Pairs Trading: A Cointegration Approach

Arlen David Schmidt
Supervisor: Dr Maxwell Stevenson
University of Sydney
Finance Honours Thesis
November 2008

Submitted in partial fulfillment of the requirement for the award of Bachelor of Economics (Honours) degree

Copyright Statement:

Student Permission
I hereby grant to The University of Sydney or its agents the right to archive and to make available my Honours thesis in the Sydney eScholarship Repository in all forms of media, now or hereafter known. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this Honours thesis.

Name: ____

Date: ____

Signature: ____

Supervisor Permission
I hereby grant to The University of Sydney or its agents the right to archive and to make available the Honours thesis of my student _________________________, in the Sydney eScholarship Repository in all forms of media, now or hereafter known.

________________________,retains all proprietary rights, such as patent rights and also retains the right to use in future works (such as articles or books) all or part of this Honours Thesis.

Name: ANDREW LEPONE

Date: 3 November, 2008

Signature: ____

Acknowledgements

First and foremost, I would like to acknowledge the one and true God, my Lord Jesus Christ for blessing me with the ability and guidance to succeed this year. I sincerely thank my supervisor, Dr Maxwell Stevenson, for providing me with his profound knowledge, direction and unrelenting encouragement. His thoughts were always valued and this study would not have been possible without his significant input. To my parents, I would like to acknowledge and thank them for their support, both financial and otherwise. I am truly appreciative. An honourable mention must go out to Michael Ng, whose thoughts and comments both contributed to, and improved, this study. Finally, to the rest of the honours cohort, both students and staff, I thank them for making this the most enjoyable and academically productive year of my university life, thus far.

Abstract

This study uses the Johansen test for cointegration to select trading pairs for use within a pairs trading framework. A long-run equilibrium price relationship is then estimated for the identified trading pairs, and the resulting mean-reverting residual spread is modeled as a Vector-Error-Correction model (VECM). The study uses 5 years of daily stock prices starting from the beginning of July, 2002. The search for trading pairs is restricted to 17 financial stocks listed on the ASX200. The results show that two cointegrated stocks can be combined in a certain linear combination so that the dynamics of the resulting portfolio are governed by a stationary process. Although a trading rule is not employed to access the profitability of this trading strategy, plots of the residual series show a high rate of zero crossings and large deviations around the mean. This would suggest that this strategy would likely be profitable. It can also be concluded that in the presence of cointegration, at least one of the speed of adjustment coefficients must be significantly different from zero.

Table of Contents

1. Introduction

Pairs trading is a statistical arbitrage hedge fund strategy designed to exploit short-term deviations from a long-run equilibrium pricing relationship between two stocks. Traditional methods of pairs trading have sought to identify trading pairs based on correlation and other non-parametric decision rules. However, as we will show, these approaches are inferior to the technique applied in this study because they do not guarantee the single most important statistical property which is fundamental to a profitable pairs trading strategy, namely, mean reversion. This study selects trading pairs based on the presence of a cointegrating relationship between the two stock price series. The presence of a cointegrating relationship then enables us to combine the two stocks in a certain linear combination so that the combined portfolio is a stationary process. The portfolio is formed by longing the relative under-valued stock and shorting the relative over-valued stock. If two cointegrated stocks share a long-run equilibrium relationship, then deviations from this equilibrium are only short-term and are expected to return to zero in future periods. To profit from this relative mis-pricing, a long position in the portfolio is opened when its value falls sufficiently below its long-run equilibrium and is closed out once the value of the portfolio reverts to its expected value. Similarly, profits may be earnt when the portfolio is trading sufficiently above its equilibrium value by shorting the portfolio until it reverts to its expected value.

Typical questions which must be answered when developing a pairs trading strategy include (1) how to identify trading pairs, (2) when is the combined portfolio sufficiently away from its equilibrium value to open a trading position, and (3) when do we close the position. From a risk management perspective, it is also important to specify maximum allowable time to maintain open trading positions, maximum allowable Value at Risk (VaR), and further possible risk reducing measures such as stop-loss triggers.

