k-1)} \delta_s \ y_{t-d}^s \right) + R \\
&= \theta^2 \sum_{s=0}^{2(k-1)} \delta_s \ y_{t-d}^{s+2} + R, \qquad (6.42)
\end{aligned}
$$

where R is the remainder, $\delta_0 = (\prod_{i=2}^{k} e_i)^2$ and $\delta_{2(k-1)} = 1$.

Substituting (6.42) into (6.40),

$$
\Delta y_t = \sum_{j=1}^{p-1} \theta_{1,j}^* \ \Delta y_{t-j} + \sum_{s=0}^{2(k-1)} \gamma_{1,s} \ y_{t-1} \ y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \sum_{j=1}^{p-1} \gamma_{2,sj} \ y_{t-d}^{s+2} \Delta y_{t-j} + \epsilon_t^*, \quad (6.43)
$$

where $\gamma_{1,s} = \theta^2 \delta_s \sum_{j=1}^{p} \theta_{2,j}$; $\gamma_{2,sj} = \theta^2 \delta_s \theta_{2,j}^*$; $s = 0, 1, \cdots, 2(k-1)$; $j = 1, 2, \cdots, (p-1)$ and $\epsilon_t^* = \epsilon_t + R \left[\sum_{j=1}^{p} \theta_{2,j} \ y_{t-j} \right]$. If $\theta = 0$, y_t is linear in term of y_{t-j}, $j = 1, 2, \cdots, p$ and $\epsilon_t^* = \epsilon_t$ since the remainder $R \equiv 0$.

Testing the null hypothesis of a unit root against alternative of a globally stationary k-ESTAR(p) model is equivalent to test,

$$
H_0 : \gamma_{1,s} = \gamma_{2,sj} = 0, \qquad \text{for all } s \text{ and } j \text{ in (6.43)} .
$$

Under the null hypothesis H_0, (6.43) becomes

$$
\begin{aligned}
\Delta y_t &= \sum_{j=1}^{p-1} \theta_{1,j}^* \ \Delta y_{t-j} + \epsilon_t, \qquad (6.44) \\
\left(1 - \sum_{j=1}^{p-1} \theta_{1,j}^* \ L^j \right) \Delta y_t &= \epsilon_t, \\
\Delta y_t &= \left(1 - \sum_{j=1}^{p-1} \theta_{1,j}^* \ L^j \right)^{-1} \epsilon_t, \\
&= \sum_{j=0}^{\infty} c_j \ \epsilon_{t-j}, \\
&= C(L) \epsilon_t = \eta_t, \qquad (6.45)
\end{aligned}
$$

where L is the lag operator, i.e. $L y_t = y_{t-1}$. We assume that the sequence $\{\eta_t\}$ satisfies the following assumption:

Assumption 6.1 *Assumptions for* $\{\eta_t\}$ *:*

- $\eta_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j} = C(L)\epsilon_t$, *where* $\{\epsilon_t\}$ *is a stationary and ergodic martingale differences sequence (MDS) with natural filtration* $\mathcal{F}_t = \sigma(\{\epsilon_i\}_{-\infty}^{t})$, *variance* σ_ϵ^2, *and* $E|\epsilon_t|^{6+r} < \infty$ *for some* $r > 0$.
- $C(L) = \sum_{j=0}^{\infty} c_j L^j$ *is a one-sided moving average polynomial in the lag operator such that* $C(1) \neq 0$ *(no unit root),* $\sum_{j=0}^{\infty} c_j = C(1) < \infty$ *and* $\sum_{j=0}^{\infty} j^p |c_j|^p < \infty$ *(one-summability) for* $p \geq 1$.

Following Phillips and Solo (1992), the Beveridge-Nelson (BN for short) decomposition (see Beveridge and Nelson, 1981) will be applied. We start with the BN lemma as follows:

Lemma 6.1 *(Lemma 2.1 in Phillips and Solo, 1992). Let* $C(L) = \sum_{j=0}^{\infty} c_j L^j$. *Then*

$$C(L) = C(1) - (1 - L)\widetilde{C}(L),$$

where $\widetilde{C}(L) = \sum_{j=0}^{\infty} \widetilde{c}_j L^j$ *and* $\widetilde{c}_j = \sum_{k=j+1}^{\infty} c_k$. *If* $p \geq 1$, *then*

$$\sum_{j=1}^{\infty} j^p |c_j|^p < \infty \Rightarrow \sum_{j=0}^{\infty} |\widetilde{c}_j|^p < \infty \text{ and } |C(1)| < \infty.$$

If $p < 1$, *then*

$$\sum_{j=1}^{\infty} j |c_j|^p < \infty \Rightarrow \sum_{j=0}^{\infty} |\widetilde{c}_j|^p < \infty.$$

Before we derive the F-test statistic for the unit root test for a k-ESTAR(p) model, we present the theorem below used in the F-test statistic derivation.

Theorem 6.1 *Assume that* $\{\eta_t\}_{t=1}^{\infty}$ *and* $\{\epsilon_t\}_{t=1}^{\infty}$ *satisfy Assumption 6.1. Let* $y_t = \sum_{i=0}^{t} \eta_i$, $t = 1, 2, \ldots, T$, *with* $y_0 = 0$. *Denote* $\lambda = \sigma_\epsilon C(1)$ *and* $\gamma_j = E(\eta_t \eta_{t-j}) = \sigma_\epsilon^2 \sum_{s=0}^{\infty} c_s c_{s+j}$, $j = 0, 1, \ldots,$ *for all* t. *Then, under* H_0, *the following sums converge jointly.*

(a) $T^{-1} \sum_{t=p+1}^{T} \left(\frac{y_t}{\sqrt{T}}\right)^q \Rightarrow \lambda^q \int W^q$,

(b) $T^{-1} \sum_{t=p+1}^{T} \left(\frac{y_{t-1}}{\sqrt{T}}\right)^2 \left(\frac{y_{t-d}}{\sqrt{T}}\right)^q \Rightarrow \lambda^{q+2} \int W^{q+2}$,

(c) $T^{-1} \sum_{t=p+1}^{T} \eta_{t-i}\eta_{t-j} \Rightarrow \gamma_{|j-i|}, \quad i, j = 1, \ldots, (p-1)$,

(d) $T^{-1} \sum_{t=p+1}^{T} \left(\frac{y_{t-d}}{\sqrt{T}}\right)^q \eta_{t-i}\eta_{t-j} \Rightarrow \gamma_{|j-i|}\lambda^q \int W^q, \quad i, j = 1, \ldots, (p-1)$,

(e) $T^{-1/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-1}}{\sqrt{T}}\right) \left(\frac{y_{t-d}}{\sqrt{T}}\right)^q \frac{\eta_{t-i}}{\sqrt{T}} \Rightarrow 0, \quad i = 1, \ldots, (p-1)$,

(f) $\sum_{t=p+1}^{T} \frac{y_{t-1}}{\sqrt{T}} \left(\frac{y_{t-d}}{\sqrt{T}}\right)^q \frac{\epsilon_t}{\sqrt{T}} \Rightarrow \sigma_\epsilon \lambda^{q+1} \int W^{q+1} dW$,

(g) $T^{-1/2}\sum_{t=p+1}^{T}\eta_{t-i}\epsilon_t \Rightarrow \sqrt{\gamma_0}\sigma_\epsilon W_i(1), \quad i=1,\ldots,(p-1),$

(h) $\sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^q \frac{\eta_{t-i}\epsilon_t}{\sqrt{T}} \Rightarrow \sqrt{\gamma_0}\sigma_\epsilon\lambda^q\int W^q dW_i, \quad i=1,\ldots,(p-1),$

as $T\to\infty$. $\int W^x = \int_0^1 W(s)^x ds$ and $\int W^x dW = \int_0^1 W(s)^x dW(s)$ where $W(s)$ is the standard Brownian motion defined on $s\in[0,1]$ and x is an integer number. W is a standard Brownian motion corresponding to MDS $\{\epsilon_t\}$ and W_i a standard Brownian motion corresponding to MDS $\{\eta_{t-i}\epsilon_t\}$, $i=1,2,\ldots,(p-1)$. Note that W and W_i are independent as $Cov(\epsilon_t,\eta_{t-i}\epsilon_t)=E(\eta_{t-i}\epsilon_t^2)=0$.

Proof:

First, we prove that $y_t/\sqrt{T}\Rightarrow\lambda W(s)$ for $t\le sT<t+1$ as $T\to\infty$. Since $\{\epsilon_t\}$ follows Assumption 6.1,

$$\frac{\sum_{i=1}^{t}\epsilon_i}{\sqrt{T}} \Rightarrow N(0,s\sigma_\epsilon^2)=\sigma_\epsilon W(s), \qquad t=1,2,\ldots,T, \tag{6.46}$$

where $W(s)$ is a standard Brownian motion with variance s, $s\in[0,1]$ (see Hong and Phillips, 2010).

Let $y_t=y_{t-1}+\eta_t$ where $\eta_t=\sum_{j=0}^{\infty}c_j\epsilon_{t-j}$ where $\{\eta_t\}$ and $\{\epsilon_t\}$ follow Assumption 6.1. Using the BN decomposition,

$$\eta_t = C(1)\epsilon_t+\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t, \tag{6.47}$$

where $\widetilde{\epsilon}_t=\widetilde{C}(L)\epsilon_t=\sum_{j=0}^{\infty}\widetilde{c}_j\epsilon_{t-j}$ and $\widetilde{c}_j=\sum_{k=j+1}^{\infty}c_k$.

From (6.47),

$$\frac{y_t}{\sqrt{T}}=\frac{\sum_{i=1}^{t}\eta_i}{\sqrt{T}}=C(1)\frac{\sum_{i=1}^{t}\epsilon_i}{\sqrt{T}}+\frac{\widetilde{\epsilon}_0}{\sqrt{T}}-\frac{\widetilde{\epsilon}_t}{\sqrt{T}}. \tag{6.48}$$

Using Markov's inequality [6],

$$P\left(\frac{\widetilde{\epsilon}_t^2}{T}>a\right)<\frac{E(\widetilde{\epsilon}_t^2)}{Ta}\to 0,\ \text{as } T\to\infty, \tag{6.49}$$

for a positive real number a, because $E(\widetilde{\epsilon}_t^2)<\infty$. [7] Similar result happens for $\widetilde{\epsilon}_0$. Thus,

[6] Markov's inequality: $P(|X|\ge a)\le E(|X|)/a$ for given a random variable X and a positive real number a. However, we use the square of random variable instead of the absolute value.

[7] $E(\widetilde{\epsilon}_t^2)=\sigma_\epsilon^2\sum_{j=0}^{\infty}\widetilde{c}_j^2=\sigma_\epsilon^2\sum_{j=0}^{\infty}|\widetilde{c}_j|^2<\infty$ by Assumption 6.1 and Lemma 6.1.

$$
\begin{aligned}
\frac{y_t}{\sqrt{T}} &= \frac{\sum_{i=1}^{t}\eta_i}{\sqrt{T}} \\
&= \frac{C(1)\sum_{i=1}^{t}\epsilon_i}{\sqrt{T}} + \frac{\widetilde{\epsilon}_0}{\sqrt{T}} - \frac{\widetilde{\epsilon}_t}{\sqrt{T}} \\
&\Rightarrow C(1)\sigma_\epsilon W(s) \qquad \text{by (6.46) and (6.49)} \\
&= \lambda W(s), \quad \text{as } T \to \infty. \qquad (6.50)
\end{aligned}
$$

Given the result of (6.50), we start to prove Theorem 6.1.

(a) and (b): The proofs can be found in Venetis *et al.* (2009).

(c) Under H_0, $T^{-1}\sum_{t=p+1}^{T}\Delta y_{t-i}\Delta y_{t-j} = T^{-1}\sum_{t=p+1}^{T}\eta_{t-i}\eta_{t-j}$. Now, for given i, and j where $i, j = 1, \ldots, (p-1)$,

$$T^{-1}\sum_{t=p+1}^{T}\eta_{t-i}\eta_{t-j} \to E(\eta_{t-i}\eta_{t-j}) = \gamma_{|j-i|} \quad \text{as } T \to \infty. \qquad (6.51)$$

If $i = j$,

$$T^{-1}\sum_{t=p+1}^{T}\eta_{t-i}^2 \to E(\eta_{t-i}^2) = \gamma_0 \quad \text{as } T \to \infty. \qquad (6.52)$$

(d) Under H_0, $T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^q\Delta y_{t-i}\Delta y_{t-j} = T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^q\eta_{t-i}\eta_{t-j}$. Now, for given i and j where $i, j = 1, \ldots, (p-1)$, and $i \geq j \geq d$,

$$
\begin{aligned}
& T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^q\eta_{t-i}\eta_{t-j} \\
&= T^{-(q/2+1)}\sum_{t=p+1}^{T}\left(y_{t-i-1} + \sum_{k=0}^{i-d}\eta_{t-d-k}\right)^q\eta_{t-i}\eta_{t-j} \\
&= T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-i-1}^q\eta_{t-i}\eta_{t-j} \\
&\quad + T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=0}^{i-d}\eta_{t-d-k}\right)^s\eta_{t-i}\eta_{t-j}\right] \qquad (6.53) \\
&= \sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^q\frac{(\eta_{t-i}\eta_{t-j} - E(\eta_{t-i}\eta_{t-j}))}{T} + \frac{E(\eta_{t-i}\eta_{t-j})}{T}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^q \\
&\quad + T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=0}^{i-d}\eta_{t-d-k}\right)^s\eta_{t-i}\eta_{t-j}\right]. \qquad (6.54)
\end{aligned}
$$

Let $w_t = \left(\sum_{k=0}^{i-d} \eta_{t-d-k}\right)^s \eta_{t-i}\eta_{t-j}$. For fixed q and s,

$$
\begin{aligned}
& T^{-(q/2+1)} \sum_{t=p+1}^{T} y_{t-i-1}^{q-s} w_t && (6.55)\\
= \; & T^{-s/2} \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s} \frac{w_t}{T} \\
= \; & T^{-s/2} \sum_{t=p+1}^{T} \left[\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s} \frac{w_t - E(w_t)}{T} + \frac{E(w_t)}{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s}\right]. && (6.56)
\end{aligned}
$$

Since $s \geq 1$,

$$\frac{E(w_t)}{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s} \Rightarrow E(w_t)\lambda^{q-s}\int W^{q-s}$$

and

$$E(w_t) = E\left[\left(\sum_{k=0}^{i-d} \eta_{t-d-k}\right)^s \eta_{t-i}\eta_{t-j}\right] < \infty. \text{ [8]}$$

Thus, for any constant $a > 0$, we have

$$
\begin{aligned}
& P\left(\sum_{t=p+1}^{T} \frac{(w_t - E(w_t))^2}{T^2} > a\right) \\
\leq \; & \frac{E\left[\sum_{t=p+1}^{T}(w_t - E(w_t))^2\right]}{T^2 a}, \text{ Markov's inequality} \\
= \; & \frac{\sum_{t=p+1}^{T} E(w_t - E(w_t))^2}{T^2 a} \\
= \; & \frac{(T-p)Var(w_t)}{T^2 a} \\
\leq \; & \frac{Var(w_t)}{Ta} \\
\rightarrow \; & 0 \quad \text{as } T \rightarrow \infty \text{ and } Var(w_t) < \infty;
\end{aligned}
$$

and

[8] As w_t is a function of $\eta_{t-1}, \ldots, \eta_{t-(p-1)}$ which are stationary processes, then w_t is also a stationary process. As a stationary process, $E(w_t) < \infty$ and $Var(w_t) < \infty$.

$$
\begin{aligned}
&\left|\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q-s}\frac{w_t-E(w_t)}{T}\right| \\
&\leq \sqrt{\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{2(q-s)}}\sqrt{\sum_{t=p+1}^{T}\frac{(w_t-E(w_t))^2}{T^2}}, \text{ Cauchy-Schwarz inequality} \\
&\Rightarrow \sqrt{\int W^{2(q-s)}}\sqrt{o_p(1)} = o_p(1), \quad \text{as } T\to\infty. \qquad (6.57)
\end{aligned}
$$

Therefore,

$$
T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=0}^{i-d}\eta_{t-d-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right] = o_p(1). \qquad (6.58)
$$

As $E(\eta_{t-i}\eta_{t-j}) = \gamma_{|j-i|} < \infty$, the first term of (6.54) converges to $o_p(1)$. Thus,

$$
T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^{q}\eta_{t-i}\eta_{t-j} \Rightarrow \gamma_{|j-i|}\lambda\int W^q, \qquad T\to\infty.
$$

If $d > i \geq j$, recalling $y_{t-d} = y_{t-i-1} - \sum_{k=1}^{d-i-1}\eta_{t-i-k}$, the same result is obtained as follow:

$$
\begin{aligned}
&T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-d}^{q}\eta_{t-i}\eta_{t-j} \\
&= T^{-(q/2+1)}\sum_{t=p+1}^{T}\left(y_{t-i-1}-\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{q}\eta_{t-i}\eta_{t-j} \\
&= T^{-(q/2+1)}\sum_{t=p+1}^{T}y_{t-i-1}^{q}\eta_{t-i}\eta_{t-j} \\
&\quad +T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right] \\
&= \sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\frac{\eta_{t-i}\eta_{t-j}-E(\eta_{t-i}\eta_{t-j})}{T} + \frac{E(\eta_{t-i}\eta_{t-j})}{T}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q} \\
&\quad +T^{-(q/2+1)}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s\binom{q}{s}y_{t-i-1}^{q-s}\left(\sum_{k=1}^{d-i-1}\eta_{t-i-k}\right)^{s}\eta_{t-i}\eta_{t-j}\right] \\
&\Rightarrow \gamma_{|j-i|}\lambda^q\int W^q \qquad (6.59)
\end{aligned}
$$

as the first and the last term of (6.59) are $o_p(1)$.

(e) Under H_0,

$$T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \Delta y_{t-i} = T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \eta_{t-i}.$$

Now, for given d and i, $i = 1, \ldots, (p-1)$, and if $d \leq i$,

$$\begin{aligned}
& T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \eta_{t-i} \\
= \; & T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left(y_{t-i-1} + \sum_{j=1}^{i} \eta_{t-j} \right) \left(y_{t-i-1} + \sum_{j=0}^{i-d} \eta_{t-d-j} \right)^{q} \eta_{t-i} \\
= \; & T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} y_{t-i-1}^{q+1} \eta_{t-i} + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[y_{t-i-1}^{q} \eta_{t-i} \sum_{j=1}^{i} \eta_{t-j} \right] \\
& + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s+1} \left(\sum_{j=0}^{i-1} \eta_{t-j-1} \right)^{s} \eta_{t-i} \right] \\
& + T^{-((q+1)/2+1)} \sum_{t=p+1}^{T} \left[\sum_{j=1}^{i} \eta_{t-j} \sum_{s=1}^{q} \binom{q}{s} y_{t-i-1}^{q-s} \left(\sum_{j=0}^{i-1} \eta_{t-j-1} \right)^{s} \eta_{t-i} \right]
\end{aligned} \tag{6.60}$$

Now, we need to show that all terms in (6.60) are $o_p(1)$.

(i) As $E(\eta_t \eta_s) \neq 0$ for $t \neq s$, the BN decomposition in (6.47) is used as follows (see Hong and Phillips, 2010),

$$\begin{aligned}
& \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \frac{\eta_{t-i}}{\sqrt{T}} \\
= \; & \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \frac{\epsilon_{t-i} C(1)}{\sqrt{T}} - \sum_{t=p+1}^{T} \left(\frac{y_{t-i-1}}{\sqrt{T}} \right)^{q+1} \left(\frac{\widetilde{\epsilon}_{t-i} - \widetilde{\epsilon}_{t-i-1}}{\sqrt{T}} \right).
\end{aligned}$$

By noting that

$$
\begin{aligned}
&\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\left(\frac{\widetilde{\epsilon}_{t-i}-\widetilde{\epsilon}_{t-i-1}}{\sqrt{T}}\right) \\
= \;& \left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}}-\left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}} \\
&+\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\left(\frac{\widetilde{\epsilon}_{t-i}-\widetilde{\epsilon}_{t-i-1}}{\sqrt{T}}\right) \\
= \;& \left(\frac{y_{T-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{T-i}}{\sqrt{T}} \\
&-\sum_{t=p+1}^{T}\left[\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q+1}-\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\right]\frac{\widetilde{\epsilon}_{t-i}}{\sqrt{T}}-\left(\frac{y_{p-i}}{\sqrt{T}}\right)^{q+1}\frac{\widetilde{\epsilon}_{p-i}}{\sqrt{T}} \\
= \;& o_p(1)-\sum_{t=p+1}^{T}\left[\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q+1}-\left(\frac{y_{t-i}-\eta_{t-i}}{\sqrt{T}}\right)^{q+1}\right]\frac{\widetilde{\epsilon}_{t-i}}{\sqrt{T}}-o_p(1) \quad \text{by (6.49)} \\
\approx \;& -(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{\eta_{t-i}\widetilde{\epsilon}_{t-i}}{T} \\
= \;& -(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\left(\frac{\eta_{t-i}\widetilde{\epsilon}_{t-i}-E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T}\right) \\
&-(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T} \\
= \;& o_p(1)-(q+1)\sum_{t=p+1}^{T}\left(\frac{y_{t-i}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\widetilde{\epsilon}_{t-i})}{T} \quad \text{similar way with (6.57)} \\
\Rightarrow \;& -(q+1)\Lambda_{\eta\eta}\lambda^{q}\int W^{q} \qquad (6.61)
\end{aligned}
$$

where $E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\Lambda_{\eta\eta}$. [9]

[9] For η_t and ϵ_t satisfying Assumption 6.1, we have

$$E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=E\left(\sum_{j=0}^{\infty}c_j\epsilon_{t-i-j}\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}c_k\epsilon_{t-i-j}\right)=\sum_{j=0}^{\infty}E(\epsilon_{t-i-j}^2)c_j\sum_{k=j+1}^{\infty}c_k=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k.$$

$$\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\sum_{h=1}^{\infty}E\left(\sum_{j=0}^{\infty}c_j\epsilon_{-j}\sum_{j=0}^{\infty}c_j\epsilon_{h-j}\right)=\sum_{j=0}^{\infty}E(\epsilon_{-j}^2)c_j\sum_{h=1}^{\infty}c_{h+j}=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k.$$

Therefore, $E(\eta_{t-i}\widetilde{\epsilon}_{t-i})=\sum_{h=1}^{\infty}E(\eta_0\eta_h)=\sigma_\epsilon^2\sum_{j=0}^{\infty}c_j\sum_{k=j+1}^{\infty}c_k\le\sigma_\epsilon^2\sum_{j=0}^{\infty}|c_j|\sum_{k=j+1}^{\infty}|c_k|\le\sigma_\epsilon^2(\sum_{j=0}^{\infty}|c_j|)^2<\infty$.

Therefore,

$$\begin{aligned}
&\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\frac{\eta_{t-i}}{\sqrt{T}} \\
\Rightarrow\ & \int(\lambda W)^{(q+1)}\lambda dW-\left(-(q+1)\Lambda_{\eta\eta}\lambda^{q}\int W^{q}\right) \qquad \text{by (6.50) and (6.61)} \\
=\ & \lambda^{(q+2)}\int W^{(q+1)}dW+(q+1)\Lambda_{\eta\eta}\lambda^{q}\int W^{q},
\end{aligned}$$

and

$$T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q+1}\frac{\eta_{t-i}}{\sqrt{T}}=o_p(1).$$

(ii) For given $i=1,\ldots,(p-1)$,

$$\begin{aligned}
&T^{-((q+1)/2+1)}\sum_{t=p+1}^{T}\left[y_{t-i-1}^{q}\eta_{t-i}\sum_{j=1}^{i}\eta_{t-j}\right] \\
=\ & \sum_{j=1}^{i}T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\frac{\eta_{t-i}\eta_{t-j}}{T} \\
=\ & \sum_{j=1}^{i}T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\left[\frac{\eta_{t-i}\eta_{t-j}-E(\eta_{t-i}\eta_{t-j})}{T}+\frac{E(\eta_{t-i}\eta_{t-j})}{T}\right] \qquad (6.62) \\
\Rightarrow\ & o_p(1) \text{ as } T\to\infty.
\end{aligned}$$

Using similar method used in (6.57), the first term of (6.62) can be shown to be $o_p(1)$ and

$$\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\eta_{t-j})}{T}\Rightarrow\gamma_{|j-i|}\lambda^{q}\int W^{q}.$$

Therefore, the last term of (6.62) will be

$$T^{-1/2}\sum_{t=p+1}^{T}\left(\frac{y_{t-i-1}}{\sqrt{T}}\right)^{q}\frac{E(\eta_{t-i}\eta_{t-j})}{T}\Rightarrow o_p(1).$$

Using similar procedure, the last two terms of (6.60) are also $o_p(1)$. Therefore, (e) is hold.

(f) For a given $d \geq 1$,

$$\begin{aligned}
& T^{-(q+2)/2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{q} \epsilon_t \\
= \; & T^{-(q+2)/2} \sum_{t=p+1}^{T} y_{t-1} \left(y_{t-1} - \sum_{j=0}^{d-1} \eta_{t-d-j} \right)^{q} \epsilon_t \\
= \; & T^{-(q+2)/2} \sum_{t=d+1}^{T} y_{t-1}^{(q+1)} \epsilon_t \\
& + T^{-(q+2)/2} \sum_{t=p+1}^{T} \left[\sum_{s=1}^{q} (-1)^s \begin{pmatrix} q \\ s \end{pmatrix} y_{t-1}^{q-s+1} \left(\sum_{j=0}^{d-1} \eta_{t-d-j} \right)^{s} \epsilon_t \right] \\
= \; & \sum_{t=p+1}^{T} \left(\frac{y_{t-1}}{\sqrt{T}} \right)^{(q+1)} \frac{\epsilon_t}{\sqrt{T}} + o_p(1) \qquad \text{similar way with (6.53)} \\
\Rightarrow \; & \int (\lambda W)^{(q+1)} \sigma_\epsilon dW \qquad \text{by (6.46) and (6.50)} \\
= \; & \sigma_\epsilon \lambda^{(q+1)} \int W^{(q+1)} dW.
\end{aligned}$$

(g) For any fixed i where $i = 1, \ldots, (p-1)$, under H_0, $\sum_{t=p+1}^{T} \Delta y_{t-i} \epsilon_t = \sum_{t=p+1}^{T} \eta_{t-i} \epsilon_t$.

$$E\left(\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \eta_{t-i} \epsilon_t \right) = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} E(\eta_{t-i}) E(\epsilon_t) = 0$$

as η_{t-i} and ϵ_t are independent and

$$\begin{aligned}
& Var\left(\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \eta_{t-i} \epsilon_t \right) \\
= \; & \frac{1}{T} \sum_{t=p+1}^{T} E(\eta_{t-i}^2) E(\epsilon_t^2) \qquad \text{as } \eta_{t-i} \text{ and } \epsilon_t \text{ are independent} \\
= \; & \frac{(T-p)}{T} \gamma_0 \sigma_\epsilon^2 \qquad \text{as } \eta_{t-i}\epsilon_t \text{ are identic for each } t \\
\to \; & \gamma_0 \sigma_\epsilon^2 \qquad \text{as } T \to \infty.
\end{aligned}$$

Since $E(\eta_{t-i}\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = 0$, $\{\eta_{t-i}\epsilon_t\}$ is MDS. Using Central Limit Theorem,

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \eta_{t-i} \epsilon_t \Rightarrow N(0, \gamma_0 \sigma_\epsilon^2) = \sqrt{\gamma_0} \sigma_\epsilon W_i(1). \tag{6.63}$$

Note that,

$$\begin{aligned}
& Cov(\eta_{t-i}\epsilon_t, \eta_{t-j}\epsilon_t) \text{ for } i,\, j = 1, \ldots, (p-1),\, i \neq j, \text{ for all } t \\
= \;& E(\eta_{t-i}\eta_{t-j}\epsilon_t^2) \\
= \;& E(\eta_{t-i}\eta_{t-j})E(\epsilon_t^2) \qquad \text{as } (\eta_{t-i}\eta_{t-j}) \text{ and } (\epsilon_t^2) \text{ are independent} \\
= \;& \gamma_{|j-i|}\sigma_\epsilon^2.
\end{aligned}$$

Therefore, there is correlation between $W_i(1)$ and $W_j(1)$. Furthermore,

$$\begin{aligned}
& Cov(\eta_{t-i}\epsilon_t, \eta_{s-j}\epsilon_s) \qquad \text{for } i,\, j = 1, \ldots, (p-1),\, i \neq j, \text{ for all } t \neq s \\
= \;& E(\eta_{t-i}\eta_{s-j}\epsilon_t\epsilon_s) \\
= \;& E(\eta_{t-i}\eta_{s-j})E(\epsilon_t)E(\epsilon_s) \qquad \text{as } (\eta_{t-i}\eta_{s-j}),\, (\epsilon_t) \text{ and } (\epsilon_s) \text{ are independent} \\
= \;& 0.
\end{aligned}$$

(h) Under H_0, $T^{-(q+1)/2}\sum_{t=p+1}^{T} y_{t-d}^q \Delta y_{t-i}\epsilon_t = \sum_{t=i+1}^{T} y_{t-d}^q \eta_{t-i}\epsilon_t$. Now, for given d, and i where $i = 1, \ldots, (p-1)$, $d \geq 1$,

$$\begin{aligned}
& T^{-(q+1)/2}\sum_{t=p+1}^{T} y_{t-d}^q \eta_{t-i}\epsilon_t \\
= \;& T^{-(q+1)/2}\sum_{t=p+1}^{T}\left(y_{t-1} - \sum_{j=0}^{d-1}\eta_{t-d-j}\right)^q \eta_{t-i}\epsilon_t \\
= \;& T^{-(q+1)/2}\sum_{t=p+1}^{T} y_{t-1}^q \eta_{t-i}\epsilon_t \\
& + T^{-(q+1)/2}\sum_{t=p+1}^{T}\left[\sum_{s=1}^{q}(-1)^s \begin{pmatrix} q \\ s \end{pmatrix} y_{t-1}^{q-s}\left(\sum_{j=0}^{d-1}\eta_{t-d-j}\right)^s \eta_{t-i}\epsilon_t\right] \\
= \;& \sum_{t=p+1}^{T}\left(\frac{y_{t-1}}{\sqrt{T}}\right)^q\left(\frac{\eta_{t-i}\epsilon_t}{\sqrt{T}}\right) + o_p(1) \qquad \text{similar way with (6.53)} \qquad (6.64) \\
\Rightarrow \;& \int(\lambda W)^q(\sqrt{\gamma_0}\sigma_\epsilon)dW_i \qquad \text{by (6.46) and (6.63)} \qquad (6.65) \\
= \;& \sqrt{\gamma_0}\sigma_\epsilon\lambda^q\int W dW_i
\end{aligned}$$

Note that as $\{\eta_{t-i}\epsilon_t\}$ is MDS and $\sum_{t=1}^{T}(\eta_{t-i}\epsilon_t/\sqrt{T})^2 < \infty$, then based on Theorem 2.1 in Hansen 1992, the first term of (6.64) will converge to (6.65).

Theorem 6.2 *Let us write (6.43) as a partitioned regression model,*

$$Y = X_1 \ \mathbf{b}_1 + X_2 \ \mathbf{b}_2 + \epsilon_t^* \tag{6.66}$$

where

$$\begin{aligned}
Y &= [\Delta y_{p+1}, \Delta y_{p+2}, \cdots, \Delta y_T]' \\
X_1 &= \big[(\Delta y_p, \Delta y_{p+1}, \cdots, \Delta y_{T-1})', (\Delta y_{p-1}, \Delta y_p, \cdots, \Delta y_{T-2})', \cdots, \\
&\quad (\Delta y_2, \Delta y_3, \cdots, \Delta y_{T-(p-1)})'\big] \\
X_2 &= \big[(y_p y_{p+1-d}^2, \cdots, y_{T-1} y_{T-d}^2)', (y_p y_{p+1-d}^3, \cdots, y_{T-1} y_{T-d}^3)', \cdots, \\
&\quad (y_p y_{p+1-d}^{2k}, \cdots, y_{T-1} y_{T-d}^{2k})', (y_{p+1-d}^2 \Delta y_p, \cdots, y_{T-d}^2 \Delta y_{T-1})', \cdots, \\
&\quad (y_{p+1-d}^2 \Delta y_2, \cdots, y_{T-d}^2 \Delta y_{T-(p-1)})', (y_{p+1-d}^3 \Delta y_p, \cdots, y_{T-d}^3 \Delta y_{T-1})', \cdots, \\
&\quad (y_{p+1-d}^3 \Delta y_2, \cdots, y_{T-d}^3 \Delta y_{T-(p-1)})', \cdots, (y_{p+1-d}^{2k} \Delta y_p, \cdots, y_{T-d}^{2k} \Delta y_{T-1})', \cdots, \\
&\quad (y_{p+1-d}^{2k} \Delta y_2, \cdots, (y_{T-d}^{2k} \Delta y_{T-(p-1)})'\big] \\
\mathbf{b}_1 &= \left(\theta_{11}^*, \theta_{12}^*, \ldots, \theta_{1(p-1)}^*\right)' \\
\mathbf{b}_2 &= \left(\gamma_1, \gamma_{21}, \cdots, \gamma_{2(2k-2)}, \gamma_{31}, \cdots, \gamma_{3(p-1)}, \gamma_{411}, \cdots, \gamma_{4(2k-2)(p-1)}\right)'
\end{aligned}$$

Under the null hypothesis of $H_0 : \theta = 0$, $\epsilon_t^* = \epsilon_t$, so that an F-type test can be constructed. The asymptotic behavior of F-type statistic to test the null hypothesis of a unit root without a drift against the alternative of a globally stationary k-ESTAR(p) model is

$$F_{nl} = \frac{1}{\hat{\sigma}_{\epsilon^*}^2} \left(\hat{\mathbf{b}}_2 - \mathbf{b}_2\right)' (X_2' M_1 X_2) \left(\hat{\mathbf{b}}_2 - \mathbf{b}_2\right) \tag{6.67}$$

where $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ is the orthogonal to X_1 projection matrix and $\hat{\sigma}_{\epsilon^*}^2$ is the maximum likelihood estimator of the error variance.

$$F_{nl} \Rightarrow F_1'(W) F_2^{-1}(W) F_1(W) \tag{6.68}$$

where $F_1(W)$ and $F_2(W)$ are described below.

Let W denote a standard Brownian motion,

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \vdots \\ \int W^{(2k+1)} dW \\ \left(\int W^2 dW_1 - W_1(1) \int W^2\right) \\ \vdots \\ \left(\int W^2 dW_{(p-1)} - W_{(p-1)}(1) \int W^2\right) \\ \left(\int W^3 dW_1 - W_1(1) \int W^3\right) \\ \vdots \\ \left(\int W^3 dW_{(p-1)} - W_{(p-1)}(1) \int W^3\right) \\ \vdots \\ \left(\int W^{2k} dW_1 - W_1(1) \int W^{2k}\right) \\ \vdots \\ \left(\int W^{2k} dW_{(p-1)} - W_{(p-1)}(1) \int W^{2k}\right) \end{bmatrix},$$

$$F_2(W) = \begin{bmatrix} F_{21}(W) & \mathbf{0} \\ \mathbf{0} & F_{22}(W) \end{bmatrix},$$

$$F_{21}(W) = \begin{bmatrix} \int W^6 & \cdots & \int W^{(2k+4)} \\ \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \cdots & \int W^{(4k+2)} \end{bmatrix},$$

$$F_{22}(W) = \begin{bmatrix} \left(\int W^4 - (\int W^2)^2\right) \mathbf{\Pi} & \cdots & \left(\int W^{2k+2} - \int W^2 \int W^{2k}\right) \mathbf{\Pi} \\ \vdots & \ddots & \vdots \\ \left(\int W^{2k+2} - \int W^2 \int W^{2k}\right) \mathbf{\Pi} & \cdots & \left(\int W^{4k} - (\int W^{2k})^2\right) \mathbf{\Pi} \end{bmatrix}$$

and

$$\mathbf{\Pi} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}_{(p-1)\times(p-1)}, \tag{6.69}$$

where $\int W^x = \int_0^1 W(s)^x ds$ and $\int W^x dW = \int_0^1 W(s)^x dW(s)$ where $W(s)$ is the standard Brownian motion defined on $s \in [0,1]$ and x in an integer number. W is a standard Brownian motion corresponding to m.d.s. $\{\epsilon_t\}$ and W_i a standard Brownian motion corresponding to MDS $\{\eta_{t-i}\epsilon_t\}$, $i = 1, 2, \ldots, (p-1)$. Note that W and W_i are independent as $Cov(\epsilon_t, \eta_{t-i}\epsilon_t) = E(\eta_{t-i}\epsilon_t^2) = 0$ and also note that $\{\eta_{t-i}\epsilon_t\}$ is MDS ρ_i, $i = 1, \ldots, (p-2)$, are constants corresponding to correlation between Δy_t and Δy_{t-i}.

Proof:

Consider the asymptotic distribution of the F-test statistics in (6.67) by testing the null hypothesis $H_0 : R\mathbf{b} = 0$ where $R = [\mathbf{0}\ \ \mathbf{I}]$ and $\mathbf{b} = (\mathbf{b}_1\ \ \mathbf{b}_2)'$. Under the null hypothesis of $H_0 : \theta = 0$, $\epsilon_t^* = \epsilon_t$, so that the sampling error of $\hat{\mathbf{b}}_2$ is given by

$$\hat{\mathbf{b}}_2 - \mathbf{b}_2 = (X_2'M_1X_2)^{-1}X_2'M_1\epsilon$$

where

$$X_2'M_1X_2 = X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$$

and

$$X_2'M_1\epsilon = X_2'\epsilon - X_2'X_1(X_1'X_1)^{-1}X_1'\epsilon$$

Let

$$D_T = diag\left(T^{-4/2}, T^{-5/2}, \cdots, T^{-(2k+2)/2}, \underbrace{T^{-3/2}, \cdots, T^{-3/2}}_{p-1}, \underbrace{T^{-4/2}, \cdots, T^{-4/2}}_{p-1}, \cdots, \underbrace{T^{-(2k+1)/2}, \cdots, T^{-(2k+1)/2}}_{p-1}\right).$$

Thus, the F_{nl} statistics in (6.67) becomes

$$F_{nl} = \frac{1}{\hat{\sigma}^2_{\epsilon^*}}(D_TX_2'M_1\epsilon)'(D_TX_2'M_1X_2D_T)^{-1}(D_TX_2'M_1\epsilon). \tag{6.70}$$

In the following, we derive the asymptotic distribution of F_{nl}. Firstly, we consider the asymptotic distributions of $D_TX_2'M_1X_2D_T$ and $D_TX_2'M_1\epsilon$.

(i) Rewrite $D_TX_2'M_1X_2D_T$ as following

$$D_TX_2'M_1X_2D_T = D_TX_2'X_2D_T - \frac{1}{\sqrt{T}}D_TX_2'X_1\left(\frac{1}{T}X_1'X_1\right)^{-1}X_1'X_2D_T\frac{1}{\sqrt{T}}. \tag{6.71}$$

Let us define $\frac{1}{T}X_1'X_1 =$

$$\begin{bmatrix} \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-1})^2 & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-2} & \cdots & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-(p-1)} \\ \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-2} & \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-2})^2 & \cdots & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-2}\Delta y_{t-(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-1}\Delta y_{t-(p-1)} & \frac{1}{T}\sum_{t=p+1}^T\Delta y_{t-2}\Delta y_{t-(p-1)} & \cdots & \frac{1}{T}\sum_{t=p+1}^T(\Delta y_{t-(p-1)})^2 \end{bmatrix}.$$

Under the null hypothesis and by using the results in Theorem 6.1,

$$\frac{1}{T}X_1'X_1 \Rightarrow \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-2} & \gamma_{p-3} & \cdots & 1 \end{bmatrix} = \gamma_0 \mathbf{\Pi} \tag{6.72}$$

where $\rho_i = \gamma_i/\gamma_0$, for $i = 1, \ldots, (p-2)$.

Let

$$D_T X_2' X_2 D_T = \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix}$$

where

$$A_1 = \begin{bmatrix} \frac{1}{T^4}\sum_{t=p+1}^{T} y_{t-1}^2 y_{t-d}^4 & \cdots & \frac{1}{T^{(2k+6)/2}}\sum_{t=p+1}^{T} y_{t-1}^2 y_{t-d}^{2k+2} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+6)/2}}\sum_{t=p+1}^{T} y_{t-1}^2 y_{t-d}^{2k+2} & \cdots & \frac{1}{T^{(4k+4)/2}}\sum_{t=p+1}^{T} y_{t-1}^2 y_{t-d}^{4k} \end{bmatrix},$$

$$A_2 = \begin{bmatrix} A_{21} & A_{22} & \cdots & A_{2(2k-1)} \end{bmatrix}$$

with

$$A_{2i} = \begin{bmatrix} \frac{1}{T^{(6+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+3} \Delta y_{t-1} & \cdots & \frac{1}{T^{(6+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+3} \Delta y_{t-(p-1)} \\ \frac{1}{T^{(7+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+4} \Delta y_{t-1} & \cdots & \frac{1}{T^{(7+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+4} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+4+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+2k+1} \Delta y_{t-1} & \cdots & \frac{1}{T^{(2k+4+i)/2}}\sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{i+2k+1} \Delta y_{t-(p-1)} \end{bmatrix},$$

and

$$A_3 = \begin{bmatrix} A_{31} & A_{32} & A_{33} & \cdots & A_{3(2k-1)} \\ A_{32} & A_{33} & A_{34} & \cdots & A_{3(2k)} \\ A_{33} & A_{34} & A_{35} & \cdots & A_{3(2k+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{3(2k-1)} & A_{3(2k)} & A_{3(2k+1)} & \cdots & A_{3(4k-3)} \end{bmatrix}_{[(p-1)(2k-1)]\times[(p-1)(2k-1)]}$$

with

$$A_{3i} = \begin{bmatrix} \frac{1}{T^{(5+i)/2}}\sum_{t=p+1}^{T} y_{t-d}^{i+3} (\Delta y_{t-1})^2 & \cdots & \frac{1}{T^{(5+i)/2}}\sum_{t=p+1}^{T} y_{t-d}^{i+3} \Delta y_{t-1} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(5+i)/2}}\sum_{t=p+1}^{T} y_{t-d}^{i+3} \Delta y_{t-1} \Delta y_{t-(p-1)} & \cdots & \frac{1}{T^{(5+i)/2}}\sum_{t=p+1}^{T} y_{t-d}^{i+3} (\Delta y_{t-(p-1)})^2 \end{bmatrix}.$$

Under the null hypothesis and by using the results in Theorem 6.1,

$$A_1 \Rightarrow \begin{bmatrix} \lambda^6 \int W^6 & \lambda^7 \int W^7 & \cdots & \lambda^{(2k+4)} \int W^{(2k+4)} \\ \lambda^7 \int W^7 & \lambda^8 \int W^8 & \cdots & \lambda^{(2k+5)} \int W^{(2k+5)} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda^{(2k+4)} \int W^{(2k+4)} & \lambda^{(2k+5)} \int W^{(2k+5)} & \cdots & \lambda^{(4k+2)} \int W^{(4k+2)} \end{bmatrix}_{(2k-1)\times(2k-1)}, \tag{6.73}$$

$$A_2 \Rightarrow \mathbf{0}_{(2k-1)\times(p-1)(2k-1)}$$

and

$$A_3 \Rightarrow \begin{bmatrix} \mathbf{\Pi}\gamma_0\lambda^4 \int W^4 & \mathbf{\Pi}\gamma_0\lambda^5 \int W^5 & \cdots & \mathbf{\Pi}\gamma_0\lambda^{(2k+2)} \int W^{(2k+2)} \\ \mathbf{\Pi}\gamma_0\lambda^5 \int W^5 & \mathbf{\Pi}\gamma_0\lambda^6 \int W^6 & \cdots & \mathbf{\Pi}\gamma_0\lambda^{(2k+3)} \int W^{(2k+3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Pi}\gamma_0\lambda^{(2k+2)} \int W^{(2k+2)} & \mathbf{\Pi}\gamma_0\lambda^{(2k+3)} \int W^{(2k+3)} & \cdots & \mathbf{\Pi}\gamma_0\lambda^{(4k)} \int W^{(4k)} \end{bmatrix}.$$

Let us define

$$\frac{1}{\sqrt{T}} D_T X_2' X_1 = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

where

$$B_1 = \begin{bmatrix} \frac{1}{T^{5/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^2 \Delta y_{t-1} & \cdots & \frac{1}{T^{5/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^2 \Delta y_{t-(p-1)} \\ \frac{1}{T^{6/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^3 \Delta y_{t-1} & \cdots & \frac{1}{T^{6/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^3 \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(2k+3)/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^{2k} \Delta y_{t-1} & \cdots & \frac{1}{T^{(2k+3)/2}} \sum_{t=p+1}^T y_{t-1} y_{t-d}^{2k} \Delta y_{t-(p-1)} \end{bmatrix}$$

and

$$B_2 = \begin{bmatrix} B_{21} \\ \vdots \\ B_{2(2k-1)} \end{bmatrix}$$

with

$$B_{2i} = \begin{bmatrix} \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} (\Delta y_{t-1})^2 & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} \Delta y_{t-1} \Delta y_{t-(p-1)} \\ \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} \Delta y_{t-2} \Delta y_{t-1} & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} \Delta y_{t-2} \Delta y_{t-(p-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} \Delta y_{t-(p-1)} \Delta y_{t-1} & \cdots & \frac{1}{T^{(3+i)/2}} \sum_{t=p+1}^T y_{t-d}^{i+1} (\Delta y_{t-(p-1)})^2 \end{bmatrix}.$$

Under the null hypothesis and by using the results in Theorem 6.1,

$$\begin{aligned} B_1 &\Rightarrow \mathbf{0}_{(2k-1)\times(p-1)}, \\ B_2 &\Rightarrow \begin{bmatrix} \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\gamma_0\lambda^{2k} \int W^{2k} \end{bmatrix}_{(p-1)(2k-1)\times(p-1)}. \end{aligned}$$

Thus,

$$\frac{1}{\sqrt{T}} D_T X_2' X_1 \Rightarrow \begin{bmatrix} \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\lambda^{2k} \int W^{2k} \end{bmatrix}. \tag{6.74}$$

$$\begin{aligned} & \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1'X_1}{T}\right)^{-1} X_1' X_2 D_T \frac{1}{\sqrt{T}} \\ \Rightarrow & \begin{bmatrix} \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi}\gamma_0\lambda^{2k} \int W^{2k} \end{bmatrix} \frac{1}{\gamma_0}\mathbf{\Pi}^{-1} \begin{bmatrix} \mathbf{0}_{(p-1)\times(2k-1)} & \mathbf{\Pi}\gamma_0\lambda^2 \int W^2 & \cdots & \mathbf{\Pi}\gamma_0\lambda^{2k}\int W^{2k} \end{bmatrix} \\ = & \begin{bmatrix} \mathbf{0}_{(2k-1)\times(2k-1)} & \mathbf{0}_{(2k-1)\times(p-1)} & \cdots & \mathbf{0}_{(2k-1)\times(p-1)} \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^4(\int W^2)^2\mathbf{\Pi} & \cdots & \gamma_0\lambda^{2k+2}\int W^2 \int W^{2k}\mathbf{\Pi} \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^5\int W^2 \int W^3\mathbf{\Pi} & \cdots & \gamma_0\lambda^{2k+3}\int W^3 \int W^{2k}\mathbf{\Pi} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{(p-1)\times(2k-1)} & \gamma_0\lambda^{2k+2}\int W^2 \int W^{2k}\mathbf{\Pi} & \cdots & \gamma_0\lambda^{4k}(\int W^{2k})^2\mathbf{\Pi} \end{bmatrix}. \end{aligned}$$

Thus,

$$\begin{aligned} D_T X_2' M_1 X_2 D_T &= D_T X_2' X_2 D_T - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1'X_1}{T}\right)^{-1} X_1' X_2 D_T \frac{1}{\sqrt{T}} \\ &\Rightarrow \begin{bmatrix} C_1 & \mathbf{0} \\ \mathbf{0} & C_2 \end{bmatrix} \\ &= \triangle F_2(W) \triangle \end{aligned} \tag{6.75}$$

where C_1 is the asymptotic distribution of A_1 in (6.73), C_2 is the asymptotic distribution of $A_3 - B_2$,

$$\triangle = diag\left(\lambda^3, \lambda^4, \cdots, \lambda^{2k+1}, \underbrace{\sqrt{\gamma_0}\lambda^2, \cdots, \sqrt{\gamma_0}\lambda^2}_{p-1}, \cdots, \underbrace{\sqrt{\gamma_0}\lambda^{2k}, \cdots, \sqrt{\gamma_0}\lambda^{2k}}_{p-1}\right) \tag{6.76}$$

and $F_2(W)$ is defined in Theorem 6.2.

(ii) $D_T X_2' M_1 \epsilon$ can be written as

$$D_T X_2' M_1 \epsilon = D_T X_2' \epsilon - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{1}{T} X_1' X_1 \right)^{-1} \frac{1}{\sqrt{T}} X_1' \epsilon. \tag{6.77}$$

The first term of (6.77) is

$$D_T X_2' \epsilon = \begin{bmatrix} E_1 \\ E_{21} \\ \vdots \\ E_{2(2k-1)} \end{bmatrix}$$

where

$$E_1 = \begin{bmatrix} T^{-2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^2 \epsilon_t \\ \vdots \\ T^{-(2k+2)/2} \sum_{t=p+1}^{T} y_{t-1} y_{t-d}^{2k} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_\epsilon \lambda^3 \int W^3 dW \\ \vdots \\ \sigma_\epsilon \lambda^{(2k+1)} \int W^{(2k+1)} dW \end{bmatrix}, \tag{6.78}$$

$$E_{2i} = \begin{bmatrix} T^{-(1+i/2)} \sum_{t=p+1}^{T} y_{t-2}^{i+1} \Delta y_{t-1} \epsilon_t \\ \vdots \\ T^{-(1+i/2)} \sum_{t=p+1}^{T} y_{t-1}^{i+1} \Delta y_{t-(p-1)} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} dW_1 \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} dW_{(p-1)} \end{bmatrix}$$

For the second term of (6.77),

$$\frac{1}{\sqrt{T}} X_1' \epsilon = \begin{bmatrix} T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-1} \epsilon_t \\ T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-2} \epsilon_t \\ \vdots \\ T^{-1/2} \sum_{t=p+1}^{T} \Delta y_{t-(p-1)} \epsilon_t \end{bmatrix} \Rightarrow \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon W_1(1) \\ \sqrt{\gamma_0} \sigma_\epsilon W_2(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon W_{(p-1)}(1) \end{bmatrix} \tag{6.79}$$

Thus, gathering (6.74), (6.72) and (6.79),

$$\begin{aligned} & \tfrac{1}{\sqrt{T}} D_T X_2' X_1 \left(\tfrac{X_1' X_1}{T} \right)^{-1} X_1' \epsilon \tfrac{1}{\sqrt{T}} \\ \Rightarrow & \begin{bmatrix} \mathbf{0} \\ \mathbf{\Pi} \gamma_0 \lambda^2 \int W^2 \\ \vdots \\ \mathbf{\Pi} \gamma_0 \lambda^{2k} \int W^{2k} \end{bmatrix} \tfrac{1}{\gamma_0} \mathbf{\Pi}^{-1} \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon W_1(1) \\ \sqrt{\gamma_0} \sigma_\epsilon W_2(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon W_{(p-1)}(1) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ D_1 \\ \vdots \\ D_{(2k-1)} \end{bmatrix} \end{aligned} \tag{6.80}$$

where

$$D_i = \begin{bmatrix} \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} W_1(1) \\ \vdots \\ \sqrt{\gamma_0} \sigma_\epsilon \lambda^{i+1} \int W^{i+1} W_{(p-1)}(1) \end{bmatrix}$$

Thus,

$$\begin{aligned}
D_T X_2' M_1 \epsilon &= D_T X_2' \epsilon - \frac{1}{\sqrt{T}} D_T X_2' X_1 \left(\frac{X_1' X_1}{T} \right)^{-1} X_1' \epsilon \frac{1}{\sqrt{T}} \\
&= \begin{bmatrix} G_1 \\ G_{21} \\ \vdots \\ G_{2(2k-1)} \end{bmatrix} \Rightarrow \sigma_\epsilon \triangle F_1(W) \qquad (6.81)
\end{aligned}$$

where G_1 has the same asymptotic distribution with E_1 in (6.78),

$$G_{2i} = E_{2i} - D_i \Rightarrow \begin{bmatrix} \sqrt{\gamma_0}\sigma_\epsilon \lambda^{i+1} \left(\int W^{i+1} dW_1 - W_1(1) \int W^{i+1} \right) \\ \vdots \\ \sqrt{\gamma_0}\sigma_\epsilon \lambda^{i+1} \left(\int W^{i+1} dW_{(p-1)} - W_{(p-1)}(1) \int W^{i+1} \right) \end{bmatrix},$$

$\triangle$ and $F_1(W)$ are defined in (6.76) and Theorem 6.2 respectively.

Thus, under H_0, the asymptotic distribution of F_{nl} can be determined using the following results,

$$\begin{aligned}
F_{nl} &= \frac{1}{\hat{\sigma}^2_{\epsilon^*}} \left(\hat{\mathbf{b}}_2 - \mathbf{b}_2 \right)' (X_2' M_1 X_2) \left(\hat{\mathbf{b}}_2 - \mathbf{b}_2 \right) \\
&= \frac{1}{\hat{\sigma}^2_{\epsilon^*}} (D_T X_2' M_1 \epsilon)' (D_T X_2' M_1 X_2 D_T)^{-1} (D_T X_2' M_1 \epsilon) \\
&\Rightarrow \frac{1}{\hat{\sigma}^2_{\epsilon^*}} (\sigma_\epsilon \triangle F_1(W))' (\triangle F_2(W) \triangle)^{-1} (\sigma_\epsilon \triangle F_1(W)) \\
&= \frac{\sigma^2_\epsilon}{\hat{\sigma}^2_{\epsilon^*}} (F_1(W))' \triangle \triangle^{-1} (F_2(W))^{-1} \triangle^{-1} \triangle (F_1(W)) \\
&= (F_1(W))' (F_2(W))^{-1} F_1(W). \qquad (6.82)
\end{aligned}$$

The final result in (6.82) is obtained because under H_0, $\epsilon^* = \epsilon$ and $\widehat{\sigma}^2_{\epsilon^*}$ is a consistent estimator of $\sigma^2_{\epsilon^*}$. Thus, $\widehat{\sigma}^2_{\epsilon^*}$ is also a consistent estimator of σ^2_ϵ.

6.5 Unit Root Test Analysis for a k-ESTAR(2) Model

In this section, a unit root test analysis for a k-ESTAR(2) model is considered. For this model, we can resolve the problem of singularity in Venetis *et al.* (2009). We will show that for this model, the test in (6.68) does not involve the nuisance parameter $\mathbf{\Pi}$. We compare the performance of our approach and other unit root tests.

Now, consider (6.39) for a k-ESTAR(2) model as follows,

$$\begin{aligned} y_t &= \theta_{1,0} + \sum_{j=1}^{2} \theta_{1,j}\ y_{t-1} \theta_{1,1}^*\ \Delta y_{t-1} \\ &+ \left[\theta_{2,0} + \sum_{j=1}^{2} \theta_{2,j}\ y_{t-1} + \theta_{2,1}^*\ \Delta y_{t-1} \right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ &\qquad t = 1, 2, \ldots, T \end{aligned} \tag{6.83}$$

where $\theta_{i,1}^* = -\sum_{k=j+1}^{p} \theta_{i,k}$, and $i = 1, 2$.

Using the same assumptions for (6.40), (6.83) becomes,

$$\begin{aligned} \Delta y_t &= \theta_{1,1}^*\ \Delta y_{t-1} + \left[(\theta_{2,1} + \theta_{2,2}) y_{t-1} + \theta_{2,1}^*\ \Delta y_{t-1} \right] G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ & t = 1, 2, \ldots, T \end{aligned} \tag{6.84}$$

where $\theta > 0$ and $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2 and $G^*(\theta, \mathbf{e}, y_{t-d})$ is the same as (6.41).

Recalling the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (6.42),

$$G^*(\theta, \mathbf{e}, y_{t-d}) \approx \theta^2 \sum_{s=0}^{2(k-1)} \delta_s y_{t-d}^{s+2} + R,$$

where R is the remainder, $\delta_0 = (\prod_{i=2}^{k} e_i)^2$ and $\delta_{2(k-1)} = 1$.

Thus, (6.84) becomes,

$$\Delta y_t = \theta_{1,1}^*\ \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ y_{t-1} y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ y_{t-d}^{s+2} \Delta y_{t-1} + \epsilon_t^* \tag{6.85}$$

where $\gamma_{1,s} = \theta^2 \delta_s (\theta_{2,1} + \theta_{2,2})$, $\gamma_{2,s} = \theta^2 \delta_s \theta_{2,1}^*$ and $\epsilon_t^* = \epsilon_t + R \left[\sum_{j=1}^{2} \theta_{2,j}\ y_{t-j} \right]$. If $\theta = 0$, y_t in (6.84) is linear in term of y_{t-1} and y_{t-2} and $\epsilon_t^* = \epsilon_t$ since the remainder $R \equiv 0$.

Testing the null hypothesis of a unit root ($H_0 : \theta = 0$) against alternative of a globally stationary k-ESTAR(2) model is equivalent to test,

$$H_0 : \gamma_{1,s} = \gamma_{2,s} = 0, \qquad \text{for all } s \text{ in (6.85)} .$$

Under the null hypothesis H_0, $\epsilon_t^* = \epsilon_t$, thus, (6.85) becomes

$$\Delta y_t = \theta_{1,1}^*\ \Delta y_{t-1} + \epsilon_t = \sum_{j=0}^{\infty} c_j\ \epsilon_{t-j} = C(L) \epsilon_t = \eta_t, \tag{6.86}$$

where L is the lag operator, i.e. $L y_t = y_{t-1}$, and $\theta_{1,1}^* = -\theta_{1,2}$.

Following the results from Section 6.4, for $p = 2$, $\mathbf{\Pi}$ in (6.69) will become a constant

1, thus $F_1(W)$ and $F_2(W)$ for the F_{nl} statistics in (6.82) becomes

$$F_{nl} = (F_1(W))'(F_2(W))^{-1}F_1(W) \tag{6.87}$$

where

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \vdots \\ \int W^{(2k+1)} dW \\ (\int W^2 dW_1 - W_1(1) \int W^2) \\ \vdots \\ (\int W^{2k} dW_1 - W_1(1) \int W^{2k}) \end{bmatrix},$$

$$F_2(W) = \begin{bmatrix} F_{21}(W) & \mathbf{0} \\ \mathbf{0} & F_{22}(W) \end{bmatrix},$$

$$F_{21}(W) = \begin{bmatrix} \int W^6 & \cdots & \int W^{(2k+4)} \\ \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \cdots & \int W^{(4k+2)} \end{bmatrix}$$

and

$$F_{22}(W) = \begin{bmatrix} (\int W^4 - (\int W^2)^2) & \cdots & (\int W^{2k+2} - \int W^2 \int W^{2k}) \\ \vdots & \ddots & \vdots \\ (\int W^{2k+2} - \int W^2 \int W^{2k}) & \cdots & (\int W^{4k} - (\int W^{2k})^2) \end{bmatrix}.$$

Note that even if the limit distribution of F_{nl} for k-ESTAR(2) model in (6.87) does not depend on any nuisance parameters, special attention is needed for values of $\theta_{1,2}$ close to -1 or 1. Under the null hypothesis, y_t is a function of $\theta_{1,2}$, as is seen from (6.86). Thus, the time series Δy_t is then close to having a unit root or becoming nonstationary. In these situations the test may reject the null hypothesis too often.

In comparison with Venetis *et al.* (2009) (we call it VPP for short), their approach will consider,

$$\begin{aligned} \Delta y_t &= \theta^*_{1,1}\, \Delta y_{t-1} + [\theta_{2,1}\, y_{t-1} + \theta_{2,2}\, y_{t-2}]G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ & \quad t = 1, 2, \ldots, T, \end{aligned} \tag{6.88}$$

rather than (6.84). Substituting the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (6.42) into (6.88) make it becomes

$$\Delta y_t = \theta^*_{1,1}\, \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\, y_{t-1}\, y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \gamma_{2,s}\, y_{t-2}\, y_{t-d}^{s+2} + \epsilon^*_t \tag{6.89}$$

rather than (6.85). ϵ^*_t is defined the same as in (6.85). Since $y_{t-1}y_{t-d}^{s+2}$ are collinear

Table 6.1: Asymptotic critical values of F test statistics for k-ESTAR(2) model.

	F_{nl}			F_{VPP}		
Sig. Level	0.1	0.05	0.01	0.1	0.05	0.01
k						
1	5.49	6.94	10.37	3.73	4.88	7.73
2	13.83	15.98	20.80	9.54	11.36	15.47
3	20.44	23.18	28.61	13.64	15.70	19.94
4	26.64	29.65	36.64	17.06	19.38	28.61

Note: Simulations were based on samples size T=10,000 and 50,000 replications.

with $y_{t-2}y_{t-d}^{s+2}$, for $s = 0, 1, \ldots, 2(k-1)$, asymptotically, Venetis *et al.* (2009) rearrange (6.89) to becomes

$$\Delta y_t = \theta^*_{1,1}\, \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\, y_{t-1}\, y_{t-d}^{s+2} + \upsilon_t \tag{6.90}$$

where $\upsilon_t = \sum_{s=0}^{2(k-1)} \gamma_{2,s}\, y_{t-2}\, y_{t-d}^{s+2} + \epsilon^*_t$. Thus, they added some regressors to the error term.

Using (6.90), $F_1(W)$ and $F_2(W)$ for the asymptotic F-test statistics for VPP will be

$$F_{VPP} = (F_1(W))'(F_2(W))^{-1}F_1(W) \tag{6.91}$$

where

$$F_1(W) = \begin{bmatrix} \int W^3 dW \\ \int W^4 dW \\ \vdots \\ \int W^{(2k+1)} dW \end{bmatrix} \text{ and } F_2(W) = \begin{bmatrix} \int W^6 & \int W^7 & \cdots & \int W^{(2k+4)} \\ \int W^7 & \int W^8 & \cdots & \int W^{(2k+5)} \\ \vdots & \vdots & \ddots & \vdots \\ \int W^{(2k+4)} & \int W^{(2k+5)} & \cdots & \int W^{(4k+2)} \end{bmatrix}.$$

The expressions of $F_1(W)$ and $F_2(W)$ in F_{nl} and F_{VPP} are different. It is due to the fact that the VPP has added some regressors into the errors term so that our result has additional terms compare to the VPP. The advantage of our approach is that it has solved the singularity problem for this case without the need to add some regressors into the errors term.

Asymptotic critical values for F-type statistics from F_{nl} in (6.87) and F_{VPP} in (6.91) with $k = 1, \ldots, 4$ are obtained via stochastic simulations and presented in Table 6.1.[10] Computer program for Table 6.1 is provided in Appendix A.3.1.

