Cointegration with a time trend and pairs trading strategy: empirical study on the S&P 500 future and spot index prices

Heni Puspaningrum
Australian Tax Office

Yan-Xia Lin
University of Wollongong

Chandra Gulati
University of Wollongong

Cointegration with a time trend and pairs trading strategy: empirical study on the S&P 500 future and spot index prices

Abstract

The efficiency and relationship between future and spot index prices has been investigated in literature by using different approaches. In this paper, the efficiency and relationship between the log future prices and the log spot prices is investigated for each contract based on cointegration model with a time trend. A new pairs trading strategy for stock prices cointegrated with time trend is proposed in this paper. Formulas used to evaluate the trading strategy are derived and the necessary conditions for making profit out of selected pairs of assets are discussed. S&P500 future contracts prices of Mar98, Jun98 and Sep98 as well as their corresponding spot index prices are considered in this paper. Our empirical studies show that the strategy proposed in this paper works very well and produces very significant high return for those selected pairs. The average return per trade for all periods and all those pairs are above 10% .

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

10-14

Cointegration with a Time Trend and Pairs Trading Strategy:
Empirical Study on the S&P 500 Future and Spot Index Prices

Heni Puspaningrum, Yan-Xia Lin and Chandra Gulati

Cointegration with a Time Trend and Pairs Trading Strategy: Empirical Study on the S&P 500 Future and Spot Index Prices

Heni Puspaningrum[1]**, Yan-Xia Lin**[2] **and Chandra Gulati**[2]

[1]Australian Taxation Office, Canberra, ACT 2600

[2]National Institute for Applied Statistics Research Australia

School of Mathematics and Applied Statistics, University of Wollongong,

Wollongong, NSW, 2500, AUSTRALIA

yanxia@uow.edu.au

ABSTRACT

The efficiency and relationship between future and spot index prices has been investigated in literature by using different approaches. In this paper, the efficiency and relationship between the log future prices and the log spot prices is investigated for each contract based on cointegration model with a time trend. A new pairs trading strategy for stock prices cointegrated with time trend is proposed in this paper. Formulas used to evaluate the trading strategy are derived and the necessary conditions for making profit out of selected pairs of assets are discussed. S&P500 future contracts prices of Mar98, Jun98 and Sep98 as well as their corresponding spot index prices are considered in this paper. Our empirical studies show that the strategy proposed in this paper works very well and produces very significant high return for those selected pairs. The average return per trade for all periods and all those pairs are above 10% .

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, tests for market efficiency in a semi-strong form sense were employed by subsequent researchers. These tests are used to determine whether future prices fully reflect all publicly available information at the time of contract. An econometric model is applied to compare 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; Gupta and Mayer, 1981) found evidences 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 requisite 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 (1.1):

$$E(\ln S_T | \mathcal{F}_t) = \ln F_{t,T} \tag{1.1}$$

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 (1.1) is based on estimates of (1.2):

$$\ln S_T = \alpha_0 + \alpha_1 \ln F_{t,T} + \upsilon_t \tag{1.2}$$

where α_0 and α_1 are constants, 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 consists 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 used to test the efficiency of future markets is cost-carry model. Consider a market containing an asset, a stock index, whose price S_t at time t, under the equivalent martingale measure evolves according to:

$$dS_t = S_t \ (\bar{r} - \bar{d}) \ dt + \sigma \ S_t \ dW_t, \tag{1.3}$$

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. Using the Ito's lemma,

$$E\left(S_t\right) = S_0 \ \exp(R \, t), \tag{1.4}$$

where $R = \bar{r} - \bar{d}$.

Standard derivatives pricing theory gives the future price $F_{t,T}$ at time t for a maturity at time $T \geq t$ as

$$F_{t,T} = E(S_T | \mathcal{F}_t), \tag{1.5}$$

where E denotes the mathematical expectation with the respect to the martingale measure P and $\mathcal{F}_t$ denotes the information set updated to time t (e.g. see Karatzas and Shreve, 1998). Given (1.4) and (1.5), the future price has the well-known formula:

$$F_{t,T} = S_t \ \exp(R(T-t)). \tag{1.6}$$

Taking natural log on both sides of (1.6) yields

$$\ln F_{t,T} = \ln S_t + R(T-t). \tag{1.7}$$

Monoyios (2002) defined basis b_t at time t as

$$b_t = f_t - s_t, \tag{1.8}$$

where $f_t = \ln F_{t,T}$ and $s_t = \ln S_t$.