The purpose of this study is to develop a method for implementing a pairs trading strategy - it does not attempt to access its profitability. Firstly, we illustrate a method for the identification of trading pairs using the Johansen test for cointegration. We can then estimate the cointegrating relationship between our pairs by regressing one on the other. This study uses Granger Causality to specify the order of regression. The residual series from the cointegrating equation describes the dynamics of the mean-reverting portfolio which we then choose to model as a Vector-Error-Correction model (VECM). This is a natural choice since we know through the "Granger Representation Theorem" (Engle and Granger, 1987) that in the presence of a cointegrating relationship this is an equivalent representation. We choose to model the residual series to extract as much information as we can about the co-movement between our trading pairs, which would be valuable in developing a trading rule. For example, estimating the "speed of adjustment" coefficients within the VECM provide us with some idea of how quickly the system reverts to equilibrium following a short-term deviation as well as which stock is responsible for the "error-correction" function. Lastly, a natural extension of estimating our VECM is to plot impulse response functions and conduct variance decomposition analysis. These variance analysis tools provide us with knowledge of how each stock price series responds to shocks to itself and the other stock sequence, as well as the degree to which each stock series evolves independently of the other.

This study proceeds in the following 5 sections. In section 2 we discuss some key issues surrounding pairs trading and its implementation. Section 3 consists of a brief outline of the existing pairs trading approaches and a summary of the limited academic literature. Section 4 and 5 describe the data used in this study and the method the study employs. We present our findings in section 6 and conclude in section 7.

2. Discussion of key issues

2.1 Long/Short Equity Investing: Profit from both Winner and Losers

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 incorporating 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 stock 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) categorize long/short equity strategies as market neutral, equitized, 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 the portfolio. This approach eliminates net equity market exposure so that the returns 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 *equitized* strategy, in addition to holding 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 equitized portfolio has a full equity market exposure at all times. Profits are made from the long/short spread, which is contingent upon the investors' stock selection abilities, as well as profits made from the portfolios exposure to systematic 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 stock index futures. Once again, profits are made from the long/short spread as well as the exposure to the changing stock index futures 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 the full long/short spread and then obtaining the desired market exposure through the use of stock index futures.

Do, Faff and Hamza (2006) structure the different approaches to long/short equity investing slightly differently to Jacobs and Levy (1993) and, in doing so, attempt to clarify the position of a pairs trading strategy amongst other seemingly related hedge fund strategies. Do et al (2006) note that due to the 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 declare 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 long/short 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). A pairs trading strategy also requires the identification of mis-priced securities. However, it seeks to identify *relative* mispricing, where the prices of two stocks are away from some known long-run equilibrium relationship.

Both classes of 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 p.1). Debate continues amongst academics and practitioners alike regarding the role, if any, of market neutrality for a successful pairs trading strategy. We will discuss this debate in section 2.3.

2.2 Why Long/Short Investing Strategies?

Jacobs and Levy (1993) and Jacobs, Levy and Starer (1999) suggest 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.

Jacobs and Levy (1993) describe the major benefit of long/short strategies with the following analogy. Suppose you expect the Yankees to win their game and the Mets to lose theirs. If you wager on baseball you would certainly not *just* bet on the Yankees to win. You would also “short” the Mets. The same logic can be applied to equity investing. Why only bet on winners? Why avail yourself to only half the opportunity? Profits can be earnt from both winning and losing stocks simultaneously, earning the full performance spread, or what Alexander and Dimitriu (2002) refer to as “double alpha”.

The fact that long/short equity strategies ensure a more efficient use of information than long-only strategies is the result of not restricting the weights of the undervalued assets to zero[1]. By allowing portfolio returns to be borne by both the short set under-performing the market and the long set over-performing the market, the strategy generates double alpha.

Another benefit of long/short investing is that, potentially, short positions provide greater opportunities than long positions. The search for undervalued stocks 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.

[1] In a long-only portfolio, the investor is only able to under-weight over-valued or poor-quality stocks. Effectively, the investor cannot profit from this mis-pricing, they can only reduce losses. A long/short strategy allows the investor to explicitly profit from this type of mis-pricing.