As suggested by Venetis *et al.* (2009), for computational purposes F_{nl} and F_{VPP} can be easily calculated following the steps below:

[10]Critical values for the VPP in this table are quite different to the values in Table 2b in Venetis *et al.* (2009) as they did not assume that $\theta_{2,0} = 0$ when $e_i = 0$ for a certain i, $i = 1, \ldots, k$.

(i) Estimate the unrestricted model on (6.85) for F_{nl} or (6.90) for F_{VPP} and keep the sum of squared residuals SSR_U.

(ii) Estimate (6.86) as the restricted model implied by the null hypothesis and keep the sum of squared residuals SSR_R. Note that based on the null hypothesis, F_{nl} and F_{VPP} have the same restricted model.

(iii) Calculate the ratio $F = T(SSR_R - SSR_U)/SSR_U$ where T denotes the number of observations in the restricted model and then compare with the critical values in Table 6.1.

6.5.1 Sufficient Conditions for Stationarity of a k-ESTAR(2) Model

For a k-ESTAR(2) model, we determine a set of sufficient conditions for parameter combinations corresponding to a stationary series. Knowing the conditions will be useful in doing simulation study presented in the next subsection.

Let us rearrange y_t in (6.84) as follows,

$$y_t = (\theta_{1,1} + \theta_{2,1}G^*(\theta, \mathbf{e}, y_{t-d}))\, y_{t-1} + (\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d}))\, y_{t-2} + \epsilon_t. \tag{6.92}$$

Considering (6.92) as an AR(2) process, the necessary stationarity conditions for this process (see Box and Jenkins, 1976, p. 58) are,

$$(\theta_{1,1} + \theta_{2,1}\, G^*(\theta, \mathbf{e}, y_{t-d})) + (\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{6.93}$$

$$(\theta_{1,2} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) - (\theta_{1,1} + \theta_{2,1}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{6.94}$$

$$-1 < \theta_{12} + \theta_{2,2}\, G^*(\theta, \mathbf{e}, y_{t-d})) < 1 \tag{6.95}$$

Note that from (6.41), $0 < G^*(\theta, \mathbf{e}, y_{t-d}) < 1$, and under the null hypothesis of a unit root in the linear term, $\theta_{1,1} + \theta_{1,2} = 1$ (see the discussion in Section 6.4). Thus, from (6.93) we obtain,

$$(\theta_{2,1} + \theta_{2,2})G^*(\theta, \mathbf{e}, y_{t-d}) < 0. \tag{6.96}$$

The stationarity conditions in (6.96) will be fulfilled if $(\theta_{2,1} + \theta_{2,2}) < 0$.

From (6.94), we have

$$(\theta_{1,2} - \theta_{1,1}) + (\theta_{2,2} - \theta_{2,1})G^*(\theta, \mathbf{e}, y_{t-d}) < 1.$$

Under the null hypothesis of a unit root, $\theta_{1,1} + \theta_{1,2} = 1$. Thus,

$$((1 - \theta_{1,1}) - \theta_{1,1}) + (\theta_{2,2} - \theta_{2,1})G^*(\theta, \mathbf{e}, y_{t-d}) < 1$$

$$(\theta_{2,2} - \theta_{2,1})G^*(\theta, \mathbf{e}, y_{t-d}) < 2\theta_{1,1} \tag{6.97}$$

The stationarity condition in (6.97) will be fulfilled if $0 \le (\theta_{2,2} - \theta_{2,1}) < 2\theta_{1,1}$ or $(\theta_{2,2} - \theta_{2,1}) \le 0 < 2\theta_{1,1}$ and $\theta_{1,1} \ge 0$.

To fulfill the stationarity condition in (6.95), $\theta_{1,2}$ should be $-1 < \theta_{1,2} < 1$, so that

$$-1 - \theta_{1,2} < \theta_{2,2} G^*(\theta, \mathbf{e}, y_{t-d}) < 1 - \theta_{1,2}.$$

Thus, the parameters should satisfy $-1 - \theta_{1,2} < \theta_{2,2} \leq 0$ or $0 \leq \theta_{2,2} < 1 - \theta_{1,2}$ to fulfill the stationarity conditions.

To summarise, one set of the sufficient stationarity conditions for (6.92) is

$$\begin{aligned} (\theta_{2,1} + \theta_{2,2}) &< 0 \\ \theta_{1,1} &\geq 0 \\ 0 \leq (\theta_{2,2} - \theta_{2,1}) < 2\theta_{1,1} \text{ or } (\theta_{2,2} - \theta_{2,1}) &\leq 0 \\ -1 < \theta_{1,2} &< 1 \\ -1 - \theta_{1,2} < \theta_{2,2} \leq 0 \text{ or } 0 \leq \theta_{2,2} &< 1 - \theta_{1,2}. \end{aligned} \tag{6.98}$$

6.5.2 Small Sample Properties of F_{nl} Test for a k-ESTAR(2) Model

In this subsection, Monte Carlo experiments of small sample size and power performance of F_{nl} Test for a k-ESTAR(2) Model are undertaken. For comparison, we include F_{VPP}, the augmented KSS test (denoted as AKSS, Kapetanios *et al.*, 2003) and the augmented Dickey-Fuller test (denoted as ADF, Fuller, 1976). For the AKSS test, we only consider for Case 1 because based on Venetis *et al.* (2009), generally this case has more power than the other cases.

The calculated F statistics from the F_{nl} and F_{VPP} are compared with the critical values in Table 6.1. The critical value for the t-test of AKSS test is -2.22 obtained from Table 1 in Kapetanios *et al.* (2003). The critical value for the t-test of ADF test is -1.95 obtained in Fuller (1976). For each experiment, the rejection probability (in percentage) of the null hypothesis are computed with the nominal size of the tests is set at 5%. The sample size is considered for $T = 50, 100, 200$ with the number of replications at 10,000.

The size of Alternative Tests

To obtain the size of alternative tests, we generate samples from the null model, i.e. :

$$\Delta y_t = \theta^*_{1,1} \Delta y_{t-1} + \epsilon_t \tag{6.99}$$

where $\theta^*_{1,1} = -\theta_{1,2}$ and ϵ_t is drawn from the standard normal distribution. In particular, we choose a broad range of parameter values for $\theta_{1,2} = \{-0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8\}$.

For computational purposes, the regression model in (6.99) becomes the restricted model for F_{nl} and F_{VPP}. Furthermore, the unrestricted models for F_{nl} and F_{VPP} are the regression models in (6.85) and (6.89) respectively. For the AKSS test and the ADF test, we include the lagged first difference (Δy_{t-1}) to overcome the autocorrelation [11], so that the regression model for the AKSS test is

$$\Delta y_t = \delta_1 \ y_{t-1} y_{t-d}^2 + \delta_2 \ \Delta y_{t-1} + \epsilon_t \tag{6.100}$$

and the regression model for the ADF test is

$$\Delta y_t = \delta_1 \ y_{t-1} + \delta_2 \ \Delta y_{t-1} + \epsilon_t. \tag{6.101}$$

The hypothesis for the AKSS test and ADF test is

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0 \tag{6.102}$$

Then, the calculated t-test for δ_1 are compared with the critical values of the AKSS test and ADF test. The null hypothesis for the AKSS test and the ADF test conclude that y_t has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that y_t is a globally stationary 1-ESTAR(1) model while the ADF test concludes that y_t is a stationary linear ARMA model. The size of the alternative tests are presented in Table 6.2. Computer program for Table 6.2 is provided in Appendix A.3.1.

As Venetis *et al.* (2009) noted, the F tests (F_{nl} and F_{VPP}) resemble the familiar $\mathcal{X}^2$ test when under the null hypothesis the process is stationary. For this reason, F_{nl} and F_{VPP} may suffer from size problems when the number of restrictions is large and the time series is short. As F_{nl} has larger number of restrictions than F_{VPP}, the distortion becomes larger for F_{nl} than F_{VPP} for the same conditions. Table 6.2 shows that F_{nl} and F_{VPP} are oversized for large values of k and $\theta_{1,2} = -0.8$. If $\theta_{1,2}$ is close to -1, $\theta_{1,1}^*$ in (6.86) will be close to 1. It means that Δy_t will be close to has an explosive unit root. Generally, if the values of $\theta_{1,2}$ is close to 0 and the sample size increases from $T = 50$ to $T = 200$, F_{nl} and F_{VPP} statistics turn close to the nominal size of 5%. In addition to the F_{nl} and F_{VPP} tests for $k = 1, 2, 3, 4$, the rejection probabilities of the null hypothesis for the AKSS and the ADF are also computed in the last two columns in Table 6.2. For all cases, the rejection probabilities for the AKSS are less than 5%. Therefore, the AKSS test has more power to detect the null hypothesis than the other methods. It happens because this test involves less estimation parameters and deals with one sided alternatives of stationarity. On the other hand, even though the ADF test also involves less estimation parameters and deals with one sided alternatives of stationarity, the rejection probabilities for the ADF tests are close to or slightly higher than 5%. Even for $\theta_{1,2} = 0$ and $T = 200$, the rejection probability of the null hypothesis for the ADF test is 5.24%.

[11]Therefore we call the tests as the augmented KSS test and the augmented DF test

Table 6.2: The size of alternative tests (in percentage)

	F_{nl}				F_{VPP}				AKSS	ADF
	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
$\theta_{1,2} = -0.8$										
T=50	7.36	16.05	25.43	36.84	5.95	9.32	11.73	14.32	4.61	5.79
T=100	5.79	9.66	12.59	16.71	5.37	6.54	7.65	8.17	4.54	5.25
T=200	5.25	7.15	7.62	9.10	5.16	5.91	5.96	6.27	4.62	5.12
$\theta_{1,2} = -0.5$										
T=50	5.91	11.15	17.38	26.27	5.53	7.39	8.67	10.54	4.38	5.39
T=100	5.13	7.01	8.12	10.00	5.15	5.76	5.97	6.28	4.45	4.98
T=200	4.94	6.00	5.42	6.19	5.04	5.57	5.10	5.20	4.60	5.26
$\theta_{1,2} = -0.2$										
T=50	5.26	9.96	15.06	23.73	5.39	6.20	7.77	9.39	4.23	5.25
T=100	4.76	5.87	7.11	8.69	4.92	4.95	4.99	5.53	4.34	4.95
T=200	4.72	5.33	4.98	5.23	5.07	5.19	4.45	4.46	4.70	5.21
$\theta_{1,2} = 0$										
T=50	5.23	9.27	14.48	23.42	5.07	5.32	7.24	8.93	4.05	5.18
T=100	4.72	5.49	6.56	8.45	4.72	4.62	4.43	4.94	4.16	4.93
T=200	4.71	5.03	4.77	5.08	4.93	4.74	4.05	4.01	4.52	5.24
$\theta_{1,2} = 0.2$										
T=50	5.45	8.96	14.38	23.42	4.95	5.01	6.66	8.68	3.88	5.18
T=100	4.80	5.45	6.21	8.53	4.55	4.30	3.94	4.62	4.01	4.91
T=200	4.84	5.00	4.85	4.81	4.85	4.31	3.67	3.69	4.46	5.21
$\theta_{1,2} = 0.5$										
T=50	6.07	8.89	13.95	22.83	4.50	4.63	6.08	8.19	3.51	5.06
T=100	5.26	5.57	6.18	8.20	4.43	3.82	3.53	4.39	3.78	4.89
T=200	5.06	5.16	4.55	4.91	4.71	3.87	3.29	3.41	4.30	5.24
$\theta_{1,2} = 0.8$										
T=50	6.74	9.28	14.13	23.78	4.28	4.56	5.61	7.88	3.46	4.93
T=100	5.89	6.13	6.58	8.82	3.95	3.49	3.65	4.33	3.39	4.83
T=200	5.59	5.29	4.91	5.14	4.07	3.22	3.11	3.28	3.76	5.27

The Power of Alternative Tests

To evaluate the power of tests against globally stationary k-ESTAR(2) model, samples were simulated from model in (6.84) with $d = 1$ and ϵ_t is drawn from a standard normal distributed. We calculate the rejection probabilities of the null hypothesis (in percentage) given that the y_t is an k-ESTAR(2) model. The simulation results are summarised in Tables 6.3-6.8. Computer program for Tables 6.3-6.8 is provided in Appendix A.3.1.

The data for Tables 6.3-6.5 are simulated with $k = 1$, i.e. $e_1 = 0$. From the three tables, the rejection probabilities increase as k increases for the F_{nl} and F_{VPP} with $T = 50$. It is happened due to large number of restrictions and short time series. Therefore, even though the rejection probabilities for the F_{nl} test with $k = 4$ are quite high (around 22%-31%), we do not recommend the results from very small sample. For $k = 1$, the rejection probabilities increase as T increases for the F_{nl} and F_{VPP}. Furthermore, for large sample ($T = 200$), the probabilities for $k = 1$ are the highest compared to other k. Apparently, with large sample, the F_{nl} and F_{VPP} tests are able to recognise the true number of equilibrium (for this case, $k = 1$). From the three tables, the F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta_{1,2}$ close to 1. For example, for $(\theta_{1,2}, \theta_{2,1}, \theta_{2,2}) = (0.9, 0, -0.9)$, $T = 200$ and $k = 1$, the F_{nl} test can detect almost 60% while the F_{VPP}, the AKSS and the ADF are around 32%, 31% and 41% respectively.

The data for Tables 6.6-6.8 are simulated with $k = 2$, i.e. $e_1 = 0, e_2 = 3$. Generally, the patterns are similar to $k = 1$. For small sample ($T = 50$), the rejection probabilities increase as k increases for the F_{nl} and F_{VPP}. For large sample ($T = 200$), for $(\theta_{2,1}, \theta_{2,2}) = (0, -0.9)$ in Table 6.6 and $(\theta_{2,1}, \theta_{2,2}) = (0.4, -0.9)$ in Table 6.8, the probabilities for $k = 2$ are the highest compared to other k when we use the F_{nl} tests while the F_{VPP} tests still have the highest probabilities with $k = 1$. Apparently, with large sample, the F_{nl} tests are more able to recognise the true number of equilibriums (for this case, $k = 2$) compared to the F_{VPP} tests. From the three tables, the F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta_{1,2}$ close to 1. For example, for $(\theta_{1,2}, \theta_{2,1}, \theta_{2,2}) = (0.9, 0.4, -0.5)$, $T = 200$ and $k = 2$, the F_{nl} test can detect around 72% while the F_{VPP}, the AKSS and the ADF are only around 5%, 4% and 19% respectively.

Table 6.3: The power of alternative tests (in percentage), $k = 1$, $e_1 = 0$ and $\theta^2 = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0, -0.9)										
T=50	30.96	18.35	21.68	31.67	39.80	12.26	11.81	14.50	35.88	26.30
T=100	74.01	32.57	24.40	25.73	85.73	32.69	20.41	19.08	84.22	78.31
T=200	99.56	81.01	58.91	49.18	99.90	90.81	69.91	57.36	99.89	99.99
(0.2, 0, -0.9)										
T=50	21.03	14.18	18.64	28.51	24.27	7.99	8.98	11.95	21.38	16.45
T=100	52.76	21.12	17.08	19.04	66.32	17.16	12.10	11.83	63.68	52.86
T=200	96.33	56.92	37.04	30.67	98.93	64.55	39.77	31.31	98.74	99.13
(0.5, 0, -0.9)										
T=50	14.45	11.73	15.96	25.72	13.14	5.55	7.20	9.66	11.37	11.11
T=100	32.14	13.40	11.84	13.94	36.84	8.44	7.04	7.56	34.51	27.17
T=200	81.29	32.36	20.38	17.71	88.00	28.26	16.97	14.67	86.93	86.64
(0.7, 0, -0.9)										
T=50	12.96	11.39	16.68	25.24	9.64	4.77	6.32	8.64	8.38	9.43
T=100	25.01	11.68	10.58	12.94	24.23	6.04	5.67	5.90	22.47	18.82
T=200	66.93	24.00	15.51	13.69	67.23	15.58	10.80	9.46	65.80	67.32
(0.9, 0, -0.9)										
T=50	14.70	12.05	17.22	26.91	6.69	3.64	5.20	7.26	5.77	7.65
T=100	24.55	12.70	12.03	13.64	12.22	3.77	3.89	4.56	11.46	12.84
T=200	59.38	22.90	16.41	14.72	32.35	6.81	5.06	5.06	31.27	41.69

Table 6.4: The power of alternative tests (in percentage), $k = 1$, $e_1 = 0$ and $\theta^2 = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.5)										
T=50	9.79	9.79	14.13	23.04	8.93	4.92	5.98	8.15	7.53	8.52
T=100	17.95	9.69	9.13	10.86	16.41	5.21	4.45	4.90	15.04	12.89
T=200	47.03	18.61	12.07	10.51	48.75	9.57	5.60	5.04	46.92	38.51
(0.2, 0.4, -0.5)										
T=50	8.21	9.22	13.88	22.66	6.64	4.06	5.55	7.72	5.77	7.44
T=100	13.13	7.64	7.82	9.06	10.08	3.55	3.59	4.04	9.11	9.64
T=200	31.09	12.87	8.90	8.00	26.91	5.31	3.79	3.71	25.45	20.66
(0.5, 0.4, -0.5)										
T=50	7.81	8.58	13.94	22.41	4.84	3.41	4.89	7.37	4.12	6.18
T=100	10.48	6.64	6.81	8.66	6.07	2.61	2.90	3.74	5.48	7.10
T=200	20.66	9.54	7.08	6.53	11.12	2.77	2.64	3.08	10.49	10.93
(0.7, 0.4, -0.5)										
T=50	8.09	8.88	13.69	22.54	4.09	3.42	5.21	7.35	3.47	5.44
T=100	10.71	6.55	7.35	9.22	4.19	2.08	2.80	3.63	3.87	6.19
T=200	19.95	9.35	7.03	6.63	6.01	1.88	2.35	2.44	5.63	7.55
(0.9, 0.4, -0.5)										
T=50	12.12	10.72	16.33	25.72	2.74	3.34	5.40	7.88	2.13	4.53
T=100	16.01	9.55	9.87	12.06	2.02	2.04	3.22	4.10	1.87	4.14
T=200	29.28	14.26	10.94	10.10	1.80	1.26	2.20	2.68	1.76	4.41

Table 6.5: The power of alternative tests (in percentage), $k = 1$, $e_1 = 0$ and $\theta^2 = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.9)										
T=50	21.28	14.29	18.27	28.21	23.81	7.96	8.60	10.99	20.66	16.38
T=100	52.82	21.66	17.49	19.17	65.04	16.71	10.61	10.96	62.60	52.67
T=200	95.93	57.26	36.76	31.32	98.58	62.72	35.05	27.43	98.43	98.94
(0.2, 0.4, -0.9)										
T=50	15.01	11.94	16.54	25.52	14.39	5.66	7.20	9.46	12.36	11.56
T=100	34.80	14.18	12.47	14.17	41.08	8.60	6.69	7.29	38.29	29.88
T=200	83.82	36.17	22.17	18.97	91.07	31.30	16.38	14.10	90.16	89.81
(0.5, 0.4, -0.9)										
T=50	11.33	10.47	14.67	23.73	8.06	4.33	5.97	8.97	6.92	8.47
T=100	21.13	10.12	9.90	11.37	18.98	4.64	4.47	5.03	17.43	15.76
T=200	57.20	20.26	13.00	12.05	61.43	10.19	7.25	6.74	59.49	55.33
(0.7, 0.4, -0.9)										
T=50	11.34	9.81	14.60	23.36	5.84	3.57	5.34	7.61	5.10	7.28
T=100	17.77	9.36	9.50	11.21	10.66	3.15	3.97	4.52	9.69	11.54
T=200	44.37	16.13	11.23	10.33	34.81	5.06	4.65	4.93	33.08	34.11
(0.9, 0.4, -0.9)										
T=50	14.88	12.02	17.56	26.86	3.39	3.30	5.53	8.08	3.04	5.92
T=100	22.92	11.96	12.14	13.85	3.84	2.53	3.14	4.13	3.44	6.84
T=200	47.91	20.84	15.55	13.58	8.38	2.47	2.75	3.40	7.72	16.31

Table 6.6: The power of alternative tests (in percentage), $k = 2$, $e_1 = 0$, $e_2 = 3$ and $\theta^2 = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0, -0.9)										
T=50	57.89	64.69	62.93	68.05	67.81	56.91	53.21	52.78	64.54	51.92
T=100	91.47	93.38	89.57	86.51	96.50	93.08	90.63	87.12	96.14	91.53
T=200	99.94	99.98	99.94	99.83	99.99	99.99	99.97	99.96	99.98	99.99
(0.2, 0, -0.9)										
T=50	46.72	50.93	50.41	57.34	52.36	41.09	37.70	38.74	49.23	38.67
T=100	77.40	82.05	75.50	72.13	87.90	77.64	73.57	68.54	86.80	76.10
T=200	99.34	99.73	99.07	97.83	99.88	99.63	99.40	98.76	99.88	99.76
(0.5, 0, -0.9)										
T=50	37.60	38.02	39.42	47.59	36.47	25.58	24.26	25.99	33.40	30.95
T=100	60.37	65.17	57.72	55.07	70.28	51.32	47.58	43.77	68.27	54.33
T=200	94.07	96.44	92.40	88.02	97.50	90.87	91.12	87.00	97.29	95.92
(0.7, 0, -0.9)										
T=50	35.83	34.58	37.06	45.00	26.85	18.53	18.24	20.38	24.39	26.49
T=100	55.52	58.64	51.31	49.85	56.23	35.19	32.36	30.64	54.53	46.71
T=200	89.44	92.28	87.50	81.52	92.59	71.87	73.71	68.13	92.20	89.75
(0.9, 0, -0.9)										
T=50	44.28	41.88	43.43	51.12	16.35	11.03	12.75	15.19	14.87	20.08
T=100	66.74	66.34	62.16	59.51	33.92	17.50	17.57	18.22	32.42	46.99
T=200	93.37	93.45	91.88	88.59	69.14	38.00	36.98	36.77	68.17	88.04

Table 6.7: The power of alternative tests (in percentage), $k = 2$, $e_1 = 0$, $e_2 = 3$ and $\theta^2 = 0.01$.

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.5)										
T=50	27.71	20.64	24.47	34.59	19.57	6.67	7.53	9.63	17.11	18.65
T=100	57.72	34.56	31.02	32.58	48.45	11.39	7.62	8.01	45.72	48.51
T=200	93.05	70.70	63.23	59.87	89.27	39.46	20.66	16.74	88.50	95.16
(0.2, 0.4, -0.5)										
T=50	22.01	17.14	21.71	31.02	11.70	4.77	6.19	8.49	10.17	12.28
T=100	42.61	26.06	24.19	25.51	27.33	5.76	5.32	6.10	25.20	29.04
T=200	82.71	54.69	48.92	46.03	72.41	16.69	9.62	9.08	71.03	74.78
(0.5, 0.4, -0.5)										
T=50	19.68	16.32	20.39	29.36	6.08	3.52	5.67	8.30	4.97	8.35
T=100	34.31	23.05	20.99	22.56	11.46	3.48	4.47	5.19	10.44	15.95
T=200	66.34	47.18	42.46	38.45	35.18	6.13	5.25	5.25	33.48	43.33
(0.7, 0.4, -0.5)										
T=50	21.52	17.42	21.83	31.40	3.98	3.27	5.78	8.43	3.28	6.06
T=100	37.15	26.86	24.26	25.13	6.02	2.72	3.98	5.29	5.39	10.44
T=200	65.76	52.44	47.30	42.56	16.43	3.98	4.40	5.14	15.55	30.72
(0.9, 0.4, -0.5)										
T=50	30.09	25.74	29.00	38.24	2.08	3.79	6.92	10.05	1.74	3.99
T=100	50.09	41.48	37.26	37.33	2.24	3.51	5.63	7.13	1.99	5.69
T=200	78.64	72.02	69.01	64.69	4.50	5.43	6.84	7.64	4.17	19.13

Table 6.8: The power of alternative tests (in percentage), $k = 2$, $e_1 = 0$, $e_2 = 3$ and $\theta^2 = 0.01$ (continue).

	F_{nl}				F_{VPP}				AKSS	ADF
$(\theta_{1,2}, \theta_{2,1}, \theta_{2,2})$	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4		
(0, 0.4, -0.9)										
T=50	53.08	52.95	53.80	60.75	59.90	33.01	30.58	34.33	56.51	46.60
T=100	88.13	84.25	79.54	76.75	94.64	74.33	65.30	62.75	93.99	85.80
T=200	99.89	99.83	99.45	98.67	99.98	99.77	99.24	98.52	99.98	99.96
(0.2, 0.4, -0.9)										
T=50	39.96	41.15	42.63	50.49	39.45	19.28	19.52	23.50	36.15	34.17
T=100	69.42	69.06	63.01	60.72	80.34	44.90	39.49	38.32	78.55	65.04
T=200	98.65	97.97	95.79	92.89	99.59	95.63	90.28	86.56	99.55	98.68
(0.5, 0.4, -0.9)										
T=50	32.68	32.54	33.96	42.94	20.14	10.85	12.42	16.53	17.61	24.45
T=100	48.60	53.30	48.07	47.17	49.13	19.32	18.37	19.85	46.70	43.92
T=200	87.57	89.57	83.60	78.50	92.90	53.87	52.25	48.22	92.28	84.59
(0.7, 0.4, -0.9)										
T=50	33.65	32.49	35.20	43.73	10.84	7.80	10.09	13.51	9.54	17.65
T=100	46.56	52.95	47.85	46.85	28.49	11.88	13.11	15.43	26.33	39.06
T=200	77.65	86.19	81.50	75.98	73.33	26.06	30.67	30.03	71.93	70.84
(0.9, 0.4, -0.9)										
T=50	46.28	45.29	47.03	53.87	4.69	7.76	10.88	14.82	4.05	12.01
T=100	62.05	68.43	65.82	63.41	8.65	4.92	12.64	14.83	7.95	37.03
T=200	83.79	93.43	92.66	89.81	24.18	19.51	23.01	25.51	22.99	60.14

6.6 Unit Root Test Analysis for a k-ESTAR(p) model

Unlike k-ESTAR(2) model, the F_{nl} test for k-ESTAR(p) model in (6.68) involves nuisance parameters $\mathbf{\Pi}$. To overcome this problem, we propose two methods. The first is a bootstrap method as an approximation to the asymptotic distribution of F_{nl} and the second is approximation of critical values obtained by assuming $\mathbf{\Pi} = \mathbf{I}_{(p-1)\times(p-1)}$.

6.6.1 Bootstrap Method

A bootstrap approximation can be used to calculate critical values and p-values. For a survey on bootstrapping time series, see Li and Maddala (1996) and for bootstrap applications as approximations of the asymptotic distributions of unit root test, see Caner and Hansen (2001) and Eklund (2003b). Caner and Hansen (2001) analysed a unit root test for a threshold autoregressive (TAR) model involving a nuisance parameter function and suggested a bootstrap method to approximate the null distribution. Eklund (2003b) analysed a unit root test for a 2-LSTAR(2) model. Finding a difficulty to obtain the asymptotic null distribution of the test statistic due to large inverse matrices, he followed the bootstrap method in Caner and Hansen (2001). Using the bootstrap method, Caner and Hansen (2001) and Eklund (2003b) found fairly good results both in size and power tests. Having similarity of STAR models and using a F test statistic as in Eklund (2003b), thus, in this section, we also follow the bootstrap method in Eklund (2003b).

Bootstrap method for k-ESTAR(p) models:

(B1) Calculate the F_{nl} statistic from the sample data based on (6.43) as an unrestricted model and (6.44) as a restricted model (see the calculation of the F_{nl} test statistic in Section 6.5 in the case of k-ESTAR(2) models).

(B2) Under the null hypothesis, y_t has a unit root as in (6.44), i.e.:

$$\Delta y_t = \sum_{j=1}^{p-1} \theta^*_{1,j}\, \Delta y_{t-j} + \epsilon_t, \qquad t = 1, \cdots, (T-p). \tag{6.103}$$

Let $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}^*_{1,1}, \ldots, \hat{\theta}^*_{1,(p-1)})'$ and $N(\hat{\mu}_\epsilon, \hat{\sigma}^2_\epsilon)$ be the estimation of $\boldsymbol{\theta}^* = (\theta^*_{1,1}, \ldots, \theta^*_{1,(p-1)})$ and $N(\mu_\epsilon, \sigma^2_\epsilon)$ which is the distribution of the errors ϵ_t in (6.103) imposing the null hypothesis.

(B3) Let ϵ^b_t be a random draw from $N(\hat{\mu}_\epsilon, \hat{\sigma}^2_\epsilon)$ and generate the bootstrap time series

$$y^b_t = y^b_{t-1} + \sum_{j=1}^{p-1} \hat{\theta}^*_{1,j}\, \Delta y^b_{t-j} + \epsilon^b_t, \qquad t = 1, \cdots, (T-p). \tag{6.104}$$

Initial values for the resampling can be set to sample values of the de-meaned series. The distribution of the series y^b_t is called the bootstrap series distribution of the data. The F_{nl} test statistic is calculated from the resampled series y_t^b as in item (B1).

(B4) Repeating this resampling operation J times yields the empirical distribution of F_{nl}^b, which is the bootstrap distribution of F_{nl}, completely determined by $\hat{\boldsymbol{\theta}}^*$ and $N(\hat{\mu}_\epsilon, \hat{\sigma}_\epsilon^2)$. For a large sample number of independent F_{nl}^b tests, estimated from J resampled series, the bootstrap p-value, defined by $p^b = P(F_{nl}^b > F_{nl})$ can be approximated by the frequency of simulated F_{nl}^b that exceeds the observed value of F_{nl}.

6.6.2 Approximation of Critical Values Assuming $\mathbf{\Pi} = \mathbf{I}_{(p-1)\times(p-1)}$

Finding a difficulty to obtain the asymptotic null distribution of the test statistic, Eklund (2003a) and Eklund (2003b) also suggested to obtain critical values by assuming the parameter in the null hypothesis equal zero. Under the null, his model is

$$\Delta y_t = \delta_1 \Delta y_{t-1} + \epsilon_t.$$

Assuming $\delta_1 = 0$ means that under the null, Δy_t are uncorrelated. Using the same argument, $\mathbf{\Pi} = \mathbf{I}$ means that under the null, Δy_t are uncorrelated as $(\rho_1, \ldots, \rho_{(p-2)}) = \mathbf{0}$. If $(\rho_1, \ldots, \rho_{(p-2)})$ are not far away from $\mathbf{0}$, these critical values would be good approximation of critical values for the asymptotic null distribution. As an example, assuming $\mathbf{\Pi} = \mathbf{I}_{2\times2}$, the critical values based on the asymptotic null distribution in (6.68) for k-ESTAR(3) models are tabulated in Table 6.9. Computer program for Table 6.9 is provided in Appendix A.3.2. We only consider k-ESTAR(3) models as for $p > 3$, we can follow the same procedure.

Table 6.9: Asymptotic critical values of F_{nl} test statistics for k-ESTAR(3) models with $\mathbf{\Pi} = \mathbf{I}_{2\times2}$.

	Significance Level		
	0.1	0.05	0.01
k=1	7.124863	8.758735	12.306371
k=2	17.82701	20.35429	25.65715
k=3	26.86799	29.96162	36.30965

Note: Simulations were based on samples size T=10,000 and 50,000 replications.

6.6.3 Monte Carlo Experiments

Monte Carlo experiments are conducted for k-ESTAR(3) models to compare the power of the F_{nl} test to detect non-linearity with other tests, i.e. F_{VPP}, AKSS and ADF. We only consider k-ESTAR(3) models as for $p > 3$, we can follow the same procedure.

Consider (6.39) for a k-ESTAR(3) model as follow,

$$\begin{aligned} y_t &= \theta_{1,0} + \sum_{j=1}^{3} \theta_{1,j}\, y_{t-1} + \sum_{j=1}^{2} \theta^*_{1,j}\, \Delta y_{t-j} \\ &+ \left[\theta_{2,0} + \sum_{j=1}^{3} \theta_{2,j}\, y_{t-1} + \sum_{j=1}^{2} \theta^*_{2,j}\, \Delta y_{t-j}\right] G(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \\ &\quad t = 1, 2, \ldots, T \end{aligned} \tag{6.105}$$

where $\theta^*_{i,j} = -\sum_{k=j+1}^{p} \theta_{i,k}$, $j = 1, 2$ and $i = 1, 2$.

Using the same assumptions for (6.40), (6.105) can be arranged to become

$$\Delta y_t = \sum_{j=1}^{2} \theta^*_{1,j}\, \Delta y_{t-j} + \left(\sum_{j=1}^{3} \theta_{2,j} y_{t-1} + \sum_{j=1}^{2} \theta^*_{2,j}\, \Delta y_{t-j} \right) G^*(\theta, \mathbf{e}, y_{t-d}) + \epsilon_t, \tag{6.106}$$

where $\theta^*_{i,1} = -(\theta_{i,2} + \theta_{i,3})$, $\theta^*_{i,2} = -\theta_{i,3}$, $i = 1, 2$.

Recalling the Taylor approximation for $G^*(\theta, \mathbf{e}, y_{t-d})$ around $\theta = 0$ in (6.42), (6.106) becomes,

$$\Delta y_t = \sum_{j=1}^{2} \theta^*_{1,j}\, \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\, y_{t-1}\, y_{t-d}^{s+2} + \sum_{s=0}^{2(k-1)} \sum_{j=1}^{2} \gamma_{2,sj}\, y_{t-d}^{s+2}\, \Delta y_{t-j} + \epsilon^*_t, \tag{6.107}$$

where $\epsilon^*_t = \epsilon_t + R$. If $\theta = 0$, y_t in (6.105) is linear in term of y_{t-1}, y_{t-2} and y_{t-3} and $\epsilon^*_t = \epsilon_t$ since the remainder $R \equiv 0$.

Testing the null hypothesis of a unit root ($H_0 : \theta = 0$) against alternative of a globally stationary k-ESTAR(3) model is equivalent to test,

$$H_0 : \gamma_{1,s} = \gamma_{2,sj} = 0, \qquad \text{for all } s \text{ and } j \ \text{ in (6.107)} .$$

The Size of Alternative Tests

In this simulation study, we want to know the probability of the propose unit root test F_{nl} to reject H_0 with a pre-set significance level if the true underlying series is a linear unit root AR(3) model. That probability can be defined as the size of alternative test. In this simulation study, we use 5% significance level. If the size of alternative test is around 5% or less, it means the test is good in detecting the true underlying series. We also compare the results with other tests, i.e. F_{VPP}, AKSS and ADF. To obtain the size of alternative tests, we generate the null model of k-ESTAR(3) models, i.e. :

$$\Delta y_t = \theta^*_{1,1}\, \Delta y_{t-1} + \theta^*_{1,2}\, \Delta y_{t-2} + \epsilon_t \tag{6.108}$$

where $\theta^*_{1,1} = -(\theta_{1,2} + \theta_{1,3})$, $\theta^*_{1,2} = -\theta_{1,3}$ and ϵ_t is drawn from the standard normal distribution. In particular, we choose a broad range of parameter values for $\theta^*_{1,1}$ and

$\theta^*_{1,2}$ so that Δy_t in (6.108) follows an AR(2) model. To fulfill the stationarity conditions of an AR(2) model, the parameters $\theta^*_{1,1}$ and $\theta^*_{1,2}$ should be: (i) $-1 < \theta^*_{1,2} < 1$, (ii) $\theta^*_{1,1} + \theta^*_{1,2} < 1$ and (iii) $\theta^*_{2,1} - \theta^*_{1,2} < 1$.

For computational purposes, the regression model in (6.108) becomes the restricted model for F_{nl} and F_{VPP}. The unrestricted model for F_{nl} is the regression models in (6.107) while the unrestricted model for F_{VPP} is:

$$\Delta y_t = \sum_{j=1}^{2} \theta^*_{1,j} \, \Delta y_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s} \, y_{t-1} \, y_{t-d}^{s+2} + \epsilon^*_t. \tag{6.109}$$

The bootstrap method is quite time consuming. Furthermore, when we try the F_{nl} tests for $k > 1$, sometimes they do not work due to singularity problem. It happens because under the null hypothesis, some nonlinear parts in (6.43) will be virtually zero. Therefore, for bootstrap method we only report for $k = 1$. For the second method in Section 6.6.2 assuming $\mathbf{\Pi} = \mathbf{I}_{2\times 2}$, the F_{nl} statistic is compared to the critical values in Table 6.9.

For the F_{VPP} test, as it does not depend on p, the critical values for k-ESTAR(3) models are the same as the critical values for k-ESTAR(2) models in Table 6.1. For the AKSS test and the ADF test, we include Δy_{t-1} and Δy_{t-2} to overcome the autocorrelation, so that the regression model for the AKSS test is

$$\Delta y_t = \delta_1 y_{t-1} y_{t-d}^2 + \delta_2 \Delta y_{t-1} + \delta_3 \Delta y_{t-2} + error \tag{6.110}$$

and the regression model for the ADF test is

$$\Delta y_t = \delta_1 y_{t-1} + \delta_2 \Delta y_{t-1} + \delta_3 \Delta y_{t-2} + error. \tag{6.111}$$

The hypothesis for the AKSS test and ADF test is

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0. \tag{6.112}$$

Then, the calculated t-test for δ_1 are compared with the critical values of the AKSS test and ADF test. The null hypothesis for the AKSS test and the ADF test conclude that y_t has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that y_t is a globally stationary 1-ESTAR(1) model while the ADF test concludes that y_t is a stationary linear ARMA model. The size of the alternative tests based on 5% significant level are presented in Table 6.10.

In Table 6.10, F^b_{nl} denotes the F_{nl} test statistic with $k = 1$ using the bootstrap method described in Section 6.6.1. The data is generated from (6.108) with $T = 250$. The rejection of the null hypothesis percentages are based on critical values from 500 bootstrap series and then the simulations are repeated by 500 independent replications. For the other F_{nl} test statistics are based on the second method. For

F_{nl}, F_{VPP}, AKSS and ADF tests, the data are generated from (6.108) with $T = 250$ and the rejection of the null hypothesis percentages are based on 10,000 independent replications. Computer program for Table 6.10is provided in Appendix A.3.2.

Similar to the size of alternative test for k-ESTAR(2) model, for all cases, the rejection probabilities for the AKSS are less than or around 5%. Therefore, the AKSS test has more power to detect the null hypothesis than other methods. It is followed by the ADF test with the rejection probabilities are close to or slightly higher than 5%. If we compare the results of F_{nl} tests using the bootstrap method in the second column and using the second method for $k = 1$ in the third column, generally the second method seems produce better results as its values are close to or slightly higher than 5%. Furthermore, its highest value is 5.39 for parameter values (-0.9,-0.9) while the highest value from the F_{nl}^{b} is 6.4 for parameter values (0,-0.7). Comparing the F_{nl} tests and the F_{VPP} tests results, generally F_{VPP} tests are better than the F_{nl} tests. It is not a surprise as the F_{VPP} tests involve less variables derived from the nonlinear term than the F_{nl} tests.

Table 6.10: The size of alternative tests (in percentage)

$(\theta^*_{1,1}, \theta^*_{1,2})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
(0,0)	5.1	4.89	5.11	5.21	5.14	5.00	4.34	4.66	4.97
(-0.2,-0.2)	5.4	4.94	5.12	5.12	5.0	3.96	3.61	4.53	4.98
(-0.5,-0.5)	5.2	5.17	4.67	4.77	4.25	3.03	3.07	3.75	5.00
(-0.7,-0.7)	5.6	5.19	4.85	4.97	3.76	2.59	2.50	3.33	4.98
(-0.9,-0.9)	4.9	5.39	5.73	5.83	3.23	2.12	2.23	2.80	5.03
(0.2,0.2)	4.8	4.85	5.29	5.74	5.35	6.01	5.42	4.88	4.93
(-0.3,0)	5.3	4.93	5.29	5.51	5.16	4.32	3.84	4.72	4.96
(-0.5,-0.2)	5.5	4.97	5.13	5.08	4.89	3.64	3.48	4.31	5.00
(-0.8,-0.5)	4.9	4.86	4.79	4.63	3.94	3.08	2.89	3.49	5.00
(-1,-0.7)	4.5	4.8	4.52	4.73	3.50	2.61	2.67	3.04	5.03
(-1.2,-0.9)	3.0	5.02	4.28	4.60	2.85	1.97	1.88	2.43	5.09
(-0.1,0.2)	5.6	4.84	5.29	5.58	5.30	5.31	4.54	4.77	4.99
(0.2,0.5)	4.8	5.11	5.82	7.50	5.54	6.47	6.85	4.95	4.72
(-0.7,0)	4.7	5.21	5.63	5.53	4.74	3.42	3.08	4.28	4.99
(-0.9,-0.2)	4.6	5.21	5.25	4.94	4.40	3.19	2.89	3.81	5.08
(-1.2,-0.5)	4.7	5.27	4.71	4.41	3.75	2.64	2.46	3.36	5.00
(-1.4,-0.7)	5.6	4.72	4.00	4.39	3.17	2.14	1.99	2.77	5.04
(-1.6,-0.9)	5.3	4.43	3.60	3.87	2.27	1.39	1.49	1.90	5.17
(-0.5,0.2)	4.7	5.05	6.01	6.10	5.08	4.14	3.67	4.69	5.02
(-0.2,0.5)	3.6	4.89	6.21	6.72	5.64	5.53	4.81	5.01	4.86
(0.3,0)	4.9	4.85	5.24	5.44	5.25	5.51	4.88	4.74	4.92
(0.1,-0.2)	5.3	4.90	5.12	5.15	5.12	4.39	4.20	4.58	4.92
(-0.2,-0.5)	6.2	5.18	4.83	4.92	4.59	3.33	3.03	4.10	4.98
(-0.4,-0.7)	5.7	5.13	4.74	4.91	4.03	2.94	2.61	3.54	5.00
(-0.6,-0.9)	5.1	4.95	4.77	5.05	3.09	2.16	2.08	2.69	5.06
(0.5,0.2)	4.7	5.05	5.60	6.83	5.39	6.36	6.78	4.82	4.84
(0.7,0)	4.6	5.0	5.72	6.71	5.48	6.00	6.48	4.83	4.89
(0.5,-0.2)	5.0	4.97	5.32	5.39	5.09	5.22	4.72	4.69	4.89
(0.2,-0.5)	5.9	5.01	4.97	5.01	4.61	3.84	3.48	4.17	4.91
(0,-0.7)	6.4	5.25	4.72	5.12	4.07	3.10	2.88	3.66	5.08
(-0.2,-0.9)	5.7	5.30	4.88	5.10	3.66	2.25	2.11	3.21	5.14

The Power of Alternative Tests

In this simulation study, we want to know the probability of the proposed unit root test F_{nl} to reject H_0 with a pre-set significance level if the true underlying series is a globally stationary nonlinear k-ESTAR(3) model. This probability can be defined as the power of alternative test. In this simulation study, we use 5% significance level. If the power of alternative test is almost 100%, it means the test is very good in detecting the true underlying series. We also compare the results with other tests, i.e. F_{VPP}, AKSS and ADF. To evaluate the power of tests against globally stationary k-ESTAR(3) model, samples from model in (6.105) are generated with ϵ_t is drawn from a standard normal distribution. The procedure to calculate the rejection probabilities of the null hypothesis (in percentage) given that the y_t is a k-ESTAR(3) model is the same as the procedure in obtaining the size of the alternative tests. The simulation results are summarised in Tables 6.11 and 6.12. Figures 6.3 and 6.4 show the plots of data from Tables 6.11 and 6.12, respectively. Computer program for Tables 6.11 and 6.12 is provided in Appendix A.3.2.

The data for Table 6.11 are simulated with $k = 1$, i.e. $e_1 = 0$. From Table 6.11 and Figure 6.3, comparing the F^b_{nl} test statistics and the F_{nl} test statistics for $k = 1$, we see that the two statistics have similar values. Thus, we conclude that the power of both methods are equal. The probabilities for the F_{nl} and F_{VPP} tests with $k = 1$ are the highest compared to other k. Apparently, with large sample, the tests are able to recognise the true number of equilibrium (for this case, $k = 1$). The F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta^*_{1,2}$ close to -1. For example, for $(\theta^*_{1,1}, \theta^*_{1,2}) = (-0.9, -0.9)$, $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$, and $k = 1$, the F_{nl} test can detect almost 30% while the F_{VPP}, the AKSS and the ADF are around 0.55%, 0.49% and 2.24% respectively.

Table 6.11: The power of alternative tests (in percentage), $\theta^2 = 0.01$, $e_1 = 0$, $d = 1$ and $T = 250$.

$(\theta^*_{1,1}, \theta^*_{1,2})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$									
(0,0)	99.8	99.62	84.53	66.23	100	96.28	83.20	100	100
(-0.2,-0.2)	90.2	89.30	43.57	29.19	97.08	53.93	30.67	96.67	98.47
(-0.5,-0.5)	53.2	49.89	18.51	13.84	64.17	13.82	8.97	62.49	60.04
(-0.7,-0.7)	38.4	36.03	14.60	11.54	39.27	77.75	5.52	37.75	32.75
(-0.9,-0.9)	40.0	41.61	19.21	14.72	20.42	4.19	3.18	19.31	20.27
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$									
(0,0)	98.6	97.59	69.38	50.82	99.45	76.92	49.55	99.33	99.89
(-0.2,-0.2)	74.0	71.97	29.25	20.12	82.13	20.37	10.19	80.60	84.98
(-0.5,-0.5)	35.8	32.46	13.18	10.68	29.49	4.34	3.36	27.70	26.16
(-0.7,-0.7)	25.0	24.84	11.12	9.29	11.93	2.26	2.24	11.17	12.50
(-0.9,-0.9)	34.6	35.90	17.47	14.18	2.70	1.25	1.46	2.51	5.30
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$									
(0,0)	70.0	67.00	35.98	26.09	60.26	10.16	5.21	58.22	63.75
(-0.2,-0.2)	31.6	30.98	14.09	10.75	15.20	2.94	2.37	13.97	13.56
(-0.5,-0.5)	15.0	16.72	8.27	6.80	3.90	1.41	1.63	3.65	5.71
(-0.7,-0.7)	14.6	15.59	8.06	7.26	2.04	1.02	1.34	1.91	4.19
(-0.9,-0.9)	26.8	27.77	14.46	12.67	0.55	0.87	1.34	0.49	2.24

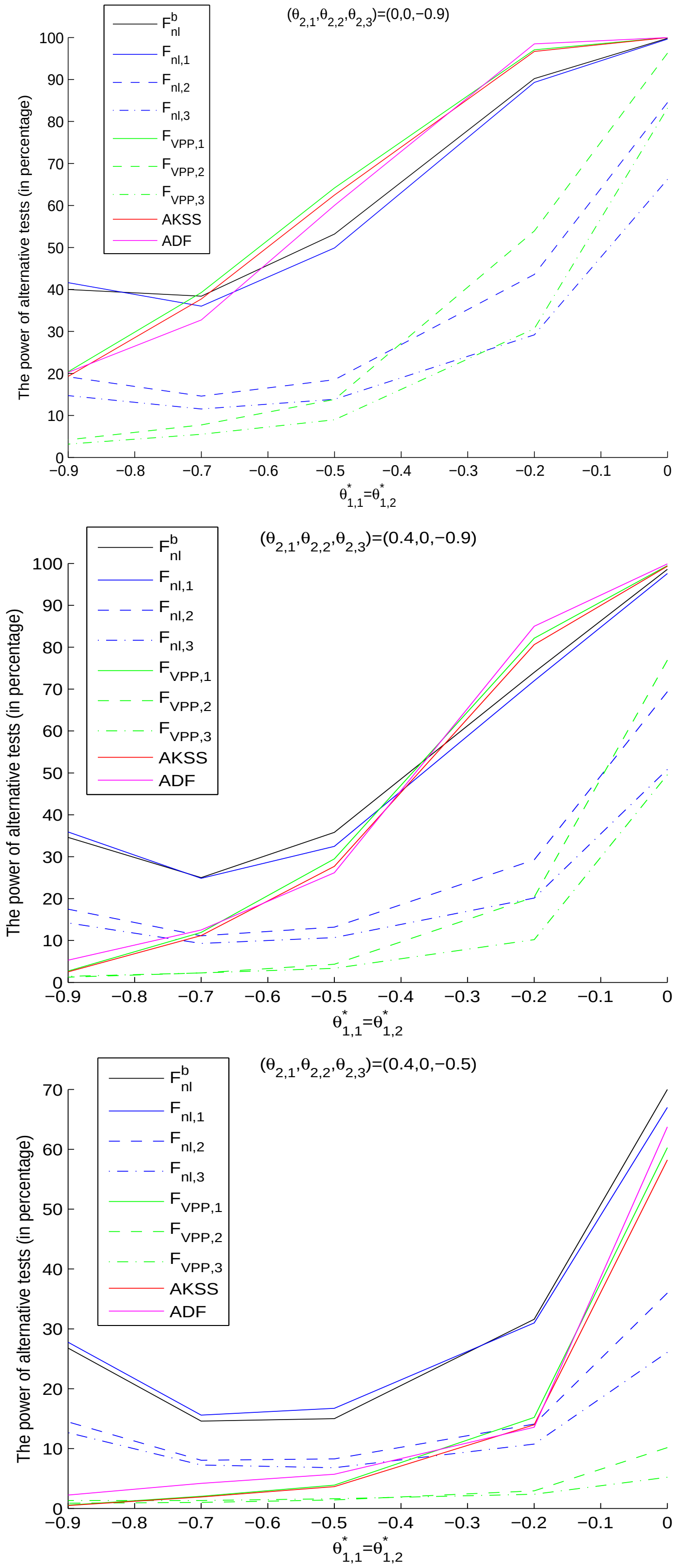


Figure 6.3: The power of alternative tests (in percentage), $\theta^2 = 0.01$, $e_1 = 0$, $d = 1$ and $T = 250$

Table 6.12: The power of alternative tests (in percentage), $\theta^2 = 0.01$, $e_1 = 0$, $e_2 = 3$, $d = 1$ and $T = 250$.

$(\theta^*_{1,1}, \theta^*_{1,2})$	F^b_{nl}	F_{nl}			F_{VPP}			AKSS	ADF
		k=1	k=2	k=3	k=1	k=2	k=3		
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$									
(0,0)	100	99.98	99.98	99.96	99.98	99.98	99.98	99.98	100
(-0.2,-0.2)	98.8	98.26	99.11	98.47	99.79	98.34	98.36	99.78	99.59
(-0.5,-0.5)	77.4	79.22	90.24	85.68	90.44	74.66	75.24	89.81	85.72
(-0.7,-0.7)	69.2	71.82	85.99	81.49	76.69	54.04	53.60	75.54	73.22
(-0.9,-0.9)	100	83.99	93.62	93.17	53.82	27.44	27.38	52.55	75.31
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$									
(0,0)	97.6	97.32	97.44	97.46	97.19	97.27	97.13	97.15	100
(-0.2,-0.2)	99.4	98.56	98.44	97.73	99.59	87.28	78.27	99.51	98.46
(-0.5,-0.5)	76.0	74.37	88.63	85.50	73.04	28.86	28.28	71.44	72.26
(-0.7,-0.7)	71.2	70.14	87.22	85.09	39.06	17.00	17.60	37.60	58.36
(-0.9,-0.9)	85.4	86.57	96.19	96.27	9.50	14.20	16.64	9.04	56.65
$(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$									
(0,0)	100	99.76	98.81	98.22	97.38	82.69	66.58	97.09	99.87
(-0.2,-0.2)	92.8	92.66	79.78	80.03	64.49	15.72	9.13	62.73	81.79
(-0.5,-0.5)	70.2	72.17	62.62	62.42	11.43	4.39	4.26	10.70	32.67
(-0.7,-0.7)	68.6	72.05	67.61	65.01	2.62	3.52	4.39	2.33	12.61
(-0.9,-0.9)	85.6	85.27	85.68	84.15	0.26	6.17	8.21	0.18	1.73

The data for Table 6.12 and Figure 6.4 are simulated with $k = 2$, i.e. $e_1 = 0$, $e_2 = 3$. Generally, the patterns are similar to $k = 1$. The results from the F^b_{nl} test and the F_{nl} test for $k = 1$, are still not much different. Generally, for $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0, 0, -0.9)$ and $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.9)$, the probabilities for $k = 2$ are the highest compared to other k when we use the F_{nl} tests while for $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$ the probabilities for $k = 1$ are the highest. The F_{VPP} tests still have the highest probabilities with $k = 1$ for all three combinations of $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3})$. Apparently, the F_{nl} tests are more able to recognise the true number of equilibriums (for this case, $k = 2$) compared to the F_{VPP} tests. The F_{nl} test shows more power to detect the alternative compared to the other methods when $\theta^*_{1,2}$ close to -1. For example, for $(\theta^*_{1,1}, \theta^*_{1,2}) = (-0.9, -0.9)$, $(\theta_{2,1}, \theta_{2,2}, \theta_{2,3}) = (0.4, 0, -0.5)$, and $k = 2$, the F_{nl} test can detect almost 85.7% while the F_{VPP}, the AKSS and the ADF are around 6.17%, 0.18% and 1.73% respectively.

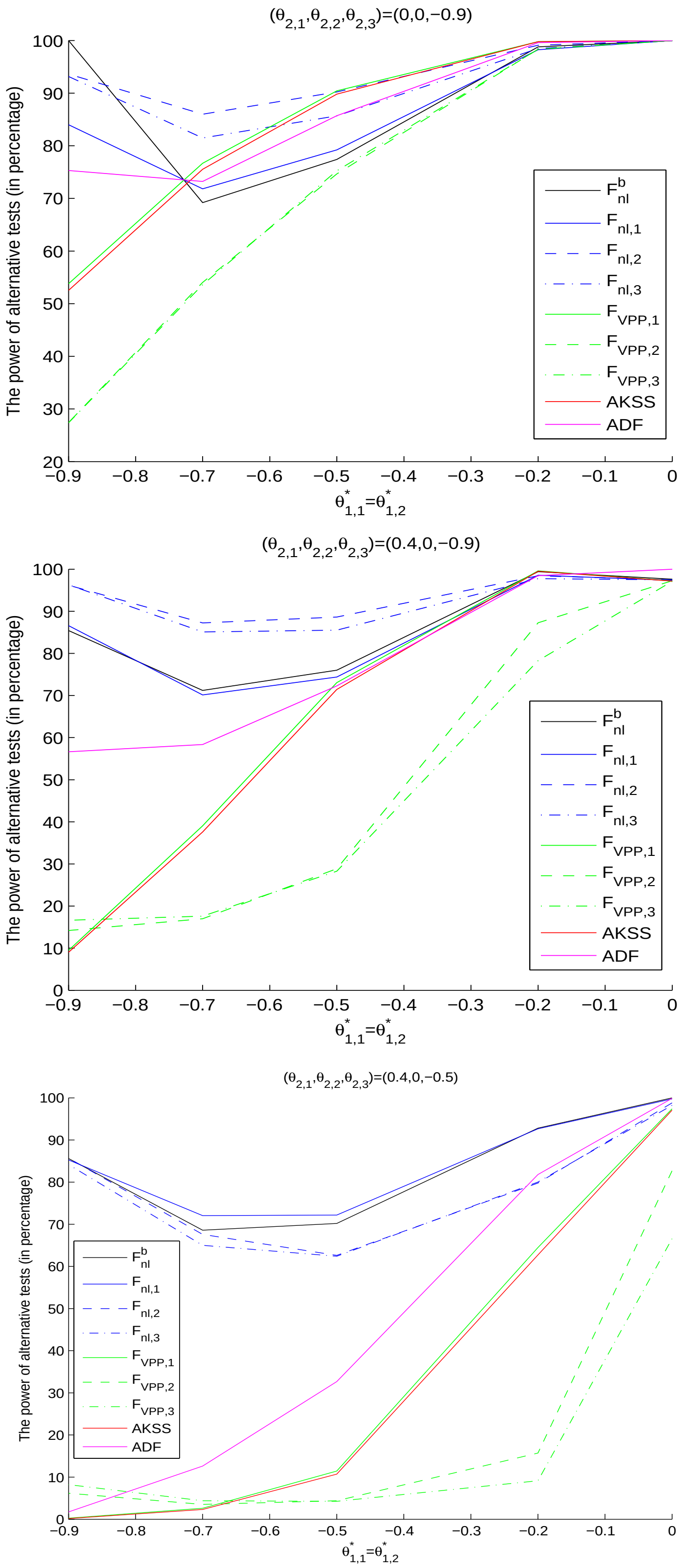


Figure 6.4: The power of alternative tests (in percentage), $\theta^2 = 0.01$, $e_1 = 0$, $e_2 = 3$, $d = 1$ and $T = 250$.

6.7 Conclusion

This chapter has extended the work of Kapetanios *et al.* (2003) and Venetis *et al.* (2009) by considering a unit root test for a k-ESTAR(p) model with a different approach. By using this approach, the singularity problem in Venetis *et al.* (2009) can be avoided without adding the collinear regressors into the error term. However, for a k-ESTAR(p) model, $p > 1$, a problem with nuisance parameters emerges. To solve the problem, we suggest 2 methods, i.e. a bootstrap method and critical values approximation method assuming there is no autocorrelation between Δy_t. From Monte Carlo simulations for k-ESTAR(3) models, the bootstrap method is time consuming and if the underlying series is actually a linear unit root AR(3) model (under the null hypothesis), it may result in a singularity problem. Therefore, we favour to critical values approximation method than the bootstrap method. For some cases, where the parameters are close to a unit root, simulation results show that our approach are better than the approach of Venetis *et al.* (2009), Kapetanios *et al.* (2003) and Dickey and Fuller (1979, 1981) in identifying the nonlinearity.

Chapter 7

Pairs Trading for an ESTAR Model Cointegration Errors

7.1 Introduction

In Chapter 5, empirical examples of pairs trading simulations with cointegration errors following an AR(1) model were presented. Even though no empirical data for pairs trading with cointegration errors following an ESTAR model were found, it is possible as some papers suggest, that some finance and economic data follow ESTAR models, see for examples Michael *et al.* (1997), Taylor *et al.* (2001), Paya *et al.* (2003), Terasvirta and Elliasson (2001), Sarno *et al.* (2002) and Monoyios and Sarno (2002). Monoyios and Sarno (2002) found the basis which is the difference between the log of future index price and the log of spot index price follows an ESTAR model. In other words, the log of future index price and the log of spot index price are cointegrated with cointegration errors follow an ESTAR model. If this is correct, we could make pairs trading strategy between future index and spot index. However, as we present in the next chapter, using currently available data for the S&P 500 future index and spot index prices, we cannot confirm that they are cointegrated with cointegration errors follow an ESTAR model. Therefore, in this chapter we use simulation data to perform pairs trading.

In Chapter 4, we presented the using integral equation approach to determine trading duration and inter-trade interval. Then, it was used to estimate number of trades and the optimal threshold for pairs trading for two cointegrated assets with cointegration errors assumed to follow an AR(1) model and an AR(2) model. Now, the question is whether the integral equation approach can also be used to determine trading duration and inter-trade interval for an ESTAR model? To answer this question, Section 7.2 will discuss the mean of first-passage time for an ESTAR model using the integral equation approach. Sections 7.3 and 7.4 will discuss trading duration, inter-trade interval and number of trades for a 1-ESTAR(1) model and a 1-ESTAR(2) model respectively. Section 7.5 presents pairs trading simulations with simulation data generated from an ESTAR model and the last section concludes this chapter.

7.2 The Mean First-passage Time for an ESTAR Model Using Integral Equation Approach

Consider that ϵ_t follows a k-ESTAR(p) model as in (6.1) with $\theta_{1,0} = \theta_{2,0} = 0$ so that the mean of $\epsilon_t = 0$, i.e.:

$$\epsilon_t = \sum_{j=1}^{p} \theta_{1,j}\ \epsilon_{t-j} + \left[\sum_{j=1}^{p} \theta_{2,j}\ \epsilon_{t-j}\right] G(\theta, \mathbf{e}, \epsilon_{t-d}) + \eta_t \tag{7.1}$$

where

$$G(\theta, \mathbf{e}, \epsilon_{t-d}) = 1 - \exp\left[-\theta^2 \left(\prod_{i=1}^{k}(\epsilon_{t-d} - e_i)\right)^2\right], \tag{7.2}$$

$\eta_t \sim$ IID $N(0, \sigma_\eta^2)$, $p > d \geq 1$ and $\mathbf{e} = (e_1, e_2, \ldots, e_k)'$.

We assume that the parameters in (7.1), i.e. $\theta_{1,j}$ and $\theta_{2,j}$, $j = 1, \ldots, p$, fulfill the globally stationary conditions of a k-ESTAR(p) model. We also assume that there is no correlation between η_t and $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_0$.

Consider again the mean first-passage time using integral equation approach for $\{y_t\}$ following an AR(p) model in Section 4.2, especially equation (4.11), i.e.:

$$E(\mathcal{T}_{a,b}(\mathbf{y}_0)) = \int_a^b E(\mathcal{T}_{a,b}(\mathbf{y}_1)) f(u|\mathbf{y}_0) du + 1 \tag{7.3}$$

where

$$f(u|\mathbf{y}_0) = \frac{1}{\sqrt{2\pi}\sigma_\xi} \exp\left(-\frac{(u - \boldsymbol{\phi}'\ \mathbf{y}_0)^2}{2\sigma_\xi^2}\right), \tag{7.4}$$

$$y_t = \phi_1\ y_{t-1} + \cdots + \phi_p\ y_{t-p} + \xi_t, \tag{7.5}$$

$\xi_t \sim$ IID $N(0, \sigma_\xi^2)$, $\mathbf{y}_0 = (y_{-p+1}, \ldots, y_{-1}, y_0)'$, $\mathbf{y}_1 = (y_{-p+2}, \ldots, y_0, u)'$ and $\boldsymbol{\phi} = (\phi_p, \ldots, \phi_1)'$.

$f(u|\mathbf{y}_0)$ in (7.4) denotes the probability of $y_1 = u$ given $\mathbf{y}_0 = (y_{-p+1}, \ldots, y_{-1}, y_0)'$. In general, $f(y_t = u|\mathbf{y}_{t-1})$ denotes the probability of $y_t = u$ given $\mathbf{y}_{t-1} = (y_{t-p}, \ldots, y_{t-2}, y_{t-1})'$. For an AR(p) model in (7.5), as ξ_t has identical distribution for all $t = 1, \ldots, T$, $f(y_t|\mathbf{y}_{t-1}) \sim f(\xi_t) \sim N(0, \sigma_\xi^2)$. Now, we are going to show that for ϵ_t in (7.1), $f(\epsilon_t|\boldsymbol{\epsilon}_{t-1}) \sim f(\eta_t) \sim N(0, \sigma_\eta^2)$, where $\boldsymbol{\epsilon}_{t-1} = (\epsilon_{t-p}, \ldots, \epsilon_{t-2}, \epsilon_{t-1})'$.