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 paired 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 strong evidence of nonlinearity in the relationship between future and spot prices of S&P 500 index (see Puspaningrum, 2012).

We concern about the way the basis is constructed. By pairing up the spot price with the future contract with the nearest maturity, the basis defined in (1.8) 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 1 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, instead of constructing one future prices series for the whole data, cointegration relationship of the log future prices and the log spot prices for each contract with a time trend is analysed in this paper.

In this paper, by assuming that R in (1.7) is a constant during the contract period we employ a cost-carrying model for each contract as follow:

$$f_t = \mu + \beta \, s_t + \delta \, t^* + \epsilon_t \tag{1.9}$$

where $t^* = (T - t)$ is the time to maturity and ϵ_t is the error term. Theoretically, based on (1.7), $\mu = 0$ and $\beta = 1$, thus δ is the estimate of R in (1.7). Using Engle-Granger method, we would like to find out whether the log future prices, the log spot prices and the time trend are cointegrated for each future contract.

Lin *et al.* (2006) and Puspaningrum *et al.* (2010) developed a pairs trading mechanism for two cointegrated assets by taking advantages of stationarity of the cointegration errors. Pairs
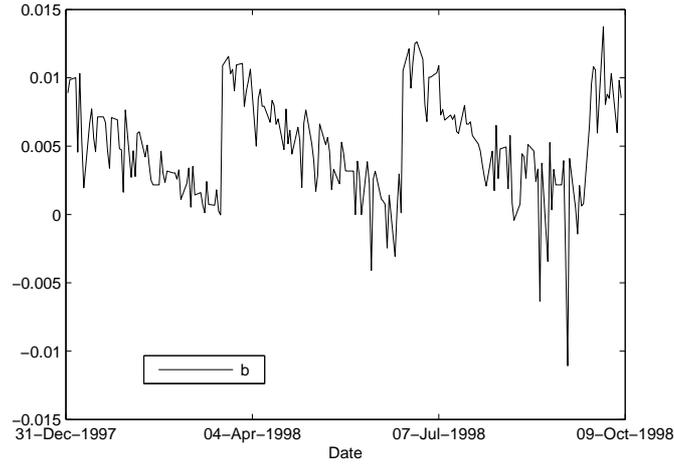
Figure 1: Plot of b_t from January 1, 1998 to October 19, 1998.

trading works by taking the arbitrage opportunity of temporary anomalies between prices of related assets which have long-run equilibrium. When such an anomaly event occurs, one asset will be overvalued relatively 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 these 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 (1.9) 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), the log future contract prices and the log spot index prices in (1.9) are cointegrated with the time trend t^*. Furthermore, future and spot index have different trading mechanisms (see Section 3). 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 (Contract for Difference) a stock index. There is uncertainty about the amount of interest and dividend received or paid. The pairs trading mechanism developed by Puspaningrum *et al.* (2010) cannot be applied to (1.9) directly. New technique for the pairs of log prices is desirable.

The rest of the sections will be organised as follows. Section 2 discusses cointegration analysis of the log future contract prices and the log spot prices with a time trend for each future contract. Section 3 discusses pairs trading mechanism between future and spot prices and shows empirical pairs trading studies using S&P500 future contracts prices of Mar98, Jun98 and Sep98 as well as their corresponding spot index prices. The last section is conclusion.

2   Cointegration Analysis with a Time Trend between Future Contract and Spot Index Prices of S&P 500

In order to determine whether cointegration with time trend exists between the log of spot prices and the log of future prices, there are two approaches commonly used, i.e.: the Engle-Granger two-step approach, developed by Engle and Granger (1987), and the Johansen's approach by Johansen (1988, 1994). In this paper, we apply the Engle-Granger two-step approach because of its simplicity and its minimisation of the variance of cointegration error which is good for pair trading. On the other hand, the Johansen approach emphases cointegration as a multivariate system.