2.3 A role for market neutrality in a pairs trading strategy?

Do et al (2006) categorize the set of long/short equity strategies as either belonging to a "market neutral" or a "pairs 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 the 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, McCrae and Gulati (2006) and Nath (2003) implicitly describe pairs trading as an implementation of market neutral investing. Both sets of authors repeatedly describe 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 immunize trading outcomes against systematic market-wide movements in prices that may work against uncovered positions.

To some extent both authors are correct, however as Alexander and Dimitriu (2002) explain 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. To illustrate, in a recent paper Brooks and Kat (2001) find evidence of significant correlation of classic long/short equity hedge funds indexes with equity market indexes such as S&P500, DJIA, Russell 2000 and NASDAQ, correlation which may still be under-estimated due to the auto-correlation of returns. Alexander and Dimitriu (2002) provide 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 (i.e. pairs trading), ensure that over a given time horizon the equities will reach an assumed equilibrium pricing relationship. In this case, the portfolio does not require a beta of zero to immunize it against systematic risk. This is handled by the assumed equilibrium pricing relationship, for example, a cointegrating relationship. To summarize, Lin et al (2006) and Nath (2003) were correct in describing pairs trading as a market neutral investment strategy, however, this market neutrality is derived from proven interdependencies within the chosen stocks combined with a portfolio beta equal to zero. Simply holding a combination of long and short positions is not sufficient to guarantee market neutrality.

What are the implications for portfolio risk if the proven interdependent relationship between the paired stocks does not hold into the future? Any investment strategy is faced with certain risks. The fundamental risk facing pair trading strategies is that the long-run equilibrium, mean-reverting relationship on which profitability is contingent upon does not hold into the future. If investors were faced with this occurrence, possibly due to a certain structural change in one of the stocks, then the portfolio would face both systematic and firm-specific risks. A key aim of any risk-adverse investor is to minimize risk for a given return, and it can be reasonably expected that through holding sufficient trading pairs the firm-specific risk component can be diversified away. The portfolio will however, remain subject to systematic risk factors, since its beta is unlikely to be zero by default (Alexander and Dimitriu, 2002). We propose that a pairs trading portfolio with zero beta can be used as a risk-management device to minimize the adverse effects of systematic risk in the case of structural change.

2.4 Cointegration and correlation in long/short strategies

Following the seminal work of Markowitz (1959), Sharpe (1964), Lintner (1965), and Black (1972), the fundamental statistical tool for traditional portfolio optimization is correlation analysis of asset 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 transaction 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) note, 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 stationarity 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 the 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, 2001).

The fundamental remark justifying the application of the 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 of 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 a cointegrating 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 price series 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. It is important to note that the identification of cointegrated pairs is not a fundamental requirement for a successful pairs trading strategy, indeed several approaches outlined in the next section make no mention of cointegration. However, pairs trading approaches which are based on cointegration can guarantee mean reversion, which is the single most important feature of a successful pairs trading strategy. No other approach can guarantee this property.

In section 3 we will provide a brief review of the different approaches to pairs trading proposed in the literature, both from the non-parametric and cointegrating frameworks.

3. Literature Review

In this section we introduce four studies which collectively describe the main approaches used to implement pairs trading, which we label: the distance method, the stochastic spread method, the more extensive stochastic residual spread method and the cointegration method. The non-parametric distance method is adopted by Gatev, Goetzmann and Rouwenhorst (1999) and Nath (2003) for empirical testing. The stochastic spread and the stochastic residual spread approaches are proposed more recently by Elliot, Van Der Hoek and Malcolm (2005) and Do, Faff and Hamza (2006), respectively. Finally, the cointegration approach is outlined in Vidyamurthy (2004). These latter approaches represent an attempt to parameterize pairs trading by explicitly modeling the mean-reverting behaviour of the spread.

3.1 The distance approach

Under the distance 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 (1999) construct 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 belong to the same broad industry categories defined by S&P. This acts as a test for robustness of any net profits identified using the unrestricted sample of pair trades.