As we assume that there is no correlation between η_t and $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_0$, the joint probability density function of $(\eta_t, \epsilon_{t-1}, \ldots, \epsilon_{t-p})$ is

$$f(\eta_t, \epsilon_{t-1}, \ldots, \epsilon_{t-p}) = f(\eta_t) f(\epsilon_{t-1}, \ldots, \epsilon_{t-p}). \tag{7.6}$$

To get $f(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-p})$, we need to obtain the determinant of Jacobian transformation. Consider

$$\eta_t = \epsilon_t - \sum_{j=1}^{p} \theta_{1,j}\ \epsilon_{t-j} - \left[\sum_{j=1}^{p} \theta_{2,j}\ \epsilon_{t-j}\right] G(\theta, \mathbf{e}, \epsilon_{t-d}).$$

Then, the determinant of Jacobian transformation is

$$
\begin{aligned}
|J| &= \begin{vmatrix} \partial\eta_t/\partial\epsilon_t & \partial\eta_t/\partial\epsilon_{t-1} & \partial\eta_t/\partial\epsilon_{t-2} & \dots & \partial\eta_t/\partial\epsilon_{t-p} \\ \partial\epsilon_{t-1}/\partial\epsilon_t & \partial\epsilon_{t-1}/\partial\epsilon_{t-1} & \epsilon_{t-1}/\partial\epsilon_{t-2} & \dots & \epsilon_{t-1}/\partial\epsilon_{t-p} \\ \partial\epsilon_{t-2}/\partial\epsilon_t & \partial\epsilon_{t-2}/\partial\epsilon_{t-1} & \epsilon_{t-2}/\partial\epsilon_{t-2} & \dots & \epsilon_{t-2}/\partial\epsilon_{t-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \partial\epsilon_{t-p}/\partial\epsilon_t & \partial\epsilon_{t-p}/\partial\epsilon_{t-1} & \epsilon_{t-p}/\partial\epsilon_{t-2} & \dots & \epsilon_{t-p}/\partial\epsilon_{t-p} \end{vmatrix} \\
&= \begin{vmatrix} 1 & \partial\eta_t/\partial\epsilon_{t-1} & \partial\eta_t/\partial\epsilon_{t-2} & \dots & \partial\eta_t/\partial\epsilon_{t-p} \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix} = 1. \qquad (7.7)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& f(\epsilon_t|\epsilon_{t-1},\dots,\epsilon_{t-p}) \qquad (7.8) \\
&= \frac{f(\epsilon_t,\epsilon_{t-1},\dots,\epsilon_{t-p})}{f(\epsilon_{t-1},\dots,\epsilon_{t-p})} = \frac{f(\eta_t)f(\epsilon_{t-1},\dots,\epsilon_{t-p})|J|}{f(\epsilon_{t-1},\dots,\epsilon_{t-p})} = f(\eta_t) \\
&= \frac{1}{\sqrt{2\pi}\sigma_\eta}\exp\left\{-\frac{\left[\epsilon_t - \sum_{j=1}^{p}\theta_{1,j}\ \epsilon_{t-j} - \left(\sum_{j=1}^{p}\theta_{2,j}\ \epsilon_{t-j}\right)G(\theta,\mathbf{e},\epsilon_{t-d})\right]^2}{2\sigma_\eta^2}\right\} \qquad (7.9)
\end{aligned}
$$

Thus, the mean first-passage time using integral equation approach for ϵ_t following an k-ESTAR(p) model is similar to (7.3), i.e.:

$$E(\mathcal{T}_{a,b}(\boldsymbol{\epsilon}_0)) = \int_a^b E(\mathcal{T}_{a,b}(\boldsymbol{\epsilon}_1))f(u|\boldsymbol{\epsilon}_0)\ du + 1 \qquad (7.10)$$

where $f(u|\boldsymbol{\epsilon}_0) = f(\epsilon_1 = u|\boldsymbol{\epsilon}_0)$, $\boldsymbol{\epsilon}_0 = (\epsilon_{-p+1},\dots,\epsilon_{-1},\epsilon_0)'$, $\boldsymbol{\epsilon}_1 = (\epsilon_{-p+2},\dots,\epsilon_0,u)'$. Since η_t has identical distribution for all $t = 1,\dots,T$, $f(\epsilon_1 = u|\boldsymbol{\epsilon}_0)$ will has the distribution as in (7.9). The same as for an AR(p) model, through iterating process p times, we will get a Fredholm integral equation of order p and we can calculate $E(\mathcal{T}_{a,b}(\boldsymbol{\epsilon}_0))$ for any given a, b, and an initial state vector $\boldsymbol{\epsilon}_0 \in [a,b]^p$.

7.3 Pairs Trading for a 1-ESTAR(1) Model

Consider ϵ_t follow a 1-ESTAR(1) model as in (7.1) with the equilibrium $e_1 = 0$ and $d = 1$, i.e.

$$\epsilon_t = \left[\theta_1 + \theta_2\left(1 - \exp(-\theta^2\epsilon_{t-1}^2)\right)\right]\ \epsilon_{t-1} + \eta_t \qquad (7.11)$$

Let TD_U denote the expected trade duration corresponding to the pre-set upper-bound U. Using (7.10) and (7.11), TD_U is defined as follows:

$$\begin{aligned} TD_U &:= E(\mathcal{T}_{0,\infty}(U)) \\ &= \lim_{b\to\infty} \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_0^b E(\mathcal{T}_{0,b}(s)) \exp\left\{ -\frac{\left[s-(\theta_1+\theta_2(1-\exp(-\theta U^2)))U\right]^2}{2\sigma_\eta^2} \right\} ds \\ &\quad + 1. \end{aligned} \tag{7.12}$$

Let IT_U denote the expected inter-trade interval for the pre-set upper-bound U.

$$\begin{aligned} IT_U &:= E(\mathcal{T}_{-\infty,U}(0)) \\ &= \lim_{b\to-\infty} \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_b^U E(\mathcal{T}_{-b,U}(s)) \exp\left(-\frac{s^2}{2\sigma_\eta^2} \right) ds + 1. \end{aligned} \tag{7.13}$$

As a 1-ESTAR(1) model is globally stationary and symmetric, the number of trades expectation during a trading period will be the same as the number of trades expectation for an AR(p) model as in (4.28), except that TD_U and IT_U are modified for a k-ESTAR(p). Thus,

$$\frac{T}{TD_U + IT_U} + 1 \geq E(N_{UP}) + 1 \geq E(N_{UT}) \geq E(N_{UP}) > \frac{T}{TD_U + IT_U} - 1. \tag{7.14}$$

where T is number of observation during the trading period, $E(N_{UT})$ denotes the expected number of upper trades, $E(N_{UP})$ denotes the expected number of periods for upper trades, TD_U and IT_U are defined as in (7.12) and (7.13), respectively.

Table 7.1 shows the comparison between simulations and integral equation approach in estimating trading duration, inter-trade interval and number of trades for ϵ_t following a 1-ESTAR(1) model in (7.11) with $e_1 = 0$. Equation (7.11) is used to generate ϵ_t for simulations with some combination parameters. 50 independent simulations are carried out and each simulation has 1000 observations. We use a pre-set upper-bound $U = 1.5$ for the simulations. Under certain assumptions pairs trading strategy as mentioned in 2.3.4, if $\epsilon_t \geq U$, we will open a pair trade, and then close the trade if $\epsilon_t \leq 0$. If there is an up-crossing of upper-bound U, but the previous open trade has not been closed, we can not open another pair trade. We calculate the average of trade duration and inter-trade interval and number of trades for each simulation and then we calculate again the average from the 50 simulations. The simulation results are compared to the values from integral equation approach. Numerical scheme to calculate the mean first-passage time based on integral equation approach in Subsection 4.3.1 is used to calculate trade duration in (7.12) and inter-trade interval in (7.13). Then, number of trades is calculated by

$$NT_U = \frac{T}{TD_U + IT_U} - 1, \tag{7.15}$$

where $T = 1000$, TD_U and IT_U are the estimates of trading duration and inter-trade interval using integral equation approach with $U = 1.5$. Computer program for Table 7.1

is provided in Appendix A.4.1.

Table 7.1: Trade duration, inter-trade interval and number of trades from simulations and using integral equation approach for a 1-ESTAR(1) model with pre-set upper-bound $U = 1.5$ and $\theta^2 = 0.01$.

σ_η	Trade Duration		Inter-trade Interval		Number of Trades	
	IE	Sim	IE	Sim	IE	Sim
$\theta_1 = -\theta_2 = -0.5$						
0.75	1.27	1.19(0.09)	25.45	22.83(4.30)	36.42	41.84(6.93)
1	1.37	1.26(0.06)	10.68	9.54(0.97)	81.91	92.18(8.12)
2	1.57	1.43(0.05)	3.68	3.13(0.21)	189.22	218.70(10.72)
$\theta_1 = -\theta_2 = 0.5$						
0.75	3.78	3.99(0.52)	31.87	31.20(5.02)	27.05	28.14(3.77)
1	3.48	3.75(0.36)	14.56	15.42(1.55)	54.45	51.72(3.97)
2	2.92	3.18(0.25)	5.40	5.93(0.52)	119.03	109.44(6.59)
$\theta_1 = 1, \theta_2 = -0.2$						
0.75	27.23	29.03(6.60)	37.66	45.88(20.12)	14.41	13.46(2.96)
1	19.82	24.00(5.93)	26.12	30.49(10.76)	20.77	18.26(3.34)
2	10.07	13.19(2.70)	12.09	15.22(3.76)	44.12	35.24(5.25)
$\theta_1 = 1, \theta_2 = -0.5$						
0.75	20.61	24.14(4.92)	31.17	33.73(13.46)	18.31	17.00(3.40)
1	15.06	17.88(3.34)	21.42	26.07(7.99)	26.41	22.76(3.83)
2	7.66	9.49(1.21)	9.70	12.21(1.99)	56.58	45.80(4.85)
$\theta_1 = 1, \theta_2 = -0.8$						
0.75	17.77	19.74(4.49)	28.44	34.23(10.33)	20.64	18.56(3.94)
1	13.03	14.93(2.93)	19.43	22.74(4.83)	29.81	26.34(4.10)
2	6.64	8.09(1.16)	8.68	11.21(1.92)	64.28	52.08(4.38)

Note: IE denotes integral equation; Sim denotes simulation; values in the parentheses are the standard deviations.

From Table 7.1, generally, the values from integral equation approach are quite close to the values from simulation as they are still within ± 2 standard deviation of the simulation values. The exceptions are for the number of trades from $\theta_1 = -\theta_2 = -0.5, \sigma_\eta = 2$ and $\theta_1 = 1, \theta_2 = -0.8, \sigma_\eta = 2$. The values from integral equation approach are significantly outside ± 2 standard deviation of the value simulation. The differences are due to a slight difference in the framework underpinning the theory of integral equations and that for simulation from the real data as by using the real data we usually open a trade at slightly higher than U, not exactly at U.

Consider again the pair trading strategy using the cointegration coefficients weighted (CCW) rule in Section 2.3.4. For two cointegrated assets $S1$ and $S2$ whose prices P_{S1} and P_{S2}, a cointegration relationship can be constructed as follows:

$$P_{S1,t} - \beta\ P_{S2,t} = \epsilon_t, \qquad \beta > 0, \tag{7.16}$$

where ϵ_t is assumed to be a stationary zero mean series and it is called cointegration errors. However, for this time, we assume that the stationarity is nonlinear following a k-ESTAR(p) model.

As a 1-ESTAR(1) model is symmetric, the lower-bound L can be set as $L = -U$, where U is the upper-bound. Furthermore, as a k-ESTAR(p) model is still stationary, the derivation of minimum profit per trade using the CCW rule will be the same as in 7.16. Thus, for a given upper-bound U, the minimum profit per trade is U.

Using the facts that profit per trade $\geq U$ and in (7.15), $E(N_{UT}) \geq \frac{T}{TD_U+IT_U} - 1$, the minimum total profit within the time horizon [0,T] will be defined as follow:

$$MTP(U) := \left(\frac{T}{TD_U + IT_U} - 1 \right) U. \tag{7.17}$$

where TD_U and IT_U are defined as in (7.12) and (7.13), respectively. The numerical algorithm in Section 4.3.4 using the MTP in (7.17) can be used to calculate the optimal pre-set upper-bound.

7.4 Pairs Trading for a 1-ESTAR(2) Model

Consider a 1-ESTAR(2) process for ϵ_t with equilibrium $e_1 = 0$ as follows:

$$\epsilon_t = \theta_{1,1}\ \epsilon_{t-1} + \theta_{1,2}\ \epsilon_{t-2} + (\theta_{2,1}\ \epsilon_{t-1} + \theta_{2,2}\ \epsilon_{t-2}) \left(1 - \exp\left(-\theta^2 \epsilon_{t-d}^2\right)\right) + \eta_t, \tag{7.18}$$

where $\eta_t \sim$ IID $N(0, \sigma_\eta^2)$ and $1 \leq d \leq 2$.

Following the mean first-passage time for an AR(2) model in Subsection 4.4.1, the mean first-passage time for a 1-ESTAR(2) model is:

$$\begin{aligned} E\left(\mathcal{T}_{a,b}(\epsilon_0, \epsilon_{-1})\right) &= \int_a^b \int_a^b E\left(\mathcal{T}_{a,b}(u_1, u_2)\right) K(\epsilon_0, \epsilon_{-1}, u_1, u_2) du_1 du_2 \\ &\quad +1 + \int_a^b \psi(\epsilon_0, \epsilon_{-1}, u_1) du_1 \end{aligned} \tag{7.19}$$

where for $d = 1$,

$$\begin{aligned} &K(\epsilon_0, \epsilon_{-1}, u_1, u_2) = \\ &\quad \frac{1}{2\pi\sigma_\eta^2} \exp\left(-\frac{\left(u_1 - \theta_{1,1}\ \epsilon_0 - \theta_{1,2}\ \epsilon_{-1} - (\theta_{2,1}\ \epsilon_0 + \theta_{2,2}\ \epsilon_{-1})(1 - \exp(-\theta^2 \epsilon_0^2))\right)^2}{2\sigma_\eta^2} \right) \times \\ &\qquad \exp\left(-\frac{\left(u_2 - \theta_{1,1}\ u_1 - \theta_{1,2}\ \epsilon_0 - (\theta_{2,1}\ u_1 + \theta_{2,2}\ \epsilon_0)(1 - \exp(-\theta^2 u_1^2))\right)^2}{2\sigma_\eta^2} \right), \end{aligned} \tag{7.20}$$

$$\begin{aligned} &\psi(\epsilon_0, \epsilon_{-1}, u_1) = \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{\left(u_1 - \theta_{1,1}\ \epsilon_0 - \theta_{1,2}\ \epsilon_{-1} - (\theta_{2,1}\ \epsilon_0 + \theta_{2,2}\ \epsilon_{-1})(1 - \exp(-\theta^2 \epsilon_0^2))\right)^2}{2\sigma_\eta^2} \right), \end{aligned} \tag{7.21}$$

for $d = 2$,

$$
\begin{aligned}
&K(\epsilon_0, \epsilon_{-1}, u_1, u_2) = \\
&\quad \frac{1}{2\pi\sigma_\eta^2} \exp\left(-\frac{\left(u_1 - \theta_{1,1}\ \epsilon_0 - \theta_{1,2}\ \epsilon_{-1} - (\theta_{2,1}\ \epsilon_0 + \theta_{2,2}\ \epsilon_{-1})(1 - \exp(-\theta^2 \epsilon_{-1}^2))\right)^2}{2\sigma_\eta^2} \right) \times \\
&\qquad\qquad \exp\left(-\frac{\left(u_2 - \theta_{1,1}\ u_1 - \theta_{1,2}\ \epsilon_0 - (\theta_{2,1}\ u_1 + \theta_{2,2}\ \epsilon_0)(1 - \exp(-\theta^2 \epsilon_0^2))\right)^2}{2\sigma_\eta^2} \right), \qquad (7.22) \\
&\psi(\epsilon_0, \epsilon_{-1}, u_1) = \\
&\quad \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{\left(u_1 - \theta_{1,1}\ \epsilon_0 - \theta_{1,2}\ \epsilon_{-1} - (\theta_{2,1}\ \epsilon_0 + \theta_{2,2}\ \epsilon_{-1})(1 - \exp(-\theta^2 \epsilon_{-1}^2))\right)^2}{2\sigma_\eta^2} \right). \qquad (7.23)
\end{aligned}
$$

Following numerical procedures for an AR(2) model in Subsection 4.4.2, (7.19) can be approximated by:

$$E_n(\mathcal{T}_{a,b}(\epsilon_0, \epsilon_{-1})) \approx \sum_{j=1}^{m} \omega_j K(\epsilon_0, \epsilon_{-1}, u_{1j}, u_{2j}) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) + 1 + \sum_{j=1}^{m} \omega_j^* \psi(\epsilon_0, \epsilon_{-1}, u_{1j}), \tag{7.24}$$

where $\epsilon_0 \in [a, b]$, n is number partition in $[a, b]$, $m = 6n^2 + 4n + 1$, ω_j is the weight of the jth node for the double integral, ω_j^* is the weight of the jth node for the single integral.

We can then evaluate the values of $E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j}))$ for $j = 1, \ldots, m$ by solving the following system of linear equations:

$$\left(I_{m\times m} - \sum_{j=1}^{m} \omega_j K(u_{1i}, u_{2i}, u_{1j}, u_{2j}) \right) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) = \psi^*(u_{1i}, u_{2i}) \tag{7.25}$$

where $I_{m\times m}$ is an identity matrix with $m \times m$ dimension , $i = j = 1, \ldots, m$ and

$$\psi^*(u_{1i}, u_{2i}) = 1 + \sum_{j=1}^{m} \omega_j^* \psi(u_{1i}, u_{2i}, u_{1j}).$$

Finally, we can use the Nystrom interpolation formula to obtain the value of $E_n(\mathcal{T}_{a,b}(\epsilon_0, \epsilon_{-1}))$, i.e. :

$$E_n(\mathcal{T}_{a,b}(\epsilon_0, \epsilon_{-1})) \approx \sum_{j=1}^{m} \omega_j K(\epsilon_0, \epsilon_{-1}, u_{1j}, u_{2j}) E_n(\mathcal{T}_{a,b}(u_{1j}, u_{2j})) + \psi^*(\epsilon_0, \epsilon_{-1}). \tag{7.26}$$

Following the mean first-passage time to calculate trade duration in (4.44) and inter-trade interval in (4.45) for an AR(2) process, we define the trade duration TD_U and inter-trade interval IT_U for a k-ESTAR(2) model with an upper-bound$U$ are as follow:

$$\begin{aligned} TD_U &:= E\left(\mathcal{T}_{0,\infty}(0,U)\right) \\ &= \lim_{b\to\infty}\int_0^b\int_0^b K(0,U,u_1,u_2)\ du_1\ du_2 + 1 + \lim_{b\to\infty}\int_0^b \psi(0,U,u_1)\ du_1. \end{aligned} \tag{7.27}$$

$$\begin{aligned} IT_U &:= E\left(\mathcal{T}_{-\infty,U}(0,0)\right) \\ &= \lim_{b\to-\infty}\int_b^U\int_b^U K(0,0,u_1,u_2)\ du_1\ du_2 + 1 + \lim_{b\to-\infty}\int_b^U \psi(0,0,u_1)\ du_1. \end{aligned} \tag{7.28}$$

Table 7.2: Trade duration, inter-trade interval and number of trades from simulations and using integral equation approach for a 1-ESTAR(2) model with pre-set upper-bound $U = 1.5$, $\theta_{1,1} = 0.1$, $\theta_{1,2} = 0.9$ and $\theta^2 = 0.01$.

σ_η^2	Trade Duration		Inter-trade Interval		Number of Trades	
	IE	Sim	IE	Sim	IE	Sim
$\theta_{21} = 0.4$, $\theta_{22} = -0.5$						
1	11.3852	19.9154(14.4161)	15.7117	22.0735(23.1014)	35.9046	23.8800(8.7216)
2	16.5187	16.3732(10.1506)	19.3566	24.5963(47.8173)	26.8743	31.6400(13.3764)
$\theta_{21} = 0.4$, $\theta_{22} = -0.9$						
1	11.7921	10.7229(4.3761)	16.9153	14.0765(6.5533)	33.8342	39.9800(11.8537)
2	12.9790	7.5257(2.5878)	18.5036	11.0346(5.4537)	30.7635	55.3600(13.7615)
$\theta_{21} = 0$, $\theta_{22} = -0.9$						
1	10.9758	6.6337(1.8126)	16.8814	10.9044(4.2987)	34.8973	58.0400(15.2409)
2	8.8289	4.9411(1.0477)	13.5005	7.0751(2.3332)	43.7840	84.2400(15.9828)

Note: IE denotes integral equation; Sim denotes simulation; values in the parentheses are the standard deviations.

Table 7.2 shows the comparison between simulations and integral equation approach in estimating trading duration, inter-trade interval and number of trades for ϵ_t following a 1-ESTAR(2) model in (7.18). 50 independent simulations are carried out and each simulation has 1000 observations. We use a pre-set upper-bound $U = 1.5$ for the simulations. Under certain assumptions pairs trading strategy as mentioned in 2.3.4, if $\epsilon_t \geq U$, we will open a pair trade, and then close the trade if $\epsilon_t \leq 0$. If there is an up-crossing of upper-bound U, but the previous open trade has not been closed, we can not open another pair trade. We calculate the average of trade duration and inter-trade interval and number of trades for each simulation and then we calculate again the average from the 50 simulations. The simulation results are compared to the values from integral equation approach. Numerical scheme to calculate the mean first-passage time based on integral equation approach for an AR(2) model in Subsection 4.4.2 is used to calculate trade duration in (7.27) and inter-trade interval in (7.28). Then, number of trades is calculated by

$$NT_U = \frac{T}{TD_U + IT_U} - 1 \tag{7.29}$$

where $T = 1000$, TD_U and IT_U are the estimates of trading duration and inter-trade interval using integral equation approach with $U = 1.5$. Computer program for Table 7.2 is provided in Appendix A.4.2.

If we compare integral equation approach and simulation results in Table 7.2 briefly, there are significant differences between the two. However, as the simulations generally have high standard deviations, the integral equation results are generally within or closely to the range ± 2 standard deviation of the simulation values.

7.5 Pairs Trading Simulation

As we cannot find empirical data for pairs trading with cointegration error follows a 1-ESTAR(2) model, simulations have been done. Figure 7.1(a) shows the plots of $S1$ and $S2$ stock prices, denoted as P_{S1} and P_{S2}. These series are generated as follow:

$$\begin{aligned}
\epsilon_t &= 0.1\epsilon_{t-1} + 0.9\epsilon_{t-2} + (0.4\epsilon_{t-1} - 0.5\epsilon_{t-2})\left(1 - \exp(-0.01\epsilon_{t-1}^2)\right) + \eta_t, \\
\eta_t &\sim N(0,1), \qquad \epsilon_1 = \epsilon_2 = 0, \\
P_{S2,t} &= P_{S2,(t-1)} + \epsilon_t^*, \qquad \epsilon_t^* \sim N(0,1), \qquad P_{S2,1} = 3, \qquad P_{S2,2} = 4 \\
P_{S1,t} &= P_{S2,t} + \epsilon_t, \qquad P_{S1,1} = 5, \qquad P_{S1,2} = 4, \\
t &= 1, 2, \ldots, 310.
\end{aligned} \tag{7.30}$$

The first ten observations of $P_{S1,t}$ and $P_{S2,t}$ from (7.30) are discharged, so that we have 300 observations used for the pairs trading simulation and consider these series as in-sample data. Note that the series $P_{S2,t}$ is a random walk which is not stationary, while the series $P_{S1,t}$ is generated as a function of $P_{S2,t}$ and ϵ_t.

Pretending that the series $P_{S1,t}$ and $P_{S2,t}$ are the stock prices and we do not know the underlying distribution generating these series, we would like to analyse the relationship between $P_{S1,t}$ and $P_{S2,t}$, whether they are cointegrated or not. Following Engle and Granger (1987), the series $P_{S1,t}$ is regressed with the series $P_{S2,t}$ and then test the residuals series ($\hat{\epsilon}$) whether it is stationary or not, i.e.

$$P_{S1,t} = c + \beta\ P_{S2,t} + \epsilon_t \tag{7.31}$$

where c and β are constants and ϵ_t is the error term. Table 7.3 reports the regression results of (7.31). From Table 7.3, we can see that both constants c and β are significantly different from zero and the β estimate is around 1.34.

The plot of residuals from regression in (7.31), i.e. $\hat{\epsilon}$, is shown in Figure 7.1(b), while

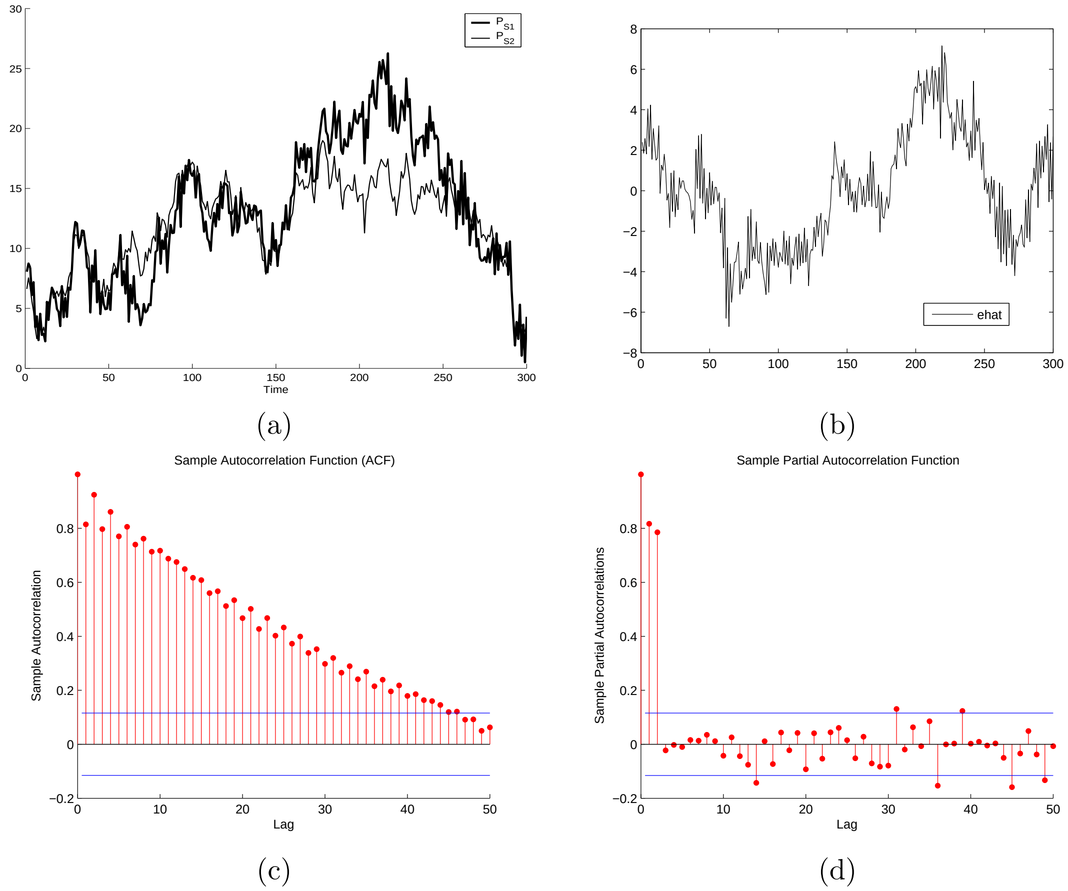


Figure 7.1: Plots of in-sample data: (a) Plots of $P_{S1,t}$ and $P_{S2,t}$ prices; (b) Plot of residuals (cointegration error), denoted as $\hat{\epsilon}$; (c) ACF plot of $\hat{\epsilon}$; (d)PACF plot of $\hat{\epsilon}$.

Table 7.3: Regression results between $P_{S1,t}$ and $P_{S2,t}$

Coefficients	Estimate	Std. Error	t value	Pr(> $\|t\|$)
c	-3.20683	0.52964	-6.055	4.23e-09 ***
β	1.33912	0.04211	31.803	¡ 2e-16 ***
RSE =	2.806	DF = 298		
R^2 =	0.7724	Adj-R^2=	0.7717	
F-stat =	1011	DF =	1 and 298	p-value: ¡ 2.2e-16

Note: Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Table 7.4: Summary Statistics

	$P_{S1,t}$	$P_{S2,t}$	$\hat{\epsilon}_t$
Minimum	0.5000	2.4769	-6.7124
Maximum	26.2691	18.9906	7.1589
Mean	12.8298	11.9754	-2.6666e-011
Variance	34.4896	14.8559	7.8493

Table 7.5: ADF Unit Root Tests

$P^{(c)}_{S1,t}$	Lags	$\Delta P_{S1,t}$	Lags
-1.5112	2	-12.934**	1
$P^{(c)}_{S2,t}$	Lags	$\Delta P_{S2,t}$	Lags
-1.6856	1	-13.358**	1
$\hat{\epsilon}_t$	Lags	$\Delta\hat{\epsilon}_t$	Lags
-1.8733	1	-16.114**	1

Notes: The statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process; (c) superscripts indicate that a constant was included in the augmented Dickey-Fuller regression; "Lags " in the second column are the lags used in the augmented Dickey-Fuller regression for $P_{S1,t}$, $P_{S2,t}$, and $\hat{\epsilon}_t$ while the last column denotes the lags used for $\Delta P_{S1,t}$, $\Delta P_{S2,t}$, and $\Delta\hat{\epsilon}_t$; * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in Fuller (1976).

its autocorrelation function plot and partial autocorrelation function plot are shown in Figure 7.1(c) and (d). From the residuals plot in Figure 7.1(b), it seems the residuals is stationary moving around zero, even though for some periods, it takes quite a long time to return to zero. From the autocorrelation plot in Figure 7.1(c), the spikes are slowly falling to zero. From the partial autocorrelation plot in Figure 7.1(d), clearly the first two spikes are very significant with the values of the spikes are around 0.8 indicating $p = 2$ is the appropriate lag for the models. To make a conclusion that the residuals series is a stationary series, more formal analysis is needed.

Table 7.4 reports summary statistics of the series $P_{S1,t}$, $P_{S2,t}$ and $\hat{\epsilon}_t$. It shows that the mean of $\hat{\epsilon}_t$ is virtually zero with variance around 7.85. Table 7.5 reports the ADF unit root tests statistics. All of the three series conclude that they have a unit root (not stationary) on level but they are stationary on the first difference series. It means that the residuals series $\hat{\epsilon}_t$ is not a stationary and the series $P_{S1,t}$ and $P_{S2,t}$ are not cointegrated. Using higher lags did not change the conclusion. However, the Johansen cointegration test between the series $P_{S1,t}$ and $P_{S2,t}$ in Table 7.6 shows that they are cointegrated with only one significant cointegrating vector which is [1, -1.3710], i.e. it means that one of $P_{S1,t}$ is cointegrated with 1.3710 of $P_{S2,t}$.

Table 7.7 reports linearity tests results for $\hat{\epsilon}_t$. The first linearity test employed is

Table 7.6: Johansen maximum likelihood cointegration results between series $P_{S1,t}$ and $P_{S2,t}$

H_0	H_1	LR	95%
Maximum Eigenvalue LR Test			
$r = 0$	$r = 1$	29.21**	14.1
$r \leq 1$	$r = 2$	3.525	3.8
Trace LR Test			
$r = 0$	$r \geq 1$	32.73**	15.4
$r \leq 1$	$r = 2$	3.525	3.8
Eigenvalues	Standardized eigenvectors		
	$P_{S1,t}$	$P_{S2,t}$	
0.0930622	1.0000	-1.3710	
0.011719	0.053609	1.0000	

Notes: LR denotes Likelihood Ration; * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in Table 1, Osterwald-Lenum (1992).

Table 7.7: Linearity tests on residuals $\hat{\epsilon}_t$

RESET Test	F(2,294) = 1.5296 [0.2183]				Lags used = 2		
Linearity tests based on Terasvirta (1994)							
d	LM^G		LM^3		LM^E		
1	2.1067 [0.0525]*		0.0431 [0.9578]		3.1592 [0.0145]**		
2	1.4760 [0.1862]		0.0091 [0.9909]		2.2246 [0.0664]*		
ESTAR unit root tests comparison							
d		F_{nl}			F_{VPP}		AKSS
	k=1	k=2	k=3	k=1	k=2	k=3	
1	14.45***	17.12**	20.61*	1.45	6.41	6.74	-1.19

a RESET test (see Ramsey (1969)) of the null hypothesis of linearity of the residuals from an AR(2) for $\hat{\epsilon}_t$ against the alternative hypothesis of general model misspecification involving a higher-order polynomial to represent a different functional form. Under the null hypothesis, the statistics is distributed as $\mathcal{X}^2(q)$ with q is equal to the number of higher-order terms in alternative model. Table 7.7 reports the result from executing RESET test statistic where the alternative model with a quadratic and a cubic terms are included. The null hypothesis cannot be rejected, suggesting that a linear AR(2) process for $\hat{\epsilon}_t$ is not misspecified.

The second linearity test is based on Terasvirta (1994) (see also in Subsection 2.1.2 for explanation of the test). The test can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. We use this test to analyse whether the series $\hat{\epsilon}_t$ is a linear AR(2) model or a nonlinear ESTAR(2) or LSTAR(2) model. If $\hat{\epsilon}_t$ follows a LSTAR(2) model, the artificial regression will be as follow:

$$\hat{\epsilon}_t = \phi_{0,0} + \sum_{j=1}^{2} \left(\phi_{0,j}\ \hat{\epsilon}_{t-j} + \phi_{1,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}_{t-d} + \phi_{2,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}_{t-d}^2 + \phi_{3,j}\ \hat{\epsilon}_{t-j}\hat{\epsilon}_{t-d}^3\right) + error \quad (7.32)$$

Keeping the delay parameter $d \leq 2$ fixed, testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = \phi_{3,j} = 0,$$

$\forall j \in \{1, 2\}$ against its complement is a general test (LM^G) of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression (i.e.: $\phi_{3,j} = 0$) if the true model is an ESTAR model, but $\phi_{3,j} \neq 0$ if the true model is an LSTAR), thus, testing the null hypothesis that

$$H_0 : \phi_{3,1} = \phi_{3,2} = 0,$$

provides a test (LM^3) of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if this restriction cannot be rejected at the chosen significance level, then a more powerful test (LM^E) for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = 0 |\ \phi_{3j} = 0,$$

$\forall j \in \{1, 2\}$. From this test, the statistics LM^G, LM^3 and LM^E with $d = 1$ are higher than the test statistics with $d = 2$, indicating that $d = 1$ is more preferred. Using $d = 1$, the LM^G test statistic is significant at 10% significant level, the LM^3 test statistic is not significant and the LM^E test statistic is significant at 5% significant level, indicating that $\hat{\epsilon}_t$ follows a nonlinear 1-ESTAR(2) model.

As in Table 7.5 the ADF unit root test does not confirm that $\hat{\epsilon}_t$ is a stationary series, we use our ESTAR unit root test F_{nl} explained in Section 6.5. For comparison, we include ESTAR unit root test of Venetis *et al.* (2009) denoted F_{VPP} and the augmented KSS test (denoted as AKSS, Kapetanios *et al.* (2003)). For the F_{nl} and F_{VPP} tests, we test for $k = 1, 2, 3$ and $d = 1$. Following assumptions in Section 6.5 for k-ESTAR(2) unit root tests. The unrestricted regression for F_{nl} will be:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ \hat{\epsilon}^{s+3}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ \hat{\epsilon}^{s+2}_{t-1}\Delta\hat{\epsilon}_{t-1} + \eta^*_t \tag{7.33}$$

where $\theta^*_{1,1} = -\theta_{1,2}$, $\gamma_{1,s} = \theta\delta_s(\theta_{2,1} + \theta_{2,2})$, $\gamma_{2,s} = \theta\delta_s\theta^*_{2,1}$, $\theta^*_{2,1} = -\theta_{2,2}$ and $\eta^* = \eta_t + R\left[\sum_{j=1}^{2} \theta_{2,j}\ y_{t-j}\right]$ and R is the remainder.

The unrestricted regression for F_{VPP} will be:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \sum_{s=0}^{2(k-1)} \gamma_{1,s}\ \hat{\epsilon}^{s+3}_{t-1} + \eta^{**}_t \tag{7.34}$$

where $\eta^{**}_t = \sum_{s=0}^{2(k-1)} \gamma_{2,s}\ \hat{\epsilon}^{s+2}_{t-1}\Delta\hat{\epsilon}_{t-1} + \eta^*_t$.

Under the null hypothesis, the F_{nl} and F_{VPP} tests will:

$$\Delta\hat{\epsilon}_t = \theta^*_{1,1}\ \Delta\hat{\epsilon}_{t-1} + \eta_t. \tag{7.35}$$

This equation will be the restricted regression for the F_{nl} and F_{VPP} tests. The calculated F tests from the F_{nl} and F_{VPP} are compared with the critical values in Table 6.1.

For the AKSS test, we only consider for case 1 because the mean of $\hat{\epsilon}_t$ is zero and there is no significant time trend in the series. For this test, the lagged first difference ($\Delta\hat{\epsilon}_{t-1}$) is included to overcome the autocorrelation, so that the regression model for the AKSS test is

$$\Delta\hat{\epsilon}_t = \delta_1\hat{\epsilon}^3_t + \delta_2\Delta\hat{\epsilon}_{t-1} + \eta_t. \tag{7.36}$$

The hypothesis for the AKSS test is

$$H_0 : \delta_1 = 0 \quad vs \quad H_1 : \delta_1 < 0 \tag{7.37}$$

Then, the calculated t-test for δ_1 in (7.36) is compared with the critical values of the AKSS test. The critical value for the t-test of AKSS test is -2.22 obtained from Table 1 in Kapetanios *et al.* (2003). The null hypothesis for the AKSS test concludes that $\hat{\epsilon}_t$ has a unit root without a drift. On the other hand, the alternative hypothesis for the AKSS test concludes that $\hat{\epsilon}_t$ is a globally stationary 1-ESTAR(2) model.

From the ESTAR tests results in Table 7.7, the F_{VPP} and AKSS tests cannot confirm that $\hat{\epsilon}_t$ is a stationary series but our test, the F_{nl} tests can identify that it is a nonlinear stationary ESTAR(2) model. As the most significant level is at $k = 1$, so

Table 7.8: Comparison of model estimations

Coeffients	AR(2)	1-ESTAR(2), d=1
$\theta_{1,1}$	0.18125(0.03774)	0.1155744(0.05612396)
$\theta_{1,2}$	0.77783(0.03770)	0.8743728(0.0480293)
$\theta_{2,1}$		11.4546619(4.059106)
$\theta_{2,2}$		-16.4558339(5.255964)
θ^2		0.0007
Tests on residuals		
SSE	294.5479	284.4167
Wald	5.3181[0.0054]**	0.68129[0.5641]
SW	0.9975[0.9433]	0.998[0.9836]
BL (25)	17.6025[0.8588]	18.8916[0.8023]

Notes: Figures in parentheses beside coefficient estimates denote the estimated standard errors. SSE is sum square error; Wald is a Wald test statistic for parameter restrictions; SW is a Shapiro-Wilk normality test for residuals; BL is a Box-Ljung autocorrelation test for residuals using 25 lags; the figures in parentheses denote the p-values.

that apparently $\hat{\epsilon}_t$ follows a 1-ESTAR(2) model with $d = 1$, i.e.:

$$\hat{\epsilon}_t = \theta_{1,0}+\theta_{1,1}\ \hat{\epsilon}_{t-1}+\theta_{1,2}\ \hat{\epsilon}_{t-2}+(\theta_{2,0}+\theta_{2,1}\ \hat{\epsilon}_{t-1}+\theta_{2,2}\ \hat{\epsilon}_{t-2})\left(1-\exp\left(-\theta^2(\hat{\epsilon}_{t-1}-e_1)^2\right)\right)+\eta_t, \tag{7.38}$$

where η_t is the errors term and e_1 is the equilibrium point.

Table 7.8 reports model comparison between a linear AR(2) model and a nonlinear 1-ESTAR(2) model with $d = 1$ for $\hat{\epsilon}_t$. The parameter estimates for a 1-ESTAR(2) model in (7.38) is obtained using the nonlinear least square method and *nlm* function in *R*. As the mean of $\hat{\epsilon}_t$ is zero, theoretically, $\theta_{1,0} = \theta_{2,0} = e_1 = 0$. This restriction produces the Wald test statistic in Table 7.8 concluding that the restriction can not be rejected at the conventional 5% significance level. The first attempt to estimate the 1-ESTAR(2) model in (7.38) with a restriction $\theta_{1,0} = \theta_{2,0} = e_1 = 0$ produces $\hat{\theta} = 0.0007$ but the last three diagonal hessian matrix are negative. Therefore we cannot obtain the estimate standard deviation of parameters. Because of this reason, parameter θ is fixed at 0.0007 and then produce the results reported in this table. A linear AR(2) is also estimated as a comparison. The Wald test statistic comparing the 1-ESTAR(2) model and the AR(2) model concludes that there is significant difference between the two models, indicating the nonlinear part is significant. Furthermore, the sum square error (SSE) from the 1-ESTAR(2) model is lower than the SSE from the linear AR(2) model. The table also shows that the Shapiro-Wilk normality test and the Box-Ljung autocorrelation test for the two models are comparably the same. Therefore, a nonlinear 1-ESTAR(2) model is more favourable than a linear AR(2) model.

From the previous ESTAR unit root test, we conclude that the residuals series

Table 7.9: Determining optimal pre-set upper bound U using integral equation approach

U	TD_U	IT_U	NT_U	MTP(U)
0.5	10.0596	11.4112	12.9725	6.4862
1.0	16.4649	21.4655	6.9092	**6.9092**
1.5	23.6235	35.4145	4.0815	6.1222
2.0	30.5155	52.4254	2.6170	5.2341
3.0	41.4703	93.1649	1.2282	3.6847
4.0	48.4276	144.6662	0.5536	2.2146

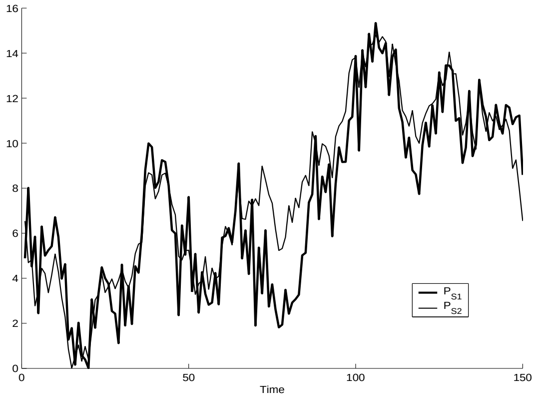
Note: U denotes an upper-bound, TD_U denotes trading duration, IT_U denotes inter-trade interval, NT_U denotes number of trades and MTP_U denotes minimum total profit using number of observations $T = 300$. The results in this table corresponds to upper-trades only. Assuming the series is symmetric, NT_U and MTP_U should be doubled if the lower-trades are also considered.

$\hat{\epsilon}_t$ of regression $P_{S1,t}$ and $P_{S2,t}$ is a stationary. Therefore, we can conclude that the series $P_{S1,t}$ and $P_{S2,t}$ are cointegrated. To perform pairs trading strategy, we need to determine the optimal bounds (upper-bound denoted as U and lower-bound denoted as L) when we open a pair trade so that over the trading period we would get the maximum MTP. Assuming a 1-ESTAR(2) model is a symmetric and the mean $\hat{\epsilon}_t = 0$, the lower-bound is set as $L = -U$. Thus, we concentrate in upper-trades to determine an optimal upper-bound U. Integral equation approach to calculate trading duration in (7.27) and inter-trade interval in (7.28) are used for several values of U, i.e. $U = \{0.5, 1, 1.5, 2, 3, 4\}$. We also use parameter estimates for 1-ESTAR(2) model in Table 7.8 and calculated $\hat{\sigma}^2_\eta = 0.99$, $a = 0$ and $b = 4\hat{\sigma}_e \approx 4 \times 3 = 12$. Number of trades NT_U is estimated using (7.29) with $T = 300$ and the minimum total profit for given U is calculated as $U \times NT_U$. The calculated results are reported in Table 7.9 showing that $U = 1$ produces the highest MTP. Therefore, we choose $U = 1$ as the optimal upper-bound and setting the lower-bound $L = -1$. Computer program for Table 7.9 is provided in Appendix A.4.2.

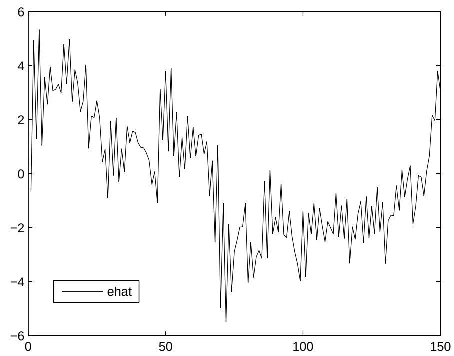
Figure 7.2 shows the plots of out-sample data generated similar to the in-sample data series, i.e.:

$$\begin{aligned}
P_{S2,t} &= P_{S2,(t-1)} + \epsilon^*_t, \qquad \epsilon^*_t \sim N(0,1) \\
\epsilon_t &= 0.1\epsilon_{t-1} + 0.9\epsilon_{t-2} + (0.4\epsilon_{t-1} - 0.5\epsilon_{t-2})\left(1 - \exp(-0.01\epsilon^2_{t-1})\right) + \eta_t, \\
\eta_t &\sim N(0,1), \\
P_{S1,t} &= P_{S2,t} + \epsilon_t, \\
t &= 301, 302, \ldots, 450.
\end{aligned} \tag{7.39}$$

From (7.39) we obtained 150 observations of series $P_{S1,t}$ and $P_{S2,t}$ as out-sample



(a)



(b)

Figure 7.2: Plots of out-sample data: (a) Plots of $P_{S1,t}$ and $P_{S2,t}$; (b) Plot of residuals (cointegration error), denoted as $\hat{e}$.

data. To calculate $\hat{\epsilon}_t$, we use regression estimates of $P_{S1,t}$ and $P_{S2,t}$ from in-sample data in Table 7.3. Thus,

$$\hat{\epsilon}_t = P_{S1,t} + 3.20683 - 1.33912 P_{S2,t}. \tag{7.40}$$

Using in-sample and out-sample data, we perform pairs trading simulations. For comparison, we use pre-set upper-bound $U = 1$ and $U = 2$. The lower-bounds are set as $-U$. The results are presented in Table 7.10. When $\hat{\epsilon}_t \geq U$, we buy 1 stock of $S1$ at the price of $P_{S1,t}$ and short-sell 1.34 stock of $S2$ at the price of $P_{S2,t}$. We unwind the position when $\hat{\epsilon}_t \leq 0$. The opposite happens when $\hat{\epsilon}_t \leq L = -U$. In this case, we short-sell 1 stock of $S1$ at the price of $P_{S1,t}$ and buy 1.34 stock of $S2$ at the price of $P_{S2,t}$ and then unwind the position when $\hat{\epsilon}_t \geq 0$. We cannot open another pair trade if the previous pair trade has not been close yet. At the end of the data period, the pair trade is closed, possibly at a loss. If the pair trade is forced to be closed at the end of period, we call this as an uncomplete trade. We record trading duration for each trade, the total profit and number of trades during in-sample and out-sample periods. Number of trades includes uncomplete trades. Average profit/trade is calculated as total profit divided by number of trades. Average trading duration is calculated as sum of trading duration from all trades divided by number of trades. Comparing the results from $U = 1$ and $U = 2$, as expected, the higher the upper-bound, the higher the average profit/trade. Furthermore, the higher the upper-bound, the lower the number of trades. Thus, even though the average profit/trade with $U = 1$ is lower than that with $U = 2$, it produces higher total profit.

Table 7.10: Pairs trading simulations

	In-sample data 300 observations	Out-sample data 150 observations
pre-set upper-bound $U = 1$		
Total profit	\$ 61.35	\$ 32.19
Average profit/trade	\$ 2.45	\$ 2.68
No. Trades	25	12
Average trading duration	9.44	10.75
pre-set upper-bound $U = 2$		
Total profit	\$ 46.43	\$ 24.56
Average profit/trade	\$ 3.09	\$ 3.07
No. Trades	15	8
Average trading duration	13.73	14

7.6 Conclusion

In this chapter, we have shown that the first-passage time using integral equation approach can be used for series following a k-ESTAR(p) model. Using this approach trading duration and inter-trade interval for pairs trading are formulated. As we have not found suitable empirical data, we generate data of 2 stock prices with cointegration errors follows the 1-ESTAR(2) model. Using these series, pairs trading simulation is performed and it shows that pairs trading strategy can work for cointegration errors following an 1-ESTAR(2) model. In the process of data analysis, we also show that our ESTAR unit root test can identify that the $\hat{\epsilon}_t$ is a nonlinear stationary where the other unit root tests, i.e. F_{VPP}, AKSS and ADF, cannot confirm that the series is stationary.

Chapter 8

Cointegration and Pairs Trading of the S&P 500 Future and Spot Index Prices

8.1 Introduction

Numerous papers have discussed market efficiency, trading methods and the relationship between future and spot index prices. A number of researchers have investigated the efficiency of futures and forward markets in commodities, foreign exchange, securities and stock index. Cargill and Rauseer (1975), Goss (1981), Kofi (1973) and Tomek and Gray (1970) applied weak form tests of efficiency. The weak form of efficiency hypothesis (unbiasedness hypothesis) defines that asset prices should not be predictable based on their own past history. Therefore, the weak form test relies on the historical sequence of prices and involves regressing spot prices at contract to maturity on previous future prices. If the intercept is zero and the slope is one, the market is regarded as efficient and the future prices are considered to be unbiased predictor of the next spot prices. However, Gupta and Mayer (1981), Burns (1983) and Garcia *et al.* (1988) criticised that these tests are not valid because the coefficient estimates are based on the ex post knowledge of the data that is not available to the agents in the market.

As an alternative, the tests for market efficiency in a semi-strong form sense were employed by subsequent researchers. These tests are determining whether future prices fully reflect all publicly available information at the time of contract. An econometric model is applied to compare and to forecast error of the model with the future prices. However, as Chowdhury (1991) noted, the results from these tests are contradictory. While several authors (Goss, 1988 and Gupta and Mayer, 1981) found evidence in support of the market efficiency hypothesis, others found rather weak results (Goss, 1983).

Ma and Hein (1990) suggested that the problem in testing market efficiency is that spot and future prices are generally not stationary as they have a unit root. Therefore, conventional statistical procedures are no longer appropriate for testing market efficiency, because they tend to bias toward incorrectly rejecting efficiency. Current developments in the theory of unit root and cointegration which can account for the nonstationary behaviour of spot and future prices, provide new methods for testing market efficiency. Hakkio and Rush (1989) and Shen and Wang (1990) demonstrated that, in an efficient market, the spot and future prices should be cointegrated. Cointegration between future and spot prices implies that they are tied together in a long-run relationship and never move far away from each other which is the requirement property of the market efficiency hypothesis.

The simple market efficiency hypothesis states that the future price is an unbiased estimator of the future spot price as in (8.1):

$$E(\ln S_T|\mathcal{F}_t) = \ln F_{t,T} \tag{8.1}$$

where S_T is the spot price at maturity time T, $F_{t,T}$ is the future price at time t with a contract maturity at T, and $E(.|\mathcal{F}_t)$ represents expectation conditional on all information up to time t. Market efficiency hypothesis test of (8.1) is based on estimates of (8.2):

$$\ln S_T = \alpha_0 + \alpha_1 \ln F_{t,T} + \upsilon_t \tag{8.2}$$

where α_0 is a constant and υ_t is the error term at time t. Assuming that the market participants are risk-neutral and expectations are rational, the hypothesis of simple efficiency involves the joint restriction that $\alpha_0 = 0$ and $\alpha_1 = 1$ (see Fujihara and Mougoue, 1997).

Recognising that future and spot prices may contain a unit root, a standard procedure in recent papers now consist of testing unit root in the spot and future prices and then applying a cointegration analysis if a unit root is found.

Another econometric model to test the efficiency of future markets is a cost-carry model and its variants. Consider a market containing an asset, a stock index, whose price S_t at time t, under the equivalent martingale measure evolves according to:

$$dS_t = S_t\,(\bar{r} - \bar{d})\,dt + \sigma\,S_t\,dW_t \tag{8.3}$$

where $\bar{r}$ is a constant risk-free interest rate, $\bar{d}$ is a constant dividend yield on the index, σ is the volatility of the index and W_t is a one-dimensional standard Brownian motion in a complete probability space. Rearrange (8.3) to become:

$$\frac{dS_t}{S_t} = r\;dt + \sigma\;dW_t \tag{8.4}$$

where $r = \bar{r} - \bar{d}$.

Let $g(S_t) = lnS_t$, by Ito's lemma,

$$\begin{aligned} d(lnS_t) &= d(g(S_t)) = \frac{\partial g(S_t)}{\partial S_t} dS_t + \frac{1}{2}\frac{\partial g(S_t)^2}{\partial^2 S_t} S_t^2 \sigma^2 dt \\ &= \frac{dS_t}{S_t} + \frac{1}{2}\left(-\frac{1}{S_t^2}\right) S_t^2 \sigma^2 dt \\ &= \frac{dS_t}{S_t} - \frac{\sigma^2}{2}\, dt. \end{aligned} \tag{8.5}$$

Thus, from (8.5),

$$\frac{dS_t}{S_t} = d(lnS_t) + \frac{\sigma^2}{2} dt. \tag{8.6}$$

Combining (8.4) and (8.6),

$$\begin{aligned} d(lnS_t) + \frac{\sigma^2}{2} dt &= r\, dt + \sigma\, dW_t \\ \int_0^t d(lnS_t) &= \int_0^t \left(r - \frac{\sigma^2}{2}\right)\, dt + \int_0^t \sigma\, dW_t \\ lnS_t - lnS_0 &= \int_0^t \left(r - \frac{\sigma^2}{2}\right)\, dt + \int_0^t \sigma\, dW_t \\ S_t &= S_0 \, \exp\left[\int_0^t \left(r - \frac{\sigma^2}{2}\right)\, dt + \int_0^t \sigma\, dW_t\right] \end{aligned} \tag{8.7}$$

To obtain $E(S_t)$, first, we consider moments of the solution:

$$\begin{aligned} E\left(S_t^p\right) &= E\left[S_0^p \, \exp\left(p \int_0^t \left(r - \frac{\sigma^2}{2}\right)\, dt + p \int_0^t \sigma\, dW_t\right)\right] \\ &= E\left(S_0^p\right) E\left[\exp\left(p \int_0^t \left(r - \frac{\sigma^2}{2}\right) dt + p \int_0^t \sigma\, dW_t\right)\right] \\ &= E\left(S_0^p\right)\left\{\exp\left[p \int_0^t \left(r - \frac{\sigma^2}{2}\right)\, dt\right] + E\left[\exp\left(p \int_0^t \sigma\, dW_t\right)\right]\right\} \end{aligned} \tag{8.8}$$

assuming S_t and W_t are independent.

Expectation in (8.8) can be computed using characteristic or moment generating functions. If Z is a Gaussian random variable with mean μ and variance σ^2, then,

$$E\left[\exp(p\, Z)\right] = \exp\left(p\, \mu + \frac{1}{2} p^2 \sigma^2\right). \tag{8.9}$$

Let $Z = \int_0^t \sigma\, dW_t$. As Z is a Gaussian with mean 0 and variance $\int_0^t \sigma^2 dt$, then,

$$E\left[\exp(p\, Z)\right] = \exp\left(p \times 0 + \frac{1}{2} p^2 \sigma^2 \int_0^t dt\right). \tag{8.10}$$

Thus, substituting (8.10) into (8.8) will become

$$\begin{aligned} E\left(S_t^p\right) &= E\left(S_0^p\right)\exp\left[p\int_0^t r\,dt + \frac{p(p-1)}{2}\,\sigma^2\int_0^t dt\right] \\ E\left(S_t\right) &= S_0\,\exp(r\,t), \qquad \text{when } p=1 \text{ and } S_0 \text{ is known.} \end{aligned} \tag{8.11}$$

Standard derivatives pricing theory gives the future price $F_{t,T}$ at time t for a maturity at time $T \geq t$ as

$$F_{t,T} = E(S_T|\mathcal{F}_t) \tag{8.12}$$

where E denotes the mathematical expectation with the respect to the martingale measure P and $\mathcal{F}_t$ denotes the information set at time t (e.g. see Karatzas and Shreve (1998)). Given (8.11) and (8.12), the future price has the well-known formula:

$$F_{t,T} = S_t\,\exp(r(T-t)). \tag{8.13}$$

Taking natural log on both sides of (8.13) yields

$$\ln F_{t,T} = \ln S_t + r(T-t). \tag{8.14}$$

Monoyios and Sarno (2002) defined basis b_t at time t as

$$b_t = f_t - s_t \tag{8.15}$$

where f_t is the log of future contract price at time t with the nearest maturity time and s_t is the log of spot index price at time t. On the same day, there are quote prices for future contracts with different maturities. For example, on July, 5, 2011, there are quote prices for future contracts S&P 500 index with maturity in September 2011, December 2011, March 2012, June 2012, September 2012, December 2012, etc. Thus, f_t on July, 5, 2011 is the log of future contract price with maturity in September 2011.

Monoyios and Sarno (2002) argued that there is significant nonlinearity in the dynamics of the basis due to the existence of transaction costs or agents heterogeneity. Using daily data time series on future contracts of the S&P 500 index and the FTSE 100 index, as well as the price levels of the corresponding underlying cash indices over the sample period from January 1, 1988 to December 31, 1998, they found that the basis follows a nonlinear stationary ESTAR (Exponential Smooth Transition Autoregressive) model. In constructing the basis, they pair up the spot price with the future contract with the nearest maturity. Similarly, using regime-switching-vector-equilibrium-correction model, Sarno and Valente (2000) also concluded that there is a nonlinearity in the basis.

However, using the same procedure as Monoyios and Sarno (2002) and S&P 500 data series from January 1, 1998 to December 31, 2009, we conclude that there is no significant difference between linear ARMA model and nonlinear ESTAR model in fitting the data. Another concern is that the way they make the future prices series. By pairing up the spot price with the future contract with the nearest maturity, it may produce artificial jumps at the time of maturity. From (8.14), given r is a constant, the basis is equal to

$$b_t = f_t - s_t = r(T - t). \tag{8.16}$$

Thus, the longer the time to maturity, the higher the difference between the log future price and the log spot price. For example for S&P 500, it has 4 maturity times during a year which are the third Friday in March, June, September and December. We find that at these times, there are jumps in the basis. Ma *et al.* (1992) argued that it may create volatility and bias in the parameter estimates. Therefore, instead of constructing one future prices series for all period of data, we analyse the cointegration relationship of the log future prices and the log spot prices for each contract with a time trend.

Crowder and Phengpis (2005) used another invariant of a cost-carrying model as follow:

$$f_t = \beta_1\, s_t + \beta_2\, c_t + \epsilon_t \tag{8.17}$$

where c_t is the cost-carry over the life of the contract or stochastic convenience yield and ϵ_t is the white noise error.

Using daily data of S&P 500 from April 21, 1982 to June 5, 2003, and the three-month Treasury bill as a proxy for the cost-carry c_t, Crowder and Phengpis (2005) found that the log future prices, the log spot prices and the cost-carry are cointegrated with β_1 estimates close to -1. They found that the long-equilibrium, i.e. ϵ_t, is a trend stationary series. However, they found that the Treasury bill is not significant in the model but it can not be excluded by the likelihood ratio (LR) test. Furthermore, the way they constructed the log future prices is the same as Monoyios and Sarno (2002) so that there are jumps in ϵ_t series.

In this thesis, by assuming that r in (8.16) is a constant during the contract period we employ a cost-carrying model for each contract as follow:

$$f_t = \mu + \beta\, s_t + \delta\, t^* + \epsilon_t \tag{8.18}$$

where $t^* = (T - t)$ is the time to maturity and ϵ_t is the error term. Theoretically, based on (8.16), $\mu = 0$ and $\beta = 1$, thus δ is the estimate of r in (8.16). Using Engle-Granger and Johansen methods, we would like to find out whether the log future prices, the log spot prices and the time trend are cointegrated.

If the log future contract prices and the log spot index prices are cointegrated, there is a possibility to develop pairs trading strategy between the two prices. Puspaningrum *et al.* (2010) developed pairs trading mechanism for two cointegrated assets by taking advantages of stationarity of the cointegration errors. Pairs trading works by taking the arbitrage opportunity of temporary anomalies between prices of related assets which have long-run equilibrium. When such an event occurs, one asset will be overvalued relative to the other asset. We can then invest in a two-assets portfolio (a pair) where the overvalued asset is sold (short position) and the undervalued asset is bought (long position). The trade is closed out by taking the opposite position of these assets after they have settled back into their long-run relationship. The profit is captured from this short-term discrepancies in the two asset prices. Since the profit does not depend on the movement of the market, pairs trading is a market-neutral investment strategy. Similarly, the stationarity properties of ϵ_t in (8.18) can be used to make pairs trading mechanism between the log future prices and the log spot prices. However, distinct from Puspaningrum *et al.* (2010), pairs trading mechanism between the log future prices and the log spot prices will involves a time trend.

The rest of the sections will be organised as follows. Section 8.2 compares nonlinearity analysis of the S&P 500 basis based on Monoyios and Sarno (2002) with currently available data. Section 8.3 provides literature review about cointegration with a time trend and discusses cointegration analysis of the log future contract prices and the log spot prices with a time trend for each future contract. Section 8.4 discusses pairs trading mechanism between future and spot prices and the last section is conclusion.

8.2 Nonlinearity of the S&P 500 Basis: Comparison to Monoyios and Sarno (2002)

Monoyios and Sarno (2002) analysed the mean reversion of S&P 500 and FTSE 100 basis with daily data spanned from January 1, 1988 to December 31, 1998. They concluded that the two basis follow ESTAR (Exponential Smooth Transition Autoregressive) models.