Consider the model in (1.9). After running the regression model in (1.9), 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 follows:

$$\Delta \hat{\epsilon}_t = \rho\,\hat{\epsilon}_{t-1} + \sum_{i=1}^{p-1} \omega_i\,\Delta \hat{\epsilon}_{t-i} + v_t \tag{2.1}$$

where $v_t \sim iid\, N(0, \sigma^2)$ and the $\hat{\epsilon}_t$ are obtained from estimating (1.9).

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 method available in Fuller (1976). However, unless the true parameter values in (1.9) 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 (1.9) 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 (1.9). 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 (1.9) is one (excluding the constant and trend), the critical values by MacKinnon (1992) will be the same as the ADF tests.

3   Pairs Trading between Future and Spot Index Prices

This section will propose a pairs trading strategy between future and spot index prices when they are cointegrated with a time trend. 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 combining these assets together and the last subsection is the empirical example of pairs trading between future and spot S&P 500 index prices.

3.1   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 movement in the future price. This is done to ensure that both parties will have sufficient funds available for mark to market. Residual credit risk exists only to the extent that (a) future prices move so dramatically that the amount required for mark to market is larger than the balance of an individual's margin account, and (b) 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.

3.2 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 proportion 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 it involves us entering a contract with a CFD issuer for a particular asset such as shares, stock indexes or 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 gives 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 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 that its current price of $25 is 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 positive return of 400% [1]. However, if the share price decreases to, say, $20, we will get a negative return of 400%. 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"

---

[1]$(\$5 \times 800)/\$1000 \times 100\%$.

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 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.

3.3 Pairs Trading between Future and Spot Index

Assume that the log of future and spot index prices are cointegrated as in (1.9). Therefore, ϵ_t from (1.9) 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 they can be excluded from the modelling[2].

Since future and spot index have different trading mechanisms and the risk and uncertainty involved in the trading mechanisms as well as the impact of time trend exist, we propose a pairs trading strategy for future and spot index below.

Pairs trading strategy:

(1) Set a pre-determined upper-bound $U > 0$ and a lower-bound $L = -U$.

(2) For upper trades, put future in "short" (sell) position and β CFD in "long" (buy) position (see (1.9) 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**. Different from Puspaningrum (2010),

---

[2]We are aware that this assumption is not realistic in practice, but as a starting point, it can give us basic understanding about pairs trading mechanism between the two assets.

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 time in 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 $\epsilon_2 < -U$.

Different from the closing rule set in the strategy proposed in Puspaningrum (2010), strategy proposed in this paper requires each opened pair trade has to be closed by next day regardless of the value of ϵ_t on next day. Therefore, the formulas developed in Puspaningrum (2010) cannot be used to estimate the number of trades and eventually estimate the optimal upper-bound U_0 for the strategy proposed in this paper.

In this section, we derive formulae for evaluating return and identify the necessary condition for making profit based on the proposed 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:

$$
\begin{aligned}
return \quad &= \quad \frac{1}{m}\left[(f_t - f_{t+1}) + \beta\,(s_{t+1} - s_t)\right] & (3.1) \\
&\approx \quad -\frac{F_{t+1} - F_t}{m\,F_t} + \beta\frac{S_{t+1} - S_t}{m\,S_t} & (3.2)
\end{aligned}
$$

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 (3.1) as follow:

$$
\begin{aligned}
return \quad &= \quad \frac{1}{m}\left[(f_t - f_{t+1}) + \beta\,(s_{t+1} - s_t)\right] \\
&= \quad \frac{1}{m}\left[f_t - \beta\,s_t - \mu - \delta\,(T - t)\right] - \frac{1}{m}\left[f_{t+1} - \beta\,s_{t+1} - \mu - \delta(T - t - 1)\right] + \frac{1}{m}\delta \\
&= \quad \frac{1}{m}\left[\epsilon_t - \epsilon_{t+1} + \delta\right]. & (3.3)
\end{aligned}
$$

Consider the scenario where ϵ_t is a stationary AR(p) model, i.e.:

$$
\epsilon_t = \theta_0 + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_p\epsilon_{t-p} + \eta_t,
$$

where the parameters $\theta_0, \theta_1, \ldots, \theta_p$ fulfill the requirements of stationary AR(p) process ; there is no correlation between η_t and $(\epsilon_{t-1}, \ldots, \epsilon_{t-p})$ and $\eta_t \sim$ IID $N(0, \sigma_\eta)$.