Gatev et al (1999) base their trading rules for opening and closing positions on a standard deviation metric. An opening long/short trade occurs when prices diverge by more than two historical deviations, as estimated during the pair formation period. Opened positions are closed-out at the next crossing of the prices.

Nath (2003) also uses a measure of distance to identify potential pair trades, although his approach does not identify mutually exclusive pairs. Nath (2003) keeps a record of distances for each pair in the universe of securities, in an empirical distribution format so that each time an observed distance crosses over the 15 percentile, a trade is opened for that pair. Contrary to Gatev et al (1999) it is possible under Nath's approach that one particular security be traded against multiple securities simultaneously. A further discrepancy between the two approaches is that Gatev et al (1999), simplistically make no attempt to incorporate any risk management measures into their trading approach. Nath (2003) incorporates a stop-loss trigger to close the position whenever the distance moves against him to hit the 5 percentile. Additionally, 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, at a price level. As Do et al (2006) notes, 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. In fact it is a common practice in existing pairs trading strategies that mispricing is measured in terms of price level.

3.2 The stochastic spread approach

Elliot et al (2005) outline an approach to pairs trading which explicitly models the mean reverting behaviour of the spread in a continuous time setting. The spread is defined as the difference between the two stock prices. The spread is driven by a latent state variable x, which is assumed to follow a Vasicek process:

$$dx_t = k(\theta - x_t)dt + \sigma dB_t \qquad (4.1)$$

where dB_t is a standard Brownian motion in some defined probability space. The state variable is known to revert to its mean θ at the rate k. By making the spread equal to the state variable plus a Gaussian noise, or:

$$y_t = x_t + H\omega_t \qquad (4.2)$$

the trader asserts that the observed spread is driven mainly by a mean reverting process, plus some measurement error where $\omega_t \sim N(0,1)$.

Elliot et al (2005) suggest that this model offers three major advantages from the empirical perspective. Firstly, it captures mean reversion which underpins pairs trading. However, according to Do et al (2006) the spread should be defined as the difference in logarithms of the prices:

$$\omega_t = \log(p_t^A) - \log(p_t^B) \tag{4.3}$$

Generally, the long term mean of the level difference in two stocks should not be constant, but widens as they increase and narrows as they decrease. The exception is when the stocks trade at similar price points. By defining the spread as log differences, this is no longer a problem. We have issues with both of these 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, whereby large deviations appear less pronounced. In effect, it gives the spread series the appearance of a 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 exhibit a long-run relationship (equivalently known as mean reversion) unless those stocks are cointegrated.

The second advantage offered by Elliot et al (2005) is that it is a continuous time model, and, as such, it is a convenient vehicle for forecasting purposes. Importantly, the trader 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 Ornstein-Uhlenbeck dynamics for which the Vasicek model is a special case, and one can easily compute the expectation $E[\mathrm{r}|x_t]$ where r denotes the first time the state variable crosses its mean θ, given its current position.

A 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 square error (MMSE). To facilitate the econometric estimation in a state space setting, one can represent equation (4.1) in a discrete time transition equation, motivated by the fact that the solution to (4.1) is Markovian:

$$x_k = E[x_k|x_{k-1}] + \varepsilon_k$$

k=1,2…, and ε is a random process with zero mean and variance equal to $\upsilon_k=VAR[x_k|x_{k-1}]$. Both conditional expectation and variance can be computed explicitly, and the above can be written as:

$$x_k = \theta(1-e^{-\kappa\Delta}) + e^{-\kappa\Delta}x_{k-1} + \varepsilon_k$$

where Δ denotes the time interval (in years) between two observations, and the variance of the random process ε happens to be a constant $\upsilon=\sigma^2/2\kappa(1-e^{-2\kappa\Delta})$. It also turns out that the conditional distribution of x_k is Gaussian. As the discrete time measurement equation becomes:

$$y_k=x_k + \omega_k$$

we now have a state space system that is linear and Gaussian in both transition and measurement equations, such that the Kalman filter recursive procedure provides optimal estimates of the parameters $\Psi=\{ \theta, \kappa, \sigma, \hbar\}$[2].