They argued that the simple cost-carry model in (8.13) should be extended to take into account the transaction costs. If a share of stock is bought at time of t for a price of S_t then the buyer's cash account is debited an amount $S_t(1+\upsilon)$, where $\upsilon > 0$ is the proportional transactions cost rate for buying the index. Similarly, a stock sale credits the seller's bank account at time t with an amount $S_t(1-\mu)$ where $\mu > 0$ is the proportional transactions cost rate for selling stock. Under these assumptions, the future price at t for delivery at T, $F_{t,T}$ must lie within the following bounds to prevent arbitrage:

$$S_t(1-\mu)\exp(r(T-t)) \leq F_{t,T} \leq S_t(1+\upsilon)\exp(r(T-t)). \tag{8.19}$$

Then, Monoyios and Sarno (2002) rewrite (8.19) to become:

$$\log(1-\mu) \leq b_{t,T} \leq \log(1+\upsilon). \tag{8.20}$$

As Monoyios and Sarno (2002) defined $b_{t,T} = \log(F_{t,T}) - \log(S_t)$, they made a mistake in (8.20). Equation (8.20) should be:

$$\log(1-\mu)+(r(T-t))\leq b_{t,T}\leq\log(1+\upsilon)+(r(T-t)), \tag{8.21}$$

or

$$\log(1-\mu)\leq b_{t,T}-(r(T-t))\leq\log(1+\upsilon). \tag{8.22}$$

Based on (8.20), Monoyios and Sarno (2002) argued that proportional transaction costs create a band for the basis within which the marginal cost of arbitrage exceeds the marginal benefit. Assuming instantaneous arbitrage at the edges of the band then implies that the bounds become reflecting barriers. Based on this argument, several studies adopt threshold-type models and provide evidences against linearity in the deviation of the basis from its equilibrium level and in favor of threshold-type behavior (e.g. Yadav *et al.*, 1994, Dwyer *et al.*, 1996, and Martens *et al.*, 1998). Monoyios and Sarno (2002) argued that even though the threshold-type models are appealing, various arguments can be made that rationalise multiple-threshold or smooth, rather than single-threshold or discrete, nonlinear adjustment of the basis toward its equilibrium value. Therefore, Monoyios and Sarno (2002) fitted the basis with ESTAR models. With an ESTAR model, the basis should become increasingly mean reverting with the size of the deviation from the equilibrium level. Transaction costs create a band of no arbitrage for the basis, but the basis can stray beyond the thresholds. Once beyond the upper or lower threshold, the basis becomes increasingly mean reverting. Within the transaction cost bands, where no trade takes places, the process driving the basis is divergent since arbitrage is not profitable.

In empirical analysis, Monoyios and Sarno (2002) defined basis b_t at time t as

$$b_t=f_t-s_t \tag{8.23}$$

where f_t is the log of future contract price at time t with the nearest maturity time and s_t is the log of spot index price at time t. Monoyios and Sarno (2002) considered a STAR model for basis b_t written as follow:

$$b_t=\theta_{1,0}+\sum_{j=1}^{p}\theta_{1,j}\ b_{t-j}+\left[\theta_{2,0}+\sum_{j=1}^{p}\theta_{2,j}\ b_{t-j}\right]G(\theta,e,b_{t-d})+\epsilon_t \tag{8.24}$$

where $\{\epsilon_t\}$ is a stationary and ergodic martingale difference sequence with variance σ_ϵ^2; $d\geq 1$ is a delay parameter; $(\theta,e)\in\{\mathbb{R}^+\times\mathbb{R}\}$ where $\mathbb{R}$ denotes the real space $(-\infty,\infty)$ and $\mathbb{R}^+$ denotes the positive real space $(0,\infty)$. The transition function $G(\theta,e,b_{t-d})$ determines the speed of adjustment to the equilibrium e. Two simple transition functions suggested by Granger and Terasvirta (1993) and Terasvirta (1994) are the logistic and exponential functions (see Chapter 6).

Monoyios and Sarno (2002) argued that an ESTAR model is more appropriate for modelling basis movement than a LSTAR model due to symmetric adjustment of the basis. Furthermore, there is fairly convincing evidence that distribution of the basis is symmetric, for example the evidence provided by Dwyer *et al.* (1996) using both parametric and nonparametric tests of symmetry applied to data for the S&P 500 index. However, Monoyios and Sarno (2002) also tested for nonlinearities arising from the LSTAR formulation, but conclude that the ESTAR model is more appropriate for modelling basis movement than a LSTAR model.

To analyse the mean-reversion of the basis, Monoyios and Sarno (2002) conducted unit root tests for the log of future prices denoted as f_t, the log of spot index prices denoted as s_t and the basis b_t using the ADF unit root tests (Fuller, 1976). They concluded that the f_t and s_t series are $I(1)$ and b_t is $I(0)$. Thus, f_t and s_t are non-stationary while the basis b_t is stationary. Then, they employed cointegration analysis between f_t and s_t using the well-known Johansen method (Johansen, 1988, 1991)and that f_t and s_t are cointegrated with cointegrating parameter equals unity confirming the cost-carry model.

Monoyios and Sarno (2002) followed Terasvirta (1994) by applying a sequence of linearity tests (see also in Subsection 2.1.2 for explanation of the test) to artificial regressions which can be interpreted as second or third-order Taylor series expansions of (8.24). Firstly, Monoyios and Sarno (2002) chose the order of regression, p, through inspection of the partial autocorrelation function (PACF). They argued that the PACF is preferred than selection criterion methods such as AIC (see Akaike, 1974) or SBIC (see Rissanen, 1978 or Schwarz, 1978) because the latter may bias by choosing order of the autoregression toward low values and any remaining serial correlation may affect the power of subsequent linearity tests. Then, Monoyios and Sarno (2002) ran the non-linearity tests which allow detection of general nonlinearity through the significance of the higher-order terms in the Taylor expansions, in which the value of delay parameter d selected as that yielding the largest value of the test statistics. The tests can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. The artificial regression of (8.24) is estimated as follow:

$$b_t = \phi_{0,0} + \sum_{j=1}^{p} \left(\phi_{0,j} b_{t-j} + \phi_{1,j} b_{t-j} b_{t-d} + \phi_{2,j} b_{t-j} b_{t-d}^2 + \phi_{3,j} b_{t-j} b_{t-d}^3 \right) + error. \tag{8.25}$$

Keeping the delay parameter d fixed, testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = \phi_{3j}, \qquad \forall j \in \{1, 2, \cdots, p\}$$

against its complement is a general test LM^G of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression(i.e., $\phi_{3,j} = 0$ if the true model is an ESTAR model, but $\phi_{3,j} \neq 0$ if the true model is an LSTAR), thus, testing the null hypothesis that

$$H_0 : \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}$$

provides a test LM^3 of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if the restrictions $\phi_{3,j} = 0$ cannot be rejected at the chosen significance level, then a more powerful test LM^E for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_0 : \phi_{1j} = \phi_{2j} |\ \phi_{3j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}.$$

In the empirical analysis, Monoyios and Sarno (2002) employed each of the tests LM^G, LM^3 and LM^E for a set of values d in the range $\{1, 2, \cdots, 10\}$. If LM^G and LM^E are both being significant, then $d = d^*$ is selected such that $LM^E(d^*) = \sup_{d \in \Im} LM^E(d)$ for $\Im = \{1, 2, \cdots, 10\}$.

Monoyios and Sarno (2002) also employed a RESET (Ramsey, 1969) test of the null hypothesis of linearity of the residuals from a linear model AR(p) for the basis against the alternative hypothesis of general model misspecification. Under the RESET test statistic, the alternative model involves a higher-order polynomial such as a quadratic and a cubic terms. Under the null hypothesis, the RESET statistic is distributed as $\mathcal{X}^2(q)$ with q equal to the number of higher-order terms in the alternative model.

Concluding that the basis follow an ESTAR model, Monoyios and Sarno (2002) fitted the basis by reparameterising the ESTAR model in (8.24) as follow:

$$\begin{aligned}\Delta b_t = \theta_{1,0} + \rho b_{t-1} + \sum_{j=1}^{p-1} \phi_{1,j}\ \Delta b_{t-j} + \\ \left[\theta_{2,0} + \rho^* b_{t-1} + \textstyle\sum_{j=1}^{p-1} \phi_{2,j}\ \Delta b_{t-j}\right] \left[1 - \exp\left(-\theta^2 (b_{t-d} - e)^2\right)\right] + \epsilon_t \end{aligned} \tag{8.26}$$

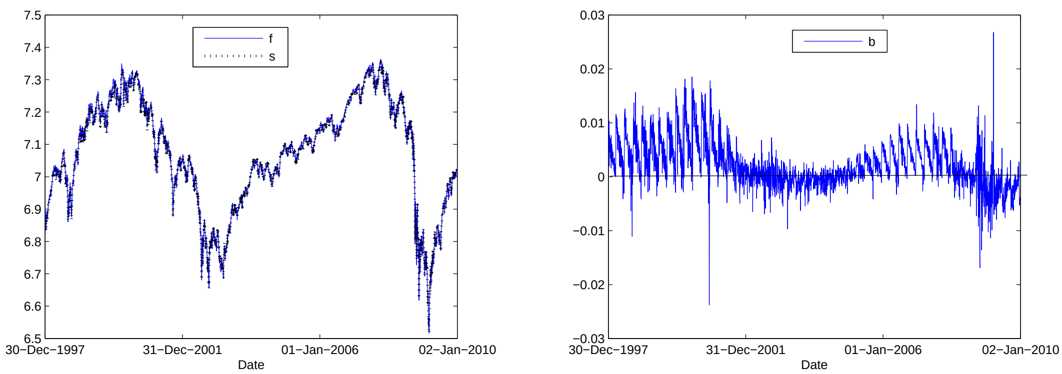
where $\Delta b_t = b_t - b_{t-1}$, $\rho = \sum_{j=1}^{p} \theta_{1,j}$, $\rho^* = \sum_{j=1}^{p} \theta_{2,j}$, $\phi_{1,j} = -\sum_{i=1+j}^{p} \theta_{1,i}$ and $\phi_{2,j} = -\sum_{i=1+j}^{p} \theta_{2,i}$.

They concluded that the restrictions $\rho = 0$, $\phi_{1,j} = -\phi_{2,j}$ and $\theta_{1,0} = \theta_{2,0} = e = 0$ could not be rejected at the 5% significance level. These restrictions imply that the equilibrium of the basis is zero. Furthermore, in the neighborhood of the equilibrium, b_t is nonstationary but it is becoming increasingly mean reverting with absolute size of the deviation from the equilibrium.

8.2.1 Empirical Analysis Using Current Data

The data set comprising daily future closing prices of the S&P 500 with contract maturity in March 1998 to December 2009, as well as the corresponding daily closing spot index prices, were obtained from Datastream database. To make comparison nonlinearity analysis of the basis to Monoyios and Sarno (2002)'s work, the same as they did, the spot index price is paired up with the log future price with the contract nearest to maturity. Using these data, "basis" is constructed as $b_t = f_t - s_t$, where f_t and s_t denote the logarithm of the future prices and the logarithm of spot index prices respectively.

A number of studies focusing on modelling intraday or short-live arbitrage have used intra-day data at different intervals.[1] However, as we want to analyse the cointegration property between the log future contract prices and the log spot index prices, the span of time series in terms of years is much more important than the number of observations per se (e.g., see Monoyios and Sarno, 2002 and Shiller and Perron, 1985). Therefore, we use daily data.



(a) (b)

Figure 8.1: (a) Plot of f_t and s_t, (b) Plot of b_t

Table 8.1: Summary Statistics

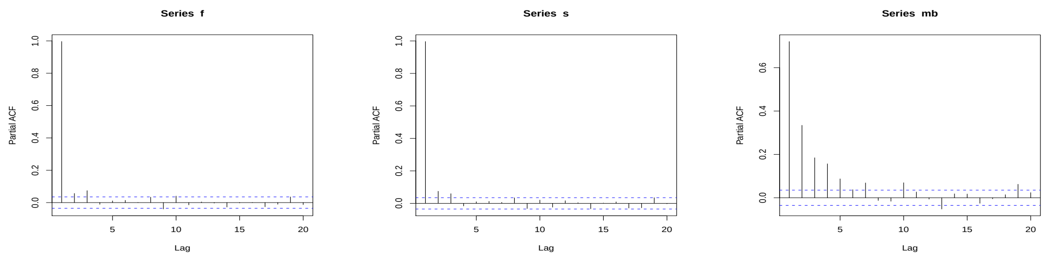
	f_t	s_t	b_t	mb_t	Δf_t	Δs_t	Δb_t
Minimum	6.5160	6.5169	-0.0237	-0.0260	-0.1040	-0.0947	-0.0336
Maximum	7.3627	7.3557	0.0267	0.0244	0.1319	0.1095	0.0264
Mean	7.0711	7.0688	0.0023	-6.60E-06	4.03E-05	4.44E-05	-4.10E-06
Variance	0.0279	0.0272	1.74E-05	1.74E-05	0.0002	0.0002	9.70E-06

Notes: f_t, s_t, b_t and mb_t denote the log of the future prices, the log of the spot index prices, the basis and the demeaned basis, respectively. The demeaned basis is defined as $mb_t = b_t - \bar{b}$, where $\bar{b}$ is the mean of the basis so that the mean of mb_t is zero. Δ is the first-difference operator.

Figure 8.1(a) shows the plots of f_t and s_t while Figure 8.1(b) shows the plot of b_t. From Figure 8.1, the plots of f_t and s_t are almost similar indicating the basis b_t which is

[1]For example, Chan (1992), Dwyer *et al.* (1996) and Miller *et al.* (1994) have used 5-,15-, and 5-min intervals, respectively.

the difference between f_t and s_t is not large. During the data period, there are 2 major financial crises. The first is in 1999-2002 due to the South American economic crisis in Argentina, Brazil and Uruguay [2] as well as the Dot-com buble crisis [3]. The second is the financial crisis of 2007 to the present triggered by the US subprime mortgage crisis[4]. Both financial crises are reflected in the fall of the log future prices and the log spot index prices. The crises are also reflected in the basis where the basis tends to has negative value during the crisis periods.



(a) (b) (c)

Figure 8.2: (a) PACF plot of f_t, (b) PACF plot of s_t, and (c) PACF plot of mb_t

Table 8.2: Lag length tests for unit root tests

	p	$\hat{\sigma}_p^2$	AIC	$\mathcal{X}^2(20)$
f_t				
	3	0.0001834	-18038.99	30.0231 [0.0694]
s_t				
	3	0.0001791	-18112.79	29.3014 [0.0820]
mb_t				
	5	6.909e-06	-28240.78^a	58.0468 [0.0000]**
	7	6.865e-06	-28256.84	41.2758 [0.0034]*
	10	6.829e-06	-28267.15	27.1426 [0.1313]
	19	6.767e-06	-28277.47	2.7141 [1.0000]

Notes: p is the number of lags in AR(p) model. $\hat{\sigma}_p^2$ is the variance estimate of AR(p) residuals. * and ** denote significant at 5% and 1% level, respectively. ^{a} superscript denotes the minimum value. $\mathcal{X}^2(20)$ is Box-Ljung autocorrelation test statistics for AR(p) residuals using 20 lags; the numbers in the parentheses are the p-values.

Table 8.1 shows some summary statistics for the log future prices f_t, the log spot index prices s_t, the basis b_t and the de-meaned basis mb_t. The PACF plots in Figure 8.2 suggest that both the future and spot index prices show significant spikes at the first

[2]See "South American economic crisis of 2002" in
http://en.wikipedia.org/wiki/South_American_economic_crisis_of_2002. Retrieved on 18/11/2010.

[3]See "Dot-com bubble" in
http://en.wikipedia.org/wiki/Dot-com_bubble. Retrieved on 18/11/2010.

[4]See "Causes of the Financial Crisis of 2007-2010" in
http://en.wikipedia.org/wiki/Causes_of_the_financial_crisis_of_2007-2010. Retrieved on 18/11/2010.

3 lags, but the first spike is very strong. The PACF plot of the basis displays a slower decay of the PACF with significant spikes at the first five lags, lag 7 , lag 10 and lag 19.

In Table 8.2, Box-Ljung autocorrelation tests statistics for AR(3) residuals using 20 lags for f_t and s_t are 30.0231 [0.0694] and 29.3014 [0.0820], respectively, where the figures in the parentheses are the p-values. Thus, we can accept the null hypothesis of no autocorrelation in residuals for f_t and s_t using an AR(3) model and then use $p = 3$ for unit root tests. Box-Ljung autocorrelation tests statistics using 20 lags on mb_t for AR(5), AR(7) AR(10) and AR(19) residuals are 58.0468 [0.0000], 41.2758 [0.0034], 27.1426 [0.1313], 2.7141 [1.0000], respectively. From these results, $p = 10$ is enough to make the residuals become unautocorrelated for mb_t.

The standard augmented Dickey-Fuller (ADF) unit root tests reported in Table 8.3 shows that both f_t and s_t are I(1) while mb_t is I(0). Using other lags does not change the conclusion.

Table 8.3: Unit Root Tests for S&P 500

Future prices	$f_t^{(c)}$	$f_t^{(c,\tau)}$	Lags	Δf_t	$\Delta f_t^{(c)}$	Lags
	-2.1139	-2.1680	2	-44.072**	-44.065**	1
Spot Index	$s_t^{(c)}$	$s_t^{(c,\tau)}$	Lags	Δs_t	$\Delta s_t^{(c)}$	Lags
prices	-2.1255	-2.1738	2	-43.824**	-43.817**	1
Basis	b_t	$b_t^{(c)}$	Lags	Δb_t	$\Delta b_t^{(c)}$	Lags
	-7.1598**	-7.1589**	9	-25.312**	-25.309**	8

Note: The statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process; (c) or (c, τ) superscripts indicate that a constant (a constant and a linear trend) was (were) included in the augmented Dickey-Fuller regression, while absence of a superscript indicates that neither a constant nor a trend were included; "Lags " in the fourth colum are the lags used in the augmented Dickey-Fuller regression for f_t, s_t, and b_t while the last colum denotes the lags used for Δf_t, Δs_t, and Δb_t, regardless the superscripts (c) or (c, τ) ; * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in Fuller (1976).

Johansen cointegration test (see Johansen, 1988, Johansen, 1991) is employed and reported in Table 8.4. The test uses a maximum likelihood procedure in a vector autoregression comprising f_t and s_t, with a lag length of 2 and an unrestricted constant term. We use a lag length of 2 because $p = 3$ is the common lag for f_t and s_t, so that in the vector autoregression, the lag length is $(p - 1) = 2$. We also try for other lags such as $(p - 1) = 6, 9, 18$, but they do not change the conclusions. Johansen likelihood ratio (LR) test statistics clearly suggests that there are two cointegrating relationships between f_t and s_t. However, the first cointegrating relationship shows

Table 8.4: Johansen Maximum Likelihood Cointegration Results for S&P 500

H_0	H_1		LR	5% Critical Value
LR Tests Based on the Maximum Eigenvalue of the Stochastic Matrix				
$r = 0$	$r = 1$		296.1**	14.07
$r \leq 1$	$r = 2$		5.105*	3.76
LR Tests Based on the Trace of the Stochastic Matrix				
$r = 0$	$r \geq 1$		301.2**	15.41
$r \leq 1$	$r = 2$		5.105*	3.76
Eigenvalues	Standardized β' eigenvectors		Standardized α' coefficients	
	f_t	s_t	f_t	s_t
0.0902856	1.0000	-1.0124	-0.19131	0.064240
0.00163014	0.37218	1.0000	-0.0023885	-0.0023786
LR-test restriction = $\mathcal{X}^2(1)$				83.801 [0.0000] **

Notes: The vector autoregression under consideration comprises the log of the future prices f_t and the log of the spot index prices s_t. H_0 and H_1 denote the null hypothesis and the alternative hypothesis, respectively. The lags used in vector error-correction model (VECM) is 2. r denotes the number of cointegrating vectors. LR denotes the LR test statistics adding the constant as an unrestricted variable. * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in Table 1 in Osterwald-Lenum (1992). The 5% critical values are reported in the last column.

much more significant than the second one. Financial theory based on the cost-carry model suggests that the cointegrating parameter equals unity, i.e. in this case, one unit price of f_t is cointegrated with one unit price of s_t or the first cointegrating vector β in the Johansen cointegration test results is [1,-1]. From Table 8.4, the first cointegrating vector, i.e. the first row of β', in the Johansen cointegration test results for the data is [1,-1.0124]. Imposing the restriction of the first row of β' equals [1,-1] produces the $\mathcal{X}^2$ statistic reported in the last row of Table 8.4. It concludes that there is not enough support for the restriction. It is quite different conclusion compared to Monoyios and Sarno (2002) where they concluded that only one cointegrating relationship exists and the cointegrating relationship with the restriction of $[1, -1]$ can be supported.

Table 8.5 reports linearity tests results. The first linearity test employed is a RESET test (see Ramsey, 1969) of the null hypothesis of linearity of the residuals from an AR(10) for mb_t against the alternative hypothesis of general model misspecification involving a higher-order polynomial to represent a different functional form. Under the null hypothesis, the statistics is distributed as $\mathcal{X}^2(q)$ with q is equal to the number of higher-order terms in alternative model. Table 8.5 reports the result from executing RESET test statistic where the alternative model with a quadratic and a cubic terms are included. We have tried the test for $d = \{1, \ldots, 5\}$ but only report for $d = \{1, 2\}$ as the results are the same. The null hypothesis is very strongly rejected considered

Table 8.5: Linearity tests on the demeaned basis mb_t

RESET Test			11.789 [0.0000] **
Lags Used			10
d	LM^G	LM^3	LM^E
p =7			
1	15.615 [0.0000] **	7.0953 [0.0000] **	19.6 [0.0000] **
2	15.615 [0.0000] **	7.0953 [0.0000] **	19.6 [0.0000] **
p =10			
1	11.027 [0.0000] **	4.5914 [0.0000] **	14.076 [0.0000] **
2	11.027 [0.0000] **	4.5914 [0.0000] **	14.076 [0.0000] **

Notes: RESET test statistics are computed considering a linear AR(p) regression with 10 lags without a constant as the constant is not significant at 5% significant level against an alternative model with a quadratic and a cubic term. The F-statistics forms are used for the RESET test, LM^G , LM^3 and LM^E and the values in parentheses are the p-values. * and ** superscripts indicate significance at 5% and 1%, respectively.

with the p-value of virtually zero, suggesting that a linear AR(10) process for mb_t is misspecified.

The second linearity tests are based on Terasvirta (1994). The tests [5] can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. The artificial regression of (8.24) is estimated. Lag lengths of 7 and 10 are considered for executing the linearity tests for mb_t using the artificial regression in (8.25). Table 8.5 shows values of the test statistics LM^G, LM^3 and LM^E. The delay parameter $d \in \{1, 2, 3, 4, 5\}$ are considered. However, the test statistics LM^G, LM^3 and LM^E show that different values of d do not affect the results. From Table 8.5, the p-values from LM^G, LM^3 and LM^E statistics are virtually zero for both $p = 7$ and $p = 10$. From the LM^G statistics, we can conclude that linearity is strongly rejected. From the LM^3 and LM^E statistics, we can conclude that a LSTAR model is much more strongly supported than an ESTAR model. It is quite different conclusion compared to Monoyios and Sarno (2002) where they made the opposite conclusion, i.e. an ESTAR model is more favoured than a LSTAR model. From Table 8.5, the LM^G, LM^3 and LM^E statistics for $p = 7$ are higher than those for $p = 10$. Therefore, we choose a LSTAR model with $p = 7$ and $d = 1$ for model estimation.

Table 8.6 reports comparison of model estimation results for a nonlinear LSTAR model with $p = 7$ and $d = 1$ and a linear AR(7) model. The nonlinear LSTAR model

[5]See also in Subsection 2.1.2 for explanation of the test.

Table 8.6: Estimation results for the demeaned basis mb_t

	LSTAR(7)	AR(7)
$\hat{\theta}_{11}(= -\hat{\theta}_{21})$	0.3642 (0.0177)	0.3672 (0.0179)
$\hat{\theta}_{12}(= -\hat{\theta}_{22})$	0.1885 (0.0189)	0.1864 (0.0190)
$\hat{\theta}_{13}(= -\hat{\theta}_{23})$	0.0856 (0.0193)	0.0902 (0.0193)
$\hat{\theta}_{14}(= -\hat{\theta}_{24})$	0.1086 (0.0192)	0.1095 (0.0192)
$\hat{\theta}_{15}(= -\hat{\theta}_{25})$	0.0556 (0.0193)	0.0604 (0.0193)
$\hat{\theta}_{16}(= -\hat{\theta}_{26})$	0.0084 (0.0189)	0.0131 (0.0190)
$\hat{\theta}_{17}(= -\hat{\theta}_{27})$	0.0713 (0.0176)	0.0708 (0.0178)
θ	-26.0070 (8.7306)	
SSE	0.0214	0.0214
LR	0.000 [1.000]	0.000 [1.0000]
SW	0.8912 [0.000]***	0.8922 [0.000]***
BL (20)	40.2877 [0.0046]*	41.3718 [0.0033]*

Notes: Figures in parentheses beside coeficient estimates denote estimated standard errors. SSE is sum square error; LR is a likelihood ratio statistics for parameter restrictions; SW is a Shapiro-Wilk normality test for residuals; BL is a Box-Ljung test for residuals using 20 lags; the figures in parentheses denote the p-values.

estimation uses a nonlinear least squares method in the form of

$$mb_t = \theta_{1,0}+\sum_{j=1}^{7}\theta_{1,j}\, mb_{t-j}+\left[\theta_{2,0}+\sum_{j=1}^{7}\theta_{2,j}\, mb_{t-j}\right]\left[\frac{1}{1+\exp\{-\theta(mb_{t-1}-e)\}}-\frac{1}{2}\right]+\epsilon_t. \tag{8.27}$$

As the mean of mb_t is zero, theoretically, $\theta_{1,0} = \theta_{2,0} = e = 0$. Further restriction of $\theta_{2,j} = -\theta_{1,j}$ for $j = 1, \cdots, 7$ produces the likelihood ratio statistic, LR, in Table 8.6 concluding that the restrictions can not be rejected at the conventional 5% significance level. A linear AR(7) is also estimated as a comparison. The LR statistic comparing the LSTAR model and the AR(7) model concludes that there is no significant difference between the two models. Furthermore, the parameter estimates of $\theta_{1,j}$, $j = 1, \cdots, 7$, for the two models are quite similar. Other statistics such as Shapiro-Wilk normality test and Box-Ljung autocorrelation test for residuals are also similar for the two models. Monoyios and Sarno (2002) did not make model comparison and they concluded that a nonlinear ESTAR model quite fits with their data.

In conclusion, using current available data, from January 1, 1998 to December 31, 2009, we examine the basis of S&P 500 following procedures in Monoyios and Sarno (2002). Even though we can conclude that there is possibility nonlinearity in the basis, there is no significant difference between a nonlinear LSTAR model and a linear autoregressive model in fitting the data. It is a different conclusion compared to Monoyios and Sarno (2002) concluding that a nonlinear ESTAR model quite fits with the data they have.

Our data has two major financial crises while the data used by Monoyios and Sarno (2002) does not have a major financial crisis. This different data characteristic may lead to different conclusions.

We also have a concern in the way the basis is constructed. By pairing up the spot price with the future contract with the nearest maturity, it may produce artificial jumps at the time of maturity. The longer the time to maturity, the higher the difference between the future price and the spot price. For example for S&P 500, it has 4 maturity times during a year which are the third Friday in March, June, September and December. We find that at those times, there are jumps in the basis. Figure 8.3 shows the plot of b_t from January 1, 1998 to October 19, 1998 with jumps on the third Friday in March, June, September 1998. Monoyios and Sarno (2002) did not discuss this issue. Ma *et al.* (1992) argued that it may create volatility and bias in the parameter estimates. Therefore, the next step of this research will examine the cointegration of f_t and s_t with a time trend for each future contract.

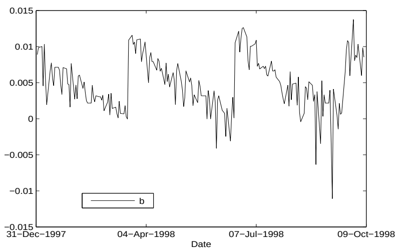


Figure 8.3: Plot of b_t from January 1, 1998 to October 19, 1998.

8.3 Cointegration Analysis with a Time Trend Between Future Contract and Spot Index Prices of S&P 500

As in Chapter 5, in order to determine whether cointegration exists between spot and future prices, we apply two techniques, e.g.: the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the technique developed by Johansen (1988). However, in this section, we include a time trend in the model and we analyse the cointegration for each future contract. Table 8.7 lists the future contract codes, and the corresponding data periods and number observations. The plots of the log of future prices (f_t) and the log of spot index prices (s_t) can be seen in panel (a) Figures 8.4-8.47.

8.3.1 Engle-Granger Approach

Consider a model

$$f_t = \mu + \beta\, s_t + \delta\, t^* + \epsilon_t \tag{8.28}$$

where f_t and s_t are the log of future and spot index prices, $t^* = (T - t)$ is a time to maturity. Literature (see for example Monoyios and Sarno, 2002 and Sarno and Valente, 2000) suggested that f_t and s_t are both $I(1)$ and if $\epsilon_t \sim I(0)$, the two series will be cointegrated of order 1.

Harris (1995) noted that estimating (8.28) using OLS achieves a consistent estimate of the long-run steady-state relationship between the variables in the model, and all dynamics and endogeneity issues can be ignored asymptotically. This arises because of what is termed the "superconsistency" property of the OLS estimator when the series are cointegrated. According to the the "superconsistency" property, if f_t and s_t are both non-stationary $I(1)$ variables and $\epsilon_t \sim I(0)$, then as sample size, T, becomes larger the OLS parameter estimates of (8.28) converges to their true values at a much faster rate than the usual OLS estimator with stationary $I(0)$ variables.

The next step of Engle-Granger approach is to test whether the residuals $\epsilon_t \sim I(1)$ against $\epsilon_t \sim I(0)$. Engle and Granger (1987) suggested the augmented Dickey-Fuller (ADF) test (Fuller, 1976) as follow:

$$\Delta\hat{\epsilon}_t = \rho\,\hat{\epsilon}_{t-1} + \sum_{i=1}^{p-1} \omega_i\,\Delta\hat{\epsilon}_{t-i} + \upsilon_t \tag{8.29}$$

where $\upsilon_t \sim IID\ N(0, \sigma^2)$ and the $\hat{\epsilon}_t$ are obtained from estimating (8.28). The plots of $\hat{\epsilon}_t$ can be seen in panel (b) Figures 8.4-8.47.

The null hypothesis of nonstationary (i.e., the series has a unit root) and thus no cointegration, $\rho = 0$, can be tested using a t-statistic with a non-normal distribution. The lag length of the augmentation terms, p, is chosen as the minimum necessary to reduce the residuals to white noise. Critical values for this test have been calculated using Monte Carlo methods available in Fuller (1976). However, unless the true parameter values in (8.28) are known, it is not possible to use the standard Dickey-Fuller tables of critical values. There are two reasons for this. First, because of the way it is constructed the OLS estimator 'chooses' the residuals in (8.28) to have the smallest sample variance, even if the variables are not cointegrated, making the residuals appear as stationary as possible. Thus, the ADF tests will tend to over-reject the null. Second, the distribution of the test statistic under the null is affected by the number of regressors (n) included in (8.28). Thus, different critical values are needed as n changes and also whether constant and/or trend are included along with the sample size. Taking into account all of these, MacKinnon (1992) has linked the critical values for particular tests to a set of parameters of a equation of the response surfaces. However, as the number of regressor in (8.28) is one (excluding the constant and trend), the critical values by MacKinnon (1992) will be the same as the ADF tests.

Other alternatives to test a unit root are the Z-tests suggested by Phillips (1987), Perron (1988) and Phillips and Perron (1988b). However, Monte Carlo work by Schwert

(1989) suggest that the tests have poor size properties (i.e., the tendency to over-reject the null when it is true) when MA terms present in many economic time series. Recently, Ng and Perron (2001) develop a unit root test that is robust to ARIMA representations with large negative MA term and also improve the power of standard ADF tests by using a GLS detrending/demeaning procedure. However, Crowder and Phengpis (2005) conclude that there are no significant difference of results from the ADF tests and Ng and Perron (2001) when they apply the tests to the S&P 500 future and spot index prices. Therefore, in this thesis only the ADF test is used.

Tables 8.8-8.11 report the ADF unit root tests for f_t and s_t as well as their first difference series. Generally, we can conclude that f_t and s_t are both $I(1)$, except for Sep00 and Sep02. For Sep00, there are significant evidences at a 5% significant level that the null hypothesis of a unit root with a constant as well as a constant and a time trend can be rejected. Thus, for this period, f_t and s_t may be considered as trend stationary series. For Sep02, the ADF unit root test statistic without a constant and a time trend is significant at a 5% significant level. However, if we look at the plots of f_t and s_t in Figure 8.22(a), there is a decreasing trend suggesting that there is a possible drift. Thus, we should use the ADF test with a constant.

Tables 8.12-8.13 report the regression model estimates based on (8.28). From these tables, the constant parameter (μ) is significant, except for Sep02, Sep04, Dec05, Mar06, Mar08 and Sep08. All future contracts show that the parameter β corresponding to the variable s_t are significant with the values close to 1. Generally, the parameter δ corresponding to the variable $t*$ are significant, except for Jun05 and Sep05. However, we are aware that these significance are only indicative because as Phillips and Durlauf (1986) noted, in finite samples, the asymptotic distribution of the OLS estimators in (8.28) and associated t-statistics are highly complicated and non-normal and thus the standard tests are invalid. These complications are confirmed in Tables 8.14-8.15 reporting the regression residuals ($\hat{\epsilon}_t$) analysis. These tables indicate that generally, the residuals violate the OLS assumptions, e.g.: uncorrelated, normally distributed and homocedasticity. Thus, the standard tests for inferences are invalid.

Tables 8.14-8.15 also report the ADF unit root tests for the regression residuals. There are mixed results of the unit root significance. For 28 out of 44 future contracts conclude that there is enough evidence at a 5% significance level to reject the null hypothesis of a unit root. Usually the regression residuals are not stationary during financial crisis. It is reasonable as during financial crisis, the economic and financial conditions are very uncertain resulting in very high volatility in economic and financial data.

Tables 8.16-8.17 report the fitted autoregressive model for regression residuals, denoted as $\hat{\epsilon}_t$. From Tables 8.14-8.15, Mar98 and Jun98 have white noise regression residuals. For other future contracts, autoregressive models with 30 lags are examined to obtain the highest significant lag. Then, using these lags, the most parsimonious autoregressive models are obtained by excluding non significance lags and leaving only significant lags at a 5% significance level. The residuals analysis of these models are provided in Tables 8.18-8.19. From these tables, the autocorrelation problem can be eliminated. However, ARCH, normality and heteroscedasticity problems are still eminent for most of the future contracts.

Table 8.18 reports nonlinear analysis for $\hat{\epsilon}_t$. Similar to nonlinear analysis for the basis of S&P 500 in Section 8.2, we apply RESET tests based on Ramsey (1969) and STAR linearity tests based on Terasvirta (1994). For the RESET tests, the null hypothesis of linearity of the residuals from autoregressive model of $\hat{\epsilon}_t$ in Table 8.16-8.17 are tested against the alternative hypothesis of general model misspecification involving a quadratic and a cubic terms. Under the null hypothesis, the statistics is distributed as $\mathcal{X}^2(3)$. From the RESET test results, the null hypothesis of linearity cannot be rejected at a 0.1% significance level, except for Dec00. It suggests that there is no strong evidence of a quadratic and a cubic terms for $\hat{\epsilon}_t$ series.

Another linearity tests is STAR test based on Terasvirta (1994) (see also in Subsection 2.1.2 for explanation of the test). The null hypothesis of this test is that $\hat{\epsilon}_t$ is a linear autoregressive model against a nonlinear ESTAR or LSTAR model. In the analysis, the delay parameter d is set as $d = 1$ as previous papers such as Monoyios and Sarno (2002) and Kapetanios *et al.* (2003) found that using $d = 1$ gave the largest value of the test statistics. The tests can also be used to discriminate between ESTAR or LSTAR models since the third-order terms disappear in the Taylor series expansion of the ESTAR transition function. Under alternative hypothesis $\hat{\epsilon}_t$ follows a nonlinear LSTAR or ESTAR model, an artificial regression is estimated as follow:

$$\hat{\epsilon}_t = \phi_{0,0} + \sum_{j=1}^{p} \left(\phi_{0,j}\, \hat{\epsilon}_{t-j} + \phi_{1,j}\, \hat{\epsilon}_{t-j}\, \hat{\epsilon}_{t-d} + \phi_{2,j}\, \hat{\epsilon}_{t-j}\, \hat{\epsilon}^2_{t-d} + \phi_{3,j}\, \hat{\epsilon}_{t-j}\, \hat{\epsilon}^3_{t-d}\right) + error \tag{8.30}$$

where d is set equal to 1 and p is the lag length chosen as the highest significance lag as in autoregressive models in Tables 8.16-8.17. Keeping the delay parameter d fixed, testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}$$

against its complement is a general test (LM^G) of the hypothesis of linearity against smooth transition nonlinearity. Given that the ESTAR model implies no cubic terms in the artificial regression(i.e., $\phi_{3,j} = 0$ if the true model is an ESTAR model, but $\phi_{3,j} \neq 0$ if the true model is an LSTAR), thus, testing the null hypothesis that

$$H_0 : \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}$$

provides a test (LM^3) of ESTAR nonlinearity against LSTAR-type nonlinearity. Moreover, if the restrictions $\phi_{3,j} = 0$ cannot be rejected at the chosen significance level, then a more powerful test (LM^E) for linearity against ESTAR-type nonlinearity is obtained by testing the null hypothesis

$$H_0 : \phi_{1,j} = \phi_{2,j} = 0 | \phi_{3,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}.$$

However, running the regression in (8.30) does not work as the values of $\hat{\epsilon}_t$ are small so that $\hat{\epsilon}_{t-j}\ \hat{\epsilon}_{t-d}^2$ and $\hat{\epsilon}_{t-j}\ \hat{\epsilon}_{t-d}^3$ for $j = 1, \ldots, p$ are virtually zero. Thus, (8.30) reduces to become:

$$\hat{\epsilon}_t \quad = \quad \phi_{0,0} + \sum_{j=1}^{p} (\phi_{0,j}\ \hat{\epsilon}_{t-j} + \phi_{1,j}\ \hat{\epsilon}_{t-j}\ \hat{\epsilon}_{t-d}) + error. \tag{8.31}$$

Using (8.31), the test that can be done is only (LM^G). It tests the null hypothesis

$$H_0 : \phi_{1,j} = 0, \qquad \forall j \in \{1, 2, \cdots, p\}$$

against its complement. If the null hypothesis can be rejected, then there is an evidence of nonlinearity in $\hat{\epsilon}_t$. However, it is difficult to differentiate between LSTAR or ESTAR model as we only have (LM^G). But it does not really matter because we are interested in whether $\hat{\epsilon}_t$ is linear or nonlinear. From Table 8.18, (LM^G) test statistics show that only 3 out of 44 future contracts (e.g. Dec00, Sep03 and Dec08) show strong evidences of nonlinearity with a 0.1% significance level. In conclusion, we can say that there is no strong evidence of nonlinearity in the relationship between future and spot prices of S&P 500 index.

8.3.2 Johansen Approach

In Chapter 2, the Johansen cointegration test was introduced. However, it does not include deterministic components (i.e.: intercept and/or time trend). The question of whether an intercept and/or time trend should enter the short- and/or long-run model is raised. Osterwald-Lenum (1992) and Johansen (1994) developed cointegration analysis to include the deterministic components. Part of this subsection is taken from Johansen (1994).

Consider the vector error correction model expanded to include an intercept and/or time trend:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{y}_{t-1} + \sum_{j=1}^{k-1} \Gamma_j\ \Delta \mathbf{y}_{t-j} + \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1\ t + \boldsymbol{\epsilon}_t \tag{8.32}$$

where $\mathbf{y}_t$ is a p-dimensional vector of variables integrated with the same order, Γ_j are $(p \times p)$ coefficient matrices, k is chosen as the minimum necessary to reduce the residuals $\boldsymbol{\epsilon}_t$ to a p-dimensional white noise error vector with covariance matrix Ω.

Johansen (1994) defined $\Gamma = I - \sum_{j=1}^{k-1}\Gamma_j$ and applied Granger's Theorem (see Engle and Granger, 1987) in the form given in Johansen (1989) stating that if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have rank r and $\boldsymbol{\alpha}_\perp^{'}\Gamma\boldsymbol{\beta}_\perp$ has full rank, where $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ are $p\times(p-r)$ matrices of full rank orthogonal to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, then $\mathbf{y}_t$ has the representation as follow:

$$\mathbf{y}_t = C\sum_{i=1}^{t}\boldsymbol{\epsilon}_i + \frac{1}{2}C\ \boldsymbol{\mu}_1\ t^2 + \boldsymbol{\tau}_1\ t + \boldsymbol{\tau}_0 + \mathbf{x}_t + \mathbf{A} \tag{8.33}$$

where $\mathbf{x}_t$ is a stationary process and $\boldsymbol{\beta}'\mathbf{A} = 0$. Here $C = \boldsymbol{\beta}_\perp\left(\boldsymbol{\alpha}_\perp^{'}\Gamma\boldsymbol{\beta}_\perp\right)^{-1}\boldsymbol{\alpha}_\perp^{'}$. The coefficients $\boldsymbol{\tau}_0$ and $\boldsymbol{\tau}_1$ are rather complicated, but can be found by inserting (8.33) into (8.32) and identifying coefficients to 1 and t, as functions of the parameters and the initial values of the model (8.32).

Note that if the cointegration rank is zero, $r=0$, then $\boldsymbol{\alpha}=\boldsymbol{\beta}=\mathbf{0}$ and we can take $\boldsymbol{\alpha}_\perp = \boldsymbol{\beta}_\perp = \mathbf{I}_{p\times p}$. In this case $\mathbf{y}_t$ does not cointegrate. If $r=p$, $\boldsymbol{\alpha}_\perp = \boldsymbol{\beta}_\perp = \mathbf{0}$ and $C=\mathbf{0}$, such that $\mathbf{y}_t$ is stationary. Also note from (8.33) that inclusion of a constant, $\boldsymbol{\mu}_0 \neq \mathbf{0}$, in (8.32) means that the underlying data generating process will include a linear trend in $\mathbf{y}_t$. Similarly, the inclusion of a time trend, $\boldsymbol{\mu}_1 \neq \mathbf{0}$, in (8.32) means that the underlying data generating process will include a quadratic trend in $\mathbf{y}_t$.

From (8.33), $\mathbf{y}_t$ has stochastic part, i.e. $\sum_{i=1}^{t}\boldsymbol{\epsilon}_i$, which is a nonstationary process with stationary differences, a so called as I(1) process. However, the reduced rank of $\Phi = \boldsymbol{\alpha}\boldsymbol{\beta}'$ implies that the stochastic part of $\boldsymbol{\beta}'\mathbf{y}_t$ is stationary. The space sp($\boldsymbol{\beta}$) is called the cointegrating space and the vectors in sp($\boldsymbol{\beta}$) are called the cointegrating vectors. The vectors $\boldsymbol{\alpha}$ are the adjustment vectors since they measure the rate of adjustment of the process $\mathbf{y}_t$ to the disequilibrium error $\boldsymbol{\beta}'\mathbf{y}_{t-1}$.

Osterwald-Lenum (1992) and Johansen (1994) stressed that it is important to understand the role played by the constant and trend terms in (8.32) because the asymptotic distributions of the tests for cointegration depend on the deterministic specifications. Osterwald-Lenum (1992) and Johansen (1994) discussed the role of the deterministic term, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$, under the assumption of reduced rank and various restrictions on $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. They suggested that the behaviour of the deterministic trend of the process $\mathbf{y}_t$ depends critically on the relation between $\boldsymbol{\mu}_t$ and the adjustment coefficient $\boldsymbol{\alpha}$. To analyse this, they decomposed the parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ in the directions of $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_\perp$ as follows:

$$\begin{aligned}\boldsymbol{\mu}_0 &= \boldsymbol{\alpha}\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_\perp\boldsymbol{\gamma}_0\\ \boldsymbol{\mu}_1 &= \boldsymbol{\alpha}\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_\perp\boldsymbol{\gamma}_1\end{aligned} \tag{8.34}$$

where $\boldsymbol{\beta}_0 = (\boldsymbol{\alpha}^{'}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}^{'}\boldsymbol{\mu}_0$ is an r-dimensional vector of intercepts in the cointegrating relations or means of the stationary variables, $\boldsymbol{\gamma}_0 = (\boldsymbol{\alpha}_\perp^{'}\boldsymbol{\alpha}_\perp)^{-1}\boldsymbol{\alpha}_\perp^{'}\boldsymbol{\mu}_0$ is a $(p-r)$-dimensional vector of linear trend slopes in the data, $\boldsymbol{\beta}_1 = (\boldsymbol{\alpha}^{'}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}^{'}\boldsymbol{\mu}_1$ is an r-

dimensional vector of linear trend coefficients in the cointegrating relations and $\boldsymbol{\gamma}_1 = (\boldsymbol{\alpha}_{\perp}^{'}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}_{\perp}^{'}\boldsymbol{\mu}_1$ is a $(p-r)$-dimensional vector of quadratic trends coefficients in the data. Restrictions on the deterministic components in (8.34) impose the restriction on the parameters $\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$.

Osterwald-Lenum (1992) and Johansen (1994) suggested that there are 5 models of the deterministic components in (8.32). The asymptotic distributions of the cointegration rank tests are not invariant under different combinations of zero-restrictions on $\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$ in the data generating process. The asymptotic distributions of the tests are given by an expression involving stochastic integrals of Brownian motions including certain deterministic components of the following form:

$$\int dBF^{'}\left[\int FF^{'}du\right]^{-1} FdB^{'}. \tag{8.35}$$

where B is a vector Brownian motion and F is a function of B and includes certain deterministic components which are different for different models. Osterwald-Lenum (1992) and Johansen (1994) developed tables of critical values for the five models:

- **Model 1.** $\boldsymbol{\mu}_t = \mathbf{0}$.
  There are no deterministic components in the data or in the cointegration relations. This is unlikely to occur in practical situations, especially as the intercept is generally needed to account for the measurement units of the variables in $\mathbf{y}_t$. The critical values for this model are provided in Table 0 in Osterwald-Lenum (1992).

- **Model 2.** $\boldsymbol{\mu}_t = \boldsymbol{\alpha}\boldsymbol{\beta}_0$.
  In this model, there are no linear trends in the levels of the data, such that the first-differenced series have a zero mean. Thus, the intercept is restricted to the long-run (the cointegration space) to take into account for the measurement units of the variables in $\mathbf{y}_t$. The critical values for this model are provided in Table 1* in Osterwald-Lenum (1992).

- **Model 3.** $\boldsymbol{\mu}_t = \boldsymbol{\alpha}\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_{\perp}\boldsymbol{\gamma}_0 = \boldsymbol{\mu}_0$.
  In this model, there are linear trends in the levels of the data, such that the first-differenced series have drifts. However, it is assumed that the intercept in the cointegration vectors is cancelled by the intercept in the short-run model, leaving only an intercept in the short-run model (i.e., $\boldsymbol{\mu}_0$). The critical values for this model are provided in Table 1 in Osterwald-Lenum (1992).

- **Model 4.** $\boldsymbol{\mu}_t = \boldsymbol{\alpha}\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_{\perp}\boldsymbol{\gamma}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}_1\ t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}\boldsymbol{\beta}_1 t$.
  In this model, there are no quadratic trends in the levels of the data, then there will be no time trend in the short-run model, but if there is some long-run linear growth which the model cannot account for, given the chosen data-set, then we

allow for a linear trend in the cointegration vectors. Thus, the cointegration space includes time as a trend-stationary variable. It is also assumed that the intercept in the cointegration vectors is canceled by the intercept in the short-run model, leaving only an intercept in the short-run model (i.e., $\boldsymbol{\mu}_0$). The critical values for this model are provided in Table 2* in Osterwald-Lenum (1992) and Table V in Johansen (1994).

- **Model 5.** $\boldsymbol{\mu}_t = \boldsymbol{\alpha}\boldsymbol{\beta}_0 + \boldsymbol{\alpha}_\perp\gamma_0 + (\boldsymbol{\alpha}\boldsymbol{\beta}_1 + \boldsymbol{\alpha}_\perp\gamma_1)\ t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1\ t.$
In this model, there are no restrictions for the deterministic components if there are quadratic trends in the levels of the data then we allow for linear trends in the short-run model. The critical values for this model are provided in Table 2 in Osterwald-Lenum (1992) and Table IV in Johansen (1994).

The question of which model should be used is not easily answered a priori. However, as in this thesis we want to confirm the cost-carry model in (8.18) and there is no indication of a quadratic trend in the future and spot index prices, then we use Model 4. Model 4 allows for r trend stationary variables and $p-r$ variables that are composed of I(1) variables and a linear trend. Thus, if $r = p$, it means all variables in $\mathbf{y}_t$ are trend stationary series.

Now, consider $\mathbf{y}_t = (f_t, s_t)'$ where f_t and s_t are the log of future and spot index prices of the S&P 500. Using deterministic Model 4, thus, (8.32) will become:

$$\begin{aligned}
\Delta\mathbf{y}_t &= \boldsymbol{\alpha}(\boldsymbol{\beta}^{'}, \boldsymbol{\beta}_1)(\mathbf{y}_{t-1}^{'}, t)^{'} + \sum_{j=1}^{k-1}\Gamma_j\ \Delta\mathbf{y}_{t-j} + \boldsymbol{\mu}_0 + \boldsymbol{\epsilon}_t \\
\begin{pmatrix} \Delta f_t \\ \Delta s_t \end{pmatrix} &= \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}\begin{pmatrix} 1 & b_1 & \beta_{1,1} \\ b_2 & 1 & \beta_{1,2} \end{pmatrix}\begin{pmatrix} f_{t-1} \\ s_{t-1} \\ \mathrm{t} \end{pmatrix} \\
&\quad + \sum_{j=1}^{k-1}\Gamma_j \begin{pmatrix} \Delta f_{t-j} \\ \Delta s_{t-j} \end{pmatrix} + \begin{pmatrix} \mu_{0,1} \\ \mu_{0,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \qquad (8.36)
\end{aligned}$$

To analyse cointegration, PcFiml software is used. Variables f_t and s_t are endogenous. A time trend t is added as an exogenous variable, so it is restricted to enter in the cointegration space (long-run) only. Constant variable is selected as an unrestricted variable.

Tables 8.21-8.24 report the Johansen cointegration analysis results for the 44 future contracts with lag= k chosen as the minimum lag so that the vector error $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ are not correlated. Figures 8.4-8.47 panel (c) show the de-meaned disequilibrium error using the first cointegration vector, i.e. $f_t + b_1\ s_t + \beta_{1,1}$ Trend. For Mar98, Jun98 and Sep98, lag=1 is enough. For other future contracts, higher lags are needed to make the error become uncorrelated. In only 19 out of 44 future contracts, it can be concluded that there is at least one cointegration vector at a 5% significant level. It may be due to limited number of observations, i.e. we only have maximum 2 years data for a future contract.

Table 8.25 compares cointegration analysis results from both approaches, i.e. the Engle-Granger approach and the Johansen approach. Generally, both approaches have the same conclusions. Only 7 out of 44 future contracts have different conclusions with the Johansen method has more rejection of the null hypothesis of cointegration. Even though there are some criticisms for the Engle-Granger approach, the Johansen approach sometimes also produce inconsistent results. For an example, Mar04, the Johansen test conclude that there is one cointegration vector. However, if we look at the plot of disequilibrium error in Figure 8.28 panel (c), it does not seem stationary. Lack of papers in theoretical ground for cointegration analysis with a time trend from both approaches give an opportunity for further research in this area to give more interpretation of the results.

The aim of this study is to determine whether f_t and s_t are cointegrated with a time trend for each future contract. However, from Table 8.25, only 19 out of 44 future contracts conclude that they are cointegrated with a time trend using both the Engle-Granger approach and the Johansen approach. Perhaps, high volatility during financial crisis in the data period affects the cointegration test. We see that the results from Mar98 - Dec99 when there was no financial crisis, we strongly conclude that f_t and s_t are cointegrated with a time trend. Furthermore, it happens may be due to limited number of observations, i.e. we only have maximum 2 years data for a future contract.

Table 8.7: Future Contracts

Future Contract Code	Data Period	No. Observations
Mar98	25/3/1997 - 19/3/1998	258
Jun98	18/6/1997 - 18/6/1998	262
Sep98	10/6/1997 - 18/9/1998	334
Dec98	18/6/1997 - 17/12/1998	392
Mar99	4/9/1997 - 18/3/1999	401
Jun99	2/7/1997 - 17/6/1999	512
Sep99	18/9/1997 - 16/9/1999	521
Dec99	18/12/1997 - 16/12/1999	521
Mar00	20/3/1998 - 16/3/2000	520
Jun00	18/6/1998 - 15/6/2000	521
Sep00	21/9/1998 - 15/9/2000	520
Dec00	22/12/1998 - 14/12/2000	518
Mar01	22/3/1999 - 15/3/2001	519
Jun01	23/6/1999 - 14/6/2001	517
Sep01	23/9/1999 - 20/9/2001	521
Dec01	30/12/1999 - 20/12/2001	516
Mar02	21/3/2000 - 15/3/2002	519
Jun02	19/6/2000 - 20/6/2002	524
Sep02	14/9/2000 - 19/9/2002	526
Dec02	9/1/2001 - 19/12/2002	508
Mar03	19/3/2001 - 21/3/2003	525
Jun03	15/6/2001 - 19/6/2003	525
Sep03	21/9/2001 - 18/9/2003	520
Dec03	26/12/2001 - 19/12/2003	518
Mar04	3/4/2002 - 18/3/2004	512
Jun04	21/6/2002 - 18/6/2004	521
Sep04	20/9/2002 - 17/9/2004	521
Dec04	20/12/2002 - 17/12/2004	521
Mar05	21/3/2003 - 18/3/2005	521
Jun05	19/6/2003 - 17/6/2005	522
Sep05	19/9/2003 - 16/9/2005	521
Dec05	19/12/2003 - 15/12/2005	520
Mar06	19/3/2004 - 16/3/2006	520
Jun06	18/6/2004 - 15/6/2006	520
Sep06	17/9/2004 - 15/9/2006	521
Dec06	17/12/2004 - 15/12/2006	521
Mar07	18/3/2005 - 15/3/2007	520
Jun07	17/6/2005 - 14/6/2007	520
Sep07	19/9/2005 - 20/9/2007	524
Dec07	16/12/2005 - 20/12/2007	525
Mar08	17/3/2006 - 19/3/2008	524
Jun08	16/6/2006 - 19/6/2008	525
Sep08	15/9/2006 - 18/9/2008	525
Dec08	15/12/2006 - 18/12/2008	525

Table 8.8: ADF Unit Root Tests of the Log of Future Contract Prices (f_t) and the Log of Spot Index Prices (s_t)

Future Code	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag
Mar98	f_t	2.1768	-1.5437	-2.2729	10	Δf_t	-4.4980**	-5.0329**	-5.0443**	9
	s_t	2.3285	-1.4064	-2.3403	10	Δs_t	-4.0468**	-4.7051**	-4.7157**	9
Jun98	f_t	1.4426	-0.99186	-1.9633	5	Δf_t	-9.5575**	-9.6875**	-9.6684**	4
	s_t	1.8321	-0.83059	-1.9169	5	Δs_t	-9.2733**	-9.4971**	-9.4785**	4
Sep98	f_t	0.36241	-1.5950	-1.5052	29	Δf_t	-3.0481**	-3.0618*	-3.2368	28
	s_t	0.64582	-1.5119	-1.4076	29	Δs_t	-3.0351**	-3.0987*	-3.2833	28
Dec98	f_t	0.62933	-1.5551	-2.7150	27	Δf_t	-2.9916**	-3.0564*	-3.0606	26
	s_t	0.87586	-1.1460	-2.6690	27	Δs_t	-2.9398**	-3.0694*	-3.0715	26
Mar99	f_t	1.2622	-0.62325	-1.7703	17	Δf_t	-5.0355**	-5.1960**	5.2205**	16
	s_t	1.6015	-0.38532	-1.7432	17	Δs_t	-4.9712**	-5.2340**	-5.2545**	16
Jun99	f_t	1.2825	-0.71837	-2.2270	17	Δf_t	-5.9441**	-6.0870**	-6.0902**	16
	s_t	1.7083	-0.50930	-2.2239	17	Δs_t	-5.8341**	-6.0930**	-6.0939**	16
Sep99	f_t	1.0564	-1.0737	-2.2119	17	Δf_t	-6.1027**	-6.1962**	-6.1894**	16
	s_t	1.4884	-0.92445	-2.1858	17	Δs_t	-6.0096**	-6.2019**	-6.1965**	16
Dec99	f_t	1.2137	-1.7541	-2.6194	19	Δf_t	-4.8449**	-5.0027**	-5.0146**	18
	s_t	1.5194	-1.4687	-2.6383	19	Δs_t	-4.7373**	-4.9887**	-5.0002**	18
Mar00	f_t	0.73206	-1.3727	-2.7698	27	Δf_t	-3.8089**	-3.8726**	-3.8683*	26
	s_t	1.0933	-1.0784	-2.7536	27	Δs_t	-3.7709**	-3.9229**	-3.9175*	26
Jun00	f_t	0.60862	-1.3137	-2.9449	27	Δf_t	-4.0057**	-4.0500**	-4.0469**	26
	s_t	0.95651	-1.0069	-2.9071	27	Δs_t	-3.9635**	-4.0788**	-4.0736**	26
Sep00	f_t	1.4145	-3.3841*	-3.5879*	29	Δf_t	-4.5090**	-5.0657**	-5.3259**	28
	s_t	1.7361	-2.8776*	-3.5676*	29	Δs_t	-4.4979**	-5.1003**	-5.3263**	28
Dec00	f_t	0.12492	-2.2510	-1.4651	24	Δf_t	-5.9423**	-5.9375**	-6.2142**	23
	s_t	0.58617	-2.2499	-1.6849	24	Δs_t	-5.9938**	-6.0197**	-6.2325**	23

Table 8.9: ADF Unit Root Tests of the Log of Future Contract Prices (f_t) and the Log of Spot Index Prices (s_t)

Future Code	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag
Mar01	f_t	-0.96447	0.20592	-0.14589	27	Δf_t	-4.4763**	-4.5627**	-4.9554**	26
	s_t	-0.59342	-0.82686	-0.27177	27	Δs_t	-4.6380**	-4.6552**	-4.9714**	26
Jun01	f_t	-0.86818	-0.52815	-1.6364	24	Δf_t	-5.1479**	-5.2174**	-5.3211**	23
	s_t	-0.48657	-1.1758	-1.6809	24	Δs_t	-5.2933**	-5.3095**	-5.3798**	23
Sep01	f_t	-1.3130	0.83280	-1.6358	13	Δf_t	-5.0927**	-5.2476**	-5.6487**	12
	s_t	-1.0021	0.34895	-1.6128	13	Δs_t	-5.1402**	-5.2207**	-5.6055**	12
Dec01	f_t	-1.2263	-0.80119	-3.0051	28	Δf_t	-3.9120**	-4.0969**	-4.0923**	27
	s_t	-0.85411	-1.0893	-3.1671	28	Δs_t	-3.9391**	-4.0261**	-4.0247**	27
Mar02	f_t	-1.3927	-1.4112	-2.3905	13	Δf_t	-5.7360**	-5.9004**	-5.9390**	12
	s_t	-0.99411	-1.6195	-2.7596	13	Δs_t	-5.8146**	-5.8942**	-5.9221**	12
Jun02	f_t	-1.9147	-1.0773	-2.4821	2	Δf_t	-17.646**	-17.790**	-17.775**	1
	s_t	-1.5450	-1.0806	-2.7054	2	Δs_t	-18.170**	-18.257**	-18.239**	1
Sep02	f_t	-2.2923*	-0.46976	-2.2027	26	Δf_t	-4.6809**	-5.2235**	-5.2133**	25
	s_t	-1.9717*	-0.18846	-1.9079	26	Δs_t	-4.8670**	-5.2600**	-5.2824**	25
Dec02	f_t	-1.5957	-1.3875	-2.9190	20	Δf_t	-4.9358**	-5.1847**	-5.1929**	19
	s_t	-1.4125	-1.2958	-2.6471	20	Δs_t	-5.0625**	-5.2531**	-5.2495**	19
Mar03	f_t	-1.2414	-0.93276	-2.6819	20	Δf_t	-5.3538**	-5.4927**	-5.4561**	19
	s_t	-1.0054	-0.94885	-2.7232	20	Δs_t	-4.8624**	-5.4945**	-5.4523**	19
Jun03	f_t	-0.83485	-1.7167	-1.2484	28	Δf_t	-4.2179**	-4.2880**	-4.4819**	27
	s_t	-0.66931	-1.6282	-1.3485	28	Δs_t	-4.2614**	-4.3038**	-4.4473**	27
Sep03	f_t	-0.24600	-1.2865	-0.84879	20	Δf_t	-5.1459**	-5.1455**	-5.2320**	19
	s_t	-0.13504	-1.2995	-0.91654	20	Δs_t	-5.2384**	-5.2342**	-5.3096**	19
Dec03	f_t	-0.29809	-1.6311	-1.0568	20	Δf_t	-5.0961**	-5.0967**	-5.4645**	19
	s_t	-0.17043	-1.5345	-1.1352	20	Δs_t	-5.2247**	-5.2196**	-5.5501**	19

Table 8.10: ADF Unit Root Tests of the Log of Future Contract Prices (f_t) and the Log of Spot Index Prices (s_t)

Future Code	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag
Mar04	f_t	-0.064028	-1.4702	-2.6544	20	Δf_t	-4.9475**	-4.9423**	-5.3582**	19
	s_t	0.099122	-1.2007	-2.6260	20	Δs_t	-5.0569**	-5.0529**	-5.3971**	19
Jun04	f_t	1.4019	-0.48175	-1.9586	21	Δf_t	-7.4418**	-7.6226**	-7.5092**	20
	s_t	1.5971	-0.59838	-1.9647	21	Δs_t	-7.3405**	-7.7053**	-7.4442**	20
Sep04	f_t	1.1449	-0.86827	-2.2175	28	Δf_t	-3.9164**	-4.0832**	-4.0413**	27
	s_t	1.2030	-0.93867	-2.1708	28	Δs_t	-3.8695**	-4.0556**	-4.0070**	27
Dec04	f_t	1.4405	-0.71847	-1.9847	19	Δf_t	-4.9443**	-5.1580**	-5.1541**	18
	s_t	1.5273	-0.74504	-1.8623	19	Δs_t	-4.9838**	-5.2225**	-5.2170**	18
Mar05	f_t	2.2326	-2.4874	-2.5876	10	Δf_t	-7.8455**	-8.1946**	-8.3447**	9
	s_t	2.2278	-2.5793	-2.6443	10	Δs_t	-7.7596**	-8.1101**	-8.2710**	9
Jun05	f_t	1.4078	-1.4502	-2.3976	13	Δf_t	-6.1427**	-6.3107**	-6.2999**	12
	s_t	1.3929	-1.3985	-2.3722	13	Δs_t	-6.1157**	-6.2806**	-6.2722**	12
Sep05	f_t	1.3488	-2.0078	-2.9675	10	Δf_t	-7.7497**	-7.8740**	-7.8891**	9
	s_t	1.4315	-1.7787	-2.8405	10	Δs_t	-7.7460**	-7.8870**	-7.8882**	9
Dec05	f_t	0.95031	-1.6735	-3.0723	1	Δf_t	-23.205**	-23.223**	-23.202**	0
	s_t	1.0314	-1.4811	-3.0187	1	Δs_t	-23.638**	-23.662**	-23.643**	0
Mar06	f_t	1.1964	-1.5237	-3.3072	1	Δf_t	-23.325**	-23.367**	-23.344**	0
	s_t	1.2281	-1.1675	-3.2873	1	Δs_t	-23.750**	-23.794**	-23.774**	0
Jun06	f_t	0.53778	-1.6747	-2.9923	1	Δf_t	-22.713**	-22.703**	-22.685**	0
	s_t	0.71900	-1.5461	-3.2500	1	Δs_t	-23.189**	-23.189**	-23.169**	0
Sep06	f_t	0.97899	-2.3134	-3.4148	1	Δf_t	-22.646**	-22.667**	-22.652**	0
	s_t	1.1074	-1.8628	-3.5467*	1	Δs_t	-23.377**	-23.409**	-23.388**	0
Dec06	f_t	1.1339	-0.28999	-2.4860	7	Δf_t	-9.6684**	-9.7376**	-9.8134**	6
	s_t	1.4091	0.22319	-2.3442	7	Δs_t	-9.5382**	-9.6506**	-9.7656**	6

Table 8.11: ADF Unit Root Tests of the Log of Future Contract Prices (f_t) and the Log of Spot Index Prices (s_t)

Future Code	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag	Data	t-adf	t-adf^c	t-adf$^{c,\tau}$	lag
Mar07	f_t	1.1584	-1.3500	-2.4573	7	Δf_t	-10.029**	-10.100**	-10.094**	6
	s_t	1.5570	-1.0323	-2.3762	7	Δs_t	-9.8971**	-10.034**	-10.024**	6
Jun07	f_t	1.5018	-0.046564	-2.3386	30	Δf_t	-4.0703**	-4.3413**	-4.4453**	29
	s_t	1.8282	0.36102	-2.1668	30	Δs_t	-3.9249**	-4.3353**	-4.4986**	29
Sep07	f_t	1.6159	-1.6182	-2.6248	17	Δf_t	-5.6320**	-5.8578**	-5.8847**	16
	s_t	1.9702	-1.2041	-2.4877	17	Δs_t	-5.3887**	-5.7419**	-5.7504**	16
Dec07	f_t	0.74363	-1.6429	-1.9903	10	Δf_t	-7.8823**	-7.9151**	-7.9409**	9
	s_t	1.1271	-1.3985	-1.9725	10	Δs_t	-7.6771**	-7.7636**	-7.7822**	9
Mar08	f_t	-0.33304	-1.1650	-0.39522	17	Δf_t	-5.1107**	-5.1157**	-5.3782**	16
	s_t	0.019872	-1.4122	-0.33952	17	Δs_t	-5.0971**	-5.0915**	-5.3526**	16
Jun08	f_t	-0.10046	-1.4026	-1.5553	10	Δf_t	-7.6189**	-7.6102**	-7.7992**	9
	s_t	0.33591	-1.9583	-1.5007	10	Δs_t	-7.6850**	-7.6846**	-7.9016**	9
Sep08	f_t	-0.88232	-0.29315	-2.0017	17	Δf_t	-4.5249**	-4.5970**	-4.8721**	16
	s_t	-0.61420	-0.79529	-1.8316	17	Δs_t	-4.5867**	-4.6123**	-4.9263**	16
Dec08	f_t	-1.1966	0.055992	-1.6616	29	Δf_t	-3.1629**	-3.3815*	-3.6839*	28
	s_t	-1.0762	-0.029232	-1.4094	29	Δs_t	-3.1263**	-3.3057*	-3.6839*	28

Note: The statistics are augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process; (c) or (c, τ) superscripts indicate that a constant (a constant and a linear trend) was (were) included in the augmented Dickey-Fuller regression, while absence of a superscript indicates that neither a constant nor a trend were included; "lag " denoted the lag used in the augmented Dickey-Fuller regression, regardless the superscripts (c) or (c, τ) ; * and ** superscripts indicate significance at 5% and 1%, respectively, based on critical values in Fuller (1976).