Let $f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t)$ denote the probability density function of ϵ_{t+1} given $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$. Thus,

$$f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t) = f(\eta_{t+1}) \sim N(0, \sigma_\eta^2) = \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^p \theta_j \epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right). \quad (3.4)$$

Then, given $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$,

$$
\begin{aligned}
E(return|\boldsymbol{\epsilon}_t) &= \frac{1}{m} \int_{-\infty}^{\infty} (\epsilon_t - \epsilon_{t+1} + \delta) f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t) d\epsilon_{t+1} \\
&= \frac{1}{m} \int_{-\infty}^{\infty} (\epsilon_t + \delta) f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t) d\epsilon_{t+1} - \frac{1}{m} \int_{-\infty}^{\infty} \epsilon_{t+1} f(\epsilon_{t+1}|\boldsymbol{\epsilon}_t) d\epsilon_{t+1} \\
&= \frac{1}{m} (\epsilon_t + \delta - c) \quad (3.5)
\end{aligned}
$$

where $c = \theta_0 + \sum_{j=1}^p \theta_j \epsilon_{t-j+1}$.

Two scenarios are considered, i.e.: $\delta > 0$ and $\delta < 0$, as we only consider the situation where a time trend exists.

If $\delta > 0$, from (??), 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 $\epsilon_{t+1} > (\epsilon_t + \delta)$. From (??), for given δ and c, the higher the value of ϵ_t, the higher the expected return. A positive expected return will be achieved if

$$\epsilon_t > c - \delta.$$

For a simple case where ϵ_t is a white noise process (i.e., $\epsilon_t = \eta_t$) with zero mean, then $c = 0$ and $\sigma_\eta = \sigma_\epsilon$. For this case, a positive expected return will be achieved if

$$\epsilon_t > U > 0 > -\delta \quad (3.6)$$

where U is the pre-set upper-bound. Since $\delta > 0$ and $U > 0$, it will mean whenever $\epsilon_t > 0$, the trade can be opened. However, there is still a possibility to get loss. Since the pair trading has to be closed on the next day after it is opened, from (??), unlike the pairs trading strategy in Puspaningrum (2010), there is no guaranty of minimal profit for each pair trade when the trading is closed. 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 $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$ and $\epsilon_t > U$ is

$$
\begin{aligned}
P(Loss|\boldsymbol{\epsilon}_t) &= \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{(\epsilon_t+\delta)}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^p \theta_j \epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\
&\leq \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{\epsilon_t}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^p \theta_j \epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\
&\leq \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{U}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^p \theta_j \epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\
&= 1 - \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{-\infty}^{U} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^p \theta_j \epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1}. \quad (3.7)
\end{aligned}
$$

For the simple case where ϵ_t is a white noise process with zero mean, the distribution of ϵ_{t+1} is the same as the distribution of ϵ_t. Thus, for this case,

$$P(Loss|\boldsymbol{\epsilon}_t) \leq 1 - \Phi\left(\frac{U}{\sigma_\eta}\right) \tag{3.8}$$

where $\Phi(.)$ denotes the cumulative distribution of a standard normal distribution. From (3.8), the higher the pre-set upper-bound U, the lower the probability of loss. Thus, choosing U depends on the risk aversion of investors. If $U = 1.65\sigma_\epsilon$,

$$P(Loss|\boldsymbol{\epsilon}_t) \leq 5\%. \tag{3.9}$$

Similarly for $\delta < 0$, from (**??**), we will have a positive return or profit if $\epsilon_{t+1} \leq \epsilon_t - |\delta|$. From (**??**), a positive expected return will be achieved if $\epsilon_t > c + |\delta|$. For a simple case where ϵ_t is a white noise process, a positive expected return will be achieved if

$$\epsilon_t > U > |\delta| \tag{3.10}$$

where U is the pre-set upper-bound. From (**??**), 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 $\boldsymbol{\epsilon}_t = (\epsilon_t, \ldots, \epsilon_{t-p+1})$ and $\epsilon_t > U$ is

$$
\begin{aligned}
P(Loss) &= \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{(\epsilon_t - |\delta|)}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\
&\leq \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{U-|\delta|}^{\infty} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1} \\
&= 1 - \frac{1}{\sqrt{2\pi}\sigma_\eta} \int_{-\infty}^{U-|\delta|} \exp\left(-\frac{(\epsilon_{t+1} - \theta_0 - \sum_{j=1}^{p}\theta_j\epsilon_{t-j+1})^2}{2\sigma_\eta^2}\right) d\epsilon_{t+1}. \quad (3.11)
\end{aligned}
$$