Despite the several advantages, this approach does have a 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.[3] 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

[2] For an introduction to the state space model and Kalman filter, see Durbin and Koopman (2001).
[3] Do, Faff and Hamza (2006) p.8 provide a proof of this.

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, with notable examples including Unilever 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 for 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 form being via American Depository Receipts (ADRs). Companies may also cross list within different exchanges within a country, such as the NASDAQ and NYSE in America[4].

3.3 The stochastic residual spread

Do, Faff and Hanmza (2006) propose a pairs trading strategy which differentiates itself from existing approaches by modeling 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 previous studies.

This approach begins with the assumption that there exists some equilibrium in the relative valuation of the two stocks measured by some spread. Mispricing is therefore construed as the state of disequilibrium which is quantified by a *residual spread function*

[4] See Badi and Tennant (2002) for more information on DLC's and cross listing.

$G(R_t^A, R_t^B, U_t)$ where U denotes some exogenous vector potentially present in formulating the equilibrium. The term "residual spread" emphasizes that the function captures any excess over and above some long term spread and may take non-zero values depending on the formulation of the spread. Market forces are assumed to play an important role in the process of mean-reversion of the spread in the long-run. Similar to previous studies, trading positions are opened once the disequilibrium is sufficiently large and the expected correction time is sufficiently short.

The proposed model adopts the same modeling and estimation framework as Elliot et al (2005). It utilises a one factor stochastic model to describe the state of mispricing or disequilibrium and to let noise contaminate its actual observation being measured by the above specification function G. To recap, let x be the state of mispricing, or residual spread, with respect to a given equilibrium relationship whose dynamic is governed by a Vasicek process:

$$dx_t = k(\theta - x_t)dt + \sigma dB_t \tag{4.4}$$

The observed mispricing is:

$$y_t = G_t = x_t + \omega_t \tag{4.5}$$

These two equations constitute a state space model of relative mispricing, defined with respect to some equilibrium relationship between two assets. The equilibrium relationship, or alternatively, the residual spread function G, is motivated by the Arbitrage Pricing Theory (APT) model (Ross, 1976). The APT model asserts that the return on a risky asset, over and above a risk free rate, should be the sum of risk premiums multiplied by their exposure. The specification of the risk factors is flexible, and may, for instance, take the form of the Fama-French 3-factor model:

$$R^i = R_f + \beta r^m + \eta^i$$

where $\beta = [\beta_1^i\beta_2^i\ldots\beta_n^i)$ and $r_m = [(R^1\text{-}r_f)(R^2\text{-}r_f)\ldots(R^n\text{-}r_f)]^T$, with R^i denoting 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. A "relative" APT on two stocks A and B can then be written as:

$$R^A = R^B + \Gamma r^m + e$$

where $\Gamma = [(\beta_1^A\text{-}\beta_1^B)\ (\beta_2^A\text{-}\beta_2^B)\ldots(\beta_n^A\text{-}\beta_n^B)]$ is a vector of exposure differentials and e is a residual noise term. In addition, it is assumed that the above relationship holds true in all time periods, such that:

$$R_t^A = R_t^B + \Gamma r_t^m + e_t$$

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

$$G_t = G(p_t^A, p_t^B, U_t) = R_t^A - R_t^B - \Gamma r_t^m \qquad (4.6)$$

If Γ is known (and r_t^m is specified), G_t is completely observable and a completely tractable model of mean-reverting relative pricing for two stocks A and B exists, which is then ready to be used for pairs trading. Similar to the Elliot et al (2005) formulation, this model may be reproduced in a state space form where the transition equation is represented by (4.4) and the measurement equation is represented by (4.5) where G_t is specified in equation (4.6). In a discrete time format, we have:

The transition equation:

$$x_k = \theta(1-e^{-\kappa\Delta}) + e^{-\kappa\Delta}x_{k-1} + \varepsilon_k \qquad (4.7)$$

The measurement equation:

$$y_k = x_k + H\omega_t \qquad (4.8)$$

This model can be reconciled to that model presented in Elliot et al (2005) when Γ is a zero vector. This state space model remains problematic with the observation function G_k being still unobserved as Γ is unknown. One may estimate Γ first using a standard linear regression with the dependent variable being (R^A-R^B) and the regressors being the excess return factors. The residual spread time series are then constructed using the calculated residuals from the regression. This time series becomes the observation for the above state space model.