Table 8.12: Regression Modelling Between f_t, s_t and t^*

Future Code	Parameter Estimates					
	$\hat{\mu}$	t-value	$\hat{\beta}$	t-value	$\hat{\delta}$	t-value
Mar98	-0.098130	-3.721**	1.0141	267.592**	-0.00017616	40.270**
Jun98	-0.12788	-4.461**	1.0181	249.270**	-0.00018065	44.726**
Sep98	-0.085167	-4.356**	1.0119	364.365**	-0.00017650	74.088**
Dec98	-0.043931	-2.410*	1.0060	389.164**	-0.00017334	90.136**
Mar99	-0.10972	-6.101**	1.0152	402.057**	-0.00018023	85.168**
Jun99	-0.19279	-10.362**	1.0264	396.440**	-0.00019458	91.637**
Sep99	-0.27491	-12.812**	1.0375	349.898**	-0.00020033	81.559**
Dec99	-0.30555	-12.670**	1.0418	313.733**	-0.00019009	74.303**
Mar00	-0.35258	-13.218**	1.0485	286.419**	-0.00018327	70.732**
Jun00	-0.34046	-12.056**	1.0472	271.562**	-0.00016818	60.701**
Sep00	-0.39977	-12.991**	1.0558	251.848**	-0.00015628	58.856**
Dec00	-0.50326	-11.834**	1.0703	183.663**	-0.00016033	72.501**
Mar01	-0.52566	-16.231**	1.0737	240.193**	-0.00016425	107.495**
Jun01	-0.49604	-20.399**	1.0695	315.689**	-0.00018138	120.116**
Sep01	-0.46084	-16.831**	1.0643	275.932**	-0.00019645	88.968**
Dec01	-0.24128	-6.540**	1.0324	196.214**	-0.00023460	61.877**
Mar02	-0.61610	-11.573**	1.0853	142.124**	-0.00021247	37.154**
Jun02	-1.0241	-15.356**	1.1439	118.998**	-0.00016616	23.536**
Sep02	-0.0093152	-0.120	0.99755	88.375**	-0.00023547	25.141**
Dec02	0.20050	3.926**	0.96774	128.667**	-0.00019325	28.673**
Mar03	0.11900	2.390*	0.97977	132.531**	-0.00015942	24.097**
Jun03	-0.36764	-9.432**	1.0523	181.734**	-0.00008346	17.471**
Sep03	-0.44888	-35.950**	1.0645	577.849**	-0.00005998	45.087**
Dec03	-0.60580	-53.942**	1.0864	662.009**	-0.00007809	70.143**

Table 8.13: Regression Modelling Between f_t, s_t and t^*

Future Code	Parameter Estimates					
	$\hat{\mu}$	t-value	$\hat{\beta}$	t-value	$\hat{\delta}$	t-value
Mar04	-0.73372	-54.026**	1.1035	565.544**	-0.00010338	78.123**
Jun04	-0.64981	-20.293**	1.0910	240.100**	-0.00009263	29.054**
Sep04	-0.065707	-1.784	1.0087	193.900**	-0.000017901	4.828**
Dec04	0.078731	4.250**	0.98888	379.224**	0.000015036	-8.505**
Mar05	0.063668	3.419**	0.99139	378.793**	0.000018681	-12.960**
Jun05	-0.23235	-8.385**	1.0334	265.000**	0.000001097	-0.707
Sep05	-0.093868	-2.629**	1.0140	202.004**	-0.000003123	1.897
Dec05	-0.014782	-0.308	1.0030	148.636**	-0.000008841	4.777**
Mar06	-0.058506	-0.957	1.0092	118.069**	-0.000022267	8.014**
Jun06	-0.41691	-7.220**	1.0592	131.598**	-0.00004943	17.950**
Sep06	-0.48802	-7.662**	1.0692	120.483**	-0.00005929	23.217**
Dec06	0.34425	6.471**	0.95333	129.181**	-0.000047102	19.429**
Mar07	0.33244	6.248**	0.95510	130.192**	-0.00005938	19.860**
Jun07	0.56276	12.965**	0.92362	155.212**	-0.00005839	21.381**
Sep07	0.24842	6.281**	0.96663	179.104**	-0.000090135	34.422**
Dec07	0.16235	5.990**	0.97811	265.058**	-0.00011005	64.748**
Mar08	-0.0053790	-0.408	1.0005	554.829**	-0.00013296	172.553**
Jun08	-0.051771	-3.725**	1.0062	527.231**	-0.00014603	200.570**
Sep08	-0.0064940	-0.264	0.99943	292.389**	-0.00014966	110.913**
Dec08	0.22881	12.405**	0.96534	367.180**	-0.00017439	66.670**

Note: * and ** superscripts indicate significance at 5% and 1%, respectively, based on the standard t-distribution.

Table 8.14: Regression Residuals Analysis

Future	ADF Unit Root	Autocorrelation(20)	ARCH(1)	Normality	Heteroscedasticity	
Code					squares	squares & cross-products
Mar98	-14.889** (0)	0.52399 [0.9550]	2.5976 [0.1083]	31.359 [0.0000]**	3.0664 [0.0172]*	2.5402 [0.0290]*
Jun98	-4.2996** (8)	0.87154 [0.6236]	1.2514 [0.2643]	19.403 [0.0001] **	14.132 [0.0000] **	11.292 [0.0000] **
Sep98	-3.3148** (22)	1.8879 [0.0129] *	8.364 [0.0041] **	21.253 [0.0000] **	22.678 [0.0000] **	20.158 [0.0000] **
Dec98	-2.5507* (21)	5.4478 [0.0000] **	6.3573 [0.0121] *	29.415 [0.0000] **	12.747 [0.0000] **	10.242 [0.0000] **
Mar99	-3.5128** (15)	6.1492 [0.0000] **	6.8219 [0.0093] **	29.473 [0.0000] **	7.9056 [0.0000] **	6.5177 [0.0000] **
Jun99	-3.8845** (6)	9.5269 [0.0000] **	22.438 [0.0000] **	13.448 [0.0012] **	8.4379 [0.0000] **	7.3969 [0.0000] **
Sep99	-2.7895** (9)	20.751 [0.0000] **	8.3006 [0.0041] **	11.706 [0.0029] **	13.054 [0.0000] **	10.748 [0.0000] **
Dec99	-3.1895** (10)	33.016 [0.0000] **	59.815 [0.0000] **	6.3094 [0.0427] *	10.676 [0.0000] **	9.4003 [0.0000] **
Mar00	-1.5428 (24)	48.84 [0.0000] **	43.02 [0.0000] **	20.266 [0.0000] **	11.396 [0.0000] **	9.4109 [0.0000] **
Jun00	-2.7487** (9)	40.207 [0.0000] **	122.17 [0.0000] **	0.34939 [0.8397]	37.757 [0.0000] **	32.173 [0.0000] **
Sep00	-1.5357 (24)	41.617 [0.0000] **	87.441 [0.0000] **	0.71568 [0.6992]	7.8726 [0.0000] **	7.0589 [0.0000] **
Dec00	-1.8604 (9)	40.839 [0.0000] **	9.6138 [0.0020] **	111.49 [0.0000] **	6.3019 [0.0001] **	5.0835 [0.0001] **
Mar01	-2.4902*(9)	50.194 [0.0000] **	25.576 [0.0000] **	21.4 [0.0000] **	9.9441 [0.0000] **	9.01 [0.0000] **
Jun01	-3.6292** (26)	34.567 [0.0000] **	69.591 [0.0000] **	46.114 [0.0000] **	7.1111 [0.0000] **	6.7449 [0.0000] **
Sep01	-1.8184 (26)	41.38 [0.0000] **	5.1831 [0.0232] *	163.53 [0.0000] **	2.1311 [0.0758]	1.9035 [0.0921]
Dec01	-0.79605 (26)	100.4 [0.0000] **	89.329 [0.0000] **	6.0848 [0.0477] *	17.805 [0.0000] **	14.986 [0.0000] **
Mar02	-1.5079 (26)	155.31 [0.0000] **	476.64 [0.0000] **	10.968 [0.0042] **	23.314 [0.0000] **	23.538 [0.0000] **
Jun02	-0.74889 (14)	242.44 [0.0000] **	1517.8 [0.0000] **	23.305 [0.0000] **	13.472 [0.0000] **	23.652 [0.0000] **
Sep02	-1.3560 (22)	570.26 [0.0000] **	1742 [0.0000] **	138.08 [0.0000] **	85.157 [0.0000] **	74.698 [0.0000] **
Dec02	-1.4388 (25)	359.92 [0.0000] **	2776.1 [0.0000] **	3.8421 [0.1465]	11.549 [0.0000] **	10.545 [0.0000] **
Mar03	-1.3876 (16)	376.2 [0.0000] **	1249.4 [0.0000] **	48.129 [0.0000] **	38.752 [0.0000] **	32.811 [0.0000] **
Jun03	-2.8567** (12)	304.81 [0.0000] **	2908.4 [0.0000] **	136.3 [0.0000] **	263.5 [0.0000] **	236.32 [0.0000] **
Sep03	-3.3713** (30)	34.513 [0.0000] **	414.29 [0.0000] **	42.719 [0.0000] **	27.019 [0.0000] **	25.401 [0.0000] **
Dec03	-4.2814** (3)	39.147 [0.0000] **	195.27 [0.0000] **	1.6473 [0.4388]	25.561 [0.0000] **	26.243 [0.0000] **

Table 8.15: Regression Residuals Analysis

Future Code	ADF Unit Root	Autocorrelation(20)	ARCH(1)	Normality	Heteroscedasticity squares	Heteroscedasticity squares & cross-products
Mar04	-3.5386** (12)	41.942 [0.0000] **	352.32 [0.0000] **	13.865 [0.0010] **	21.429 [0.0000] **	18.861 [0.0000] **
Jun04	-2.4380* (16)	111.57 [0.0000] **	707.45 [0.0000] **	68.641 [0.0000] **	31.575 [0.0000] **	25.212 [0.0000] **
Sep04	-3.3323** (10)	162.78 [0.0000] **	1052 [0.0000] **	83.456 [0.0000] **	85.114 [0.0000] **	71.379 [0.0000] **
Dec04	-4.3466** (10)	41.913 [0.0000] **	1366.5 [0.0000] **	96.071 [0.0000] **	66.936 [0.0000] **	54.668 [0.0000] **
Mar05	-3.1973** (10)	41.893 [0.0000] **	215.44 [0.0000] **	7.08 [0.0290] *	19.743 [0.0000] **	16.741 [0.0000] **
Jun05	-4.0695** (3)	55.206 [0.0000] **	298.23 [0.0000] **	13.134 [0.0014] **	20.747 [0.0000] **	18.16 [0.0000] **
Sep05	-2.3379* (3)	85.742 [0.0000] **	378.83 [0.0000] **	13.479 [0.0012] **	8.7848 [0.0000] **	10.066 [0.0000] **
Dec05	-1.7189 (3)	170.55 [0.0000] **	683.8 [0.0000] **	28.267 [0.0000] **	12.246 [0.0000] **	13.018 [0.0000] **
Mar06	-2.1813* (20)	176.88 [0.0000] **	2184.6 [0.0000] **	117.47 [0.0000] **	56.478 [0.0000] **	47.145 [0.0000] **
Jun06	-1.2585 (4)	223.31 [0.0000] **	494.27 [0.0000] **	50.406 [0.0000] **	25.05 [0.0000] **	20.999 [0.0000] **
Sep06	-1.2948 (29)	224.33 [0.0000] **	925.79 [0.0000] **	67.517 [0.0000] **	70.682 [0.0000] **	69.914 [0.0000] **
Dec06	-2.0239*(19)	264.93 [0.0000] **	1048.6 [0.0000] **	45.832 [0.0000] **	30.458 [0.0000] **	24.627 [0.0000] **
Mar07	-1.6254 (3)	245.17 [0.0000] **	1140.4 [0.0000] **	16.925 [0.0002] **	12.368 [0.0000] **	17.266 [0.0000] **
Jun07	-2.2021*(19)	173.08 [0.0000] **	1251.7 [0.0000] **	16.734 [0.0002] **	41.844 [0.0000] **	35.217 [0.0000] **
Sep07	-3.1722** (19)	134.49 [0.0000] **	1097.4 [0.0000] **	188.01 [0.0000] **	106.05 [0.0000] **	104.08 [0.0000] **
Dec07	-2.5523* (19)	68.702 [0.0000] **	355.36 [0.0000] **	63.594 [0.0000] **	50.861 [0.0000] **	47.178 [0.0000] **
Mar08	-3.5510** (19)	26.49 [0.0000] **	187.81 [0.0000] **	21.114 [0.0000] **	36.878 [0.0000] **	29.446 [0.0000] **
Jun08	-2.4356* (17)	36.972 [0.0000] **	186.03 [0.0000] **	14.999 [0.0006] **	41.505 [0.0000] **	33.533 [0.0000] **
Sep08	-1.2043 (17)	136.46 [0.0000] **	784.7 [0.0000] **	26.396 [0.0000] **	76.532 [0.0000] **	62.02 [0.0000] **
Dec08	-1.7918 (30)	89.694 [0.0000] **	591.01 [0.0000]**	37.471 [0.0000] **	47.515 [0.0000] **	39.712 [0.0000] **

Notes: *', * and ** superscripts indicate significance at 10%, 5% and 1%, respectively. The statistics for unit root tests are the augmented Dickey-Fuller test statistics for the null hypothesis of a unit root process and the numbers in the brackets denote the number of lags used in tests.

Table 8.16: Autoregressive fitted models for regression residuals

Future Code	Autoregressive model
Mar98	$\hat{\epsilon}_t = \hat{\epsilon}_t$
Jun98	$\hat{\epsilon}_t = \hat{\epsilon}_t$
Sep98	$\hat{\epsilon}_t = 0.19748\hat{\epsilon}_{t-9} + 0.15992\hat{\epsilon}_{t-12} - 0.12015\hat{\epsilon}_{t-16} - 0.17715\hat{\epsilon}_{t-23}$
Dec98	$\hat{\epsilon}_t = 0.14611\hat{\epsilon}_{t-2} + 0.13405\hat{\epsilon}_{t-3} + 0.13424\hat{\epsilon}_{t-8} + 0.12933\hat{\epsilon}_{t-9} + 0.13048\hat{\epsilon}_{t-12}$
Mar99	$\hat{\epsilon}_t = 0.099840\hat{\epsilon}_{t-1} + 0.13907\hat{\epsilon}_{t-2} + 0.11639\hat{\epsilon}_{t-6} + 0.13102\hat{\epsilon}_{t-7} + 0.10119\hat{\epsilon}_{t-8} + 0.15622\hat{\epsilon}_{t-12} - 0.10224\hat{\epsilon}_{t-16}$
Jun99	$\hat{\epsilon}_t = 0.19166\hat{\epsilon}_{t-1} + 0.18391\hat{\epsilon}_{t-2} + 0.14346\hat{\epsilon}_{t-5} + 0.11304\hat{\epsilon}_{t-6} + 0.11386\hat{\epsilon}_{t-7}$
Sep99	$\hat{\epsilon}_t = 0.21466\hat{\epsilon}_{t-1} + 0.24277\hat{\epsilon}_{t-2} + 0.16101\hat{\epsilon}_{t-5} + 0.11585\hat{\epsilon}_{t-6} + 0.10972\hat{\epsilon}_{t-7}$
Dec99	$\hat{\epsilon}_t = 0.26543\hat{\epsilon}_{t-1} + 0.27758\hat{\epsilon}_{t-2} + 0.15395\hat{\epsilon}_{t-5} + 0.14057\hat{\epsilon}_{t-7} - 0.092441\hat{\epsilon}_{t-9} + 0.12192\hat{\epsilon}_{t-10}$
Mar00	$\hat{\epsilon}_t = 0.23116\hat{\epsilon}_{t-1} + 0.22273\hat{\epsilon}_{t-2} + 0.11335\hat{\epsilon}_{t-4} + 0.11550\hat{\epsilon}_{t-6} + 0.14342\hat{\epsilon}_{t-7} - 0.10887\hat{\epsilon}_{t-9} + 0.14277\hat{\epsilon}_{t-10} + 0.085879\hat{\epsilon}_{t-25}$
Jun00	$\hat{\epsilon}_t = 0.24882\hat{\epsilon}_{t-1} + 0.23654\hat{\epsilon}_{t-2} + 0.091570\hat{\epsilon}_{t-4} + 0.10482\hat{\epsilon}_{t-6} + 0.10442\hat{\epsilon}_{t-7} + 0.11871\hat{\epsilon}_{t-10}$
Sep00	$\hat{\epsilon}_t = 0.28150\hat{\epsilon}_{t-1} + 0.20308\hat{\epsilon}_{t-2} + 0.14155\hat{\epsilon}_{t-4} + 0.10496\hat{\epsilon}_{t-7} + 0.079299\hat{\epsilon}_{t-25}$
Dec00	$\hat{\epsilon}_t = 0.32509\hat{\epsilon}_{t-1} + 0.15799\hat{\epsilon}_{t-2} + 0.13317\hat{\epsilon}_{t-4} + 0.22146\hat{\epsilon}_{t-7} - 0.10334\hat{\epsilon}_{t-9} + 0.19591\hat{\epsilon}_{t-10}$
Mar01	$\hat{\epsilon}_t = 0.30703\hat{\epsilon}_{t-1} + 0.19330\hat{\epsilon}_{t-2} + 0.092983\hat{\epsilon}_{t-3} + 0.16230\hat{\epsilon}_{t-4} + 0.14674\hat{\epsilon}_{t-7} - 0.093746\hat{\epsilon}_{t-9} + 0.11373\hat{\epsilon}_{t-10}$
Jun01	$\hat{\epsilon}_t = 0.28201\hat{\epsilon}_{t-1} + 0.18169\hat{\epsilon}_{t-2} + 0.13473\hat{\epsilon}_{t-3} + 0.17794\hat{\epsilon}_{t-4} + 0.11889\hat{\epsilon}_{t-7} - 0.10987\hat{\epsilon}_{t-9} + 0.088296\hat{\epsilon}_{t-10}$
Sep01	$\hat{\epsilon}_t = 0.29641\hat{\epsilon}_{t-1} + 0.16755\hat{\epsilon}_{t-2} + 0.20912\hat{\epsilon}_{t-3} + 0.13448\hat{\epsilon}_{t-4} - 0.10003\hat{\epsilon}_{t-5} + 0.22095\hat{\epsilon}_{t-7}$
Dec01	$\hat{\epsilon}_t = 0.30754\hat{\epsilon}_{t-1} + 0.19525\hat{\epsilon}_{t-2} + 0.22829\hat{\epsilon}_{t-3} + 0.28813\hat{\epsilon}_{t-7} - 0.13927\hat{\epsilon}_{t-13} + 0.18039\hat{\epsilon}_{t-20} - 0.088214\hat{\epsilon}_{t-27}$
Mar02	$\hat{\epsilon}_t = 0.35321\hat{\epsilon}_{t-1} + 0.24023\hat{\epsilon}_{t-2} + 0.21777\hat{\epsilon}_{t-3} + 0.16493\hat{\epsilon}_{t-7}$
Jun02	$\hat{\epsilon}_t = 0.46600\hat{\epsilon}_{t-1} + 0.22878\hat{\epsilon}_{t-2} + 0.25750\hat{\epsilon}_{t-3} + 0.19421\hat{\epsilon}_{t-7} - 0.14562\hat{\epsilon}_{t-8} - 0.12081\hat{\epsilon}_{t-13} + 0.10815\hat{\epsilon}_{t-15}$
Sep02	$\hat{\epsilon}_t = 0.40168\hat{\epsilon}_{t-1} + 0.23246\hat{\epsilon}_{t-2} + 0.16004\hat{\epsilon}_{t-3} + 0.20285\hat{\epsilon}_{t-7} + 0.11159\hat{\epsilon}_{t-10} - 0.11341\hat{\epsilon}_{t-13} + 0.10677\hat{\epsilon}_{t-15} - 0.11558\hat{\epsilon}_{t-23}$
Dec02	$\hat{\epsilon}_t = 0.50991\hat{\epsilon}_{t-1} + 0.28631\hat{\epsilon}_{t-2} + 0.16753\hat{\epsilon}_{t-7} + 0.12334\hat{\epsilon}_{t-10} - 0.10792\hat{\epsilon}_{t-17}$
Mar03	$\hat{\epsilon}_t = 0.51633\hat{\epsilon}_{t-1} + 0.28570\hat{\epsilon}_{t-2} + 0.15250\hat{\epsilon}_{t-7} + 0.11922\hat{\epsilon}_{t-10} - 0.089791\hat{\epsilon}_{t-17}$
Jun03	$\hat{\epsilon}_t = 0.44826\hat{\epsilon}_{t-1} + 0.18000\hat{\epsilon}_{t-2} + 0.15905\hat{\epsilon}_{t-3} + 0.12355\hat{\epsilon}_{t-5} + 0.12330\hat{\epsilon}_{t-7} - 0.068518\hat{\epsilon}_{t-13}$
Sep03	$\hat{\epsilon}_t = 0.41106\hat{\epsilon}_{t-1} + 0.21360\hat{\epsilon}_{t-2} + 0.14273\hat{\epsilon}_{t-5} + 0.10873\hat{\epsilon}_{t-7} - 0.10688\hat{\epsilon}_{t-9} + 0.11475\hat{\epsilon}_{t-10} - 0.055447\hat{\epsilon}_{t-31}$
Dec03	$\hat{\epsilon}_t = 0.41638\hat{\epsilon}_{t-1} + 0.23423\hat{\epsilon}_{t-2} + 0.095554\hat{\epsilon}_{t-3} + 0.11743\hat{\epsilon}_{t-4}$

Table 8.17: Autoregressive fitted models for regression residuals

Future Code	Autoregressive model
Mar04	$\hat{\epsilon}_t = 0.46020\hat{\epsilon}_{t-1} + 0.27062\hat{\epsilon}_{t-2} + 0.12513\hat{\epsilon}_{t-3}$
Jun04	$\hat{\epsilon}_t = 0.44710\hat{\epsilon}_{t-1} + 0.23416\hat{\epsilon}_{t-2} + 0.1161\hat{\epsilon}_{t-3} + 0.13113\hat{\epsilon}_{t-5} + 0.10852\hat{\epsilon}_{t-12} - 0.092389\hat{\epsilon}_{t-17}$
Sep04	$\hat{\epsilon}_t = 0.41103\hat{\epsilon}_{t-1} + 0.25499\hat{\epsilon}_{t-2} + 0.13921\hat{\epsilon}_{t-4} + 0.13949\hat{\epsilon}_{t-11}$
Dec04	$\hat{\epsilon}_t = 0.42536\hat{\epsilon}_{t-1} + 0.26436\hat{\epsilon}_{t-2} + 0.12970\hat{\epsilon}_{t-3}$
Mar05	$\hat{\epsilon}_t = 0.35119\hat{\epsilon}_{t-1} + 0.25708\hat{\epsilon}_{t-2} + 0.12911\hat{\epsilon}_{t-3} + 0.15651\hat{\epsilon}_{t-4} - 0.10020\hat{\epsilon}_{t-9} + 0.096715\hat{\epsilon}_{t-11}$
Jun05	$\hat{\epsilon}_t = 0.33393\hat{\epsilon}_{t-1} + 0.25117\hat{\epsilon}_{t-2} + 0.16545\hat{\epsilon}_{t-3} + 0.14174\hat{\epsilon}_{t-4}$
Sep05	$\hat{\epsilon}_t = 0.33733\hat{\epsilon}_{t-1} + 0.27745\hat{\epsilon}_{t-2} + 0.19888\hat{\epsilon}_{t-3} + 0.13385\hat{\epsilon}_{t-4}$
Dec05	$\hat{\epsilon}_t = 0.37575\hat{\epsilon}_{t-1} + 0.29111\hat{\epsilon}_{t-2} + 0.21665\hat{\epsilon}_{t-3} + 0.088250\hat{\epsilon}_{t-4}$
Mar06	$\hat{\epsilon}_t = 0.40982\hat{\epsilon}_{t-1} + 0.26678\hat{\epsilon}_{t-2} + 0.22053\hat{\epsilon}_{t-3} + 0.11895\hat{\epsilon}_{t-8} - 0.056890\hat{\epsilon}_{t-21}$
Jun06	$\hat{\epsilon}_t = 0.38145\hat{\epsilon}_{t-1} + 0.15304\hat{\epsilon}_{t-2} + 0.18761\hat{\epsilon}_{t-3} + 0.13634\hat{\epsilon}_{t-4} + 0.12291\hat{\epsilon}_{t-5}$
Sep06	$\hat{\epsilon}_t = 0.40114\hat{\epsilon}_{t-1} + 0.11449\hat{\epsilon}_{t-2} + 0.21843\hat{\epsilon}_{t-3} + 0.16686\hat{\epsilon}_{t-4} + 0.11055\hat{\epsilon}_{t-11} + 0.11133\hat{\epsilon}_{t-19} - 0.11678\hat{\epsilon}_{t-20}$
	$-0.12216\hat{\epsilon}_{t-26} + 0.095771\hat{\epsilon}_{t-30}$
Dec06	$\hat{\epsilon}_t = 0.51688\hat{\epsilon}_{t-1} + 0.15327\hat{\epsilon}_{t-2} + 0.18668\hat{\epsilon}_{t-3} + 0.11013\hat{\epsilon}_{t-4} + 0.10405\hat{\epsilon}_{t-19} - 0.10025\hat{\epsilon}_{t-20}$
Mar07	$\hat{\epsilon}_t = 0.57412\hat{\epsilon}_{t-1} + 0.15314\hat{\epsilon}_{t-2} + 0.24685\hat{\epsilon}_{t-3}$
Jun07	$\hat{\epsilon}_t = 0.55021\hat{\epsilon}_{t-1} + 0.12132\hat{\epsilon}_{t-2} + 0.21324\hat{\epsilon}_{t-3} + 0.11149\hat{\epsilon}_{t-11} + 0.10851\hat{\epsilon}_{t-19} - 0.14383\hat{\epsilon}_{t-20}$
Sep07	$\hat{\epsilon}_t = 0.40562\hat{\epsilon}_{t-1} + 0.21166\hat{\epsilon}_{t-3} + 0.11024\hat{\epsilon}_{t-4} + 0.12222\hat{\epsilon}_{t-6} + 0.13162\hat{\epsilon}_{t-10} + 0.12340\hat{\epsilon}_{t-19} - 0.17234\hat{\epsilon}_{t-20}$
Dec07	$\hat{\epsilon}_t = 0.40215\hat{\epsilon}_{t-1} + 0.24658\hat{\epsilon}_{t-3} + 0.15464\hat{\epsilon}_{t-6} + 0.11394\hat{\epsilon}_{t-16}$
Mar08	$\hat{\epsilon}_t = 0.29911\hat{\epsilon}_{t-1} + 0.13307\hat{\epsilon}_{t-2} + 0.16512\hat{\epsilon}_{t-3} + 0.11250\hat{\epsilon}_{t-6} + 0.12891\hat{\epsilon}_{t-7}$
Jun08	$\hat{\epsilon}_t = 0.34668\hat{\epsilon}_{t-1} + 0.11034\hat{\epsilon}_{t-2} + 0.21733\hat{\epsilon}_{t-3} + 0.10891\hat{\epsilon}_{t-4} + 0.13072\hat{\epsilon}_{t-9} + 0.11532\hat{\epsilon}_{t-17} - 0.13290\hat{\epsilon}_{t-18}$
Sep08	$\hat{\epsilon}_t = 0.421788\hat{\epsilon}_{t-1} + 0.15283\hat{\epsilon}_{t-2} + 0.25412\hat{\epsilon}_{t-3} + 0.13531\hat{\epsilon}_{t-6} + 0.13361\hat{\epsilon}_{t-17} - 0.12374\hat{\epsilon}_{t-18}$
Dec08	$\hat{\epsilon}_t = 0.46176\hat{\epsilon}_{t-1} + 0.18098\hat{\epsilon}_{t-2} + 0.12237\hat{\epsilon}_{t-5} + 0.16297\hat{\epsilon}_{t-6} + 0.12025\hat{\epsilon}_{t-11} + 0.10106\hat{\epsilon}_{t-15} - 0.24424\hat{\epsilon}_{t-16}$
	$+0.15721\hat{\epsilon}_{t-26} - 0.16230\hat{\epsilon}_{t-28} + 0.15457\hat{\epsilon}_{t-29} - 0.10589\hat{\epsilon}_{t-31}$

Table 8.18: Autoregressive fitted model residuals analysis

Future Code	Autocorrelation(20)	ARCH(1)	Normality	Heteroscedasticity squares	Heteroscedasticity squares & cross-products
Mar98					
Jun98					
Sep98	0.90068 [0.5866]	6.5774 [0.0108] *	29.709 [0.0000] **	1.9266 [0.0558]	1.2608 [0.2311]
Dec98	1.2369 [0.2207]	3.2147 [0.0738]	57.255 [0.0000] **	1.698 [0.0794]	2.6146 [0.0002] **
Mar99	1.0279 [0.4278]	4.4679 [0.0352] *	46.933 [0.0000] **	1.1542 [0.3092]	1.5344 [0.0306] *
Jun99	0.86132 [0.6374]	12.32 [0.0005] **	39.562 [0.0000] **	1.7158 [0.0743]	1.8659 [0.0130] *
Sep99	0.82972 [0.6777]	8.039 [0.0048] **	146.42 [0.0000] **	0.49936 [0.8906]	1.1651 [0.2801]
Dec99	1.0343 [0.4192]	7.0267 [0.0083] **	118.43 [0.0000] **	2.6508 [0.0019] **	2.1847 [0.0006] **
Mar00	0.73043 [0.7955]	13.905 [0.0002] **	21.961 [0.0000] **	0.91747 [0.5490]	1.4007 [0.0509]
Jun00	0.94729 [0.5265]	2.2094 [0.1378]	80.611 [0.0000] **	1.2001 [0.2796]	0.9385 [0.5559]
Sep00	1.0811 [0.3660]	7.9973 [0.0049] **	21.937 [0.0000] **	0.88415 [0.5632]	0.94499 [0.5464]
Dec00	1.1114 [0.3333]	3.1761 [0.0753]	589.05 [0.0000] **	0.87179 [0.5760]	2.1576 [0.0008] **
Mar01	0.59457 [0.9173]	7.508 [0.0064] **	323.11 [0.0000] **	1.0307 [0.4208]	3.311 [0.0000] **
Jun01	0.7275 [0.7988]	6.9061 [0.0089] **	301.02 [0.0000] **	0.98561 [0.4667]	3.013 [0.0000] **
Sep01	0.8925 [0.5971]	0.9918 [0.3198]	955.28 [0.0000] **	19.069 [0.0000] **	13.876 [0.0000] **
Dec01	1.3375 [0.1497]	13.392 [0.0003] **	177.79 [0.0000] **	1.3235 [0.1891]	3.0491 [0.0000] **
Mar02	1.0137 [0.4437]	14.77 [0.0001] **	215.68 [0.0000] **	1.0796 [0.3759]	2.2685 [0.0053] **
Jun02	1.1381 [0.3062]	11.869 [0.0006] **	121.96 [0.0000] **	3.5748 [0.0000] **	3.3228 [0.0000] **
Sep02	1.1721 [0.2738]	27.863 [0.0000] **	140.26 [0.0000] **	4.8729 [0.0000] **	3.794 [0.0000] **
Dec02	0.43635 [0.9850]	16.704 [0.0001] **	37.768 [0.0000] **	3.5228 [0.0002] **	2.9578 [0.0000] **
Mar03	0.67451 [0.8525]	15.142 [0.0001] **	31.976 [0.0000] **	2.3652 [0.0097] **	1.9166 [0.0100] **
Jun03	0.62316 [0.8965]	18.655 [0.0000] **	14.633 [0.0007] **	1.6565 [0.0733]	2.1191 [0.0010] **
Sep03	0.59872 [0.9143]	2.4514 [0.1181]	6.2348 [0.0443] *	2.6709 [0.0009] **	1.7353 [0.0068] **
Dec03	0.9824 [0.4822]	3.9164 [0.0484] *	14.694 [0.0006] **	4.9197 [0.0000] **	5.5097 [0.0000] **

Note: Residuals from Mar98 and Jun98 are white noise series (see Table 8.14)

Table 8.19: Autoregressive fitted model residuals analysis

Future Code	Autocorrelation(20)	ARCH(1)	Normality	Heteroscedasticity squares	Heteroscedasticity squares & cross-products
Mar04	1.3901 [0.1209]	8.9389 [0.0029] **	21.835 [0.0000] **	6.4871 [0.0000] **	5.9729 [0.0000] **
Jun04	1.325 [0.1570]	3.1653 [0.0758]	49.565 [0.0000] **	5.5568 [0.0000] **	4.7745 [0.0000] **
Sep04	1.3075 [0.1678]	3.4104 [0.0654]	29.17 [0.0000] **	3.3482 [0.0010] **	2.2528 [0.0057] **
Dec04	1.2533 [0.2055]	2.627 [0.1057]	17.772 [0.0001] **	5.2648 [0.0000] **	4.5131 [0.0000] **
Mar05	1.4545 [0.0924]	2.5631 [0.1100]	2.848 [0.2408]	1.142 [0.3234]	1.4392 [0.0729]
Jun05	1.1614 [0.2835]	4.3701 [0.0371] *	3.9926 [0.1358]	0.71275 [0.6804]	2.5893 [0.0013] **
Sep05	1.1928 [0.2549]	8.036 [0.0048] **	12.657 [0.0018] **	1.3549 [0.2140]	2.5991 [0.0012] **
Dec05	0.96032 [0.5099]	8.0878 [0.0046] **	44.87 [0.0000] **	1.1748 [0.3123]	3.0033 [0.0002] **
Mar06	1.4448 [0.0964]	21.365 [0.0000] **	31.533 [0.0000] **	6.0267 [0.0000] **	4.6388 [0.0000] **
Jun06	0.94664 [0.5274]	1.8775 [0.1712]	8.3249 [0.0156] *	1.8924 [0.0439] *	1.5879 [0.0510]
Sep06	1.4026 [0.1152]	0.94075 [0.3326]	1.4872 [0.4754]	0.71165 [0.8002]	0.86553 [0.7389]
Dec06	0.88277 [0.6097]	7.4727 [0.0065] **	22.365 [0.0000] **	1.9676 [0.0254] *	1.8282 [0.0074] **
Mar07	1.0736 [0.3740]	6.3964 [0.0117] *	74.125 [0.0000] **	0.7901 [0.5779]	1.9395 [0.0445] *
Jun07	0.75929 [0.7631]	0.61833 [0.4320]	79.81 [0.0000] **	0.6549 [0.7947]	0.68027 [0.8877]
Sep07	0.77203 [0.7483]	1.0115 [0.3150]	97.688 [0.0000] **	2.5402 [0.0016] **	2.0341 [0.0006] **
Dec07	1.6563 [0.0370] *	0.68331 [0.4088]	70.245 [0.0000] **	0.52075 [0.8411]	0.86107 [0.6020]
Mar08	1.3219 [0.1587]	2.8068 [0.0945]	42.161 [0.0000] **	0.97248 [0.4665]	1.7178 [0.0274] *
Jun08	0.78547 [0.7323]	6.4094 [0.0117] *	34.329 [0.0000] **	3.1237 [0.0001] **	2.2756 [0.0001] **
Sep08	0.63943 [0.8835]	3.9955 [0.0462] *	54.094 [0.0000] **	4.4043 [0.0000] **	2.5877 [0.0000] **
Dec08	1.2569 [0.2033]	39.948 [0.0000] **	254.28 [0.0000] **	4.6679 [0.0000] **	8.1464 [0.0000] **

Notes: *', * and ** superscripts indicate significance at 10%, 5% and 1%, respectively.

Table 8.20: Nonlinear Analysis for Autoregressive Fitted Model Residuals

Future Code	RESET(3)	LM^G
Sep98	0.065253 [0.9368]	1.594 [0.0474]*
Dec98	0.66583 [0.5145]	1.6433 [0.0780]
Mar99	1.0721 [0.3433]	1.598 [0.0669]
Jun99	0.061577 [0.9403]	1.5029 [0.1639]
Sep99	3.015 [0.0499]*	0.48993 [0.8421]
Dec99	2.825 [0.0602]	1.0886 [0.3690]
Mar00	1.7185 [0.1804]	0.9634 [0.5161]
Jun00	0.010124 [0.9899]	0.34752 [0.9674]
Sep00	1.5169 [0.2204]	0.77162 [0.7791]
Dec00	6.6409 [0.0014]**	2.7781 [0.0024]**
Mar01	2.2866 [0.1027]	1.4735 [0.1459]
Jun01	0.45491 [0.6348]	1.4424 [0.1670]
Sep01	0.44952 [0.6382]	1.7997 [0.0852]
Dec01	0.82735 [0.4378]	0.68402 [0.8791]
Mar02	0.21939 [0.8031]	2.3928 [0.0205]*
Jun02	0.63042 [0.5328]	1.608 [0.0677]
Sep02	2.164 [0.1160]	1.5685 [0.0462]*
Dec02	0.086451 [0.9172]	1.1344 [0.3173]
Mar03	2.0782 [0.1262]	1.085 [0.3653]
Jun03	0.24312 [0.7843]	1.5212 [0.1056]
Sep03	0.7928 [0.4532]	1.7821 [0.0069]**
Dec03	1.3977 [0.2481]	1.8221 [0.1232]
Mar04	0.082043 [0.9212]	3.2155 [0.0227]*
Jun04	0.87767 [0.4164]	1.8642 [0.0191]*
Sep04	2.2584 [0.1056]	1.3723 [0.1824]
Dec04	0.26204 [0.7696]	0.74291 [0.5268]
Mar05	1.1597 [0.3144]	0.91915 [0.5215]
Jun05	1.6328 [0.1964]	0.21202 [0.9318]
Sep05	0.10377 [0.9015]	0.56751 [0.6863]
Dec05	0.88373 [0.4139]	0.11283 [0.9780]
Mar06	0.93084 [0.3949]	1.1181 [0.3246]
Jun06	4.3711 [0.0131]*	0.14799 [0.9806]
Sep06	2.8808 [0.0571]	1.4966 [0.0469]*
Dec06	0.84836 [0.4287]	1.1281 [0.3164]
Mar07	0.93723 [0.3924]	1.1415 [0.3318]
Jun07	1.1173 [0.3280]	0.70666 [0.8207]
Sep07	1.2644 [0.2833]	1.2069 [0.2431]
Dec07	0.0688 [0.9335]	1.0495 [0.4024]
Mar08	0.079534 [0.9236]	0.88521 [0.5178]
Jun08	0.46882 [0.6260]	1.7306 [0.0314]*
Sep08	1.5287 [0.2178]	1.0707 [0.3791]
Dec08	2.359 [0.0956]	3.724 [0.0000]**

Notes: * and ** superscripts indicate significance at 5% and 1%, respectively. Note: Residuals from Mar98 and Jun98 are white noise series (see Table 8.14)

Table 8.21: Johansen Cointegration Analysis

Future Code	Ho:rank=r	LR tests		standardized β' eigenvectors			lag used	vector error
		Max Eigenvalue	Trace	f_t	s_t	t		autocorrelation(20)
Mar98	r = 0	181.8**	187.7**	1.0000	-1.0113	0.00017354	1	1.0859 [0.3012]
	r <= 1	5.906	5.906	-0.42727	1.0000	-0.00067888		
Jun98	r = 0	188.3**	197.4**	1.0000	-1.0139	0.00017678	1	1.2014 [0.1300]
	r <= 1	9.097	9.097	-0.49025	1.0000	-0.00054120		
Sep98	r = 0	219**	224.3**	1.0000	-1.0077	0.00017262	1	1.2664 [0.0691]
	r <= 1	5.302	5.302	-0.57908	1.0000	-0.00030354		
Dec98	r = 0	43.35**	48.62**	1.0000	-1.0018	0.00017076	3	1.4302 [0.0111] *
	r <= 1	5.272	5.272	-0.64695	1.0000	-0.00032604		
Mar99	r = 0	85.09**	90.18**	1.0000	-1.0122	0.00017769	2	1.2968 [0.0490] *
	r <= 1	5.087	5.087	-0.55763	1.0000	-0.00044177		
Jun99	r = 0	58.1**	65.76**	1.0000	-1.0269	0.00019428	3	1.3079 [0.0413] *
	r <= 1	7.666	7.666	1.2361	1.0000	-0.0015227		
Sep99	r = 0	40.42**	48.36**	1.0000	-1.0412	0.00020164	3	1.3941 [0.0155] *
	r <= 1	7.937	7.937	-5.4856	1.0000	0.0023939		
Dec99	r = 0	19.39*	25.81*	1.0000	-1.0678	0.00020338	5	1.2178 [0.1013]
	r <= 1	6.423	6.423	-1.1328	1.0000	-9.6489e-005		
Mar00	r = 0	11.05	17.87	1.0000	-1.0885	0.00020578	5	1.3217 [0.0356] *
	r <= 1	6.82	6.82	-1.1093	1.0000	-9.3423e-005		
Jun00	r = 0	19.54*	27.58*	1.0000	-1.0718	0.00018174	4	1.2834 [0.0534]
	r <= 1	8.042	8.042	-1.2714	1.0000	-1.6127e-006		
Sep00	r = 0	18.19	24.35	1.0000	-1.1556	0.00020890	6	1.3956 [0.0153] *
	r <= 1	6.168	6.168	-1.0119	1.0000	-0.00014114		
Dec00	r = 0	12.59	17.54	1.0000	-1.2161	0.00019560	8	1.2932 [0.0485] *
	r <= 1	4.949	4.949	-0.95930	1.0000	-0.00023883		

Table 8.22: Johansen Cointegration Analysis

Future Code	Ho:rank=r	LR tests		standardized β' eigenvectors			lag used	vector error
		Max Eigenvalue	Trace	f_t	s_t	t		autocorrelation(20)
Mar01	r = 0	11.73	18.6	1.0000	-1.1120	0.00016148	6	1.2702 [0.0612]
	r <= 1	6.866	6.866	-1.0548	1.0000	-0.00025779		
Jun01	r = 0	19.47*	25.69*	1.0000	-1.0930	0.00018059	5	1.4086 [0.0130]*
	r <= 1	6.214	6.214	-1.3050	1.0000	-0.00044932		
Sep01	r= 0	12.42	16.76	1.0000	-0.97349	0.00027595	7	1.2425 [0.0804]
	r <= 1	4.344	4.226	-0.91913	1.0000	-0.00016021		
Dec01	r = 0	9.82	11.85	1.0000	-1.4808	-6.9385e-005	6	1.3944 [0.0155] *
	r <= 1	2.026	2.026	-0.99979	1.0000	-0.00024533		
Mar02	r = 0	11.97	13.54	1.0000	-1.3297	5.2010e-005	5	1.2665 [0.0634]
	r <= 1	1.567	1.567	-0.99206	1.0000	-0.00014236		
Jun02	r = 0	20.63*	24.59	1.0000	-1.6214	-0.00027843	5	1.4142 [0.0121]*
	r <= 1	3.96	3.96	0.26544	1.0000	0.00047555		
Sep02	r = 0	15.12	21.51	1.0000	0.53603	0.0019500	4	1.2768 [0.0570]
	r <= 1	6.387	6.387	-1.1715	1.0000	-0.00032451		
Dec02	r = 0	14.26	21.73	1.0000	1.4285	0.0032200	5	1.2318 [0.0892]
	r <= 1	7.465	7.465	-1.2243	1.0000	-0.00031229		
Mar03	r = 0	11.24	15.76	1.0000	-2.6029	-0.0014571	5	1.3831 [0.0176] *
	r <= 1	4.517	4.517	-1.1501	1.0000	-0.00019996		
Jun03	r = 0	10.86	14.64	1.0000	-1.0991	1.0327e-005	3	1.1882 [0.1317]
	r <= 1	3.783	3.783	-0.097755	1.0000	0.00039678		
Sep03	r = 0	36.1**	38.01**	1.0000	-1.0746	5.0933e-005	4	1.3874 [0.0168]*
	r <= 1	1.907	1.907	-0.55096	1.0000	-0.00019795		
Dec03	r = 0	22.02*	27.02*	1.0000	-1.0976	7.6291e-005	4	1.3122 [0.0394]*
	r <= 1	4.993	4.993	-1.0317	1.0000	0.00017907		

Table 8.23: Johansen Cointegration Analysis

Future Code	Ho:rank=r	LR tests		standardized β' eigenvectors			lag used	vector error
		Max Eigenvalue	Trace	f_t	s_t	t		autocorrelation(20)
Mar04	r = 0	24**	40.98**	1.0000	-1.0485	2.9274e-005	5	1.3306 [0.0324]*
	r <= 1	16.98**	16.98**	-0.89051	1.0000	-9.9935e-005		
Jun04	r = 0	14.73	21.28	1.0000	-0.91725	-0.00011570	29	1.335 [0.0319]*
	r <= 1	6.544	6.544	-0.82429	1.0000	-0.00012646		
Sep04	r = 0	26.05**	36.6**	1.0000	-1.0312	-1.4929e-005	45	1.3857 [0.0188]*
	r <= 1	10.55	10.55	-1.1684	1.0000	9.9154e-005		
Dec04	r = 0	23.47*	29.65*	1.0000	-0.97460	-2.8184e-005	16	1.3033 [0.0439] *
	r <= 1	6.181	6.181	-0.79796	1.0000	-8.4084e-005		
Mar05	r = 0	34.27**	43.76**	1.0000	-0.99896	-1.6235e-005	2	1.4223 [0.0109]*
	r <= 1	9.489	9.489	-1.5034	1.0000	0.00021445		
Jun05	r = 0	21.73*	29.42*	1.0000	-1.0192	-2.2046e-006	3	1.1406 [0.1951]
	r <= 1	7.686	7.686	-0.70966	1.0000	-8.3167e-005		
Sep05	r = 0	12.39	19.44	1.0000	-0.90038	-1.9144e-005	3	1.3823 [0.0178]*
	r <= 1	7.047	7.047	-0.87635	1.0000	-4.0854e-005		
Dec05	r = 0	9.848	14.43	1.0000	-0.66074	-6.6649e-005	3	1.0152 [0.4450]
	r <= 1	4.585	4.585	-0.93078	1.0000	-3.9935e-005		
Mar06	r = 0	14.81	18.67	1.0000	0.46281	-0.00050229	35	1.3641 [0.0234]*
	r <= 1	3.866	3.866	-0.94383	1.0000	-5.2764e-005		
Jun06	r = 0	10.36	14.06	1.0000	-2.7027	0.00059863	19	1.3877 [0.0172] *
	r <= 1	3.71	3.71	-1.0751	1.0000	-0.00036784		
Sep06	r = 0	15.21	27.05*	1.0000	-0.93623	7.7387e-005	6	1.2879 [0.0510]
	r <= 1	11.84	11.84	-0.44973	1.0000	-0.00017627		
Dec06	r = 0	10.73	17.04	1.0000	-0.56383	-0.00010337	3	1.3957 [0.0152] *
	r <= 1	6.31	6.31	-1.0664	1.0000	-6.7977e-005		

Table 8.24: Johansen Cointegration Analysis

Future Code	Ho:rank=r	LR tests		standardized β' eigenvectors			lag used	vector error
		Max Eigenvalue	Trace	f_t	s_t	t		autocorrelation(20)
Mar07	r = 0	8.761	13.53	1.0000	-0.64367	-4.9711e-005	3	1.2155 [0.1033]
	r <= 1	4.772	4.772	-0.93707	1.0000	-0.00011641		
Jun07	r = 0	12.57	21.72	1.0000	-0.90656	6.8500e-005	4	1.3501 [0.0259]*
	r <= 1	9.154	9.154	-1.9222	1.0000	0.00029727		
Sep07	r = 0	24.27**	30.92**	1.0000	-0.95199	0.00010593	11	1.3735 [0.0199]*
	r <= 1	6.65	6.65	0.81229	1.0000	-0.00078032		
Dec07	r = 0	11.51	16.81	1.0000	-0.94658	0.00010273	6	1.3704 [0.0205]*
	r <= 1	5.296	5.296	-0.84924	1.0000	-0.00016385		
Mar08	r = 0	28.24**	30.84**	1.0000	-0.99570	0.00013229	4	1.3893 [0.0164]*
	r <= 1	2.592	2.592	-0.85837	1.0000	-7.7412e-005		
Jun08	r = 0	13.94	19.84	1.0000	-1.0063	0.00014324	4	1.3415 [0.0285]*
	r <= 1	5.895	5.895	-1.3174	1.0000	-0.00024832		
Sep08	r = 0	9.015	14.17	1.0000	-0.90640	0.00019775	3	1.2268 [0.0929]
	r <= 1	5.153	5.153	-0.98744	1.0000	-0.00012815		
Dec08	r = 0	14.59	21.4	1.0000	-1.0547	0.00028238	23	1.3725 [0.0206]*
	r <= 1	6.802	6.802	-1.0968	1.0000	-0.00023847		

Notes: *', * and ** superscripts indicate significance at 10%, 5% and 1%, respectively. a Lagrange-Multiplier (LM) test for serial correlation is used for lag up to 20. F-statistics are shown. This test is done through the auxiliary regression of the residuals on the original variables and lagged residuals (missing lagged residuals at the start of the sample are replaced by zero, so no observations are lost). unrestricted variables are included in the auxiliary regression. The null hypothesis is no autocorrelation, which would be rejected if the test statistic is too high.

Table 8.25: Comparison of cointegration analysis result using Engel-Granger and Johansen

Future Code	Engel-Granger	Johansen	Future Code	Engel-Granger	Johansen
Mar98			Mar04		
Jun98			Jun04		×
Sep98			Sep04		
Dec98			Dec04		
Mar99			Mar05		
Jun99			Jun05		
Sep99			Sep05		×
Dec99			Dec05	×	×
Mar00	×	×	Mar06		×
Jun00			Jun06	×	×
Sep00	×	×	Sep06	×	×
Dec00	×	×	Dec06		×
Mar01		×	Mar07	×	×
Jun01			Jun07		×
Sep01	×	×	Sep07		
Dec01	×	×	Dec07		×
Mar02	×	×	Mar08		
Jun02	×	×	Jun08		×
Sep02	×	×	Sep08	×	×
Dec02	×	×	Dec08	×	×
Mar03	×	×			
Jun03		×			
Sep03					
Dec03					

Notes: × denotes f_t and s_t are not cointegrated at a 5% significant level.

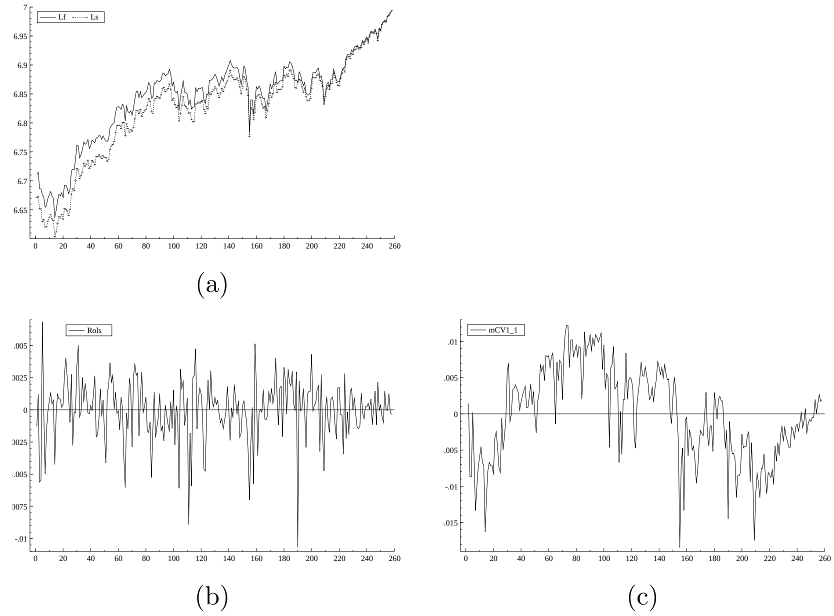


Figure 8.4: Future contract Mar98 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag =1 .

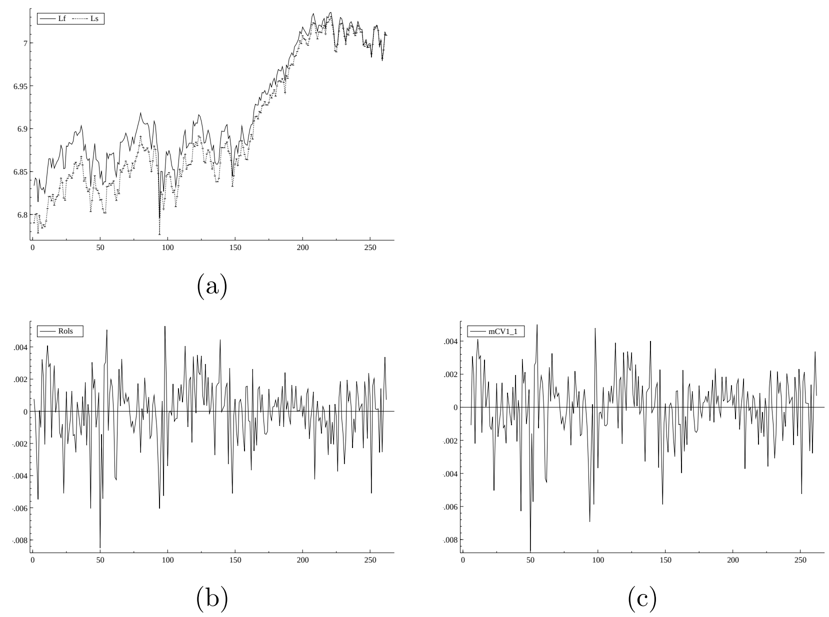


Figure 8.5: Future contract Jun98 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag =1 .

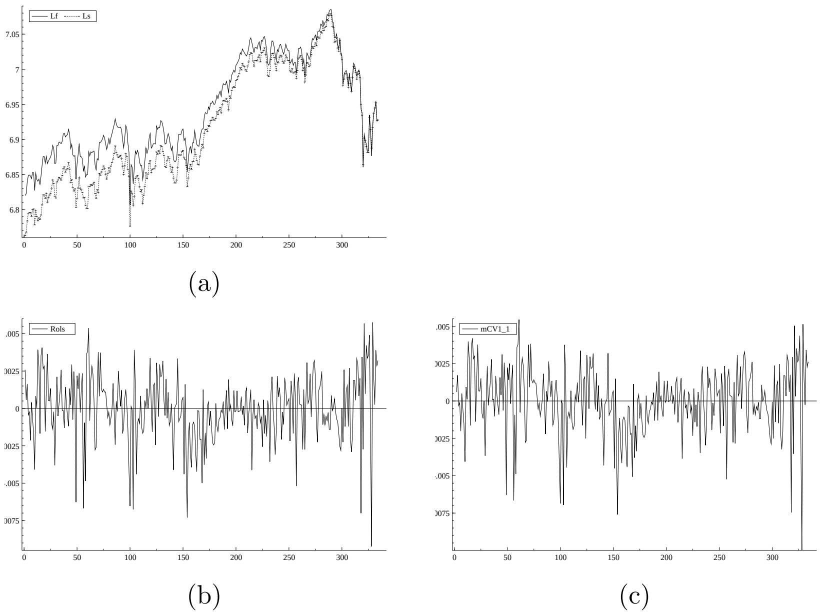


Figure 8.6: Future contract Sep98 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag =1 .

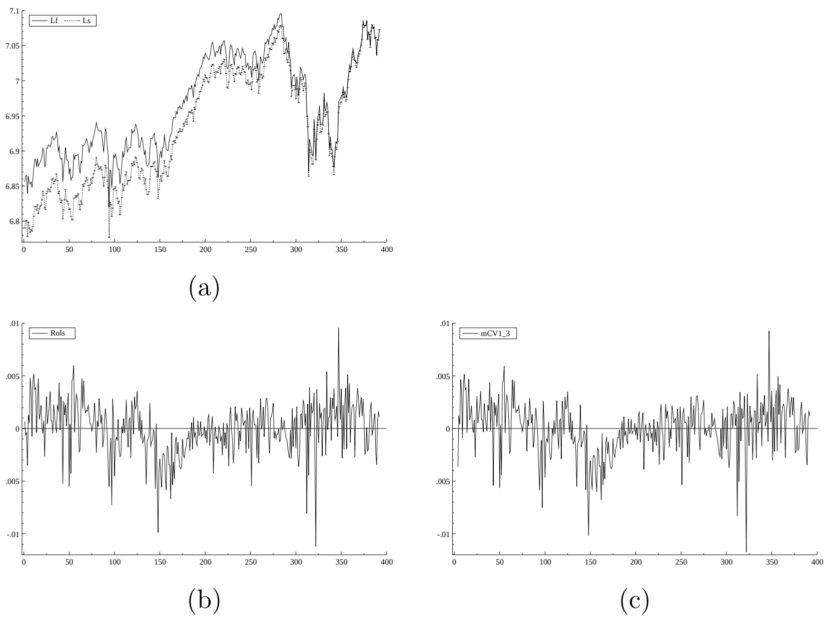
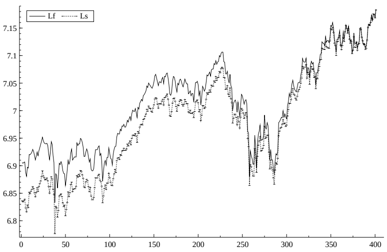
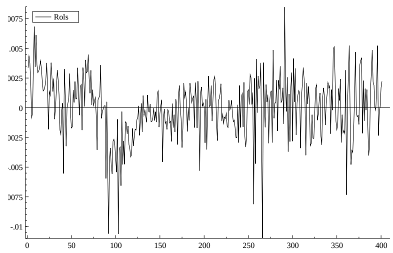
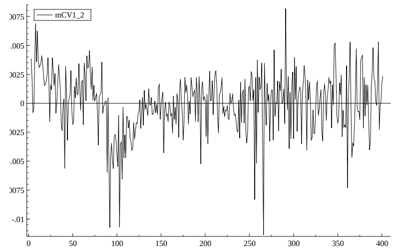


Figure 8.7: Future contract Dec98 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



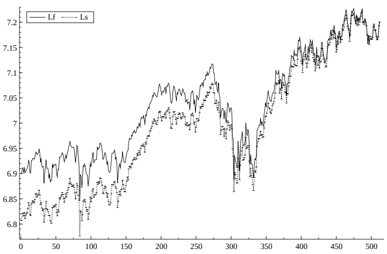
(a)



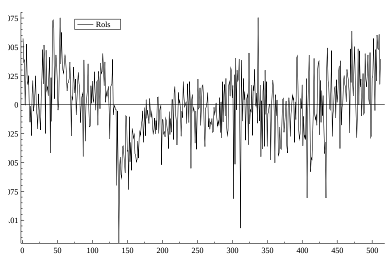


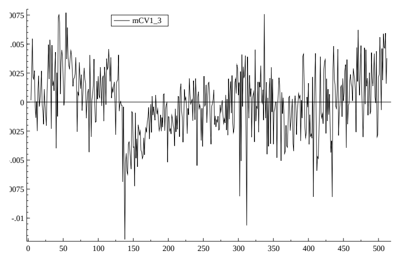
(b) (c)

Figure 8.8: Future contract Mar99 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 2.