For the simple case where ϵ_t is a white noise process,

$$P(Loss|\boldsymbol{\epsilon}_t) \leq 1 - \Phi\left(\frac{U - |\delta|}{\sigma_\eta}\right) \tag{3.12}$$

3.4 Empirical Pairs Trading Studies Using S&P 500 Stock Index Data

Daily prices data of future contracts Mar98, Jun98, Sep98 and Dec98 as well as their corresponding spot index prices are used to make pairs trading study. From Table **??**, $\hat{\epsilon}_t$ in training period for Mar98, Jun98 and Sep98 can be considered as white noise process because the autocorrelation with 20 lags can be rejected at a 1% significant level. Using future contracts prices and corresponding spot index prices during training periods, a regression model in (1.9) is formed. Regression modelling results between f_t, s_t and t^* using (1.9) can be seen in Table **??**.

Regression residuals from (1.9) will be:

$$\hat{\epsilon}_t = f_t - \hat{\mu} - \hat{\beta}\, s_t - \hat{\delta}\, t^*. \tag{3.13}$$

Regression residuals analysis can be seen in Table **??**. If $\hat{\epsilon}_t$ in (3.13) 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 3.3 is performed for training data to see whether the pairs trading strategy is profitable or not.

As Mar98, Jun98 and Sep98, $\hat{\epsilon}_t$ are accepted as white noise and the value of $\delta < 0$, we choose the pre-set upper-bound U as $1.65\sigma_{\hat{\epsilon}} + |\delta|$ and the pre-set lower-bound as $L = -U$. From (3.9), such U gives that the maximum of the probability of a trade having loss will not be greater than 5%.

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. (3.13). Using the pre-set upper-bound U and the lower-bound $L = -U$ from training period, the pairs trading strategy described in Subsection 3.3 is also performed. Table **??** reports pairs trading results during training and trading periods. Total return, average return per trade and number of trades are recorded. Total return is defined as follow:

$$\text{Total return} = \sum_{i=1}^{TN} \text{return}_i \tag{3.14}$$

where TN is the total number of pair trades during the period. Return for each pair trade is calculated based on (3.1) with the percentage margin $m = 3\%$ and β is replaced by $\hat{\beta}$. 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 results in Table **??** show that the pairs trading strategy works very well and produces very significant high return. The average return per trade for all periods is above 10% . The average positive return per trade for all periods is only slightly higher than the average return per trade. This indicates that the losses do not have significant impact on the total return. As seen from Table **??**, we only have a few losses.

4 Conclusion

In this paper, we use cointegration analysis with a time trend based on the Engle-Granger to determine whether the log of future prices and the log of spot prices are cointegrated and then use the cointegration relationship to perform pairs trading strategy. A new pairs trading strategy is suggested for future and spot index prices. The proposed strategy involves the

decision of opening position, which is determined by pre-determined bound U and underlying cointegrating errors. Since a time trend is involved in the cointegration relationship between log of future prices and the log of spot prices, to reduce trading risk, the decision of closing rule in the proposed pairs trading strategy is that each trade must be closed by next day regardless of profit achieved or not. It is shown that, if the cointegration error ϵ_t follows AR model, the probability of profit loss in one trade can be evaluated and controlled if the pre-determined bound U is appropriate. The pairs trading strategy is applied to future contracts Mar98, Jun98 and Sep98 where they show strong cointegration relationship and the regression residuals (or we can also say the the cointegration errors) follow a white noise process. The empirical results show that the pairs trading strategy proposed in this paper works very well and produces very significant high return during training periods and trading periods.

For a cointegration with time trend system, the longer a pair-trading stays in the system, the higher the probability of loss will be. Therefore, "next day closing" is proposed in this paper. By adopting this strategy, trading loss may incur. Finding an optimal strategy to remove this restriction on closing trade is of interest and it will be investigated in future.