An alternative solution is to redefine the observation $y=R^A-R^B$ such that the measurement equation is rewritten as:

$$y_k = x_k + \Gamma r_k^m + H\omega_k \tag{4.9}$$

This formulation allows the mispricing dynamics and the vector of exposure factor differentials Γ to be identified simultaneously by estimating the state space model, and helps avoid doubling up estimation errors from the two step procedure. Equation (4.7) and (4.9) constitute a model of stochastic residual spread for a pairs trading stratgey. This is a linear and Gaussian state space model, which can be estimated by Maximum Likelihood Estimation (MLE) where the likelihood function is of a *error prediction decomposition* form.[5]

To summarize, Do et al (2006) formulate a continuous time model of mean reversion in the relative pricing between two assets where the relative pricing model has been adopted from the APT model of single asset pricing. An econometric framework, similar to that proposed in Elliot et al (2005) has also been formulated to aid in the estimation process. It is important to note that this model does not make any assumptions regarding the validity of the APT model. Rather it adapts the factor structure of the APT to derive a relative pricing framework without requiring the validity of the APT to the fullest sense. Therefore, whereas a strict application of the APT may mean the long-run level of mispricing, or θ, should be close to zero, a non-zero estimate does not serve to invalidate the APT or the pairs trading model as a whole. Rather it may imply that there is a firm specific premium commanded by one company relative to

[5] See Durbin and Koopman (2001)

another, which could reflect such things such as managerial superiority. This could easily be incorporated into the model by simply adding or subtracting a constant term in the equilibrium function, G_t.

3.4 The Cointegration Approach

The cointegration approach outlined in Vidyamurthy (2004) is an attempt to parameterise pairs trading, by exploring the possibility of cointegration (Engle and Granger, 1987). Cointegration is a statistical relationship where two time series that are both integrated of same order *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 I(1) stock price series are combined to produce a stationary, or I(0), portfolio time series. This is desirable from the forecasting perspective, since regression of non-stationary variables results in *spurious regression*[6] (Lim and Martin, 1995). Cointegration incorporates mean reversion into a 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. Cointegrated time series can equivalently be represented in a Vector Error Correction model ("*Granger Representation Theory*") in which the dynamics of one time series is modeled as a function of its own lags, the lags of its cointegrated pair, and an error-correction component which corrects for deviations from the equilibrium relationship in the previous period. The significance of this is that forecasts can be made based on historical information.

To test for cointegration Vidyamurthy (2004) adopts the Engle and Granger's 2-step approach (Engle and Granger, 1987) in which the log price of stock A is first regressed against log price of stock B in what we refer to as the *cointegrating equation:*

$$\log(p_t^A) - \gamma \log(p_t^B) = \mu + \varepsilon_t \tag{4.10}$$

[6] Spurious regression arises from the static regression of non-stationary processes. There are two types of spurious regression: type 1 that involves falsely rejecting an existent relation, and type 2 that falsely accepts a non-existent relation. This study is concerned with avoiding type 2 spurious regressions which occur when variables are differenced to make the time series stationary. For more information refer to Chiarella et al (2008).

where γ is the "cointegrating coefficient" and the constant term μ captures some sense of "premium" in stock A versus stock B. The estimated residual series is then tested for stationarity using the augmented Dickey-Fuller test (ADF). Under this procedure, results are sensitive to the ordering of the variables. For example, if instead $log(p_t^{\boldsymbol{B}})$ is regressed against $log(p_t^{\boldsymbol{A}})$ then a different residual series will be estimated from the same sample. This issue can be resolved using the t-statistics from Engle and Yoo (1987).