(a)





(b) (c)

Figure 8.9: Future contract Jun99 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.

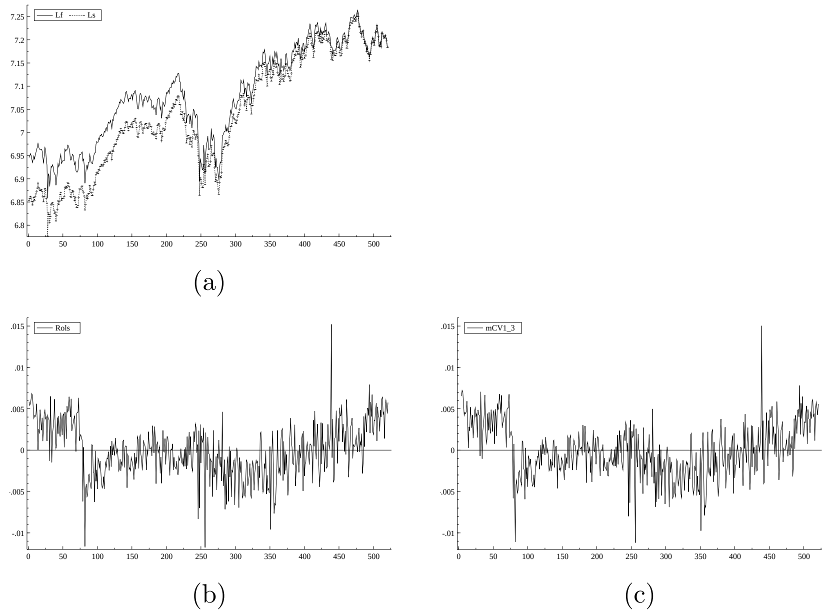


Figure 8.10: Future contract Sep99 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.

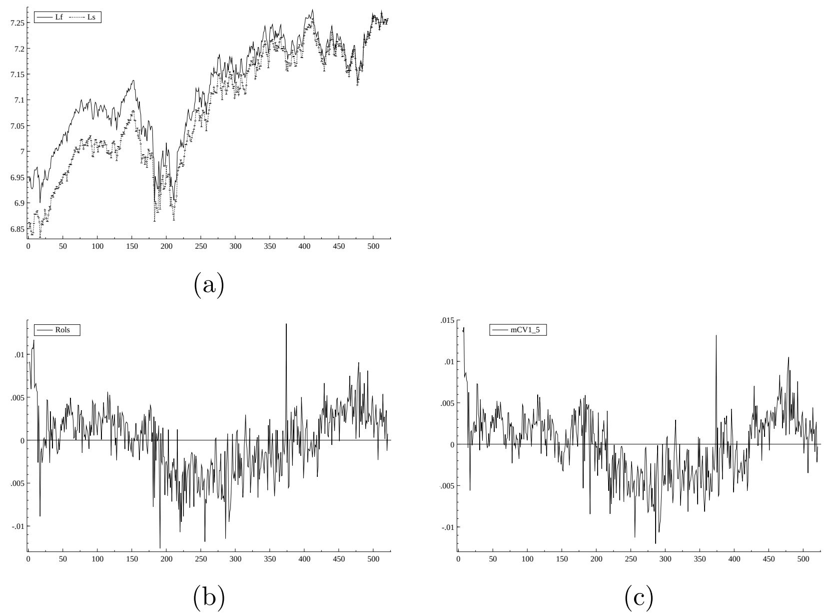


Figure 8.11: Future contract Dec99 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.

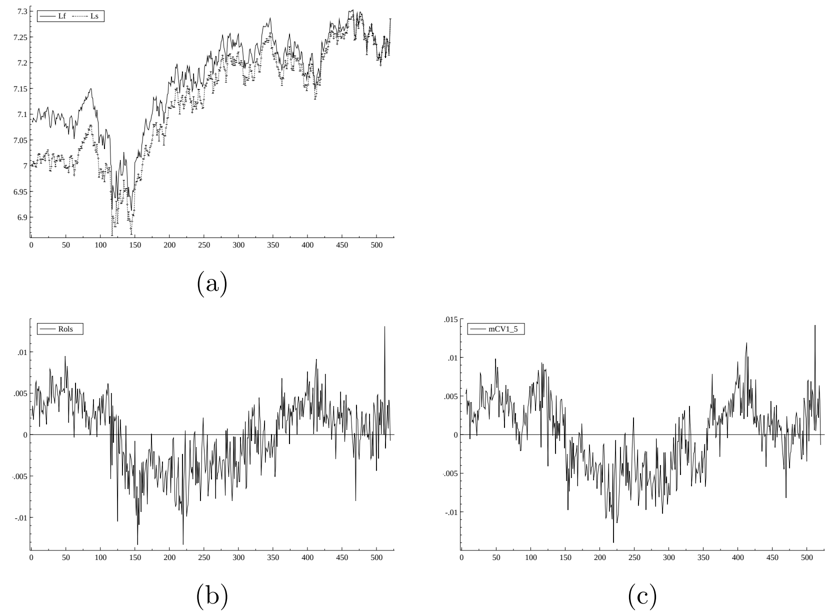


Figure 8.12: Future contract Mar00 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5 .

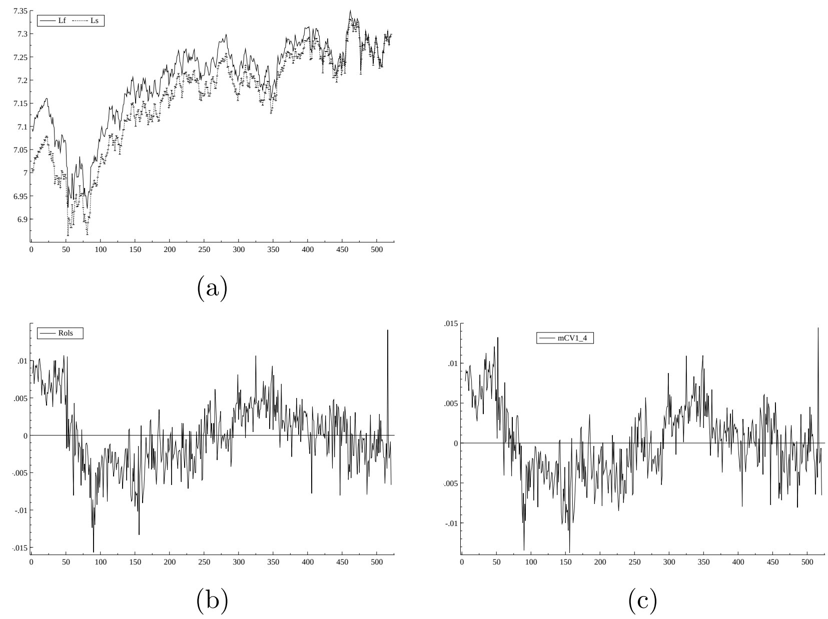
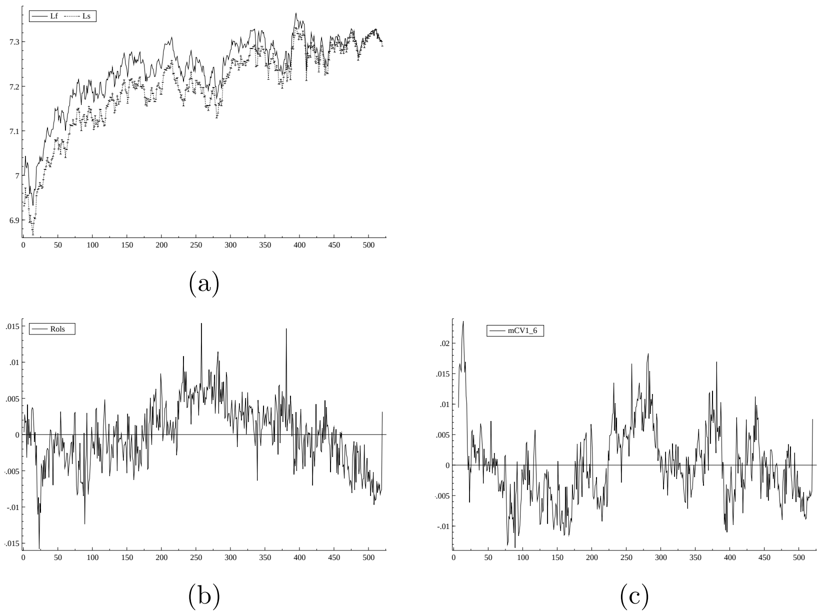


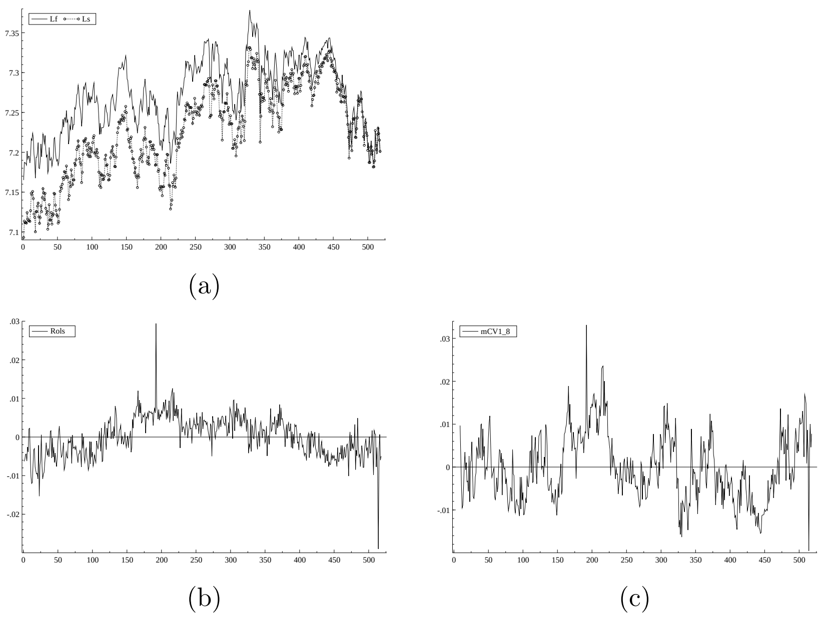
Figure 8.13: Future contract Jun00 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.



(a)

(b) (c)

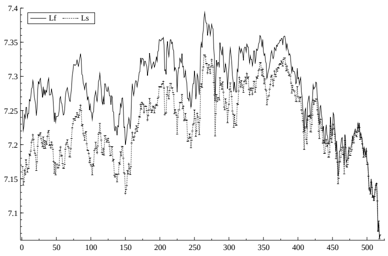
Figure 8.14: Future contract Sep00 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 6.



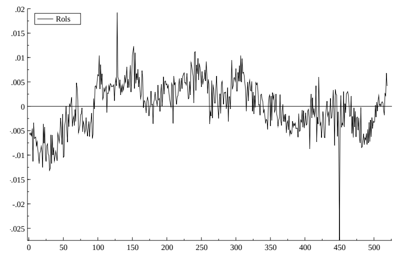
(a)

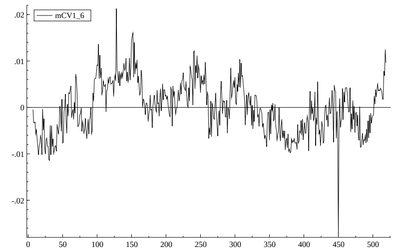
(b) (c)

Figure 8.15: Future contract Dec00 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 8.



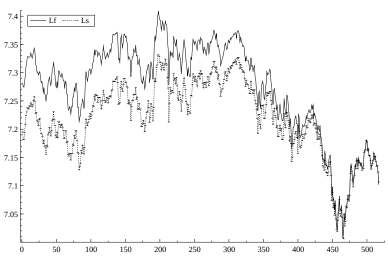
(a)



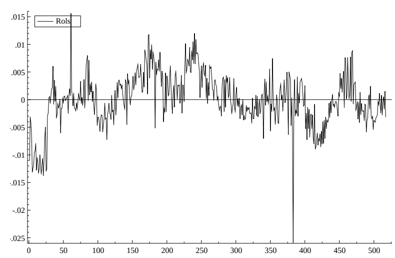


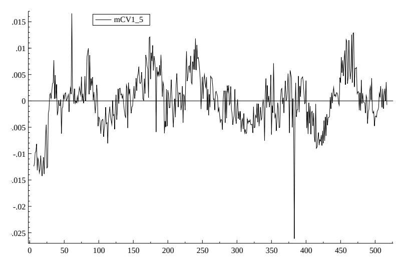
(b) (c)

Figure 8.16: Future contract Mar01 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 6.



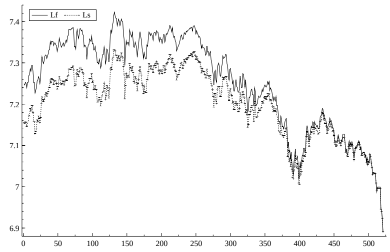
(a)



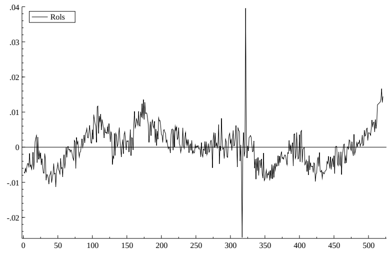


(b) (c)

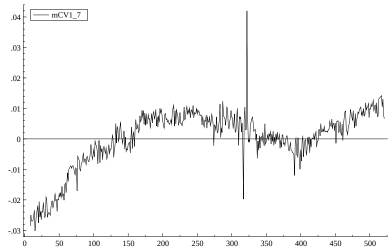
Figure 8.17: Future contract Jun01 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.



(a)

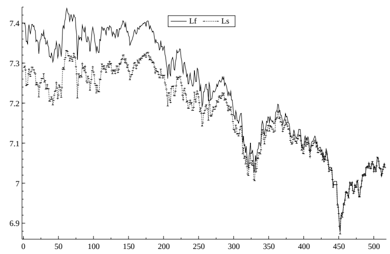


(b)

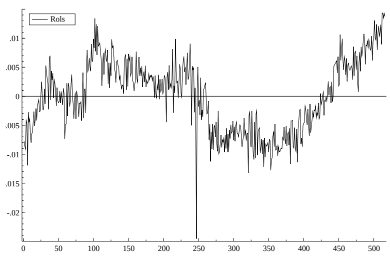


(c)

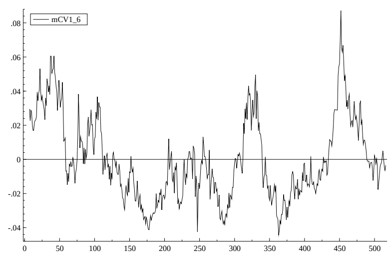
Figure 8.18: Future contract Sep01 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 7.



(a)

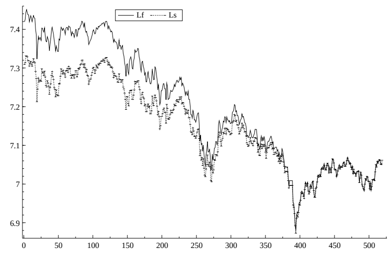


(b)

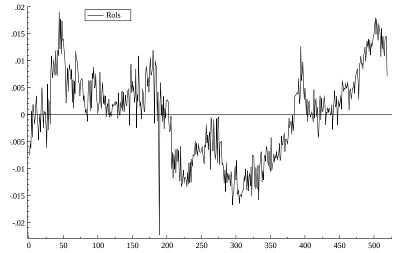


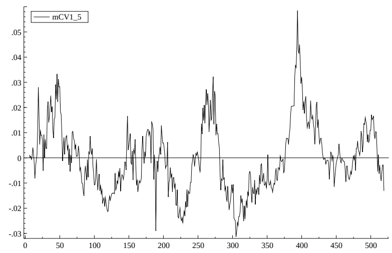
(c)

Figure 8.19: Future contract Dec01 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 6.



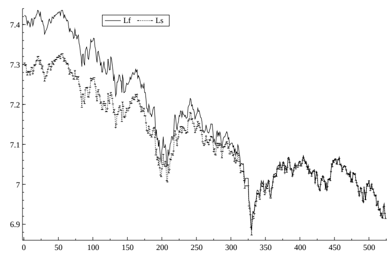
(a)



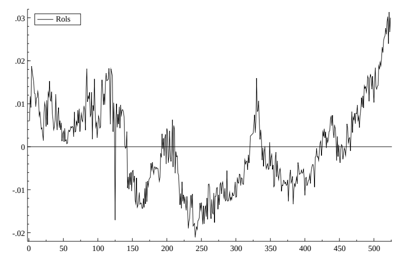


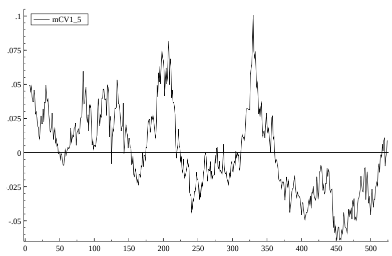
(b) (c)

Figure 8.20: Future contract Mar02 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.



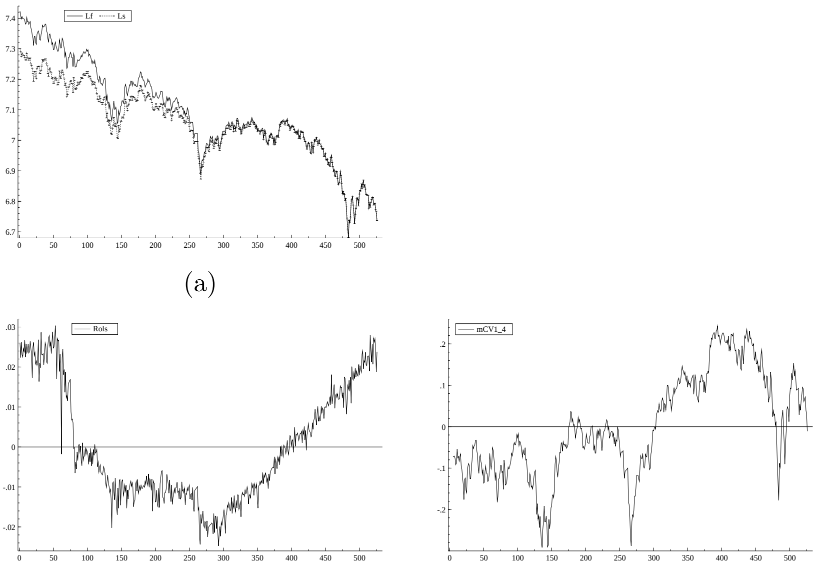
(a)





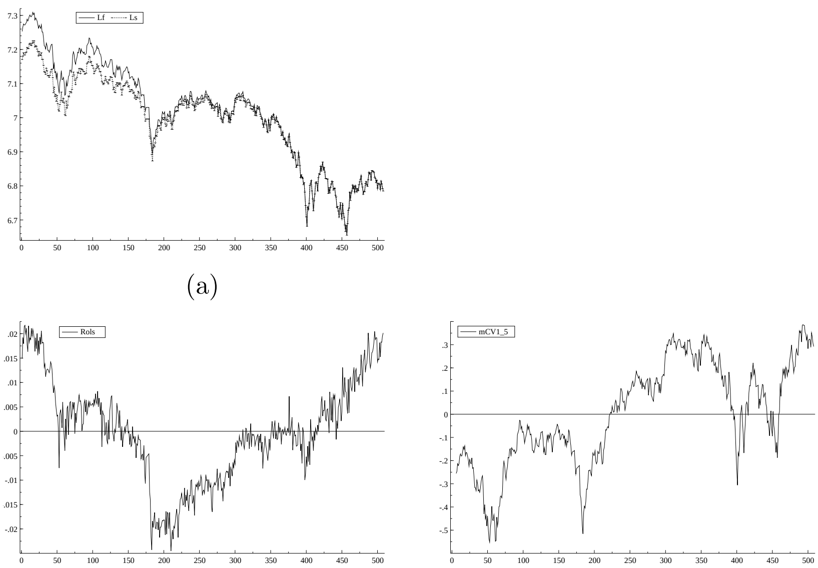
(b) (c)

Figure 8.21: Future contract Jun02 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.



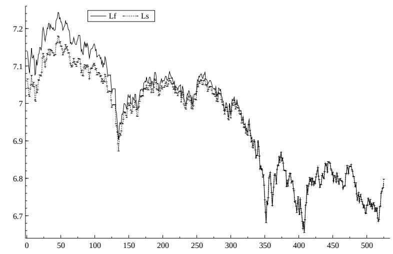
(a) (b) (c)

Figure 8.22: Future contract Sep02 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.

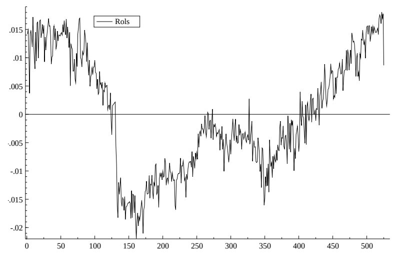


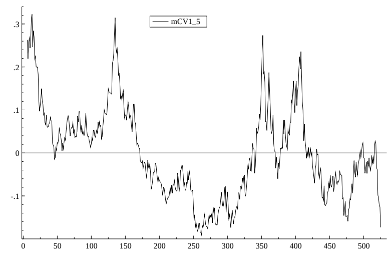
(a) (b) (c)

Figure 8.23: Future contract Dec02 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.



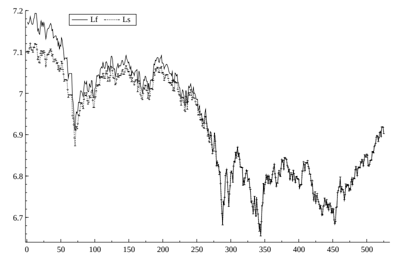
(a)



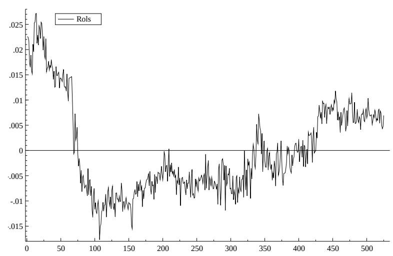


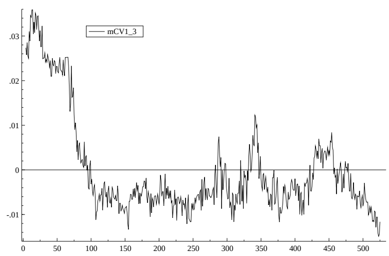
(b) (c)

Figure 8.24: Future contract Mar03 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.



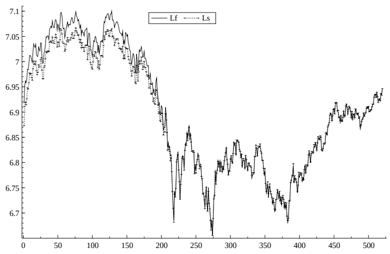
(a)



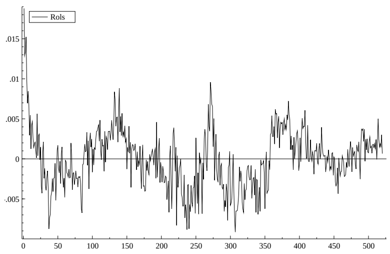


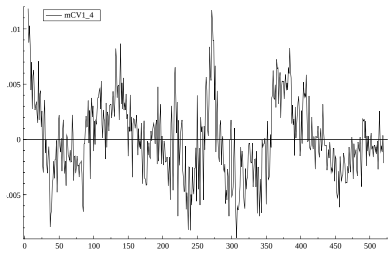
(b) (c)

Figure 8.25: Future contract Jun03 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



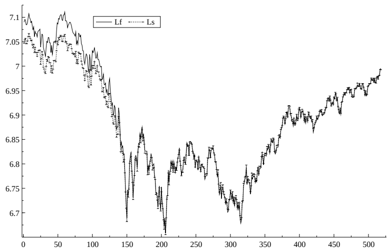
(a)



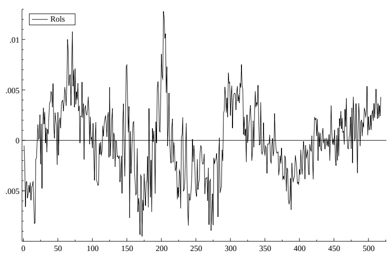


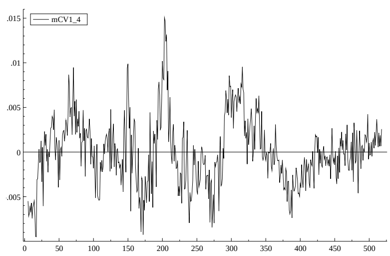
(b) (c)

Figure 8.26: Future contract Sep03 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.



(a)





(b) (c)

Figure 8.27: Future contract Dec03 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.

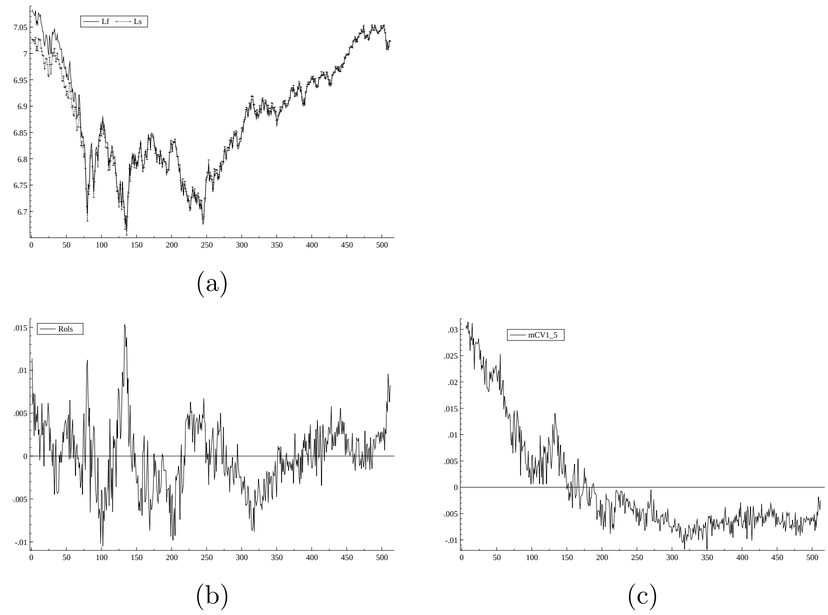


Figure 8.28: Future contract Mar04 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 5.

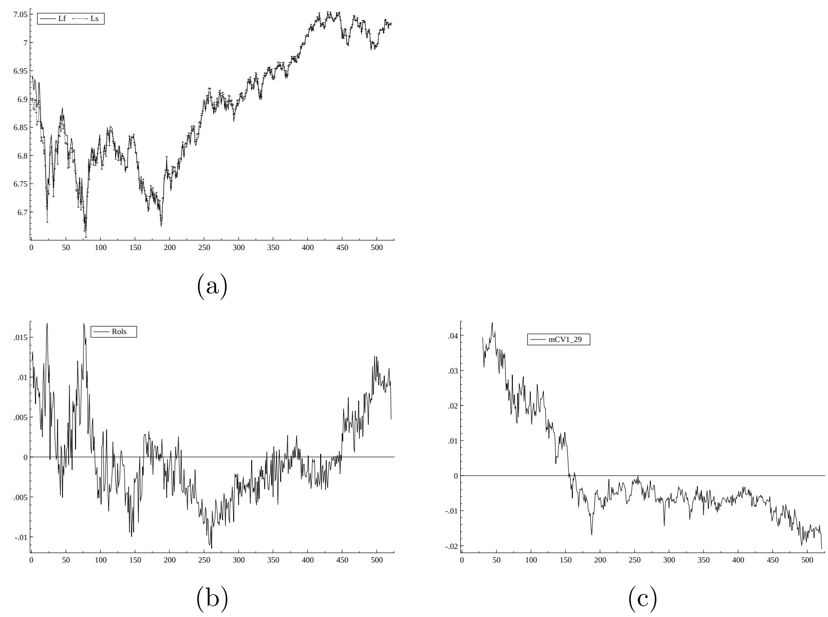


Figure 8.29: Future contract Jun04 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 29.

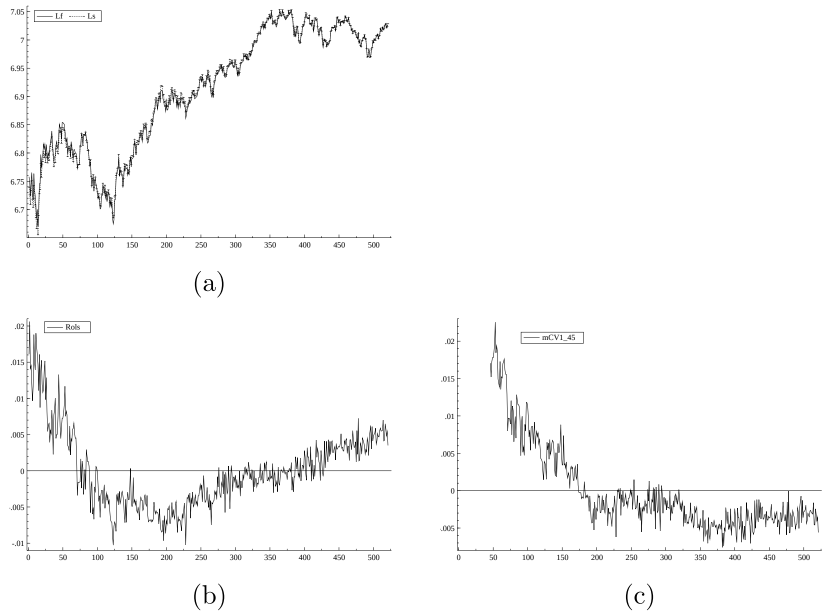


Figure 8.30: Future contract Sep04 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 45.

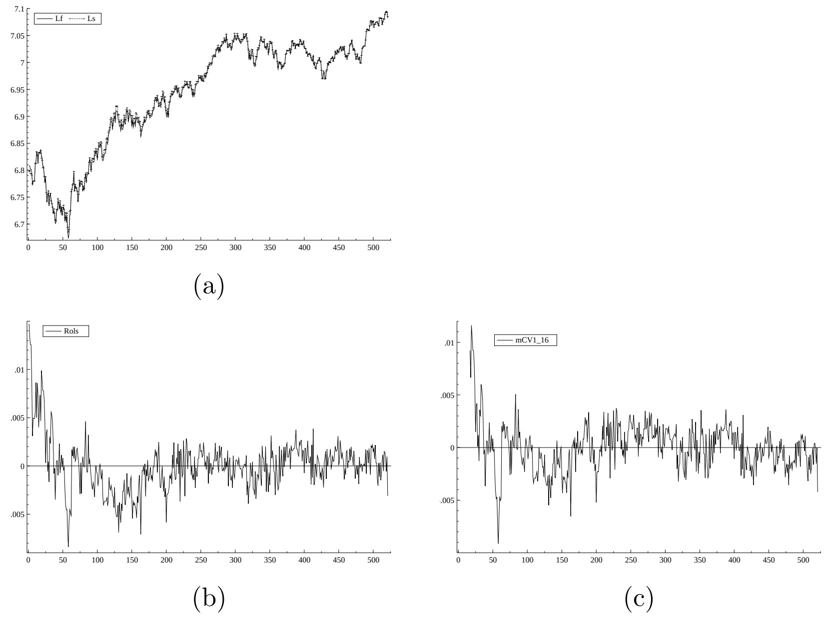
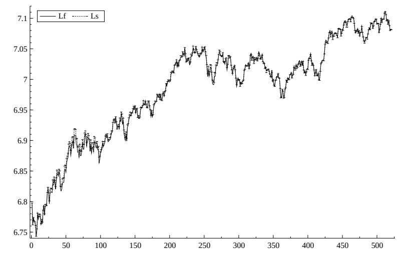
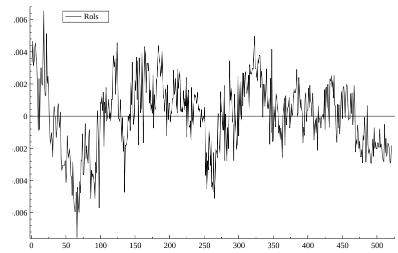
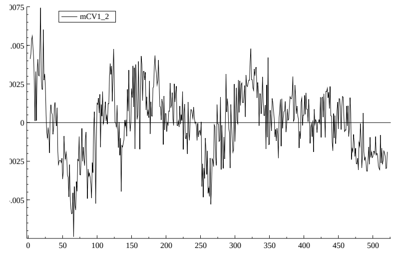


Figure 8.31: Future contract Dec04 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 16.



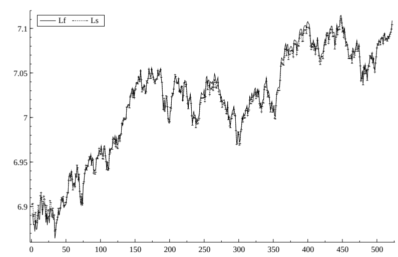
(a)



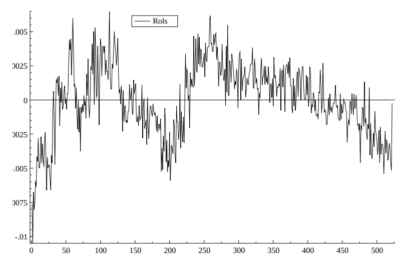


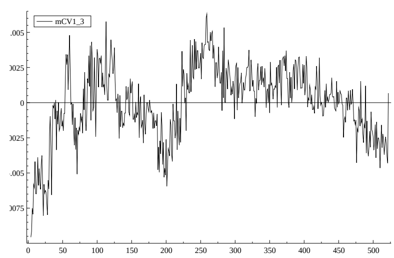
(b) (c)

Figure 8.32: Future contract Mar05 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 2.



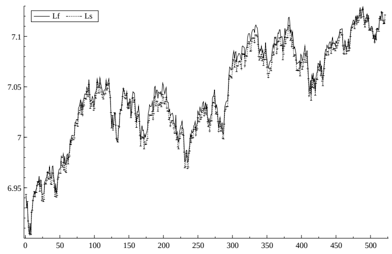
(a)



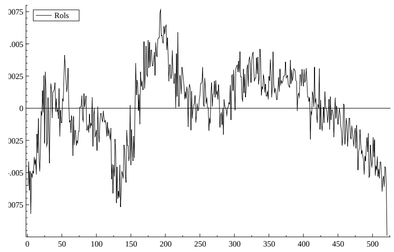


(b) (c)

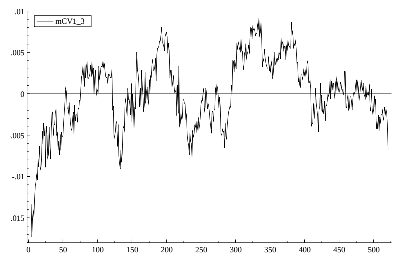
Figure 8.33: Future contract Jun05 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



(a)

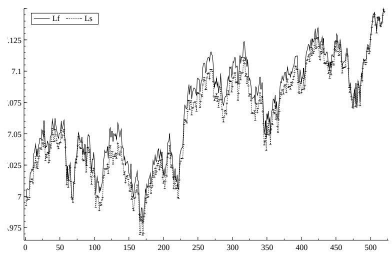


(b)

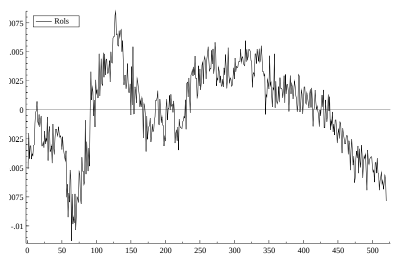


(c)

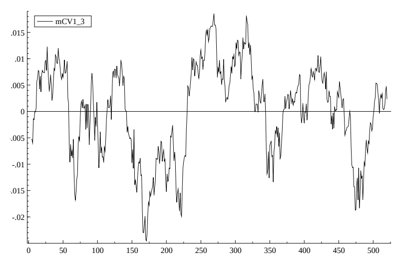
Figure 8.34: Future contract Sep05 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



(a)



(b)



(c)

Figure 8.35: Future contract Dec05 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.

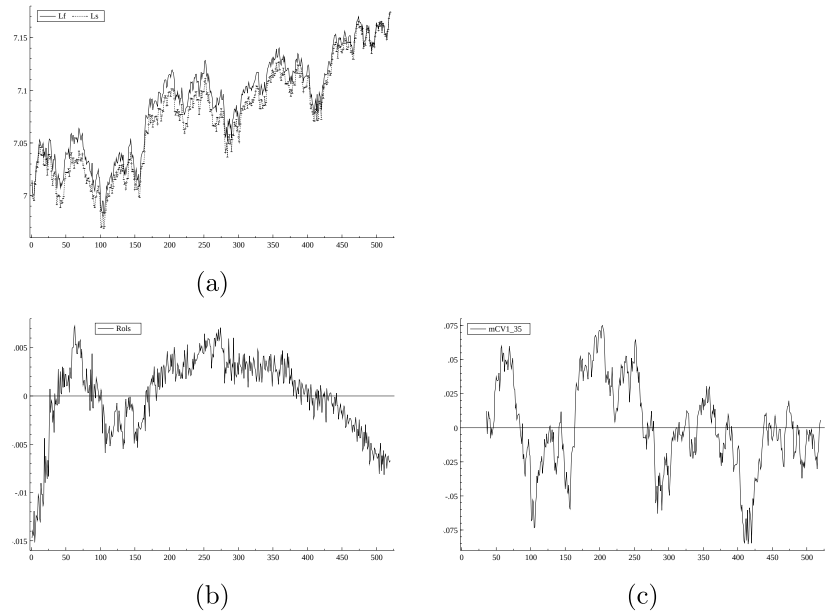


Figure 8.36: Future contract Mar06 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 35.

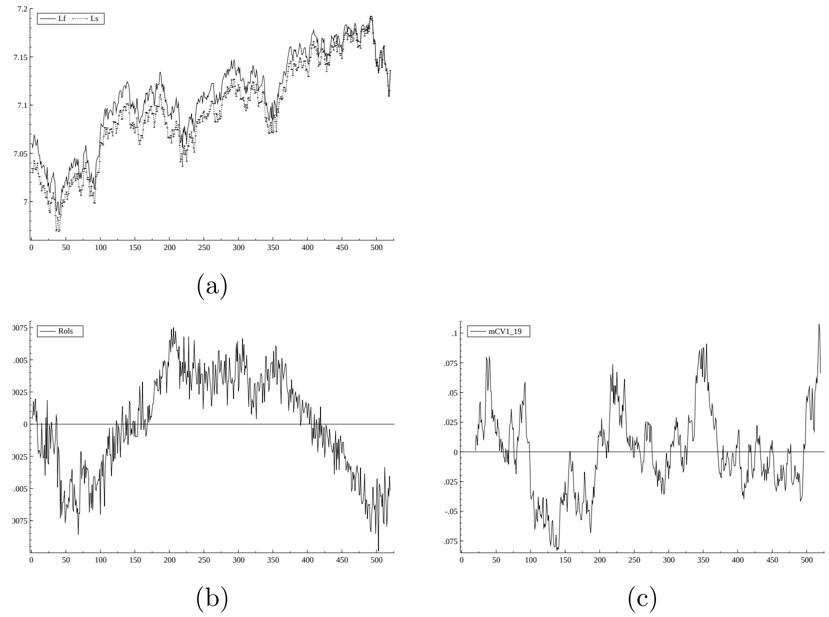
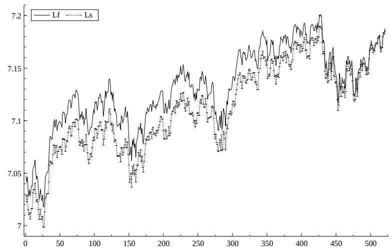
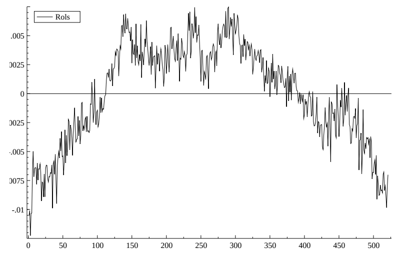
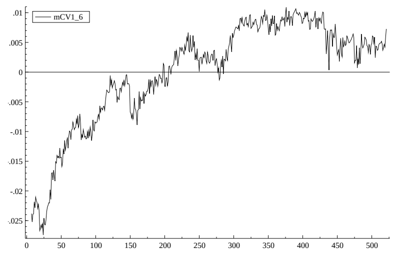


Figure 8.37: Future contract Jun06 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 19.



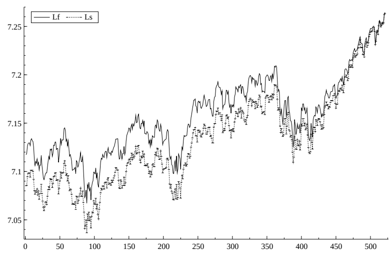
(a)



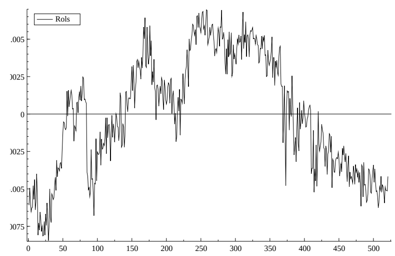


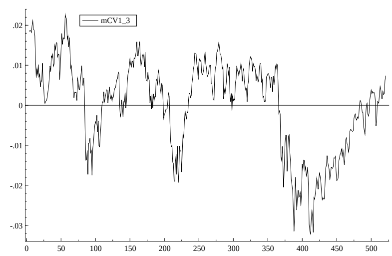
(b) (c)

Figure 8.38: Future contract Sep06 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 6.



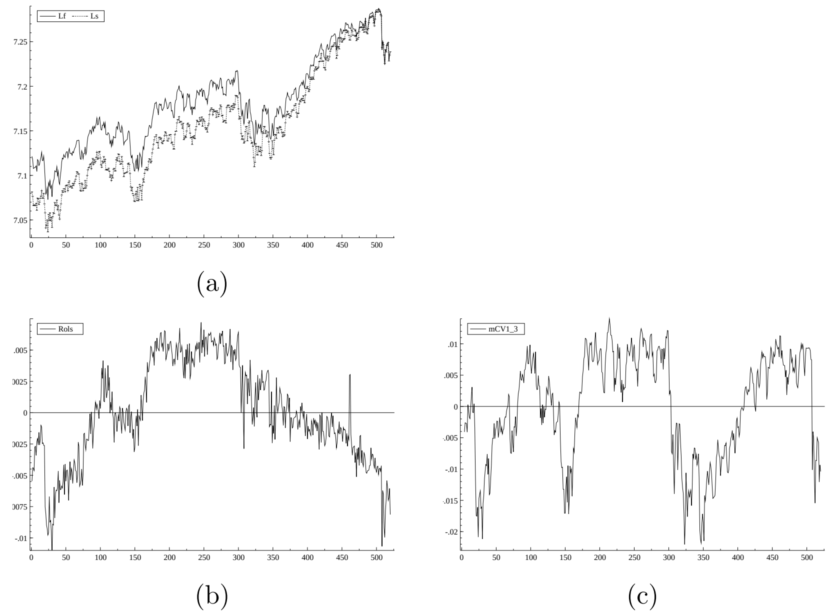
(a)





(b) (c)

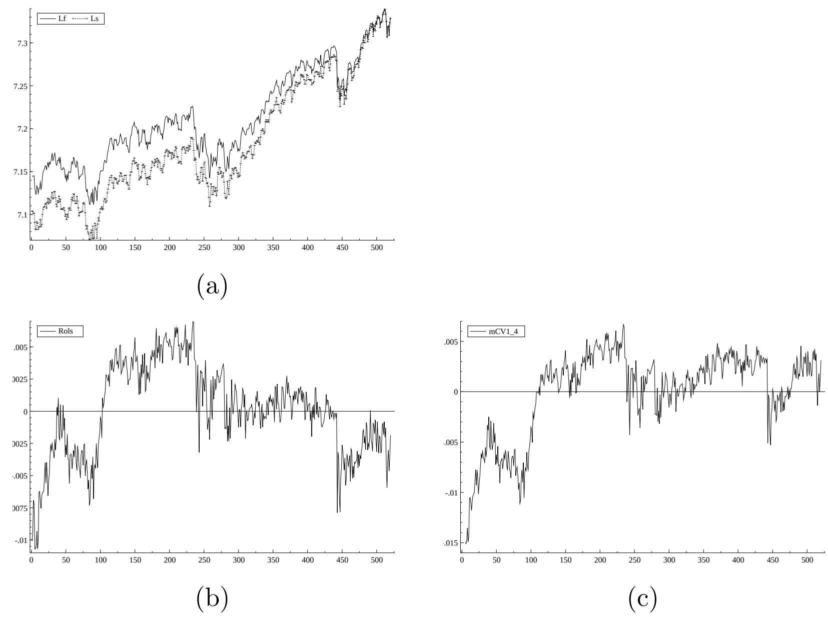
Figure 8.39: Future contract Dec06 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



(a)

(b) (c)

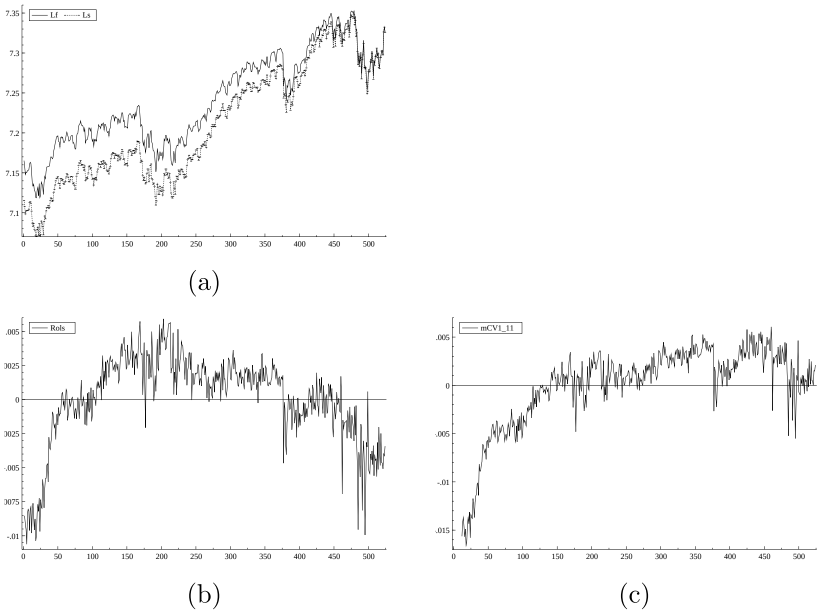
Figure 8.40: Future contract Mar07 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



(a)

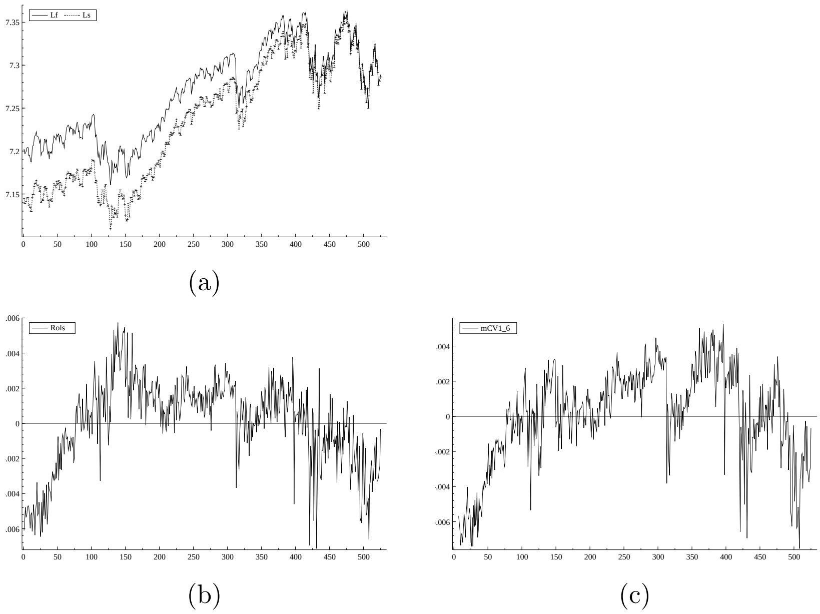
(b) (c)

Figure 8.41: Future contract Jun07 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.



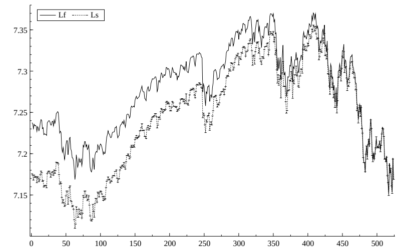
(a) (b) (c)

Figure 8.42: Future contract Sep07 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 11.

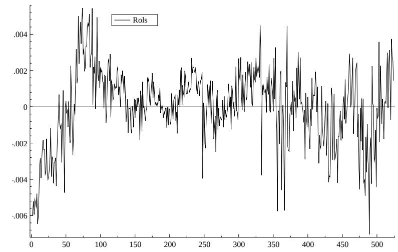


(a) (b) (c)

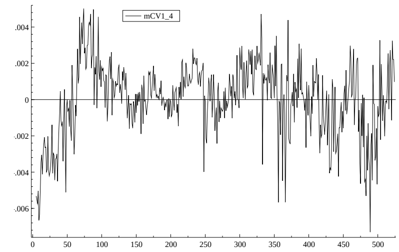
Figure 8.43: Future contract Dec07 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 6.



(a)

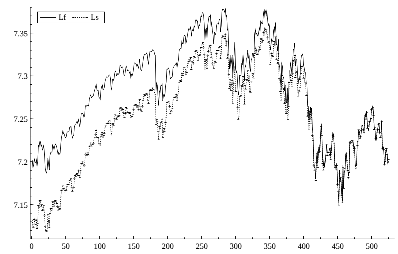


(b)

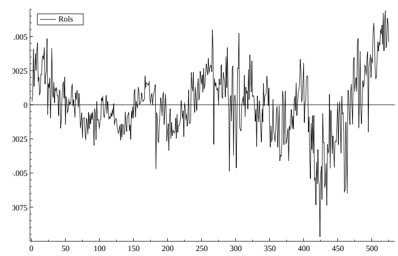


(c)

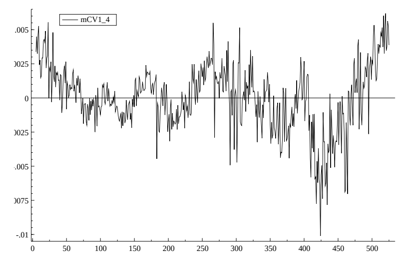
Figure 8.44: Future contract Mar08 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.



(a)

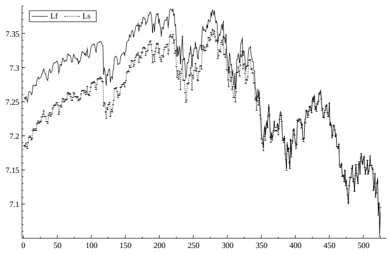


(b)

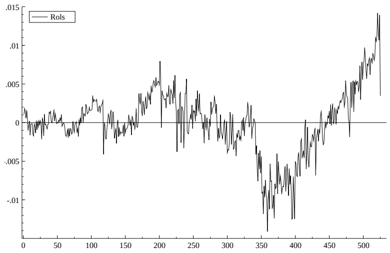


(c)

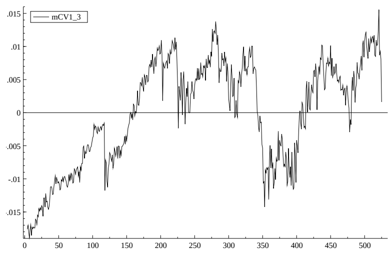
Figure 8.45: Future contract Jun08 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 4.



(a)

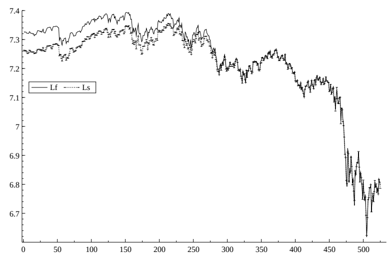


(b)

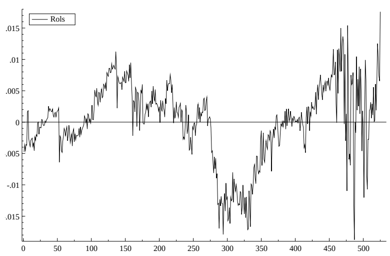


(c)

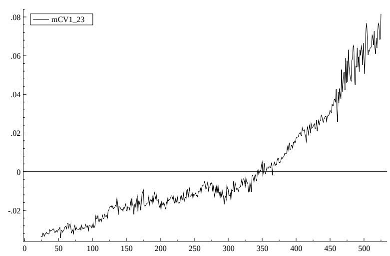
Figure 8.46: Future contract Sep08 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 3.



(a)



(b)



(c)

Figure 8.47: Future contract Dec08 : (a) Plots of f_t and s_t, (b) Regression Residuals, (c) The first disequilibrium error (demeaned) with lag = 23.

8.4 Pairs Trading between Future and Spot Index Prices

This section will explain pairs trading strategy between future and spot index prices of the S&P 500. The first subsection describes how the future contracts are traded in the market while the second subsection explains how the spot indexes are traded in the market. The third subsection shows the pairs trading strategy by putting these assets together while the last subsection gives an empirical example of pair trading simulations.

8.4.1 The Mechanism of Future Contracts

Future contracts are traded on organised exchanges with standardised terms. For examples, S&P 500 future contracts are traded on the Chicago Mercantile Exchange (CME) and FTSE 100 future contracts are traded on the London International Financial Futures Exchange (LIFFE). Stock index futures were introduced in Australia in 1983 in the form of Share Price Index (SPI) futures which are based on the Australian Stock Exchange's (ASX) All Ordinaries Index which is the benchmark indicator of the Australian stock market.

There is a feature known as "marking to market" for future contracts. It means that the intermediate gains or losses are given by the difference between today's future price and yesterday's future price. This concept is standard across all major future contracts. Contracts are marked to market at the close of trading day until the contract expires. At expiration, there are two different mechanisms for settlement. Most financial futures (such as stock index, foreign exchange and interest rate futures) are cash settled, whereas most physical futures (such as agricultural, metal and energy futures) are settled by delivery of the physical commodity.

Another feature for future contracts is "margin". Although future contracts require no initial investment, future exchanges require both the buyer and seller to post a security deposit known as "margin". Margin is typically set at an amount that is larger than usual one-day moves in the future price. This is done to ensure that both parties will have sufficient funds available to mark to market. Residual credit risk exists only to the extend that (1) future prices move so dramatically that the amount required to mark to market is larger than the balance of an individual's margin account, and (2) the individual defaults on payment of the balance. In this case, the exchange bears the loss so that participants in futures markets bear essentially zero credit risk. Margin rules are stated in terms of "initial margin" which must be posted when entering the contract and "maintenance margin" which is the minimum acceptable balance in the margin account. If the balance of the account falls below the maintenance level, the exchange makes a "margin call" upon individual, who must then restore the account to the level of initial margin before start of trading the following day.

8.4.2 Trading Index

Stock index represents the weighted average market value of all shares selected in the index. Large investment funds can build a portfolio mimicking the index as a passive investment strategy. They hold stocks underlying the index in a proposition consistent with weights set by the index. Thus, the return from the portfolio will be the same as the return of the index. Building this kind of portfolio will need a lot of money so that it is not suitable for individual investors. However, nowadays, there is a financial product called as "Contract for Difference" (or CFD for short). This hot new product was launched in Australia about 5 years ago. The key feature of CFD is that they involve us entering a contract with a CFD issuer for a particular asset such as shares, stock indexes and foreign currencies. If the price moves as we thought it would, we will get a profit as the CFD issuer pays us the difference between the initial price of the asset when we enter the contract and the price it is trading at when we close out the contract. The opposite thing happens if the price does not move as we thought. So, that is why it is named as a contract for difference.

By having CFD, we never actually own the underlying assets, only the right to get any gains from the price changes and of course the responsibility for any losses. CFD are highly geared products since we only pay a very small deposit or margin, often as little as 3% of the value of the assets we buy. As a result, CFD generally give us much more leverage than using a margin loan. This means even a small price movement in our favour can generate a large percentage gain, while a small movement against us can result in a large percentage loss. For an example, let us assume the outlook for resource stocks are very encouraging. We want to buy a big resources group XYZ Ltd because we have assessed its current price of $25 to be well below the intrinsic value. But we only have $1000 available which does not allow us to buy many shares. If, instead we use the same amount to buy CFD for XYZ Ltd shares, we would be able to take an exposure to 800 shares with 5% of the value of the underlying shares. If the share price increases to, say, $30, we will get a profit of 300% [6]. However, if the share price decreases to, say, $20, we will get a loss of 300%. As we never actually own the assets, it is possible to bet on prices falling ("going short") as well as betting on prices increasing ("going long").

Another important aspect of CFD is that they incorporate an interest. When we trade a "long" CFD , that is, one that is based on the expectation of a price rise, we will incur interest each day we hold the CFD. On the other hand, if we trade "short" CFD, that is, one that is based on the expectation of a price fall, we will receive interest each day. The rate of interest for long positions is usually around 2 percentage points above the overnight cash rate.

Similar to having real assets, if we have CFD for shares or stock indexes, we will

[6] $(\$5 \times 800 - \$1000)/\$1000 \times 100\%$.

receive dividends if we hold a long CFD which can offset the interest incurred. In contrast, the interest paid on a short position may be reduced by the dividends the underlying shares generate during the time the position is open.

The same as a margin loan, we can be called to contribute more of our own money should the underlying asset price move against our bet. This call is made to ensure the deposit we paid to buy the CFD does not fall in percentage terms as the asset price changes.

8.4.3 Pairs Trading between Future and Spot Index

As we explained in the previous subsections, future and spot index have different trading mechanisms. The mark to market aspect of futures results in a risk. The uncertainty is about the amount of daily transfers of profits or losses. Similar thing happens by trading a CFD for a stock index. There is uncertainty about the amount of interest and dividend received or paid. Furthermore, there is a time trend involved in the relationship between future and spot index prices. Because of these features, pairs trading strategy between future and spot index the S&P 500 here is different to pairs trading strategy described in Chapters 4 and 5.

Assume that the log of future and spot index prices are cointegrated as follow:

$$f_t = \mu + \beta\, s_t + \delta\,(T - t) + \epsilon_t \tag{8.37}$$

where f_t is the log of future contract price with the nearest maturity date is T, s_t is the log of spot index prices and and $(T-t)$ is a time to maturity based on trading day. As f_t and s_t are cointegrated, ϵ_t from (8.37) follows a stationary model.

We make further assumptions below to simplify our discussion:

(1) Pairs trading is done by trading future index and CFD of the index.

(2) The percentage margin for future contracts and CFD are the same.

(3) The trading costs for both assets are very small so that it can be excluded from the modelling. We aware that this assumption is not realistic in practice, but for a starting point, it can give us basic understanding about pairs trading mechanism between the two assets.

Pairs trading strategy:

(1) Set a pre-determined upper-bound U and a lower-bound $L = -U$.

(2) For upper trades, put future in “short” (sell) position and β CFD in “long” (buy) position if $\epsilon_t > U$ for an upper trade. Then, close the pair trade by taking the opposite position on the next day. For lower trades, put future in “long” (buy) position and β CFD in “short” (sell) position if $\epsilon_t < -U$ for a lower trade. Then,

close the pair trade by taking the opposite position on the next day. We impose daily trading to reduce uncertainty of mark to market outcomes and the interest. The longer we hold the assets, the higher the uncertainty, thus the higher the risks. Nowadays, many investors are doing daily trading (Hely, 2008) to speculate for asset prices movements.

(3) There is no overlap pairs trades. It means, for example, if we open a pair trade at $t = 1$ because $\epsilon_1 > U$, we have to close it on the next day and we cannot open another pair trade, even if at $t = 2$, $\epsilon_2 > U$ or $\epsilon_2 < -U$.

As we do not close a pair trade at a certain value, we cannot use the formulas developed in Chapter 4 to estimate the number of trades and eventually estimate the optimal upper-bound U_0. However, we can still use the stationary properties of ϵ_t in the pairs trading strategy.

Suppose that we open an upper trade at time t as $\epsilon_t \geq U$. So, we put future in "short" (sell) position and β CFD in "long" (buy) position. Then, we close it on the next day at time $t + 1$ by taking the opposite position. We define return for each pair trade as follow:

$$\begin{aligned} return &= \frac{1}{m}\left((f_t - f_{t+1}) + \beta\,(s_{t+1} - s_t)\right) && (8.38)\\ &\approx -\frac{F_{t+1} - F_t}{m\,F_t} + \beta\frac{S_{t+1} - S_t}{m\,S_t} && (8.39)\end{aligned}$$

where F_t and S_t are the prices of future and spot index prices, respectively, and m is the percentage margin. The percentage margins for future contract and CFD index are assumed the same. We arrange (8.38) as follow:

$$\begin{aligned} return &= \frac{1}{m}\left[(f_t - f_{t+1}) + \beta\,(s_{t+1} - s_t)\right]\\ &= \frac{1}{m}\left[f_t - \beta\,s_t - \mu - \delta\,(T - t)\right] - \frac{1}{m}\left[f_{t+1} - \beta\,s_{t+1} - \mu - \delta\,(T - t - 1)\right] + \frac{1}{m}\delta\\ &= \frac{1}{m}\left[f_t - \beta\,s_t - \mu - \delta\,(T - t)\right] - \frac{1}{m}\left[f_{t+1} - \beta\,s_{t+1} - \mu - \delta(T - t - 1)\right] + \frac{1}{m}\delta\\ &= \frac{1}{m}\left[\epsilon_t - \epsilon_{t+1} + \delta\right] \geq \frac{1}{m}\left[U - \epsilon_{t+1} + \delta\right] && (8.40)\end{aligned}$$

Suppose that we have $\delta > 0$. From (8.40), we will have a positive return or profit if $\epsilon_{t+1} \leq \epsilon_t$. In contrast, we will have a negative return or loss if $\epsilon_{t+1} > (\epsilon_t + \delta)$.

Let $f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t)$ denotes the probability density function of ϵ_{t+1} given $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$. From Section 4.1, if ϵ_t follows an AR(p) model,

$$f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t) = f(\eta_{t+1}) \sim N(0, \sigma_\eta^2) = \frac{1}{\sqrt{2\pi}\sigma_\eta}\exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right). \quad (8.41)$$

Thus, given $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$,

$$\begin{aligned} E(return) &= \frac{1}{m}\int_{-\infty}^{\infty}(\epsilon_t - \epsilon_{t+1} + \delta) f(\epsilon_{t+1}|\overrightarrow{\epsilon}_t) d\epsilon_{t+1} \\ &= \frac{1}{m}\int_{-\infty}^{\infty}(\epsilon_t + \delta) f(\epsilon_{t+1}|\overrightarrow{\epsilon}_t) d\epsilon_{t+1} - \frac{1}{m}\int_{-\infty}^{\infty}\epsilon_{t+1} f(\epsilon_{t+1}|\overrightarrow{\epsilon}_t) d\epsilon_{t+1} \\ &= \frac{1}{m}(\epsilon_t + \delta - c) \;\geq\; \frac{1}{m}(U + \delta - c) \end{aligned} \tag{8.42}$$

where $c = \theta_0 + \sum_{j=1}^{p} \theta_j \epsilon_{t-j+1}$.

From (8.42), it shows that the higher the value of ϵ_t, the higher the expected return and a positive expected return will be achieved if

$$\epsilon_t > c - \delta.$$

In a simple case where ϵ_t is a white noise process with zero mean, $c = 0$ and $\sigma_\eta = \sigma_\epsilon$. Thus, for this case, a positive expected return will be achieved if

$$\epsilon_t > U > -\delta \tag{8.43}$$

where U is the pre-set upper-bound.

From (8.40), not like the pairs trading strategy in Chapter 4, there is no guaranty of minimal profit for each pair trade. We will have a negative return or loss if $\epsilon_{t+1} > (\epsilon_t + \delta)$. The probability of getting a loss for a pair trade, given $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$ is

$$\begin{aligned} P(Loss) &= \frac{1}{\sqrt{2\pi}\sigma_\eta}\int_{(\epsilon_t+\delta)}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\ &\leq \frac{1}{\sqrt{2\pi}\sigma_\eta}\int_{\epsilon_t}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\ &\leq \frac{1}{\sqrt{2\pi}\sigma_\eta}\int_{U}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\ &= 1 - \frac{1}{\sqrt{2\pi}\sigma_\eta}\int_{-\infty}^{U} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1}. \end{aligned} \tag{8.44}$$

For a simple case where ϵ_t is a white noise process with zero mean, $\sigma_\eta = \sigma_\epsilon$ and distribution of ϵ_{t+1} is the same as the distribution of ϵ_t. Thus, for this case,

$$P(Loss) \leq 1 - \Phi\left(\frac{U}{\sigma_\epsilon}\right) \tag{8.45}$$

where $\Phi(.)$ denotes the cumulative distribution of a standard normal distribution. From (8.45), the higher the pre-set upper-bound U, the lower the probability of loss. Thus, choosing U depends on the risk averse of investors. If $U = 1.65\sigma_{\epsilon}$,

$$P(Loss) \leq 5\%. \tag{8.46}$$

8.4.4 Empirical Pairs Trading Simulations using S&P 500 Stock Index Data

Prices data of future contracts Mar98, Jun98, Sep98 and Dec98 as well as their corresponding spot index prices are used to make pairs trading simulation. Using future contracts prices and corresponding spot index prices during training periods, regression models are formed between log of future contracts prices (f_t), log of spot index prices (s_t) and a time trend $t^* = (T - t)$ defined as time to maturity, i.e.:

$$f_t = \mu + \beta\, s_t + \delta\, t^* + \epsilon_t. \tag{8.47}$$

Regression residuals from (8.47) will be:

$$\hat{\epsilon}_t = f_t - \hat{\mu} - \hat{\beta}\, s_t - \hat{\delta}\, t^*. \tag{8.48}$$

If $\hat{\epsilon}_t$ in (8.48) is stationary using the ADF unit root test, we decide the upper-bound U as well as the lower-bound $-U$ for the pairs trading. Using these bounds, pairs trading strategy described in Subsection 8.4.3 is performed for training data to see whether the pairs trading strategy is profitable or not.

Assuming the data in trading period will have the same pattern as the data in training period, $\hat{\epsilon}_t$ is calculated for trading period using the model obtained from training period, i.e. (8.48). Using the pre-set upper-bound U and the lower-bound $L = -U$ from training period, pairs trading strategy described in Subsection 8.4.3 is also performed. Total return, average return per trade and number of trades are recorded.

Table 8.26 reports pairs trading simulation results during training and trading periods. From Table 8.14, $\hat{\epsilon}_t$ in training period for Mar98, Jun98 and Sep98 can be considered as white noise processes because the autocorrelation with 20 lags can be rejected at a 1% significant level. Thus, the pre-set upper-bound U is set as $1.65\sigma_{\hat{\epsilon}}$ and the pre-set lower-bound is set as $L = -U$. The autocorrelation of $\hat{\epsilon}_t$ in training period for Dec98 cannot be rejected even at a 10% significant level. However, from the plot of $\hat{\epsilon}_t$ in Figure 8.7(b), the series still has good turn around the mean, so the pre-set upper-bound U is still set as $1.65\sigma_{\hat{\epsilon}}$ and the pre-set lower-bound is set as $L = -U$. Total return is defined as follow:

$$\text{Total return} = \sum_{i=1}^{TN} \text{return}_i \tag{8.49}$$

where TN is the total number of pair trades during the period. Return for each pair trade is calculated based on (8.38) with the percentage margin $m = 3\%$. We also report the number of positive return and the average positive return per trade as well as the number of negative return and the average negative return per trade to see the comparison between profits and losses. The pairs trading simulation results in Table 8.26 show that the pairs trading strategy works very well and produce very significant high return even for Dec98 where the series cannot be considered as a white noise process. The average return per trade for all periods are above 10% . The average positive return per trade for all periods are only slightly higher than the average return per trade. This indicates that the losses do not have significant impact to the total return. As we can see from the table, we only have a few losses. Example log of future and spot prices SP500 Mar98 pairs trading simulation in Excell is shown in Figure 8.27. For example, on 31/3/1997, $\hat{\epsilon}_t = 0.0068445$ which is higher than the pre-set upper-bound $U = 0.0038$. Therefore, we buy one future contract and one CFD S&P 500 index (as $\hat{\beta}$ is close to 1, we use $\beta = 1$. It means the number of future contract and CFD in the pair trading are the same). On the next day, $\hat{\epsilon}_t$ become close to zero and by taking the opposite position, the return of 0.7% is made.

Table 8.26: Pairs trading summary

	Training period	Trading period
Mar98	25/3/1997 - 19/3/1998	20/3/1998 - 18/6/1998
Standard Dev $\hat{\epsilon}_t = \sigma_{\hat{\epsilon}}$	0.0023	
$U_o = 1.65\sigma_{\hat{\epsilon}}$	0.0038	0.0038
Total no. trades	19	6
Total return	316.4%	78.01%
Average return/trade	16.7 %	13.0%
No. positive return	18	6
Average positive return/trade	17.497%	13.0%
No. negative return	1	0
Average negative return/trade	-14.3 %	0
Jun98	18/6/1997 - 18/6/1998	19/6/1998 - 18/9/1998
Standard Dev $\hat{\epsilon}_t = \sigma_{\hat{\epsilon}}$	0.002	
$U_o = 1.65\sigma_{\hat{\epsilon}}$	0.0033	0.0033
Total no. trades	22	9
Total return	336.5%	115.99%
Average return/trade	15.3 %	12.887%
No. positive return	21	7
Average positive return/trade	15.95 %	15.4%
No. negative return	1	2
Average negative return/trade	-1.511 %	-4.08%
Sep98	10/6/1997 - 18/9/1998	21/9/1998 - 17/12/1998
Standard Dev $\hat{\epsilon}_t = \sigma_{\hat{\epsilon}}$	0.0022	
$U_o = 1.65\sigma_{\hat{\epsilon}}$	0.0036	0.0036
Total no. trades	19	8
Total return	323.68%	122.14%
Average return/trade	17.03%	15.27 %
No. positive return	17	8
Average positive return/trade	18.36 %	15.27%
No. negative return	2	0
Average negative return/trade	-5.73 %	0
Dec98	18/6/1997 - 17/12/1998	18/12/1998 - 18/3/1999
Standard Dev $\hat{\epsilon}_t = \sigma_{\hat{\epsilon}}$	0.0025	
$U_o = 1.65\sigma_{\hat{\epsilon}}$	0.004	0.004
Total no. trades	27	8
Total return	346.98 %	101.99%
Average return/trade	11.566%	12.75%
No. positive return	23	8
Average positive return/trade	13.81%	12.75 %
No. negative return	4	0
Average negative return/trade	-7.35 %	0

Date	F	S	LF	LS	T-t	Rols	u=0.0038		beta=1	
25/3/1997	821.1	789.07	6.7106	6.6709	258	-0.001268	b=buy		return	NT
26/3/1997	824.5	790.5	6.7148	6.6727	257	0.0012046	s=sell			
27/3/1997	801.3	773.88	6.6862	6.6514	256	-0.005614	b	s		
28/3/1997	801.3	773.88	6.6862	6.6514	255	-0.005437	s	b	0	1
31/3/1997	793.25	757.12	6.6761	6.6295	254	0.0068445	s	b		
4/01/1997	790.3	759.64	6.6724	6.6328	253	-7.47E-05	b	s	0.00705	1
4/02/1997	776.3	750.11	6.6545	6.6202	252	-0.00497	b	s		
4/03/1997	779	750.32	6.658	6.6205	251	-0.001606	s	b	0.00319	1
4/04/1997	788	757.9	6.6695	6.6306	250	-0.000135				
4/07/1997	792.95	762.13	6.6758	6.6361	249	0.0006589				
4/08/1997	797.55	766.12	6.6815	6.6413	248	0.0013244				
4/09/1997	790.9	760.6	6.6732	6.6341	247	0.0004604				
4/10/1997	788.6	758.34	6.6703	6.6311	246	0.0007418				
4/11/1997	762.85	737.65	6.6371	6.6035	245	-0.004229	b	s		
14/4/1997	772.15	743.73	6.6492	6.6117	244	-0.000259	s	b	0.00391	1
15/4/1997	784.75	754.72	6.6654	6.6263	243	0.0012285				
16/4/1997	793.6	763.53	6.6766	6.638	242	0.0008503				
17/4/1997	791.6	761.77	6.6741	6.6356	241	0.0008433				
18/4/1997	795.75	766.34	6.6793	6.6416	240	0.000183				
21/4/1997	789.5	760.37	6.6714	6.6338	239	0.0004046				
22/4/1997	806.25	774.61	6.6924	6.6524	238	0.0027596				
23/4/1997	806.1	773.64	6.6922	6.6511	237	0.0040203	s	b		
24/4/1997	802.1	771.18	6.6872	6.6479	236	0.0024515	b	s	0.00179	1
25/4/1997	794.65	765.37	6.6779	6.6404	235	0.0009649				
28/4/1997	800.95	772.96	6.6858	6.6502	234	-0.000969				
29/4/1997	826.05	794.05	6.7167	6.6771	233	0.0027667				
30/4/1997	829	801.34	6.7202	6.6863	232	-0.00276				
5/01/1997	828.05	798.53	6.7191	6.6828	231	-0.000168				
5/02/1997	843.05	812.97	6.737	6.7007	230	-0.000213				
5/05/1997	864.2	830.24	6.7618	6.7217	229	0.0034256				
5/06/1997	862.8	827.76	6.7602	6.7187	228	0.005014	s	b		
5/07/1997	845.1	815.62	6.7395	6.7039	227	-0.000555	b	s	0.00595	1
5/08/1997	850.15	820.26	6.7454	6.7096	226	-0.000174				
5/09/1997	857.05	824.78	6.7535	6.7151	225	0.0025131				
5/12/1997	868.8	837.66	6.7671	6.7306	224	0.0005926				
13/5/1997	865.15	833.13	6.7629	6.7252	223	0.0020575				
14/5/1997	867.3	836.04	6.7654	6.7287	222	0.0011799				
15/5/1997	872.05	841.88	6.7708	6.7356	221	-0.000241				
16/5/1997	859.1	829.75	6.7559	6.7211	220	-0.000309				
19/5/1997	863.2	833.27	6.7606	6.7254	219	0.0003352				
20/5/1997	871.55	841.66	6.7703	6.7354	218	-2.10E-05				
21/5/1997	870.05	839.35	6.7686	6.7326	217	0.0012196				
22/5/1997	867.25	835.66	6.7653	6.7282	216	0.0026402				
23/5/1997	874.9	847.03	6.7741	6.7417	215	-0.002105				
26/5/1997	874.9	847.03	6.7741	6.7417	214	-0.001929				
27/5/1997	878.65	849.71	6.7784	6.7449	213	-0.00068				
28/5/1997	877.9	847.21	6.7775	6.7419	212	0.0016306				
29/5/1997	872.6	844.08	6.7715	6.7382	211	-0.000495				
30/5/1997	877.95	848.28	6.7776	6.7432	210	0.00076				
6/02/1997	873.5	846.36	6.7725	6.7409	209	-0.001848				
6/03/1997	870.45	845.48	6.769	6.7399	208	-0.004114	b	s		
6/04/1997	869.4	840.11	6.7678	6.7335	207	0.001316	s	b	0.00516	1
6/05/1997	873.25	843.43	6.7722	6.7375	206	0.0019112				
6/06/1997	889.95	858.01	6.7912	6.7546	205	0.0036511				
6/09/1997	893.6	862.91	6.7953	6.7603	204	0.0021456				

Table 8.27: Example log of future and spot prices SP500 Mar98 pairs trading simulation.

8.5 Conclusion

In this chapter, we have done some work related to relationship between future contract prices and spot index prices of the S&P 500. Firstly, the basis of the log of future contract prices and the log of spot index prices of the S&P 500 using currently available data is analysed based on Monoyios and Sarno (2002). Even though we can conclude that there is possibility of nonlinearity in the basis, there is no significant difference between a nonlinear LSTAR model and a linear autoregressive model in fitting the data. It is a different conclusion compared to Monoyios and Sarno (2002) concluding that a nonlinear ESTAR model quite fits with their data. A concern in the way they constructed the basis arise as by pairing up the spot price with the future contract with the nearest maturity, it produces artificial jumps at the time of maturity. Therefore, the second topic of this chapter is to examine the cointegration of f_t and s_t with a time trend for each future contract.

Cointegration analysis with a time trend based on the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the Johansen approach Johansen (1988) are used. Only 19 out of 44 future contracts conclude that they are cointegrated with a time trend using both the Engle-Granger approach and the Johansen approach. Perhaps, high volatility during financial crisis in the data period affects the cointegration test. We see that the results from Mar98 - Dec99 when there was no financial crisis, we strongly conclude that f_t and s_t are cointegrated with a time trend. Furthermore, it happens may be due to limited number of observations, i.e. we only have maximum 2 years data for a future contract.

The third topic of this chapter is about using cointegration analysis with a time trend based on the Engle-Granger to determine whether f_t and s_t are cointegrated and then use the cointegration relationship to perform pairs trading strategy. The pairs trading simulation results show that the pairs trading strategy works very well and it is able to produce very significant high return during training periods and trading periods. The pairs trading strategy is only applied for future contracts in 1998 where they show strong cointegration relationship and the regression residuals (or we can also say the the cointegration errors) follow a white noise process. As we can see from Table 8.14, for other future contracts, especially during economics crisis, the regression residuals follow more complex models involving high order autoregressive models. Thus, some may argue that this pairs trading strategy cannot be used in general practice. Even though that argument can be true, in this chapter we have given a new idea to analyse cointegration between future and spot index prices and apply the cointegration to perform pairs trading between future and spot index prices.

Chapter 9

Conclusion and Further Research

9.1 Conclusion

This thesis extends the pairs trading strategy in Lin *et al.* (2006) based on the cointegration coefficients weighted (CCW) rule. The CCW rule works by trading the number of stocks based on their cointegration coefficients to achieve a guaranteed minimum profit per trade. The minimum profit per trade corresponds to the pre-set boundaries upper-bound U and lower-bound L chosen to open trades. However, they did not give a methodolgy to choose the upper-bound U. In this thesis, the pre-set boundary values are chosen if they are the optimal boundary values. Optimality of the pre-set boundary values is determined by maximising the minimum total profit (MTP) over a specified trading horizon. The MTP is a function of the minimum profit per trade and the number of trades during the trading horizon. As the derivation of the pre-set minimum profit per trade is already provided in Lin *et al.* (2006), this thesis provides the estimated number of trades. The number of trades is also influenced by the distance of the pre-set boundaries from the long-run cointegration equilibrium. The higher the pre-set boundaries for opening trades, the higher the minimum profit per trade but the trade numbers will be lower. The opposite applies for lowering the boundary values. The number of trades over a specified trading horizon is estimated jointly by the average trade duration and the average inter-trade interval. For any pre-set boundaries, both of those values are estimated by making an analogy to the mean first-passage times for a stationary process.

In Chapter 3, the problem to determine trade duration and inter-trade interval for cointegration error following white noise or AR(1) processes in pairs trading strategy is solved using Markov chain approach. The results from simulations are actually encouraging. However, it is difficult to extend the results for general AR(p) processes, $p > 1$. Even for a simple AR(1) process, it involves a double integral to calculate $p_{i,j}$. Therefore, in Chapter 4, another approach namely integral equation approach is used.

In Chapter 4, trade duration and inter-trade interval for cointegration error following AR(1) and AR(2) processes in pairs trading strategy are evaluated by using integral equation approach. The results from simulations are encouraging. The methodology can be extended for AR(p) processes, $p > 2$. However, there is a challenge in numerical scheme for AR(p) processes, $p > 2$, in calculating trade duration and inter-trade interval as it involves more than 2 integrals. For example, to calculate trade duration and inter-trade interval for an AR(2) process, it involves 2 integrals (see (4.44) and (4.45)) and by using $n = 12$, it needs to calculate an inverse matrix of $(m \times m)$ where $m = 6n^2 + 4n + 1 = 913$. Thus, for $p > 2$ the numerical scheme will become more complex.

In Chapter 5, results of pair trading strategy using three cointegrated pair of finance stocks, e.i. WBC-BOQ, WBC-FKP and CBA-ASX, are given. Selection of trading pairs, determination of training and trading periods, and measurement of profits and returns are discussed. For the three pairs selected, trading strategy and resulting profits and trades conducted are given. From the summary trading simulation results, we can conclude that generally the results from the 6-months trading periods are better than the 3-months trading periods in terms of the average profit per trade and the average return per trade. It happens because when we use 3-months trading periods, the frequencies of making losses due to incomplete trades are higher than using 6-months trading periods. The significant impact can be seen from the pair WBC-FKP which has significant smaller profit per trade and the average return per trade when we use 3-months trading periods compared to these indicators when we use 6-months trading periods.

Chapters 3–5 explain pairs trading strategies assuming that the cointegration error ϵ_t is a linear stationary AR(p) process. However, ϵ_t may not follow a linear stationary AR(p) process but a nonlinear stationary process. For an example, Monoyios and Sarno (2002) concluded that there is a cointegration relationship between the log of future contract prices and the log of spot index S&P 500 and FTSE 100 and the basis of the indexes follow a nonlinear stationary ESTAR (Exponential Smooth Transition Autoregressive) model. But, the standard linear unit root tests such as augmented Dickey-Fuller (ADF) test sometimes cannot meet requirement for nonlinear processes. Therefore, in Chapter 6, we extend the work of Kapetanios *et al.* (2003) and Venetis *et al.* (2009) by considering a unit root test for a k-ESTAR(p) model with a different approach. By using this approach, the singularity problem in Venetis *et al.* (2009) can be avoided without adding the collinear regressors into the error term. However, for a k-ESTAR(p) model, $p > 1$, a problem with nuisance parameters emerges. To solve the problem, we suggest two methods, i.e. a bootstrap method and critical values approximation method assuming there is no autocorrelation between Δy_t. From Monte Carlo simulations for k-ESTAR(3) models, the bootstrap method is time consuming and if the underlying series is actually a linear unit root AR(3) model (under the null hypothesis), it may result in a singularity problem. Therefore, we favour to critical values approximation method than the bootstrap method. For some cases, where the parameters are close to a unit root, simulation results show that our approach are better than the results from Venetis *et al.* (2009) , Kapetanios *et al.* (2003) and the Augmented Dickey-Fuller (ADF) tests Dickey and Fuller (1979, 1981) in term of identifying the nonlinearity.

In Chapter 7, we have shown that the first-passage time using integral equation approach can be used for series following a k-ESTAR(p) model. Using this approach trading duration and inter-trade interval for pairs trading are formulated. As we have not found suitable empirical data, we generate data of 2 stock prices with cointegration errors follows the 1-ESTAR(2) model. Using these series, pairs trading simulation is performed and it shows that pairs trading strategy can work for cointegration errors following an 1-ESTAR(2) model. In the process of data analysis, we also show that our ESTAR unit root test can identify that the $\hat{\epsilon}_t$ is a nonlinear stationary where the other unit root tests, i.e. F_{VPP}, AKSS and ADF, cannot confirm that the series is stationary.

In Chapter 8, we have done some works related to relationship between future contract prices and spot index prices of the S&P 500. Firstly, the basis of the log of future contract prices and the log of spot index prices of the S&P 500 using currently available data is analysed based on Monoyios and Sarno (2002). Even though we can conclude that there is possibility nonlinearity in the basis, there is no significant difference between a nonlinear LSTAR model and a linear autoregressive model in fitting the data. It is a different conclusion compared to Monoyios and Sarno (2002) concluding that a nonlinear ESTAR model quite fits with the data they have. A concern in the way they constructed the basis arise as by pairing up the spot price with the future contract with the nearest maturity, it produces artificial jumps at the time of maturity. Therefore, the second topic of this chapter has examined the cointegration of f_t and s_t with a time trend for each future contract. Cointegration analysis with a time trend based on the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the Johansen approach Johansen (1988) are used. Only 19 out of 44 future contracts conclude that they are cointegrated with a time trend using both the Engle-Granger approach and the Johansen approach. Perhaps, high volatility during financial crisis in the data period affects the cointegration test and also limited number of observations, i.e. we only have maximum 2 years data for a future contract. Based on the Engle-Granger approach and then running nonlinearity tests for the regression residuals, we can say that there is no strong evidence of nonlinearity in the relationship between future and spot prices of S&P 500 index. The third topic of this chapter has used cointegration analysis with a time trend based on the Engle-Granger to determine whether f_t and s_t are cointegrated and then use the cointegration relationship to perform pairs trading strategy. The pairs trading simulation results show that the pairs trading strategy works very well producing very significant high return during training periods and trading periods. The pairs trading strategy is only applied for future contracts in 1998 where they show strong cointegration relationship and the regression residuals (or we can also say the the cointegration errors) follow a white noise process. As we can see from Table 8.14, for other future contracts, especially during economics crisis, the regression residuals follow more complex models involving high order autoregressive models. Thus, people may argue that this pairs trading strategy cannot be used in general practice. Even though that argument can be true, in this chapter we have given a new idea to analyse cointegration between future and spot index prices and apply the cointegration to perform pairs trading between future and spot index prices.

9.2 Further Research

In Chapters 3 and 4, we found difficulties to calculate the mean first passage-time using Markov chain or integral equation approaches with $p > 2$. Further research can be taken to develop the numerical scheme for the mean first passage-time using Markov chain or integral equation approaches with $p > 2$.

In doing pairs trading simulation for empirical data in Chapter 5, we are aware that in practice, the cointegrated stocks may not follow the assumptions given in this thesis. For example, the cointegration relationship may disappear or has different model in the trading period, or the cointegration error may not be symmetric or may not an AR(1) process. Whether the technique displayed in this paper works or not only relies on two conditions:(i) within the in-sample data, there is a linear combination of stocks to form an AR(1) series, (ii) such relationship does not significantly change in the trading period (out-sample data). In this thesis, a framework that can be applied for a cointegrated stock pair with AR(1) cointegration error is given and it is shown that the technique works quite well for the stocks which meet the conditions. Further investigations can be carried out to explore different assumptions and the practicability of the strategy in the reality. Pairs trading involving more than two assets can also be considered.

In Chapter 6, we have developed a unit root test for a nonlinear stationary k-ESTAR(p) model. Further research can be carried out to develop a unit root test for other nonlinear models, such as a k-LSTAR(p) model.

In Chapter 8, we use cointegration analysis with a time trend. Lack of papers in theoretical ground for for this topic give an opportunity for further research in this area to give more interpretation of the results. In this chapter, we have also presented pairs trading strategy between future contract and CFD stock index based on cointegration analysis with a time trend. Further research can explore development of pairs trading strategy with different assumptions, such as pairs trading between future contract and CFD involving trading costs.

Appendix A

Program Files

A.1 Chapter 3 Programs

Matlab program to calculate theoretical trade duration using the Markov chain approach in Table 3.1.

```
clear; clc;

u = 1.5; %upper-bound
phi=-0.5;
se = 0.866; %sigma-epsilon
inft = se*4; %value for infinity

c1 = 1/(2*pi*se^2*sqrt(1-phi^2));
c2 = 1/(2*se^2*(1-phi^2));

n1 =0;
for e1 = u:0.01:inft
    for e2 = u:0.01:inft
        n1 = n1 +  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end

n1
d1 = 1- normcdf(u,0,se)
p11 = n1/d1
n2 =0;

for e1 = 0:0.01:u
    for e2 = 0:0.01:u
        n2 = n2 +  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end

n2
d2 = normcdf(u,0,se) - 0.5
```

```
p22 = n2/d2
n3 =0;

for e1 = 0:0.01:u
    for e2 = u:0.01:inft
        n3 = n3+  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end

n3
p12 = n3/d1
p21 = n3/d2

TD = (1-p22+p12)/((1-p11)*(1-p22)-p12*p21)
```

Matlab program to calculate theoretical inter-trade interval using the Markov chain approach in Table 3.1.

```
clear; clc;

u = 1.5; %upper-bound
phi= -0.5;
se = 0.866; %sigma-epsilon
inft = se*4; %value for infinity

c1 = 1/(2*pi*se^2*sqrt(1-phi^2));
c2 = 1/(2*se^2*(1-phi^2));

n1 =0;

for e1 = -inft:0.01:0
    for e2 = -inft:0.01:0
        n1 = n1 +  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end

n1
d1 = 0.5;
p11 = n1/d1
n2 =0;

for e1 = 0:0.01:u
    for e2 = 0:0.01:u
        n2 = n2 +  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end
```

```
n2
d2 = normcdf(u,0,se) - 0.5
p22 = n2/d2
n3 =0;

for e1 = 0:0.01:u
    for e2 = -inft:0.01:0
        n3 = n3+  c1*0.01^2*exp(-c2*(e1^2-2*phi*e1*e2+e2^2));
    end
end

n3

p12 = n3/d1

p21 = n3/d2

IT = (1-p22+p12)/((1-p11)*(1-p22)-p12*p21)
```

Matlab program to calculate simulation trade duration, inter-trade interval and number of trades in Tables 3.1 and 4.2.

```
clear; clc;

% Case : the error cointegration follows an AR(1) process
% assumption : only one trading is allowed

for f = 1:50
ntrade(f) = 0; open = 0; TD(f) = 0; IT(f) =0;

nopen = 0; nclose = 0; phi = -0.5;
%sep = 1;
%s = sep*sqrt((1-phi^2));
s = 0.75;
m = 1000; % number of obeservations
e(1) = randn; for i = 2:m;
    a(i)= s*randn;
    e(i) = phi*e(i-1) + a(i);

    % open the trading when error cointegration exceed u= 1.5
    if (e(i) > 1.5) & (open == 0)
        open = 1;
        nopen = nopen + 1;
        topen(nopen) = i;
        if (nopen > 1)
            IT(f) = IT(f) + topen(nopen) - tclose(nclose);
        end
```

```
    end

     % close the trading when error cointegration below zero
     if (e(i) < 0) & (open == 1)
         open = 0;
         nclose = nclose + 1;
         tclose(nclose) = i;
         TD(f) = TD(f) + tclose(nclose) - topen(nopen) ;
        ntrade(f) = ntrade(f) + 1;
     end
end

mTD(f) = TD(f)/ntrade(f); mIT(f) = IT(f)/(nopen-1);

end

NT = mean(ntrade)
sNT = std(ntrade)
eTD = mean(mTD)
sTD = std(mTD)
eIT = mean(mIT)
sIT = std(mIT)
```

A.2 Chapter 4 Programs

A.2.1 AR(1) Model

Matlab program used in Table 4.1 for simulation the mean first-passage time over a=0, given $Y_0 = 1.5$.

```
function S = sinteg1(p,x0,phi,a,s2)

y(1) = phi*x0 + randn*sqrt(s2);
if (y(1) <= a);
        S = 1;
        return
end
x = x0;
for i = 2:100;
    e(i) = randn*sqrt(s2);
    x(1:p-1,1)= x(2:p,1);
    x(p,1) = y(i-1);
    y(i) = phi*x + e(i);
    if (y(i) <= a);
        S = i;
        break
    end
```

```
end


clear;
clc;

p = 1;
x0 = 1.5;
phi = -0.5;
a = 0;
s2 = 4;

for i = 1:10000
   C(i) = sinteg1(p,x0,phi,a,s2);
end
meanC = mean(C)
sC = std(C)
```

Matlab program used in Tables 4.1 and 4.2. to calculate the mean first-passage time over a=0, given $y_0 = 1.5$ using integral equation approach.

```
clear; clc;

s = 1; theta1 = 0.5;

a = 0;

var = s/(1-theta1^2); lb = ceil(6*sqrt(var)); n = ceil((lb+a)/0.05);
h = (lb+a)/n;


F = zeros(n+1,n+1); for i= 0:n
    x = i*h ;
    for j = 0:n
        y =  j*h;
         if (j==0) | (j==n)
              w = 1;
         else w = 2;
         end
         F(i+1,j+1)=h*w*0.5*exp(-(y- theta1*x)^2/(2*s^2))/(sqrt(2*pi)*s);
     end
end

F = eye(n+1,n+1) - F; F = inv(F); P = ones(n+1,1); ft = F*P;

x0 = 1.5;
 sum = 0;
```

```
 for i = 0:n
    y = i*h;
    if (i==0) | (i==n);
          w = 1;
     else w = 2;
     end
 sum = sum + h*w*0.5*exp(-(y- theta1*x0)^2/(2*s^2))
       *ft(i+1,1)/(sqrt(2*pi*s));
end
ET = sum +1 %the mean-passage time over a=0, given x0 =1.5
```

Matlab program to calculate optimal pre-set upper-bound U_o using integral equation approach in Table 4.3.

```
clear; clc;

a = 0; phi = 0.5; T=1000;

s=1; %sigma_a^2

var = s/(1-phi^2); b = ceil(5*sqrt(var));

n = ceil((b-a)/0.05);

F = zeros(n+1,n+1); h = (b-a)/n;

  for i= 0:n
    x = a+i*h;
    for j = 0:n
        y = a+j*h;
         if (j==0) | (j==n)
              w = 1;
         else w = 2;
         end
         F(i+1,j+1)=h*w*0.5*exp(-(y- phi*x)^2/(2*s))/(sqrt(2*pi*s));
     end
end

F = eye(n+1,n+1) - F; F = inv(F); P = ones(n+1,1); ft = F*P;

profit = 0; for k = 0.1:0.01:0.4
 h = (b-a)/n;
 sum = 0;
 for i = 0:n
    y = a+i*h;
    if (i==0) | (i==n);
          w = 1;
```

```
     else w = 2;
     end
    sum = sum + h*w*0.5*exp(-(y-phi*k)^2/(2*s))*ft(i+1,1)/(sqrt(2*pi*s));
end

TD = sum +1;

  %Calculate the waiting duration or inter-trade interval

    m = ceil((k-(-b))/0.05);
    F = zeros(m+1,m+1);
    h = (k-(-b))/m;
    for i= 0:m
        x = (-b)+i*h;
        for j = 0:m
            y = (-b)+j*h;
            if (j==0) | (j==m)
                w = 1;
            else w = 2;
            end
          Fw(i+1,j+1)=h*w*0.5*exp(-(y- phi*x)^2/(2*s))/(sqrt(2*pi*s));
          end
     end

 Fw = eye(m+1,m+1) - Fw;
 Fw = inv(Fw);
 Pw = ones(m+1,1);
 f = Fw*Pw;

 sum = 0;
 for i = 0:m
    y = (-b)+i*h;
    if (i==0) | (i==m);
          w = 1;
     else w = 2;
     end
    sum = sum + h*w*0.5*exp(-(y)^2/(2*s))*f(i+1,1)/(sqrt(2*pi*s));
end WD = sum +1;

D = TD+WD;

pr = ((T/D)-1)*k;

if (pr > profit)
    profit = pr;
    u = k;
end
```

```
end

profit %the maximum of MTP
u %optimal upper-bound
```

A.2.2 AR(2) Model

Matlab program to calculate the mean first-passage time an AR(2) process crossing 0 given initial values (y_{-1}, y_0) using integral equation approach in Tables 4.4 and 4.5.

Function finteg11 will be used in the main program:

```
function K=finteg11(a,b,s,phi,p);

x = zeros(2,(p+1)); x(:,1) = a; x(:,2) = [x(2,1);b(1,1)]; x(:,3) =
b;

ep = zeros(1,p); ep(1,p)=1;

sum = 0; for i = 1:p
    sum = sum + (ep*x(:,(i+1)) - phi*x(:,i))^2;
end

sum = -sum/(2*s); K = exp(sum)/(2*pi*s);
```

Function finteg12 will be used in the main program:

```
function psi = finteg12(x0,a,b,s,phi,n)

h = (b-a)/n; sum =0; for i= 0:n
    x = a+i*h;
    if (i==0) | (i==n)
        w = 1;
    else w = 2;
    end
    sum = sum + h*w*0.5*exp(-(x- phi*x0)^2/(2*s));
end
sum = sum/(sqrt(2*pi*s)); psi = 1 +sum;
```

Main program:

```
clear; clc;

p = 2; a = 0;
s = 1; %std residuals of an AR(2) process
phi=[-0.3 0.5]; r1 = phi(1,2)/(1-phi(1,1)); r2 = phi(1,1)+
(phi(1,2)^2/(1-phi(1,1))); var = s/(1-r1*phi(1,2)-r2*phi(1,1)); b =
```

```
ceil(5*sqrt(var)) n = ceil((b-a)/0.5)

h = (b-a)/n; L = 0.5*(h^2);

N = zeros((7*2*(n^2)),3); M = zeros((6*(n^2)+4*n+1),3); NT
=zeros(7,3); T = zeros(2,3);

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+(i*h) a+(i*h) a+(i+1)*h; a+j*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j)):(7*(n*i+j+1)),:) = NT;
    end
end

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+i*h a+(i+1)*h a+(i+1)*h; a+(j+1)*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
```

```
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j+n^2)):(7*(n*i+j+n^2+1)),:) = NT;
    end
end N;

v = zeros(1,2); for i= 1:(7*2*(n^2))
    v = N(i,1:2);
    for j = (i+1):(7*2*(n^2))
        if (N(j,1:2)== v)
            N(i,3) = N(i,3)+N(j,3);
            N(j:(7*2*(n^2)-1),:)=N((j+1):(7*2*(n^2)),:);
            N((7*2*(n^2)),:)= zeros(1,3);
        end
    end
end N;

m = 6*(n^2)+4*n+1; M = N(1:m,:);

F = zeros(m,m);
  for i= 1:m
    A(1,1) = M(i,1);
    A(2,1) = M(i,2);
    for j = 1:m
    B(1,1) = M(j,1);
    B(2,1) = M(j,2);
    k = M(j,3)*finteg11(A,B,s,phi,p);
    F(i,j)= k;
    end
end F; F = eye(m,m)-F;
```

```
P=zeros(m,1); for i = 1:m
    x=[M(i,1);M(i,2)];
    P(i,1) = finteg12(x,a,b,s,phi,n);
end P;

f = F\P;

     x0 = [2;1.5] %initial values
     sum = 0;
      for i = 1:m
       B(1,1)=M(i,1);
       B(2,1)=M(i,2);
       sum = sum + M(i,3)*finteg11(x0,B,s,phi,p)*f(i,1);
      end

sum = sum + finteg12(x0,a,b,s,phi,n)
```

Matlab program used in Table 4.4 for simulation the mean first-passage time over a=0, given (y_{-1}, y_0).

```
function S = sinteg1(p,x0,phi,a,s2)

y(1) = phi*x0 + randn*sqrt(s2);
if (y(1) <= a);
        S = 1;
        return
end
x = x0;
for i = 2:100;
    e(i) = randn*sqrt(s2);
    x(1:p-1,1)= x(2:p,1);
    x(p,1) = y(i-1);
    y(i) = phi*x + e(i);
    if (y(i) <= a);
        S = i;
        break
    end
end


clear;
clc;

p = 2;
phi = [0.3 0.5];
s = 1;
```

```
a = 0;

x0 = [0.5;1.5];
for i = 1:10000
   C(i) = sinteg1(p,x0,phi,a,s);
end
mC = mean(C)
sC = std(C)

x0 = [2;1.5];
for i = 1:10000
   C(i) = sinteg1(p,x0,phi,a,s);
end
mC = mean(C)
sC = std(C)
```

Matlab program for simulation in Table 4.5.

```
clear; clc;

% Case : the error cointegration follows an AR(2) process
% assumption : only one trading is allowed

for f = 1:50
ntrade(f) = 0; open = 0; TD(f) = 0; IT(f) =0;

nopen = 0; nclose = 0; phi = [0.3 0.5];
%sep = 1;
%s = sep*sqrt((1-phi^2));
s = 1;
m = 1000; % number of obeservations
e(1) = randn;
e(2) = randn;
for i = 3:m;
    a(i)= s*randn;
    e(i) = phi*[e(i-2);e(i-1)] + a(i);

    % open the trading when error cointegration exceed u= 1.5
    if (e(i) > 1.5) & (open == 0)
        open = 1;
        nopen = nopen + 1;
        topen(nopen) = i;
        if (nopen > 1)
            IT(f) = IT(f) + topen(nopen) - tclose(nclose);
        end
    end
```

```
     % close the trading when error cointegration below zero
     if (e(i) < 0) & (open == 1)
         open = 0;
         nclose = nclose + 1;
         tclose(nclose) = i;
         TD(f) = TD(f) + tclose(nclose) - topen(nopen) ;
        ntrade(f) = ntrade(f) + 1;
     end
end

mTD(f) = TD(f)/ntrade(f); mIT(f) = IT(f)/(nopen-1);

end

NT = mean(ntrade)
sNT = std(ntrade)
eTD = mean(mTD)
sTD = std(mTD)
eIT = mean(mIT)
sIT = std(mIT)
```

A.3 Chapter 6 Programs

A.3.1 k-ESTAR(2) Model

R program for k-ESTAR(2) critical values:

```
T <- 10000
end<- T + 50
rep <- 50000

f1 <- c()
f1a <- c()
f2 <- c()
f2a <- c()
f3 <- c()
f3a <- c()

f4 <- c()
f4a <- c()

x<-c()
x[1]<-0
e<-c()
e[1]<-0
```

```
F11 <- c()
F11a <- c()
F21 <- matrix(0,1,1)
F21a <- matrix(0,1,1)

F12 <- c()
F12a <- c()
F22 <- matrix(0,3,3)
F22a <- matrix(0,3,3)

F13 <- c()
F13a <- c()
F23 <- matrix(0,5,5)
F23a <- matrix(0,5,5)

F14 <- c()
F14a <- c()
F24 <- matrix(0,7,7)
F24a <- matrix(0,7,7)

for(z in 1:rep){

for (i in 2:end){
e[i]<- rnorm(1,0,1)
x[i]<- x[i-1] + e[i]

}

y <- 0
ib2 <- 0
ib3 <- 0
ib4 <- 0
ib5 <- 0
ib6 <- 0
ib7 <- 0

ib8 <- 0
ib9 <- 0
ib10 <- 0
ib11 <- 0
ib12 <- 0
ib13 <- 0

ib14 <- 0
ib15 <- 0
ib16 <- 0
ib17 <- 0
ib18 <- 0
```

```
ib3db <- 0

ib4db <- 0
ib5db <- 0
ib6db <- 0
ib7db <- 0
ib8db <- 0
ib9db <- 0
ib2db1 <- 0
ib3db1 <- 0
ib4db1 <- 0
ib5db1 <- 0
ib6db1 <- 0

ib7db1 <- 0
ib8db1 <- 0

b1 <- 0

for (i in 1:T){
y <- x[i+49]
ib2 <- ib2 + T^{-2}*y^2
ib3 <- ib3 +T^{-2.5}*y^3
ib4 <- ib4 + T^{-3}*y^4
ib5 <- ib5 + T^{-3.5}*y^5

ib6 <- ib6 + T^{-4}*y^6
ib7 <- ib7 + T^{-4.5}*y^7

ib8 <- ib8 + T^{-5}*y^8
ib9 <- ib9 + T^{-5.5}*y^9

ib10 <- ib10 + T^{-6}*y^10
ib11 <- ib11 + T^{-6.5}*y^11

ib12 <- ib12 + T^{-7}*y^12
ib13 <- ib13 + T^{-7.5}*y^13

ib14 <- ib14 + T^{-8}*y^14
ib15 <- ib15 + T^{-8.5}*y^15

ib16 <- ib16 + T^{-9}*y^16
ib17 <- ib17 + T^{-9.5}*y^17

ib18 <- ib18 + T^{-10}*y^18
ib3db <- ib3db + T^{-2}*y^3*e[i+50]
ib4db <- ib4db + T^{-2.5}*y^4*e[i+50]
```

```
ib5db <- ib5db + T^{-3}*y^5*e[i+50]

ib6db <- ib6db + T^{-3.5}*y^6*e[i+50]

ib7db <- ib7db + T^{-4}*y^7*e[i+50]

ib8db <- ib8db + T^{-4.5}*y^8*e[i+50]

ib9db <- ib9db + T^{-5}*y^9*e[i+50]

ib2db1 <- ib2db1 + T^{-1.5}*y^2*e[i+49]*e[i+50]

ib3db1 <- ib3db1 + T^{-2}*y^3*e[i+49]*e[i+50]

ib4db1 <- ib4db1 + T^{-2.5}*y^4*e[i+49]*e[i+50]

ib5db1 <- ib5db1 + T^{-3}*y^5*e[i+49]*e[i+50]

ib6db1 <- ib6db1 + T^{-3.5}*y^6*e[i+49]*e[i+50]

ib7db1 <- ib7db1 + T^{-4}*y^7*e[i+49]*e[i+50]

ib8db1 <- ib8db1 + T^{-4.5}*y^8*e[i+49]*e[i+50]

b1 <- b1 + T^{-1/2}*e[i+49]*e[i+50] }


F11[1] <- ib3db
F11a[1] <- ib2db1 - b1*ib2

F12[1] <- ib3db
F12[2] <- ib4db
F12[3] <- ib5db

F12a[1] <- ib2db1 - b1*ib2
F12a[2] <- ib3db1 - b1*ib3

F12a[3] <- ib4db1 - b1*ib4

F13[1] <- ib3db
F13[2] <- ib4db
F13[3] <- ib5db
F13[4] <- ib6db
F13[5] <- ib7db

F13a[1] <- ib2db1 - b1*ib2
F13a[2] <- ib3db1 - b1*ib3
```

```
F13a[3] <- ib4db1 - b1*ib4
F13a[4] <- ib5db1 - b1*ib5

F13a[5] <- ib6db1 - b1*ib6

F14[1] <- ib3db
F14[2] <- ib4db
F14[3] <- ib5db
F14[4] <- ib6db
F14[5] <- ib7db
F14[6] <- ib8db
F14[7] <- ib9db

F14a[1] <- ib2db1 - b1*ib2
F14a[2] <- ib3db1 - b1*ib3

F14a[3] <- ib4db1 - b1*ib4
F14a[4] <- ib5db1 - b1*ib5

F14a[5] <- ib6db1 - b1*ib6
F14a[6] <- ib7db1 - b1*ib7

F14a[7] <- ib8db1 - b1*ib8

F21[1,1] <- ib6

F21a[1,1] <- ib4 - ib2^2

F22[1,1] <- ib6
F22[1,2] <- ib7
F22[1,3] <- ib8
F22[2,1] <- ib7
F22[2,2] <- ib8
F22[2,3] <- ib9
F22[3,1] <- ib8
F22[3,2] <- ib9
F22[3,3] <- ib10

F22a[1,1] <- ib4 - ib2^2
F22a[1,2] <- ib5 - ib2*ib3

F22a[1,3] <- ib6 - ib2*ib4
F22a[2,1] <- ib5 - ib2*ib3

F22a[2,2] <- ib6 - ib3^2
F22a[2,3] <- ib7 - ib3*ib4

F22a[3,1] <- ib6 - ib4*ib2
```

```
F22a[3,2] <- ib7 - ib4*ib3

F22a[3,3] <- ib8 - ib4^2

F23[1,1] <- ib6
F23[1,2] <- ib7
F23[1,3] <- ib8
F23[1,4] <- ib9
F23[1,5] <- ib10
F23[2,1] <- ib7
F23[2,2] <- ib8
F23[2,3] <- ib9
F23[2,4] <- ib10
F23[2,5] <- ib11
F23[3,1] <- ib8
F23[3,2] <- ib9
F23[3,3] <- ib10
F23[3,4] <- ib11
F23[3,5] <- ib12
F23[4,1] <- ib9
F23[4,2] <- ib10
F23[4,3] <- ib11
F23[4,4] <- ib12

F23[4,5] <- ib13
F23[5,1] <- ib10
F23[5,2] <- ib11

F23[5,3] <- ib12
F23[5,4] <- ib13
F23[5,5] <- ib14

F23a[1,1] <- ib4 - ib2^2
F23a[1,2] <- ib5 - ib2*ib3

F23a[1,3] <- ib6 - ib2*ib4
F23a[1,4] <- ib7 - ib2*ib5

F23a[1,5] <- ib8 - ib2*ib6
F23a[2,1] <- ib5 - ib3*ib2

F23a[2,2] <- ib6 - ib3^2
F23a[2,3] <- ib7 - ib3*ib4

F23a[2,4] <- ib8 - ib3*ib5
F23a[2,5] <- ib9 - ib3*ib6

F23a[3,1] <- ib6 - ib4*ib2
```

```
F23a[3,2] <- ib7 - ib4*ib3

F23a[3,3] <- ib8 - ib4^2
F23a[3,4] <- ib9 - ib4*ib5

F23a[3,5] <- ib10 - ib4*ib6
F23a[4,1] <- ib7 - ib5*ib2

F23a[4,2] <- ib8 - ib5*ib3
F23a[4,3] <- ib9 - ib5*ib4

F23a[4,4] <- ib10 - ib5*ib5
F23a[4,5] <- ib11 - ib5*ib6

F23a[5,1] <- ib8 - ib6*ib2
F23a[5,2] <- ib9 - ib6*ib3

F23a[5,3] <- ib10 - ib6*ib4
F23a[5,4] <- ib11 - ib6*ib5

F23a[5,5] <- ib12 - ib6*ib6

F24[1,1] <- ib6
F24[1,2] <- ib7
F24[1,3] <- ib8
F24[1,4] <- ib9
F24[1,5] <- ib10
F24[1,6] <- ib11
F24[1,7] <- ib12
F24[2,1] <- ib7
F24[2,2] <- ib8
F24[2,3] <- ib9
F24[2,4] <- ib10
F24[2,5] <- ib11
F24[2,6] <- ib12
F24[2,7] <- ib13
F24[3,1] <- ib8
F24[3,2] <- ib9
F24[3,3] <- ib10
F24[3,4] <- ib11
F24[3,5] <- ib12

F24[3,6] <- ib13
F24[3,7] <- ib14
F24[4,1] <- ib9
F24[4,2] <- ib10
F24[4,3] <- ib11
F24[4,4] <- ib12
```

```
F24[4,5] <- ib13

F24[4,6] <- ib14
F24[4,7] <- ib15
F24[5,1] <- ib10

F24[5,2] <- ib11
F24[5,3] <- ib12
F24[5,4] <- ib13

F24[5,5] <- ib14
F24[5,6] <- ib15
F24[5,7] <- ib16

F24[6,1] <- ib11
F24[6,2] <- ib12
F24[6,3] <- ib13

F24[6,4] <- ib14
F24[6,5] <- ib15
F24[6,6] <- ib16

F24[6,7] <- ib17
F24[7,1] <- ib12
F24[7,2] <- ib13

F24[7,3] <- ib14
F24[7,4] <- ib15
F24[7,5] <- ib16

F24[7,6] <- ib17
F24[7,7] <- ib18

F24a[1,1] <- ib4 - ib2^2
F24a[1,2] <- ib5 - ib2*ib3

F24a[1,3] <- ib6 - ib2*ib4
F24a[1,4] <- ib7 - ib2*ib5

F24a[1,5] <- ib8 - ib2*ib6
F24a[1,6] <- ib9 - ib2*ib7

F24a[1,7] <- ib10 - ib2*ib8
F24a[2,1] <- ib5 - ib3*ib2

F24a[2,2] <- ib6 - ib3^2
F24a[2,3] <- ib7 - ib3*ib4
```

```
F24a[2,4] <- ib8 - ib3*ib5
F24a[2,5] <- ib9 - ib3*ib6

F24a[2,6] <- ib10 - ib3*ib7
F24a[2,7] <- ib11 - ib3*ib8

F24a[3,1] <- ib6 - ib4*ib2
F24a[3,2] <- ib7 - ib4*ib3

F24a[3,3] <- ib8 - ib4^2
F24a[3,4] <- ib9 - ib4*ib5

F24a[3,5] <- ib10 - ib4*ib6
F24a[3,6] <- ib11 - ib4*ib7

F24a[3,7] <- ib12 - ib4*ib8
F24a[4,1] <- ib7 - ib5*ib2

F24a[4,2] <- ib8 - ib5*ib3
F24a[4,3] <- ib9 - ib5*ib4

F24a[4,4] <- ib10 - ib5*ib5
F24a[4,5] <- ib11 - ib5*ib6

F24a[4,6] <- ib12 - ib5*ib7
F24a[4,7] <- ib13 - ib5*ib8

F24a[5,1] <- ib8 - ib6*ib2
F24a[5,2] <- ib9 - ib6*ib3

F24a[5,3] <- ib10 - ib6*ib4
F24a[5,4] <- ib11 - ib6*ib5

F24a[5,5] <- ib12 - ib6*ib6
F24a[5,6] <- ib13 - ib6*ib7

F24a[5,7] <- ib14 - ib6*ib8
F24a[6,1] <- ib9 - ib7*ib2

F24a[6,2] <- ib10 - ib7*ib3
F24a[6,3] <- ib11 - ib7*ib4

F24a[6,4] <- ib12 - ib7*ib5
F24a[6,5] <- ib13 - ib7*ib6

F24a[6,6] <- ib14 - ib7*ib7
F24a[6,7] <- ib15 - ib7*ib8
```

```
F24a[7,1] <- ib10 - ib8*ib2
F24a[7,2] <- ib11 - ib8*ib3

F24a[7,3] <- ib12 - ib8*ib4
F24a[7,4] <- ib13 - ib8*ib5

F24a[7,5] <- ib14 - ib8*ib6
F24a[7,6] <- ib15 - ib8*ib7
F24a[7,7]<- ib16 - ib8*ib8

f1[z] <- F11%*% solve(F21,F11)
f1a[z] <- f1[z] + F11a%*% solve(F21a,F11a)

f2[z] <- F12%*% solve(F22,F12)
f2a[z] <- f2[z] + F12a%*% solve(F22a,F12a)

f3[z] <- F13%*% solve(F23,F13)
f3a[z] <- f3[z] + F13a%*% solve(F23a,F13a)

f4[z] <- F14%*% solve(F24,F14)
f4a[z] <- f4[z] + F14a%*% solve(F24a,F14a)

}

q1 <- c(quantile(f1,0.9),quantile(f1,0.95),quantile(f1,0.99))
print(q1)

q1a <- c(quantile(f1a,0.9),quantile(f1a,0.95),quantile(f1a,0.99))
print(q1a)

q2 <- c(quantile(f2,0.9),quantile(f2,0.95),quantile(f2,0.99))
print(q2)

q2a <- c(quantile(f2a,0.9),quantile(f2a,0.95),quantile(f2a,0.99))
print(q2a)

q3 <- c(quantile(f3,0.9),quantile(f3,0.95),quantile(f3,0.99))
print(q3)

q3a <- c(quantile(f3a,0.9),quantile(f3a,0.95),quantile(f3a,0.99))
print(q3a)

q4 <- c(quantile(f4,0.9),quantile(f4,0.95),quantile(f4,0.99))
print(q4)

q4a <- c(quantile(f4a,0.9),quantile(f4a,0.95),quantile(f4a,0.99))
print(q4a)
```

R program simulation ESTAR(2):

```
T <- 250
end<- T + 50

rep <- 10000

teta11 <- 1
teta12 <- 0.2
teta21 <- 0
teta22 <- 0
teta20 <- 0

teta <- 0

set.seed(1)

countF1 <- 0
countF2 <- 0
countF3 <- 0
countF4 <- 0
countV1 <- 0
countV2 <- 0
countV3 <- 0
countV4 <- 0
countKSS <- 0
countADF <- 0


y <- c()
y[1] <- 0
y[2] <- 0

dy0 <- c()
dy1 <- c()
y1 <- c()
y2 <- c()
y3 <- c()
y4 <- c()

y5 <- c()
y6 <- c()
y7 <- c()
y8 <- c()
y9 <- c()
y2dy1 <- c()
y3dy1 <- c()
```

```
y4dy1 <- c()
y5dy1 <- c()
y6dy1 <- c()
y7dy1 <- c()
y8dy1 <- c()

for(z in 1:rep){

for (i in 3:end){

y[i]<- y[i-1] - teta12*(y[i-1] - y[i-2]) + rnorm(1,0,1) }

for (i in 1:T){
dy0[i] <- y[i+50] - y[i+49]

dy1[i] <- y[i+49] - y[i+48]
y1[i] <- y[i+49]
y2[i] <- y[i+49]^2
y3[i] <- y[i+49]^3
y4[i] <- y[i+49]^4
y5[i] <- y[i+49]^5

y6[i] <- y[i+49]^6
y7[i] <- y[i+49]^7
y8[i] <- y[i+49]^8

y9[i] <- y[i+49]^9
y2dy1[i] <- y2[i]*dy1[i]

y3dy1[i] <- y3[i]*dy1[i]
y4dy1[i] <- y4[i]*dy1[i]

y5dy1[i] <- y5[i]*dy1[i]
y6dy1[i] <- y6[i]*dy1[i]

y7dy1[i] <- y7[i]*dy1[i]
y8dy1[i] <- y8[i]*dy1[i] }

RF <- lm(dy0 ~ 0 + dy1)
SrF <- deviance(RF)

UF1 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 )
SuF1 <- deviance(UF1)

F1 <- (SrF-SuF1)*T/(SuF1)
if (F1 > 6.9459) countF1 <- countF1 + 1

UF2 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y3dy1 + y4dy1)
```

```
SuF2 <- deviance(UF2)
F2 <- (SrF-SuF2)*T/(SuF2)

if (F2 > 15.982) countF2 <- countF2 + 1

UF3 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y6 + y7 + y3dy1 +
y4dy1 + y5dy1 + y6dy1)

SuF3 <- deviance(UF3)
F3 <- (SrF-SuF3)*T/(SuF3)

if (F3 > 23.18338) countF3 <- countF3 + 1

UF4 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y6 + y7 + y8 + y9
+ y3dy1 + y4dy1 + y5dy1 + y6dy1 + y7dy1 + y8dy1)

SuF4 <- deviance(UF4)
F4 <- (SrF-SuF4)*T/(SuF4)

if (F4 > 29.65) countF4 <- countF4 + 1

UV1 <- lm(dy0 ~ 0 + dy1 + y3)

SuV1 <- deviance(UV1)
V1 <- (SrF-SuV1)*T/(SuV1)

if (V1 > 4.8826) countV1 <- countV1 + 1

UV2 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5)
SuV2 <- deviance(UV2)

V2 <- (SrF-SuV2)*T/(SuV2)

if (V2 > 11.364) countV2 <- countV2 + 1

UV3 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5 + y6 + y7)

SuV3 <- deviance(UV3)
V3 <- (SrF-SuV3)*T/(SuV3)

if (V3 > 15.7059 ) countV3 <- countV3 + 1

UV4 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5 + y6 + y7 + y8 + y9)

SuV4 <- deviance(UV4)
V4 <- (SrF-SuV4)*T/(SuV4)

if (V4 > 19.38346) countV4 <- countV4 + 1
```

```
ks <- lm(dy0 ~ 0 + y3)
Tks <- coef(ks)/sqrt(vcov(ks))

if (Tks < -2.22) countKSS <- countKSS + 1

adf <- lm(dy0 ~ 0 + y1 + dy1)
coefADF <- coef(adf)

covADF <- vcov(adf)
Tadf <- coefADF[1]/sqrt(covADF[1,1])

if (Tadf < - 1.95) countADF <- countADF + 1

}

pcount <- c(countF1*100/rep, countF2*100/rep, countF3*100/rep,
countF4*100/rep, countV1*100/rep, countV2*100/rep,
countV3*100/rep, countV4*100/rep, countKSS*100/rep,
countADF*100/rep)

print(pcount)
```

A.3.2 k-ESTAR(3) Model

R program for k-ESTAR(3) critical values:

```
T <- 10000
end<- T + 50
rep <- 50000

f1 <- c()
f1a <- c()
f2 <- c()
f2a <- c()
f3 <- c()
f3a <- c()

x<-c()
x[1]<-0
e <-c()
e[1]<-0

F11 <- c()
F11a <- c()
F21 <- matrix(0,1,1)
```

```
F21a <- matrix(0,2,2)

F12 <- c()
F12a <- c()
F22 <- matrix(0,3,3)
F22a <- matrix(0,6,6)

F13 <- c()
F13a <- c()
F23 <- matrix(0,5,5)

F23a <- matrix(0,10,10)

for(z in 1:rep){

for (i in 2:end){
e[i]<- rnorm(1,0,1)
x[i]<- x[i-1] + e[i] }

y <- 0
ib2 <- 0
ib3 <- 0
ib4 <- 0
ib5 <- 0
ib6 <- 0
ib7 <- 0

ib8 <- 0
ib9 <- 0
ib10 <- 0
ib11 <- 0
ib12 <- 0
ib13 <- 0
ib14 <- 0

ib3db <- 0
ib4db <- 0
ib5db <- 0
ib6db <- 0
ib7db <- 0

ib2db1 <- 0
ib3db1 <- 0
ib4db1 <- 0
ib5db1 <- 0
ib6db1 <- 0

ib2db2 <- 0
```

```
ib3db2 <- 0
ib4db2 <- 0
ib5db2 <- 0
ib6db2 <- 0

b1 <- 0
b2 <- 0

for (i in 1:T){
y <- x[i+49]
ib2 <- ib2 + T^{-2}*y^2

ib3 <- ib3 + T^{-2.5}*y^3
ib4 <- ib4 + T^{-3}*y^4

ib5 <- ib5 + T^{-3.5}*y^5
ib6 <- ib6 + T^{-4}*y^6

ib7 <- ib7 + T^{-4.5}*y^7
ib8 <- ib8 + T^{-5}*y^8

ib9 <- ib9 + T^{-5.5}*y^9
ib10 <- ib10 + T^{-6}*y^10

ib11 <- ib11 + T^{-6.5}*y^11
ib12 <- ib12 + T^{-7}*y^12

ib13 <- ib13 + T^{-7.5}*y^13
ib14 <- ib14 + T^{-8}*y^14

ib3db <- ib3db + T^{-2}*y^3*e[i+50]

ib4db <- ib4db + T^{-2.5}*y^4*e[i+50]

ib5db <- ib5db + T^{-3}*y^5*e[i+50]

ib6db <- ib6db + T^{-3.5}*y^6*e[i+50]

ib7db <- ib7db + T^{-4}*y^7*e[i+50]

ib2db1 <- ib2db1 + T^{-1.5}*y^2*e[i+49]*e[i+50]

ib3db1 <- ib3db1 + T^{-2}*y^3*e[i+49]*e[i+50]

ib4db1 <- ib4db1 + T^{-2.5}*y^4*e[i+49]*e[i+50]

ib5db1 <- ib5db1 + T^{-3}*y^5*e[i+49]*e[i+50]
```

```
ib6db1 <- ib6db1 + T^{-3.5}*y^6*e[i+49]*e[i+50]

ib2db2 <- ib2db2 + T^{-1.5}*y^2*e[i+48]*e[i+50]

ib3db2 <- ib3db2 + T^{-2}*y^3*e[i+48]*e[i+50]

ib4db2 <- ib4db2 + T^{-2.5}*y^4*e[i+48]*e[i+50]

ib5db2 <- ib5db2 + T^{-3}*y^5*e[i+48]*e[i+50]

ib6db2 <- ib6db2 + T^{-3.5}*y^6*e[i+48]*e[i+50]

b1 <- b1 + T^{-1/2}*e[i+49]*e[i+50]

b2 <- b2 + T^{-1/2}*e[i+48]*e[i+50] }


F11[1] <- ib3db

F11a[1] <- ib2db1 - b1*ib2
F11a[2] <- ib2db2 - b2*ib2

F12[1] <- ib3db
F12[2] <- ib4db
F12[3] <- ib5db

F12a[1] <- ib2db1 - b1*ib2
F12a[2] <- ib2db2 - b2*ib2

F12a[3] <- ib3db1 - b1*ib3
F12a[4] <- ib3db2 - b2*ib3

F12a[5] <- ib4db1 - b1*ib4
F12a[6] <- ib4db2 - b2*ib4

F13[1] <- ib3db
F13[2] <- ib4db
F13[3] <- ib5db
F13[4] <- ib6db
F13[5] <- ib7db

F13a[1] <- ib2db1 - b1*ib2
F13a[2] <- ib2db2 - b2*ib2

F13a[3] <- ib3db1 - b1*ib3
F13a[4] <- ib3db2 - b2*ib3

F13a[5] <- ib4db1 - b1*ib4
```

```
F13a[6] <- ib4db2 - b2*ib4

F13a[7] <- ib5db1 - b1*ib5
F13a[8] <- ib5db2 - b2*ib5

F13a[9] <- ib6db1 - b1*ib6
F13a[10] <- ib6db2 - b2*ib6

F21[1,1] <- ib6
F21a[1,1] <- ib4 - ib2^2
F21a[1,2] <- 0

F21a[2,2] <- ib4 - ib2^2
F21a[2,1] <- 0

F22[1,1] <- ib6
F22[1,2] <- ib7
F22[1,3] <- ib8
F22[2,1] <- ib7
F22[2,2] <- ib8
F22[2,3] <- ib9
F22[3,1] <- ib8
F22[3,2] <- ib9
F22[3,3] <- ib10

F22a[1,1] <- ib4 - ib2^2
F22a[1,2] <- 0
F22a[1,3] <- ib5 - ib2*ib3
F22a[1,4] <- 0
F22a[1,5] <- ib6 - ib2*ib4
F22a[1,6] <- 0

F22a[2,2] <- ib4 - ib2^2
F22a[2,1] <- 0
F22a[2,4] <- ib5 - ib2*ib3
F22a[2,3] <- 0
F22a[2,6] <- ib6 - ib2*ib4
F22a[2,5] <- 0

F22a[3,1] <- ib5 - ib2*ib3
F22a[3,2] <- 0
F22a[3,3] <- ib6 - ib3^2
F22a[3,4] <- 0
F22a[3,5] <- ib7 - ib3*ib4
F22a[3,6] <- 0

F22a[4,2] <- ib5 - ib2*ib3
F22a[4,1] <- 0
```

```
F22a[4,4] <- ib6 - ib3^2
F22a[4,3] <- 0
F22a[4,6] <- ib7 - ib3*ib4
F22a[4,5] <- 0

F22a[5,1] <- ib6 - ib4*ib2
F22a[5,2] <- 0

F22a[5,3] <- ib7 - ib4*ib3
F22a[5,4] <- 0
F22a[5,5] <- ib8 - ib4^2
F22a[5,6] <- 0
F22a[6,2] <- ib6 - ib4*ib2
F22a[6,1] <- 0

F22a[6,4] <- ib7 - ib4*ib3
F22a[6,3] <- 0
F22a[6,6] <- ib8 - ib4^2
F22a[6,5] <- 0

F23[1,1] <- ib6
F23[1,2] <- ib7
F23[1,3] <- ib8
F23[1,4] <- ib9
F23[1,5] <- ib10
F23[2,1] <- ib7
F23[2,2] <- ib8
F23[2,3] <- ib9
F23[2,4] <- ib10
F23[2,5] <- ib11
F23[3,1] <- ib8
F23[3,2] <- ib9
F23[3,3] <- ib10
F23[3,4] <- ib11
F23[3,5] <- ib12
F23[4,1] <- ib9
F23[4,2] <- ib10
F23[4,3] <- ib11
F23[4,4] <- ib12

F23[4,5] <- ib13
F23[5,1] <- ib10
F23[5,2] <- ib11

F23[5,3] <- ib12
F23[5,4] <- ib13
F23[5,5] <- ib14
```

```
F23a[1,1] <- ib4 - ib2^2
F23a[1,2] <- 0
F23a[1,3] <- ib5 - ib2*ib3
F23a[1,4] <- 0
F23a[1,5] <- ib6 - ib2*ib4
F23a[1,6] <- 0

F23a[1,7] <- ib7 - ib2*ib5
F23a[1,8] <- 0

F23a[1,9] <- ib8 - ib2*ib6
F23a[1,10] <- 0

F23a[2,2] <- ib4 - ib2^2
F23a[2,1] <- 0
F23a[2,4] <- ib5 - ib2*ib3
F23a[2,3] <- 0
F23a[2,6] <- ib6 - ib2*ib4
F23a[2,5] <- 0

F23a[2,8] <- ib7 - ib2*ib5
F23a[2,7] <- 0

F23a[2,10] <- ib8 - ib2*ib6
F23a[2,9] <- 0

F23a[3,1] <- ib5 - ib3*ib2
F23a[3,2] <- 0
F23a[3,3] <- ib6 - ib3^2
F23a[3,4] <- 0
F23a[3,5] <- ib7 - ib3*ib4
F23a[3,6] <- 0

F23a[3,7] <- ib8 - ib3*ib5
F23a[3,8] <- 0

F23a[3,9] <- ib9 - ib3*ib6
F23a[3,10] <- 0

F23a[4,2] <- ib5 - ib3*ib2
F23a[4,1] <- 0
F23a[4,4] <- ib6 - ib3^2
F23a[4,3] <- 0
F23a[4,6] <- ib7 - ib3*ib4
F23a[4,5] <- 0

F23a[4,8] <- ib8 - ib3*ib5
F23a[4,7] <- 0
```

```
F23a[4,10] <- ib9 - ib3*ib6
F23a[4,9] <- 0

F23a[5,1] <- ib6 - ib4*ib2
F23a[5,2] <- 0

F23a[5,3] <- ib7 - ib4*ib3
F23a[5,4] <- 0
F23a[5,5] <- ib8 - ib4^2
F23a[5,6] <- 0
F23a[5,7] <- ib9 - ib4*ib5
F23a[5,8] <- 0

F23a[5,9] <- ib10 - ib4*ib6
F23a[5,10] <- 0

F23a[6,2] <- ib6 - ib4*ib2
F23a[6,1] <- 0

F23a[6,4] <- ib7 - ib4*ib3
F23a[6,3] <- 0
F23a[6,6] <- ib8 - ib4^2
F23a[6,5] <- 0
F23a[6,8] <- ib9 - ib4*ib5
F23a[6,7] <- 0
F23a[6,10] <- ib10 - ib4*ib6
F23a[6,9] <- 0

F23a[7,1] <- ib7 - ib5*ib2
F23a[7,2] <- 0

F23a[7,3] <- ib8 - ib5*ib3
F23a[7,4] <- 0

F23a[7,5] <- ib9 - ib5*ib4
F23a[7,6] <- 0

F23a[7,7] <- ib10 - ib5*ib5
F23a[7,8] <- 0

F23a[7,9] <- ib11 - ib5*ib6
F23a[7,10] <- 0

F23a[8,2] <- ib7 - ib5*ib2
F23a[8,1] <- 0

F23a[8,4] <- ib8 - ib5*ib3
```

```
F23a[8,3] <- 0

F23a[8,6] <- ib9 - ib5*ib4
F23a[8,5] <- 0

F23a[8,8] <- ib10 - ib5*ib5
F23a[8,7] <- 0

F23a[8,10] <- ib11 - ib5*ib6
F23a[8,9] <- 0

F23a[9,1] <- ib8 - ib6*ib2
F23a[9,2] <- 0

F23a[9,3] <- ib9 - ib6*ib3
F23a[9,4] <- 0

F23a[9,5] <- ib10 - ib6*ib4
F23a[9,6] <- 0

F23a[9,7] <- ib11 - ib6*ib5
F23a[9,8] <- 0

F23a[9,9] <- ib12 - ib6*ib6
F23a[9,10] <- 0

F23a[10,2] <- ib8 - ib6*ib2
F23a[10,1] <- 0

F23a[10,4] <- ib9 - ib6*ib3
F23a[10,3] <- 0

F23a[10,6] <- ib10 - ib6*ib4
F23a[10,5] <- 0

F23a[10,8] <- ib11 - ib6*ib5
F23a[10,7] <- 0

F23a[10,10] <- ib12 - ib6*ib6
F23a[10,9] <- 0

f1[z] <- F11%*% solve(F21,F11)
f1a[z] <- f1[z] + F11a%*% solve(F21a,F11a)

f2[z] <- F12%*% solve(F22,F12)
f2a[z] <- f2[z] + F12a%*% solve(F22a,F12a)

f3[z] <- F13%*% solve(F23,F13)
```

```
f3a[z] <- f3[z] + F13a%*% solve(F23a,F13a)

}

q1 <- c(quantile(f1,0.9),quantile(f1,0.95),quantile(f1,0.99))
print(q1)

q1a <- c(quantile(f1a,0.9),quantile(f1a,0.95),quantile(f1a,0.99))
print(q1a)

q2 <- c(quantile(f2,0.9),quantile(f2,0.95),quantile(f2,0.99))
print(q2)

q2a <- c(quantile(f2a,0.9),quantile(f2a,0.95),quantile(f2a,0.99))
print(q2a)

q3 <- c(quantile(f3,0.9),quantile(f3,0.95),quantile(f3,0.99))
print(q3)

q3a <- c(quantile(f3a,0.9),quantile(f3a,0.95),quantile(f3a,0.99))
print(q3a)
```

R program simulation ESTAR(3):

```
T <- 200
end<- T + 50

rep <- 10000

teta11 <- 1
teta12 <- 0.9
teta21 <- 0.4
teta22 <- -0.5

teta20 <- 0

teta <- 0.01

set.seed(1)

countF1 <- 0
countF2 <- 0
countF3 <- 0
countF4 <- 0
countV1 <- 0
countV2 <- 0
```

```
countV3 <- 0
countV4 <- 0
countKSS <- 0
countADF <- 0

y <- c()
y[1] <- 0
y[2] <- 0

dy0 <- c()
dy1 <- c()
y1 <- c()
y2 <- c()
y3 <- c()
y4 <- c()

y5 <- c()
y6 <- c()
y7 <- c()
y8 <- c()
y9 <- c()
y2dy1 <- c()
y3dy1 <- c()
y4dy1 <- c()
y5dy1 <- c()
y6dy1 <- c()
y7dy1 <- c()
y8dy1 <- c()

for(z in 1:rep){

for (i in 3:end){ y[i]<- y[i-1] - teta12*(y[i-1] - y[i-2]) +
(teta21*y[i-1] +
teta22*y[i-2])*(1-exp(-teta*y[i-1]^2*(y[i-1]-3)^2))+ rnorm(1,0,1)
}

for (i in 1:T){
dy0[i] <- y[i+50] - y[i+49]

dy1[i] <- y[i+49] - y[i+48]

y1[i] <- y[i+49]
y2[i] <- y[i+49]^2
y3[i] <- y[i+49]^3

y4[i] <- y[i+49]^4
y5[i] <- y[i+49]^5
y6[i] <- y[i+49]^6
```

```
y7[i] <- y[i+49]^7
y8[i] < - y[i+49]^8
y9[i] <- y[i+49]^9

y2dy1[i] <- y2[i]*dy1[i]
y3dy1[i] <- y3[i]*dy1[i]

y4dy1[i] <- y4[i]*dy1[i]
y5dy1[i] <- y5[i]*dy1[i]

y6dy1[i] <- y6[i]*dy1[i]
y7dy1[i] <- y7[i]*dy1[i]

y8dy1[i] <- y8[i]*dy1[i] }

RF <- lm(dy0 ~ 0 + dy1)
SrF <- deviance(RF)

UF1 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 )
SuF1 <- deviance(UF1)

F1 <- (SrF-SuF1)*T/(SuF1)
if (F1 > 6.9459) countF1 <- countF1 + 1

UF2 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y3dy1 + y4dy1)
SuF2 <- deviance(UF2)
F2 <- (SrF-SuF2)*T/(SuF2)

if (F2 > 15.982) countF2 <- countF2 + 1

UF3 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y6 + y7 + y3dy1 +
y4dy1 + y5dy1 + y6dy1)

SuF3 <- deviance(UF3)
F3 <- (SrF-SuF3)*T/(SuF3)

if (F3 > 23.18338) countF3 <- countF3 + 1

UF4 <- lm(dy0 ~ 0 + dy1 + y3 + y2dy1 + y4 + y5 + y6 + y7 + y8 + y9
+ y3dy1 + y4dy1 + y5dy1 + y6dy1 + y7dy1 + y8dy1)

SuF4 <- deviance(UF4)
F4 <- (SrF-SuF4)*T/(SuF4)

if (F4 > 29.65) countF4 <- countF4 + 1

UV1 <- lm(dy0 ~ 0 + dy1 + y3)
```

```
SuV1 <- deviance(UV1)
V1 <- (SrF-SuV1)*T/(SuV1)

if (V1 > 4.8826) countV1 <- countV1 + 1

UV2 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5)
SuV2 <- deviance(UV2)

V2 <- (SrF-SuV2)*T/(SuV2)
if (V2 > 11.364) countV2 <- countV2 + 1

UV3 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5 + y6 + y7)

SuV3 <- deviance(UV3)
V3 <- (SrF-SuV3)*T/(SuV3)

if (V3 > 15.7059 ) countV3 <- countV3 + 1

UV4 <- lm(dy0 ~ 0 + dy1 + y3 + y4 + y5 + y6 + y7 + y8 + y9)

SuV4 <- deviance(UV4)
V4 <- (SrF-SuV4)*T/(SuV4)

if (V4 > 19.38346) countV4 <- countV4 + 1

KSS <- lm(dy0 ~ 0 + y3 + dy1)
coefKSS <- coef(KSS)

covKSS <- vcov(KSS)
TKSS <- coefKSS[1]/sqrt(covKSS[1,1])

if (TKSS < -2.22) countKSS <- countKSS + 1

adf <- lm(dy0 ~ 0 + y1 + dy1)
coefADF <- coef(adf)

covADF <- vcov(adf)
Tadf <- coefADF[1]/sqrt(covADF[1,1])

if (Tadf < - 1.95) countADF <- countADF + 1

}

pcount <- c(countF1*100/rep, countF2*100/rep, countF3*100/rep,
countF4*100/rep, countV1*100/rep, countV2*100/rep,
countV3*100/rep, countV4*100/rep, countKSS*100/rep,
countADF*100/rep)
```

```
print(pcount)
```

R program bootstrap 1-ESTAR(3) simulations

```
T <- 200
end<- T + 10
rep <- 1000

teta11 <- 1
teta12 <- 0
teta13 <- 0
teta21 <- 0
teta22 <- 0
teta23 <- 0
teta <- 0

set.seed(1)

count1 <- 0
count2 <- 0
count3 <- 0
countv1 <- 0
countv2 <- 0
countv3 <- 0
countKSS <- 0
countADF <- 0

for(z in 1:rep){

y<-c()
y[1]<-0
y[2] <- 0
y[3] <- 0
e<-c()
e[1]<-0
e[2] <- 0

e[3] <- 0
dy0 <- c()
dy1 <- c()
dy2 <- c()
y1 <- c()
y2 <- c()

y3 <- c()
```

```
y4 <- c()
y5 <- c()
y6 <- c()
y7 <- c()
y2dy1 <- c()
y2dy2 <- c()
y3dy1 <- c()
y3dy2 <- c()
y4dy1 <- c()
y4dy2 <- c()
y5dy1 <- c()
y5dy2 <- c()
y6dy1 <- c()
y6dy2 <- c()
y7dy1 <- c()
y7dy2 <- c()

for (i in 4:end){
e[i]<- rnorm(1,0,1)

y[i]<- teta11*y[i-1] + teta12*y[i-2] + teta13*y[i-3] +
 (teta21*y[i-1] + teta22*y[i-2]
 + teta23*y[i-3])*(1-exp(-teta*y[i-1]^2)) + e[i]
}

for (i in 1:T){
dy0[i] <- y[i+10] - y[i+9]

dy1[i] <- y[i+9] - y[i+8]
dy2[i] <- y[i+8] - y[i+7]

y1[i] <- y[i+9]
y2[i] <- y[i+9]^2
y3[i] <- y[i+9]^3

y4[i] <- y[i+9]^4
y5[i] <- y[i+9]^5
y6[i] <- y[i+9]^6

y7[i] <- y[i+9]^7
y2dy1[i] <- y2[i] *dy1[i]

y2dy2[i] <- y2[i] *dy2[i]
y3dy1[i] <- y3[i] *dy1[i]

y3dy2[i] <- y3[i] *dy2[i]
y4dy1[i] <- y4[i] *dy1[i]
```

```
y4dy2[i] <- y4[i] *dy2[i]
y5dy1[i] <- y5[i] *dy1[i]

y5dy2[i] <- y5[i] *dy2[i]
y6dy1[i] <- y6[i] *dy1[i]

y6dy2[i] <- y6[i] *dy2[i]
y7dy1[i] <- y7[i] *dy1[i]

y7dy2[i] <- y7[i] *dy2[i] }

Mr <- lm(dy0 ~ 0 + dy1 + dy2)
Sr <- deviance(Mr)
par <- coef(Mr)

Mu1 <- lm(dy0 ~ 0 + dy1 + dy2 + y3 + y2dy1 + y2dy2)

Su1 <- deviance(Mu1)
r1 <- resid(Mu1)
mr1 <- mean(r1)

sdr1 <- sqrt(var(r1))
par1 <- coef(Mu1)
f1 <- (Sr-Su1)*T/(Su1)

Mu2 <- lm(dy0 ~ 0 + dy1 + dy2 + y3 + y2dy1 + y2dy2 + y4 + y3dy1 +
y3dy2 + y5 + y4dy1 + y4dy2)
Su2 <- deviance(Mu2)

r2 <- resid(Mu2)
mr2 <- mean(r2)
sdr2 <- sqrt(var(r2))

par2 <- coef(Mu2)
f2 <- (Sr-Su2)*T/(Su2)

Mu3 <- lm(dy0 ~ 0 + dy1 + dy2 + y3 + y2dy1 + y2dy2 + y4 + y3dy1 +
y3dy2 + y5 + y4dy1 + y4dy2 + y6 + y5dy1 + y5dy2 + y7 + y6dy1 +
y6dy2)

Su3 <- deviance(Mu3)
r3 <- resid(Mu3)
mr3 <- mean(r3)

sdr3 <- sqrt(var(r3))
par3 <- coef(Mu3)
f3 <- (Sr-Su3)*T/(Su3)
```

```
Muv1 <- lm(dy0 ~ 0 + dy1 + dy2 + y3)
Suv1 <- deviance(Muv1)

fv1 <- (Sr-Suv1)*T/(Suv1)
if (fv1 > 4.88) countv1 <- countv1 +1

Muv2 <- lm(dy0 ~ 0 + dy1 + dy2 + y3 + y4 + y5)

Suv2 <- deviance(Muv2)
fv2 <- (Sr-Suv2)*T/(Suv2)

if (fv2 > 11.36) countv2 <- countv2 +1

Muv3 <- lm(dy0 ~ 0 + dy1 + dy2 + y3 + y4 + y5 + y6 + y7)

Suv3 <- deviance(Muv3)
fv3 <- (Sr-Suv3)*T/(Suv3)

if (fv3 > 15.70) countv3 <- countv3 +1

ks <- lm(dy0 ~ 0 + y3 + dy1 + dy2)
coefKS <- coef(ks)

covKS <- vcov(ks)
Tks <- coefKS[1]/sqrt(covKS[1,1])

if (Tks < -2.22) countKSS <- countKSS + 1

adf <- lm(dy0 ~ 0 + y1 + dy1 + dy2)
coefADF <- coef(adf)

covADF <- vcov(adf)
Tadf <- coefADF[1]/sqrt(covADF[1,1])

if (Tadf < - 1.95) countADF <- countADF + 1

b <- 500 #number of bootstrap sim

fb <- c()

# bootstrap for k=1

for (j in 1:b){
e<-c()
e[1]<- r[1]
e[2] <- r[2]
e[3] <- r[3]
```

```
x <- c()
x[1] <- y[11]
x[2] <- y[12]
x[3] <- y[13]
dx0 <- c()

dx1 <- c()
dx2 <- c()
x3  <- c()
x2dx1 <- c()
x2dx2 <- c()

for (i in 4:(T+3)){
e[i]<- rnorm(1,mr1,sdr1)
x[i]<- x[i-1] + par1[1]*(x[i-1]-x[i-2]) + par1[2]*(x[i-2]-x[i-3])+ e[i] }

for (i in 1:T){
dx0[i] <- x[i+3] - x[i+2]

dx1[i] <- x[i+2] - x[i+1]
dx2[i] <- x[i+1] - x[i]

x3[i] <- x[i+2]^3
x2dx1[i] <- x[i+2]^2 *dx1[i]

x2dx2[i] <- x[i+2]^2 *dx2[i] }

Mrb <- lm(dx0 ~ 0 + dx1 + dx2)
Srb <- deviance(Mrb)

Mub <- lm(dx0 ~ 0 + dx1 + dx2 + x3 + x2dx1 + x2dx2)

Sub <- deviance(Mub)
fb[j] <- (Srb-Sub)*T/(Sub) }

q1 <- quantile(fb,0.95)

if (f1 > q1) count1 <- count1 + 1


# bootstrap for k=2

for (j in 1:b){

e<-c()
e[1]<- r[1]
e[2] <- r[2]
e[3] <- r[3]
```

```
x <- c()

x[1] <- y[11]
x[2] <- y[12]
x[3] <- y[13]
dx0 <- c()
dx1 <- c()
dx2 <- c()
x2  <- c()
x3  <- c()
x4  <- c()
x5  <- c()

x2dx1 <- c()
x2dx2 <- c()
x3dx1 <- c()
x3dx2 <- c()
x4dx1 <- c()
x4dx2 <- c()

for (i in 4:(T+3)){
e[i]<- rnorm(1,mr2,sdr2)
x[i]<- x[i-1] + par2[1]*(x[i-1]-x[i-2]) + par2[2]*(x[i-2]-x[i-3])+
e[i] }

for (i in 1:T){
dx0[i] <- x[i+3] - x[i+2]

dx1[i] <- x[i+2] - x[i+1]
dx2[i] <- x[i+1] - x[i]

x2[i] <- x[i+2]^2
x3[i] <- x[i+2]^3
x4[i] <- x[i+2]^4

x5[i] <- x[i+2]^5
x2dx1[i] <- x2[i] *dx1[i]

x2dx2[i] <- x2[i] *dx2[i]
x3dx1[i] <- x3[i] *dx1[i]

x3dx2[i] <- x3[i] *dx2[i]
x4dx1[i] <- x4[i] *dx1[i]

x4dx2[i] <- x4[i] *dx2[i] }

Mrb <- lm(dx0 ~ 0 + dx1 + dx2)
Srb <- deviance(Mrb)
```

```
Mub <- lm(dx0 ~ 0 + dx1 + dx2 + x3 + x2dx1 + x2dx2 + x4 + x3dx1 +
x3dx2 + x5 + x4dx1 + x4dx2)

Sub <- deviance(Mub)

fb[j] <- (Srb-Sub)*T/(Sub)

}

q2 <- quantile(fb,0.95)

if (f2 > q2) count2 <- count2 + 1

# bootstrap for k=3

for (j in 1:b){

e<-c()
e[1]<- r[1]
e[2] <- r[2]
e[3] <- r[3]
x <- c()

x[1] <- y[11]
x[2] <- y[12]
x[3] <- y[13]
dx0 <- c()
dx1 <- c()
dx2 <- c()
x2  <- c()
x3  <- c()
x4  <- c()
x5  <- c()
x6  <- c()
x7  <- c()
x2dx1 <- c()
x2dx2 <- c()
x3dx1 <- c()
x3dx2 <- c()
x4dx1 <- c()
x4dx2 <- c()
x5dx1 <- c()
x5dx2 <- c()
x6dx1 <- c()
x6dx2 <- c()

for (i in 4:(T+3)){
```

```
e[i]<- rnorm(1,mr3,sdr3)

x[i]<- x[i-1] + par3[1]*(x[i-1]-x[i-2]) + par3[2]*(x[i-2]-x[i-3])
+ e[i] }

for (i in 1:T){
dx0[i] <- x[i+3] - x[i+2]

dx1[i] <- x[i+2] - x[i+1]
dx2[i] <- x[i+1] - x[i]

x2[i] <- x[i+2]^2
x3[i] <- x[i+2]^3
x4[i] <- x[i+2]^4

x5[i] <- x[i+2]^5
x6[i] <- x[i+2]^6
x7[i] <- x[i+2]^6

x2dx1[i] <- x2[i] *dx1[i]
x2dx2[i] <- x2[i] *dx2[i]

x3dx1[i] <- x3[i] *dx1[i]
x3dx2[i] <- x3[i] *dx2[i]

x4dx1[i] <- x4[i] *dx1[i]
x4dx2[i] <- x4[i] *dx2[i]

x5dx1[i] <- x5[i] *dx1[i]
x5dx2[i] <- x5[i] *dx2[i]

x6dx1[i] <- x6[i] *dx1[i]
x6dx2[i] <- x6[i] *dx2[i] }

Mrb <- lm(dx0 ~ 0 + dx1 + dx2)
Srb <- deviance(Mrb)

Mub <- lm(dx0 ~ 0 + dx1 + dx2 + x3 + x2dx1 + x2dx2 + x4 + x3dx1 +
x3dx2 + x5 + x4dx1 + x4dx2 + x6 + x5dx1 + x5dx2 + x7 + x6dx1 +
x6dx2)

Sub <- deviance(Mub)

fb[j] <- (Srb-Sub)*T/(Sub)
}

q3 <- quantile(fb,0.95)
```

```
if (f3 > q3) count3 <- count3 + 1 }

out <- c(count1*100/rep,count2*100/rep, count3*100/rep,
countv1*100/rep,countv2*100/rep,countv3*100/rep, countKSS*100/rep,
countADF*100/rep)

print(out)
```

A.4 Chapter 7 Programs

A.4.1 1-ESTAR(1) Model

Matlab program to calculate trade duration for 1-ESTAR(1) model using integral equation approach.

```
clear; clc;

a = 0; theta1 = 1; theta2 = -0.8; theta = 0.01;

s = 2; u = 1.5;

var = s^2/(1-0.9^2); b = ceil(5*sqrt(var)); n = ceil((b-a)/0.05);

F = zeros(n+1,n+1); h = (b-a)/n;

  for i= 0:n
    x = a+i*h;
    for j = 0:n
        y = a+j*h;
         if (j==0) | (j==n)
              w = 1;
         else w = 2;
         end
         Q = (theta1 + theta2*(1-exp(-theta*x^2)))*x;
         F(i+1,j+1)=h*w*0.5*exp(-(y- Q)^2/(2*s^2))/(sqrt(2*pi*s^2));
     end
end

F = eye(n+1,n+1) - F;
F = inv(F);
P = ones(n+1,1); ft = F*P;

 sum = 0;
 for i = 0:n
    y = a+i*h;
    if (i==0) | (i==n);
```

```
          w = 1;
     else w = 2;
    end
     Qu = (theta1 + theta2*(1-exp(-theta*u^2)))*u;
    sum = sum + h*w*0.5*exp(-(y-Qu)^2/(2*s^2))*ft(i+1,1)/(sqrt(2*pi*s^2));
end TD = sum +1
```

Matlab program to calculate inter-trades interval for 1-ESTAR(1) model using integral equation approach.

```
clear; clc;

a = 0; theta1 = -0.5; theta2 = 0.5; theta = 0.01;

s = 0.75; u = 1.5;

var = s^2/(1-0.9^2);

b = ceil(5*sqrt(var));
%n = ceil((b-a)/0.05);

%F = zeros(n+1,n+1);
%h = (b-a)/n;


  %Calculate the waiting duration

    m = ceil((u-(-b))/0.05);
    F = zeros(m+1,m+1);
    h = (u + b)/m;
    for i= 0:m
        x = (-b)+i*h;
        for j = 0:m
            y = (-b)+j*h;
            if (j==0) | (j==m)
                w = 1;
            else w = 2;
            end
            Q = (theta1 + theta2*(1-exp(-theta*x^2)))*x;
            Fw(i+1,j+1)=h*w*0.5*exp(-(y-Q)^2/(2*s^2))/(sqrt(2*pi*s^2));
          end
     end

 Fw = eye(m+1,m+1) - Fw;
 Fw = inv(Fw);
```

```
 Pw = ones(m+1,1);
 f = Fw*Pw;

 sum = 0;
 for i = 0:m
    y = (-b)+i*h;
    if (i==0) | (i==m);
          w = 1;
     else w = 2;
     end
    sum = sum + h*w*0.5*exp(-(y)^2/(2*s^2))*f(i+1,1)/(sqrt(2*pi*s^2));
end WD = sum +1 #inter-trades interval
```

```
clear; clc;

% Simulation for determining trading duration and
% inter-trades interval in pairs trading problem

% Case : the error cointegration follows an 1-ESTAR(1) process

% assumption : only one trading is allowed
for f = 1:50 ntrade(f) = 0; open = 0; TD(f) = 0; IT(f) =0;

nopen = 0; nclose = 0; theta1 = -0.5; theta2 = 0.5; theta = 0.01;
%sep = 1;
%s = sep*sqrt((1-phi^2));
s = 2;
m = 1000; % number of observations
e(1) = randn; for i = 2:m;
    a(i)= s*randn;
    e(i) = (theta1 + theta2*(1-exp(-theta*e(i-1)^2)))*e(i-1)+ a(i);

    % open the trading when error cointegration exceed u= 1.5
    if (e(i) > 1.5) & (open == 0)
        open = 1;
        nopen = nopen + 1;
        topen(nopen) = i;
        if (nopen > 1)
            IT(f) = IT(f) + topen(nopen) - tclose(nclose);
        end
    end

     % close the trading when error cointegration below zero
     if (e(i) < 0) & (open == 1)
         open = 0;
         nclose = nclose + 1;
```

```
         tclose(nclose) = i;
         TD(f) = TD(f) + tclose(nclose) - topen(nopen) ;
        ntrade(f) = ntrade(f) + 1;
     end

end


mTD(f) = TD(f)/ntrade(f); mIT(f) = IT(f)/(nopen-1);
end

NT = mean(ntrade)
sNT = std(ntrade)
eTD = mean(mTD)
sTD = std(mTD)
eIT = mean(mIT)
sIT = std(mIT)
```

A.4.2 1-ESTAR(2) Model

Matlab program to calculate trade duration for 1-ESTAR(2) model using integral equation approach.

```
clear; clc;

%p = 2;
a = 0; s = 1;

theta11 = 0.1; theta12 = 0.9; theta21 = 0.4; theta22 = -0.5;

theta = 0.01;

par=[theta11 theta12 theta21 theta22 theta];

b = ceil(5*s) n = ceil((b-a)/0.5)

h = (b-a)/n; L = 0.5*(h^2);

N = zeros((7*2*(n^2)),3); M = zeros((6*(n^2)+4*n+1),3);

NT =zeros(7,3); T = zeros(2,3);

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+(i*h) a+(i*h) a+(i+1)*h; a+j*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
```

```
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j)):(7*(n*i+j+1)),:) = NT;
    end
end

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+i*h a+(i+1)*h a+(i+1)*h; a+(j+1)*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
```

```
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j+n^2)):(7*(n*i+j+n^2+1)),:) = NT;
    end
end N;

v = zeros(1,2); for i= 1:(7*2*(n^2))
    v = N(i,1:2);
    for j = (i+1):(7*2*(n^2))
        if (N(j,1:2)== v)
            N(i,3) = N(i,3)+N(j,3);
            N(j:(7*2*(n^2)-1),:)=N((j+1):(7*2*(n^2)),:);
            N((7*2*(n^2)),:)= zeros(1,3);
        end
    end
end N;

m = 6*(n^2)+4*n+1; M = N(1:m,:);

F = zeros(m,m);
  for i= 1:m
    A(1,1) = M(i,1);
    A(2,1) = M(i,2);
    for j = 1:m
    B(1,1) = M(j,1);
    B(2,1) = M(j,2);
    k = M(j,3)*fintESTAR1(A,B,s,par);
    F(i,j)= k;
    end
end F; F = eye(m,m)-F;

P=zeros(m,1); for i = 1:m
    x0=[M(i,1);M(i,2)];
    P(i,1) = fintESTAR2(x0,a,b,s,par,n);
end P;

f = F\P;

x0 = [1.5;1.5]; sum0= 0; for i = 1:m
    B(1,1)=M(i,1);
    B(2,1)=M(i,2);
    sum0 = sum0 + M(i,3)*fintESTAR1(x0,B,s,par)*f(i,1);
```

```
end

sum0 = sum0 + fintESTAR2(x0,a,b,s,par,n)
```

Matlab program for function “K”

```
function K=fintESTAR1(A,B,s,par);


sum = (B(1,1) - par(1,1)*A(1,1)- par(1,2)*A(2,1) -
(par(1,3)*A(1,1)+ par(1,4)*A(2,1))*(1-exp(-par(1,5)*A(1,1)^2)))^2
+ (B(2,1) - par(1,1)*B(1,1) -par(1,2)*A(1,1) - (par(1,3)*B(1,1)+
par(1,4)*A(1,1))*(1-exp(-par(1,5)*B(1,1)^2)))^2;

sum = -sum/(2*s); K = exp(sum)/(2*pi*s);
```

Matlab program for function “psi”

```
function psi = fintESTAR2(x0,a,b,s,par,n)

h = (b-a)/n; sum =0; for i= 0:n
    x = a+i*h;
    if (i==0) | (i==n)
        w = 1;
    else w = 2;
    end
    sum = sum + h*w*0.5*exp(-(x - par(1,1)*x0(1,1) - par(1,2)*x0(2,1) -
    (par(1,3)*x0(1,1)+ par(1,4)*x0(2,1))
      *(1-exp(-par(1,5)*x0(1,1)^2)))^2/(2*s));
end
sum = sum/(sqrt(2*pi*s)); psi = 1 +sum;
```

Matlab program to calculate inter-trades interval for 1-ESTAR(2) model using integral equation approach.

```
clear; clc;

%p = 2;
%a = 0;
s = 2;

theta11 = 0.1; theta12 = 0.9; theta21 = 0; theta22 = -0.9; theta =
```

```
0.01;

par=[theta11 theta12 theta21 theta22 theta];

u = 1.5; b = ceil(5*s); n = ceil((u+b)/0.5);


h = (u-(-b))/n; L = 0.5*(h^2);

N = zeros((7*2*(n^2)),3); M = zeros((6*(n^2)+4*n+1),3); NT
=zeros(7,3); T = zeros(2,3);

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[-b+(i*h) -b+(i*h) -b+(i+1)*h; -b+j*h -b+(j+1)*h -b+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j)):(7*(n*i+j+1)),:) = NT;
    end
end

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[-b+i*h -b+(i+1)*h -b+(i+1)*h; -b+(j+1)*h -b+(j+1)*h -b+j*h];
        NT(1,1) = T(1,1);
```

```
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j+n^2)):(7*(n*i+j+n^2+1)),:) = NT;
    end
end N;

v = zeros(1,2); for i= 1:(7*2*(n^2))
    v = N(i,1:2);
    for j = (i+1):(7*2*(n^2))
        if (N(j,1:2)== v)
            N(i,3) = N(i,3)+N(j,3);
            N(j:(7*2*(n^2)-1),:)=N((j+1):(7*2*(n^2)),:);
            N((7*2*(n^2)),:)= zeros(1,3);
        end
    end
end N;

m = 6*(n^2)+4*n+1; M = N(1:m,:);

F = zeros(m,m);
  for i= 1:m
    A(1,1) = M(i,1);
    A(2,1) = M(i,2);
    for j = 1:m
    B(1,1) = M(j,1);
    B(2,1) = M(j,2);
```

```
    k = M(j,3)*fintESTAR1(A,B,s,par);
    F(i,j)= k;
    end
end F; F = eye(m,m)-F;

P=zeros(m,1); for i = 1:m
    x0=[M(i,1);M(i,2)];
    P(i,1) = fintESTAR2(x0,-b,u,s,par,n);
end P;

f = F\P;

x0 = [0;0]; sum0= 0; for i = 1:m
    B(1,1)=M(i,1);
    B(2,1)=M(i,2);
    sum0 = sum0 + M(i,3)*fintESTAR1(x0,B,s,par)*f(i,1);
end
sum0 = sum0 + fintESTAR2(x0,-b,u,s,par,n)
```

Matlab program to find optimal upper-bound U for 1-ESTAR(2) model.

```
clear; clc;

%p = 2;
a = 0;
s = 0.99; %variance ESTAR model residuals
ve = 8 ; %variance e


theta11 = 0.1155744; theta12 = 0.8743728; theta21 = 11.4546619;
theta22 = -16.4558339; theta = 0.0007;

par=[theta11 theta12 theta21 theta22 theta];

b = ceil(4*sqrt(ve))
n = ceil((b-a)/0.5)

h = (b-a)/n;
L = 0.5*(h^2); %triangle area

N = zeros((7*2*(n^2)),3); M = zeros((6*(n^2)+4*n+1),3);

NT =zeros(7,3); T = zeros(2,3);

% Calculate trading duration
```

```
for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+(i*h) a+(i*h) a+(i+1)*h; a+j*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j)):(7*(n*i+j+1)),:) = NT;
    end
end

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[a+i*h a+(i+1)*h a+(i+1)*h; a+(j+1)*h a+(j+1)*h a+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
```

```
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j+n^2)):(7*(n*i+j+n^2+1)),:) = NT;
    end
end N;

v = zeros(1,2); for i= 1:(7*2*(n^2))
    v = N(i,1:2);
    for j = (i+1):(7*2*(n^2))
        if (N(j,1:2)== v)
            N(i,3) = N(i,3)+N(j,3);
            N(j:(7*2*(n^2)-1),:)=N((j+1):(7*2*(n^2)),:);
            N((7*2*(n^2)),:)= zeros(1,3);
        end
    end
end N;

m = 6*(n^2)+4*n+1; M = N(1:m,:);

F = zeros(m,m);
  for i= 1:m
    A(1,1) = M(i,1);
    A(2,1) = M(i,2);
    for j = 1:m
    B(1,1) = M(j,1);
    B(2,1) = M(j,2);
    k = M(j,3)*fintESTAR1(A,B,s,par);
    F(i,j)= k;
    end
end F; F = eye(m,m)-F;

P=zeros(m,1); for i = 1:m
    x0=[M(i,1);M(i,2)];
    P(i,1) = fintESTAR2(x0,a,b,s,par,n);
end P;

f = F\P;
```

```
profit = 0;

for z = 3:0.1:6  %range of possible optimal U
   x0 = [z;z];
   sum0= 0;
   for i = 1:m
        B(1,1)=M(i,1);
        B(2,1)=M(i,2);
        sum0 = sum0 + M(i,3)*fintESTAR1(x0,B,s,par)*f(i,1);
   end
   TD = sum0 + fintESTAR2(x0,a,b,s,par,n)


 %Calculate the waiting duration

    n = ceil((z-(-b))/0.05);
    h = (z-(-b))/n;
    L = 0.5*(h^2);

    N = zeros((7*2*(n^2)),3);
    M = zeros((6*(n^2)+4*n+1),3);
    NT =zeros(7,3);
    T = zeros(2,3);

    for i = 0:(n-1);
    for j = 0:(n-1);
    T =[-b+(i*h) -b+(i*h) -b+(i+1)*h; -b+j*h -b+(j+1)*h -b+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
```

```
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j)):(7*(n*i+j+1)),:) = NT;
    end
    end

for i = 0:(n-1);
    for j = 0:(n-1);
    T =[-b+i*h -b+(i+1)*h -b+(i+1)*h; -b+(j+1)*h -b+(j+1)*h -b+j*h];
        NT(1,1) = T(1,1);
        NT(1,2) = T(2,1);
        NT(1,3) = 2*L*1/40;
        NT(2,1) = T(1,2);
        NT(2,2) = T(2,2);
        NT(2,3) = 2*L*1/40;
        NT(3,1) = T(1,3);
        NT(3,2) = T(2,3);
        NT(3,3) = 2*L*1/40;
        nt = 1/2*T(:,1) + 1/2*T(:,2);
        NT(4,1) = nt(1,1);
        NT(4,2) = nt(2,1);
        NT(4,3) = 2*L*1/15;
        nt = 1/2*T(:,1) + 1/2*T(:,3);
        NT(5,1) = nt(1,1);
        NT(5,2) = nt(2,1);
        NT(5,3) = 2*L*1/15;
        nt = 1/2*T(:,2) + 1/2*T(:,3);
        NT(6,1) = nt(1,1);
        NT(6,2) = nt(2,1);
        NT(6,3) = 2*L*1/15;
        nt = 1/3*T(:,1) + 1/3*T(:,2)+ 1/3*T(:,3);
        NT(7,1) = nt(1,1);
        NT(7,2) = nt(2,1);
        NT(7,3) = 2*L*9/40;
        N((1+7*(n*i+j+n^2)):(7*(n*i+j+n^2+1)),:) = NT;
    end
end N;

v = zeros(1,2); for i= 1:(7*2*(n^2))
    v = N(i,1:2);
    for j = (i+1):(7*2*(n^2))
        if (N(j,1:2)== v)
            N(i,3) = N(i,3)+N(j,3);
            N(j:(7*2*(n^2)-1),:)=N((j+1):(7*2*(n^2)),:);
            N((7*2*(n^2)),:)= zeros(1,3);
        end
```

```
    end
end N;

m = 6*(n^2)+4*n+1; M = N(1:m,:);

F = zeros(m,m);
  for i= 1:m
    A(1,1) = M(i,1);
    A(2,1) = M(i,2);
    for j = 1:m
    B(1,1) = M(j,1);
    B(2,1) = M(j,2);
    k = M(j,3)*fintESTAR1(A,B,s,par);
    F(i,j)= k;
    end
end F; F = eye(m,m)-F;

P=zeros(m,1); for i = 1:m
    x0=[M(i,1);M(i,2)];
    P(i,1) = fintESTAR2(x0,-b,z,s,par,n);
end P;

f = F\P;

x0 = [0;0]; sum0= 0; for i = 1:m
    B(1,1)=M(i,1);
    B(2,1)=M(i,2);
    sum0 = sum0 + M(i,3)*fintESTAR1(x0,B,s,par)*f(i,1);
end WD = sum0 + fintESTAR2(x0,-b,z,s,par,n)

NT= (300/(TD+WD))-1;

pr = NT*z;

if (pr > profit)
    profit = pr;
    u = k;
    mTD = TD;
    mWD = WD;
    mNT = NT;
end

end

profit
u
mTD
mWD
```

mNT

```
clear; clc;

% Simulation for determining trading duration and
% inter-trades interval in pairs trading problem

% Case : the error cointegration follows a k-ESTAR(2) process

% assumption : only one trading is allowed
for f = 1:50 ntrade(f) = 0; open = 0; TD(f) = 0; IT(f) =0;

nopen = 0; nclose = 0; theta11 = 0.1; theta12 = 0.9; theta21 = 0;
theta22 = -0.9; theta = 0.01;
%sep = 1;
%s = sep*sqrt((1-phi^2));
s = 1; %variance
e(1) = sqrt(s)*randn; e(2) = sqrt(s)*randn;

m = 1000; % number of obeservations

for i = 3:m;
    a(i)= sqrt(s)*randn;
    e(i) = theta11*e(i-1) + theta12*e(i-2) +
    (theta21*e(i-1)+theta22*e(i-2))*(1-exp(-theta*e(i-1)^2))+ a(i);

    % open the trading when error cointegration exceed u= 1.5
    if (e(i) > 1.5) & (open == 0)
        open = 1;
        nopen = nopen + 1;
        topen(nopen) = i;
        if (nopen > 1)
            IT(f) = IT(f) + topen(nopen) - tclose(nclose);
        end
    end

     % close the trading when error cointegration below zero
     if (e(i) < 0) & (open == 1)
         open = 0;
         nclose = nclose + 1;
         tclose(nclose) = i;
         TD(f) = TD(f) + tclose(nclose) - topen(nopen) ;
        ntrade(f) = ntrade(f) + 1;
     end

end
```

```
mTD(f) = TD(f)/ntrade(f); mIT(f) = IT(f)/(nopen-1);
end

eTD = mean(mTD)
sTD = std(mTD)
eIT = mean(mIT)
sIT = std(mIT)

NT = mean(ntrade)
sNT = std(ntrade)
```

Bibliography

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions in Automatic Control*, **AC-19**, 716–723.

Alexander, C. and Dimitriu, A. (2002) The cointegartion alpha: Enhanced index tracking and long-short equity market neutral strategies. *Discussion papers*, Finance ISMA Center, University of Reading, UK.

— (2005) Index and statistical arbitrage: Tracking error or cointegration? *Discussion papers*, Finance ISMA Center, University of Reading, UK.

Alexander, C., Giblin, I. and Weddington, W. (2001) Cointegration and asset management: A new active hedge fund strategy. *Discussion papers*, Finance ISMA Center, University of Reading, UK.

Andrade, S., Di Pietro, V. and Seasholes, M. (2005) Understanding the profitability of pairs trading. *Working paper*, UC Berkeley Haas School.

Atkinson, K. (1997) *The Numerical Solution of Integral Equations of the Second Kind.* Cambridge: Cambridge University Press.

Bair, J. and Haesbroeck, G. (1997) Monotonous stability for neutral fixed points. *Bulletin Belgium Mathematics Society*, **4**, 639–646.

Balke, N. and Fomby, T. (1997) Threshold cointegration. *International Economics Review*, **38**, 627–645.

Barnhill, T. and Jordan, J.V. ans Seale, W. (1987) Maturity and refunding effects on treasury-bond future price variance. *Journal of Financial Research*, **10**, 121–131.

Basak, G. and Ho, K.-W. (2004) Level-crossing probabilities and first-passage times for linear processes. *Advances in Applied Probability*, **36**, 643–666.

Bedi, J. and Tennant, P. (2002) Dual–listed companies. *Reserve Bank of Australian Bulletin.*

Benhabib, J., Schmitt-Grohe, S. and Uribe, M. (1999) The perils of Taylor rules. *CEPR Discussion Papers 2314.*

Berben, R. and van Dijk, D. (1999) Unit root tests and asymmetric adjustment: A reassessment. Unpublish manuscript, Tinbergen Institute, Erasmus University of Rotterdam.

Bertram, W. (2009) An overview of the financial markets and an application of financial mathematics. Presented at Industry Seminar, School of Mathematics and Applied Statistics, Wollongong University.

Beveridge, S. and Nelson, C. (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics*, **7**, 151–174.

Bhattacharya, R. and Lee, C. (1995) On geometric ergodicity of nonlinear autoregressive models. *Statistics and Probability Letters*, **22**, 311–315.

Black, F. (1972) Capital market equilibrium and restricted borrowing. *Journal of Business*, **45**, 444–454.

Bodie, Z., Kane, A. and Marcus, A. (2009) *Investments (8th edition)*. Boston: McGraw-Hill.

Box, G. and Jenkins, G. (1976) *Time Series Analysis: Forecasting and Control*. Holden-Day Inc.

Brooks, C. and Kat, H. (2002) The statistical properties of hedge fund index returns and their implications for investors. *Journal of Alternative Investments*, **5**, 26–44.

Burns, J. (1983) Futures markets and efficiency. in Streit, M.E., Ed., Futures Markets. Oxford, UK: Blackwell, 46–74.

Butterworth, D. and Holmes, P. (2001) The hedging effectiveness of stock index fixtures: evidence for the FTSE-100 and FTSE-mid250 indexes traded in the UK. *Applied Financial Economics*, **11**, 57–68.

Cagan, P. (1956) The monetary dynamics of hyperinflantion. In *Studies in the Quantity Theory of Money* (ed. M. Friedman), 25–117. Chicago: University of Chicago Press.

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

Cargill, T. and Rauseer, G. (1975) Temporal price behavior in commodity futures markets. *The Journal of Finance*, **30**, 1043–1053.

Chan, K. (1991) Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society, Series B*, **53**, 691–696.

— (1992) A futher analysis of the lead-lag relationship between the cash market and stock index future markets. *Review of Financial Studies*, **5**, 123–152.

— (1993) Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, **21**, 520–533.

Chan, K. and Tsay, R. (1998) Limiting properties of the the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, **45**, 413–426.

Chen, C. and Lee, J. (1995) Bayesian inference of threshold autoregressive model. *Journal of Time Series Analysis*, **16**, 483–492.

Chiarella, C., Gao, S. and Stevenson, M. (2008) Resolving the price-earning puzzle and related econometric issues. *Working paper*, Faculty of Economics and Business, University of Sydney.

Chowdhury, A. (1991) Futures market efficiency: Evidence from cointgration tests. *The Journal of Futures Markets*, **11**, 577–589.

Cornell, B. and French, K. (1983) The pricing of stock index futures. *Journal of Futures Markets*, **3**, 1–14.

Crowder, W. and Phengpis, C. (2005) Stability of the S&P 500 futures market efficiency condistions. *Applied Financial Economics*, **15**, 855–866.

Diamond, P. (1982) Aggregate demand management in search equilibrium. *Journal of Political Economy*, **90**, 881–894.

Dickey, D. and Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427–431.

— (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057–1072.

Do, B. and Faff, R. (2008) Does naive pairs trading still work? *Working paper*, Monash University.

Do, B., Faff, R. and Hamza, K. (2006) A new approach to modelling and estimation for pairs trading. *Working paper*, Monash University.

Doornik, J. and Hendry, D. (1997) *Modeling Dynamic Systems Using PcFiml 9.0 for Windows.* London: International Thomson Business Press.

Durbin, J. and Koopman, S. (2001) *Time Series Analysis by State Space Models.* USA: Oxford University Press.

Dwyer, G., Locke, P. and Yu, W. (1996) Index arbitrage and nonlinear dynamics between the S&P 500 future and cash. *Review of Financial Studies*, **9**, 301–332.

Ehrman, D. (2006) *The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures.* New York: John Wiley & Sons Inc.

Eklund, B. (2003a) Testing the unit root hypothesis against the logistic smooth transition autoregressive model. *SSE/EFI Working Paper Series in Economics and Finance 546*, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

— (2003b) A nonlinear alternative to the unit root hypothesis. *SSE/EFI Working Paper Series in Economics and Finance 547*, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

Elliott, R., Aggoun, L. and Moore, J. (1995) *Hidden Markov Models.* Berlin: Springer.

Elliott, R., van der Hoek, J. and W.P., M. (2005) Pairs trading. *Quantitative Finance*, **5**, 271–276.

Enders, W. and Granger, C. (1998) Unit root tests and asymmetric adjustment with an example using the trem structure of interest rates. *Journal of Business and Economics Statistics*, **16**, 304–311.

Engelberg, J., Gao, P. and Jagannathan, R. (2008) An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity. Available at SSRN: http://ssrn.com/abstract=1330689.

Engle, R. and Granger, C. (1987) Co-integration and error correction representation, estimation and testing. *Econometrica*, **55**, 251–276.

Figlewski, S. (1984a) Explaining the early discounts on stock index futures: The case for disequilibrium. *Financial Analysts Journal*, **40**, 43–47.

— (1984b) Hedging performance and basis risk in stock index futures. *Journal of Finance*, **34**, 157–170.

— (1985) Hedging with stock index futures: theory and application in a new market. *Journal of Futures Markets*, **5**, 183–199.

Finch, S. (2004) Ornstein−Uhlenbeck process. Unpublished Note. Available online at http://pauillac.inria.fr/algo/bsolve/constant/constant.html.

Fujihara, R. and Mougoue, M. (1997) Linear dependence, nonlinear dependence and petroleum futures market efficiency. *The Journal of Futures Markets*, **17**, 75–99.

Fuller, W. (1976) *Introduction to Statistical Time Series.* New York: John Wiley.

Fund, W. and Hsieh, D. (1999) A primer of hedge funds. *Journal of Empirical Finance*, **6**, 309–331.

Galenko, A., Popova, E. and Popova, I. (2007) Trading in the presence of cointegration. Available at SSRN:http://ssrn.com/abstract=1023791.

Garcia, P., Leuthold, R., Fortenbery, T. and Sarassoro, G. (1988) Pricing efficiency in the live cattle future market: Further interpretation and measurement. *American Journal of Agricultural Economics*, **70**, 162–169.

Gatev, E., Goetzmann, W. and Rouwenhorst, K. (1998) Pairs trading: Performance of a relative value arbitrage rule. *Working Papers ysm26*, Yale School of Management.

— (2006) Pairs trading: Performance of a relative value arbitrage rule. *The Review of Financial Studies*, **19**, 797–827.

Gillespie, T. and Ulph, C. (2001) Pair trades methodology: A question of mean reversion. In *8th International Conference of Forum for Interdiciplinary Mathematics.*

Goss, B. (1981) The forward pricing function of the London Metal Exchange. *Applied Economics*, **13**, 133–150.

— (1983) The semi-strong form efficiency of the London Metal Exchange. *Applied Economics*, **15**, 681–698.

— (1988) A semi-strong test of the efficiency of the aluminium and copper markets at the LME. *The Journal of Future Markets*, **8**, 67–77.

Granger, C. and Terasvirta, T. (1993) *Modelling Nonlinear Economic Relationship.* Oxford: Oxford University Press.

Gupta, S. and Mayer, T. (1981) A test of the efficiency of futures markets in commodities. *Review of World Economics*, **117**, 661–671.

Habak, C. (2002) *Pairs Trading: Applying Cointegration to a Pairs Trading Strategy.* Bachelor of mathematics and finance honours thesis, University of Wollongong.

Hakkio, C. and Rush, M. (1989) Market efficiency and cointegration: An application to the Sterling and Deutschemark exchange markets. *Journal of International Money and Finance*, **8**, 75–88.

Hansen, B. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413–430.

— (1997) Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, **1**, 119–131.

— (2000) Sample splitting and threshold estimation. *Econometrica*, **64**, 575–603.

Harris, R. (1995) *Using Cointegration Analysis in Econometric Modeling.* London: Prentice Hall.

He, C. and Sandberg, R. (2005b) Dickey–Fuller type of tests against nonlinear dynamic models. *SEE/EFI Working Paper Series in Economics and Finance 580*, Department of Economic Statistics, Stockholm School of Economics, Stockholm, Sweden.

Hely, S. (2008) Day trading. *Money*, **105**, 78–79.

Hendry, D. and Doornik, J. (1996) *Empirical Econometric Modeling Using PcGive 9.0 for Windows.* London: International Thomson Business Press.

Herlemont, D. (2008) A range trading system. Available at http://www.yats.com/doc/range-trading-en.html.

Holmes, P. (1995) Ex ante hedge ratios and the hedging effectiveness of the FTSE-100 stock index futures contract. *Applied Economics Letters*, **2**, 56–59.

Hong, G. and Susmel, R. (2003) Pairs-trading in the Asian ADR market. Available at http://www.bauer.uh.edu/rsusmel/Academic/ptadr.pdf.

Hong, S. and Phillips, P. (2010) Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business and Economic Statistics*, **28**, 96–104.

Huck, N. (2009) Pairs selection and outranking: An application to the S&P100 index. *Eurepean Journal of Operational Research*, **196**, 819–825.

— (2010) Pairs selection and outranking: the multi-step ahead forecasting case. *Eurepean Journal of Operational Research*, **207**, 1702–1716.

Jacobs, B. and Levy, K. (1993) Long/short equity investing. *Journal of Portfolio Management*, **20**, 52–64.

Jacobs, B., Levy, K. and Starer, D. (1999) Long–short portfolio management: An integrated approach. *Journal of Portfolio Management*, **25**, 23–32.

Jensen, M. (1969) The performance of mutual funds in the period 1945–1964. *Journal of Finance*, **23**, 389–416.

Johansen, S. (1988) Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.

— (1989) Likelihood based inference on cointegration. *Lecture notes*, University of Copenhagen.

— (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.

— (1994) The role of the constant and linear terms in cointegration analysis of non-stationary variables. *Ecnometric Reviews*, **13**, 205–230.

Kapetanios, G., Shin, Y. and Snell, A. (2003) Testing for a unit root in the nonlinear STAR framework. *Journala of Econometrics*, **112**, 359–379.

Karatzas, I. and Shreve, S. (1998) *Methods of Mathematical Finance*. New York: Springer.

Kawaller, I., Kock, P. and Koch, T. (1987) The temporal price relationship between S&P 500 futures and the S&P 500 index. *The Journal of Finance*, **42**, 1309–1329.

Kemeny, J. and Snell, J. (1960) *Finite Markov Chains*. New Jersey: D.Van Nostrand Company Inc., Princeton.

Kenourgios, D. and Samitas, A. (2008) Hedge ratio estimation and hedging effectiveness: The case of the S&P 500 stock index futures contract. *International Journal of Risk Assessment and Management*, **9**, 121–134.

Kinlay, A. and Ramaswamy, K. (1988) Index-futures arbitrage and the behavior of stock index futures prices. *The Review of Financial Studies*, **1**, 137–158.

Kofi, T. (1973) A framework for comparing the efficiency of future markets. *American Journal of Agricultural Economics*, **55**, 584–594.

Koutmos, G. and Tucker, M. (1996) Temporal relationships and dynamic interactions between spot and future stock markets. *The Journal of Future Markets*, **16**, 55–69.

Lai, C. (1978) First order autoregressive markov processes. *Stochastic Processes and Their Application*, **7**, 65–72.

Layard, R., Nickell, S. and Jackman, R. (1991) *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.

Li, H. and Maddala, G. (1996) Bootstrapping time series models. *Econometric Reviews*, **15**, 115–158.

Lim, G. and Martin, V. (1995) Regression-based cointegration estimators. *Journal of Economic Studies*, **22**, 3–22.

Lin, Y.-X., McCrae, M. and Gulati, C. (2006) Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences*, **2006**, 1–14. (doi: 10.1155/JAMDS/2006/73803).

Lintner, J. (1965) The valuation of risk assets and the selection of risk investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **4**, 13–37.

Lo, M. and Zivot, E. (2001) Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, **5**, 533–576.

Luukkonen, R., Saikkonen, P. and Terasvirta, T. (1988) Testing linearity against smooth transition autoregressive models. *Biometrica*, **75**, 491–499.

Ma, C. and Hein, S. (1990) Testing unbiasedness in future markets: Further evidence. *Working paper*, Texas Tech University.

Ma, C., Mercer, J. and Walker, M. (1992) Rolling over futures contracts: A note. *The Journal of Futures Markets*, **12**, 203–217.

MacKinnon, J. (1992) Critical values for cointegration tests. In *Long–run Economic Relationship: Readings in Cointegration (Advanced Texts in Econometrics)* (eds. R. Engle and C. Granger), 267–276. USA: Oxford University Press.

Markowitz, H. (1959) *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

Martens, M., Kofman, P. and Vorst, T. (1998) A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics*, **13**, 245–263.

McMillan, D. and Speight, A. (2005) Nonlinear dynamics and competing behavioral interpretations: Evidence from intra-day FTSE-100 index and futures data. *The Journal of Futures Markets*, **26**, 343–368.

Michael, P., Nobay, R. and Peel, D. (1997) Transactions costs and non-lienar adjustment in real exchange rates: An empirical investigation. *Journal of Political Economy*, **105**, 862–879.

Miller, M., Muthuswamy, J. and Whaley, R. (1994) Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance*, **49**, 479–513.

Milonas, N. (1986) Price variability and the maturity effect in futures markets. *The Journal of Futures Markets*, **6**, 443–460.

Modest, D. and Sundaresan, M. (1983) The relationship between spot and future prices in stock index futures markets: Some preliminary evidence. *Journal of Futures Markets*, **3**, 14–41.

Monoyios, M. and Sarno, L. (2002) Mean reversion in stock index futures markets: A nonlinear analysis. *The Journal of Futures Markets*, **22**, 285–314.

Mudchanatongsuk, S., Primbs, J. and Wong, W. (2008) Optimal pairs trading: A stochastic control approach. In *American Control Conference*. Washington, USA.

Nath, P. (2003) High frequency pairs trading with US treasury securities: Risks and rewards for hedge funds. *Working paper*, London Business School.

Ng, S. and Perron, P. (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, **69**, 1519–1554.

Osterwald-Lenum, M. (1992) A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, **54**, 461–472.

Papandakis, G. and Wisocky (2008) Pairs trading and accounting information. *Working paper*, MIT Sloan School of Management.

Paya, I., Venetis, I. and Peel, D. (2003) Further evidence on PPP adjustment speeds: the case of effective real exchange rates and the EMS. *Oxford Bulletin of Economics and Statistics*, **65**, 421–438.

Perron, P. (1988) Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of Economic Dynamics and Control*, **12**, 297–332.

Phillips, P. (1987) Time series regression with a unit root. *Econometrica*, **55**, 277–301.

Phillips, P. and Durlauf, S. (1986) Multiple time series regression with integrated processes. *Review of Economic Studies*, **53**, 473–495.

Phillips, P. and Perron, P. (1988a) Testing for a unit root in time series regression. *Biometrica*, **75**, 335–346.

— (1988b) Testing for a unit root in time series regression. *Biometrica*, **75**, 335–346.

Phillips, P. and Solo, V. (1992) Asymptotics for linear processes. *The Annals of Statistics*, **20**, 971–1001.

Pippenger, M. and Goering, G. (1993) A note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics*, **55**, 473–481.

Puspaningrum, H., Lin, Y.-X. and Gulati, C. (2009) Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. *Working Paper 2109*, Centre for Statitical and Survey Methodology, University of Wollongong.

— (2010) Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. *Journal of Statistical Theory and Practice*, **4**, 391–419.

Rampertshammer, S. (2007) An Ornstein-Uhlenbeck framework for pairs trading. *Tech. rep.*, Department of Mathematics and Statistics, University of Melbourne, Australia. Available at www.ms.unimelb.edu.au/publications/RampertshammerStefan.pdf.

Ramsey, J. (1969) Tests for specification errors in classical least-squares regression analysis. *Journal of the Royal Statistical Analysis, Series B,*, **13**, 350–371.

Reimers, H. (1992) Comparisons of tests for multivariate cointegration. *Statistical Papers*, **33**, 335–359.

Rice, S. (1945) Mathematical analysis of random noise. *Bell System Technical Journal*, **24**, 146–156.

Rissanen, J. (1978) Modeling by shortest data description. *Automatica*, **14**, 465–471.

Ross, S. (1976) Arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**, 341–360.

Saikonnen, P. and Luukkonen, R. (1988) Lagrange multiplier tests for testing non-linearities in time series models. *Scandinavian Journal of Statistics*, **15**, 55–68.

Sargan, J. and Bhargava, A. (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica*, **51**, 153–174.

Sargent, T. and Wallace, N. (1973) Rational expectations and the dynamics of hyper-inflantion. *International Economics Review*, **14**, 328–350.

Sarno, L., Taylor, M. and Peel, D. (2002) Nonlinear equilibrium correction in US real money balances, 1869–1997. *Journal of Money, Credit and Banking*, **35**, 787–799.

Sarno, L. and Valente, G. (2000) The cost of carry model and regime shifts in stock index futures markets: An empirical investigation. *The Journal of Futures Markets*, **20**, 603–624.

Schmidt, A. (2008) *Pairs Trading: A Cointegration Approach.* Undergraduate honours thesis, School of Finance, University of Sydney.

Schwarz, G. (1978) Estimating the dimention of a model. *The Annals of Statistics*, **4**, 461–464.

Schwert, G. (1989) Tests for unit roots. *Journal of Business and Economic Statistics*, **7**, 147–159.

Sharpe, W. (1964) Capital assets prices: A theory of market equilibrium under condition of risk. *Journal of Finance*, **19**, 425–442.

Shen, C. and Wang, L. (1990) Examining the validity of a test of futures markets efficiency: A comment. *The Journal of Futures Markets*, **10**, 195–196.

Shiller, R. and Perron, P. (1985) Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters*, **18**, 381–386.

Stock, J. and Watson, M. (1991) Variable trends in economic time series. *Journal of Economic Perspectives*, **2**, 147–174.

Taylor, M., Peel, D. and Sarno, L. (2001) Nonlinear in real exchange rates: Towards a solution of the purchasing power parity puzzles. *International Economic Review*, **42**, 1015–1042.

Terasvirta, T. (1994) Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208–218.

Terasvirta, T. and Elliasson, A. (2001) Nonlinear error correction and the UK demand for broad money, 1878–1993. *Journal of Applied Econometrics*, **16**, 277–288.

Tomek, W. and Gray, R. (1970) Temporal relationships among prices on commodity future markets: Their allocative and stabilizing roles. *American Journal of Agricultural Economics*, **52**, 372–380.

Tong, H. (1978) On a threshold model. In *Pattern Recognition and Signal Processing* (ed. C. Chen). Amsterdam: Sijhoff and Noordhoff.

Venetis, I., Paya, I. and Peel, D. (2009) ESTAR model with multiple fixed points: Testing and estimation. *Working paper*, Department od Economics, Lancaster University Management School, Lancaster, UK.

Vidyamurthy, G. (2004) *Pairs Trading: Quantitative Methods and Analysis*. New York: John Wiley.

Whistler, M. (2004) *Trading Pairs - Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies*. New York: John Wiley.

Yadav, P., Pope, P. and Paudyal, K. (1994) Threshold autoregressive modelling in finance: The price differences of equivalent assets. *Mathematical Finance*, **4**, 205–221.