Equation (4.10) says that a portfolio comprising long 1 unit of stock A and short γ units of stock B 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) process. Vidyamurthy (2004) develops trading strategies based on the assumed dynamics of the portfolio. The basic trading idea is to open a long position in the portfolio when it is sufficiently below its long-run equilibrium (μ-Δ) and similarly, short the portfolio when it is sufficiently above its long-run value (μ+Δ). Once the portfolio mean reverts to its long-run equilibrium value the position is closed and profit is earned equal to \$$\Delta$ per trade[7]. The key question when developing a trading strategy is what value of Δ is going to maximise the profit function[8]. Vidyamurthy (2004) presents both a parametric approach and a non-parametric empirical approach for conducting this analysis. The first approach models the residuals as an ARMA process and then uses Rice's formula (Rice, 1945) to calculate the rate of zero crossings and level crossings for different values of Δ in order to plot the profit function. The value Δ which maximises the profit function is chosen as the trading trigger. The alternative non-parametric approach constructs 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 estimation 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_t continue into the future.

[7] This is gross profit assuming no transaction costs, market impact costs or costs associated with illiquidity. Any proprietary implementation of this strategy must account for these costs.
[8] Profit function (π)=Δ*n where n=f(Δ) is the expected number of trading opportunities over a given period of time for a given Δ.

This approach appears to be favored by Vidyamurthy (2004) due to its simplicity and avoidance of model mis-specification.

Apart from being rather ad hoc, 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. To overcome these problems this study uses the more rigorous *Johansen test* for cointegration which is based on a Vector-Error-Correction model (VECM).

4. Data

The data used for this study comprises of the daily stock prices of 17 financial stocks listed on the Australian Stock Exchange (ASX). All of the 17 stocks are listed on the ASX200 which means that they are amongst the largest and most actively traded stocks in Australia. This feature is important for pairs trading since illiquidity on both the long and short side of the market is a fundamental risk when implementing this trading strategy. We believe that only searching for pairs from the most actively traded stocks on the ASX will ensure that we remain "price-takers".

Table 1: A list of financial stocks used in this study

Trading name (stock code)
Australia and New Zealand Bank (ANZ)
Westpac Banking Corporation (WBC)
Bank of Queensland (BOQ)
Lendlease Corporation (LLS)
Suncorp Metway (SUN)
National Australia Bank (NAB)
Perpetual (PPT)
QBE Group (QBE)
Commonwealth Bank of Australia (CBA)
St George Bank (SBG)
Bendigo Bank (BEN)
FKP Property Group (FKP)
Macquarie Group (MAC)
AXA Asia Pacific Holdings (AXA)
AMP Group (AMP)
Australian Stock Exchange (ASX)
Insurance Australia Group (IAG)

The sample itself consists of daily stock prices over a 5 year period starting from 1st July, 2002. The sample included Saturday and Sunday price observations which had to be removed from the sample prior to its analysis. Since the market is not open on the weekends, including price observations for both Saturday and Sunday could bias our results. Once the weekend price observations were removed from the sample we were left with 1285 daily price observations which could be used to identify our trading pairs.

There are several motivations for restricting our search for trading pairs to only stocks from within the same industry classification. Traditionally pairs trading has been seen as a market neutral strategy by construction. However unless the portfolio is actually constructed to have a zero beta it is likely that it will inhibit some market risk. Ideally we would like to immunize our combined portfolio against all systematic risk so that all returns generated by our positions are those that arise from convergence of the residual spread. We do not want to earn profits from holding long positions in a bull market since this would defeat the purpose of developing a trading strategy that was not conditioned upon the absolute value of the stocks traded. By restricting our trading pairs to stocks from within the same industry we assume that it is likely that those stocks will have similar exposures to systematic risk, or beta. Thus, the resulting portfolio should have a beta close to zero. Ideally we would choose stocks with the same betas so that the combined portfolio had a beta of exactly zero, but because we are working with a limited sample the current constraint will suffice.

The second motivation for choosing stocks belonging to the same industry classification stems from an attempt to align those observed statistical relationships with some theoretical reasoning. Although cointegration does indeed offer some very attractive properties with which to develop a trading strategy, it is necessary to understand what is driving the fundamental relationship between the trading pairs. Stocks that are cointegrated must be driven by the same underlying factors so that they share a long-run equilibrium relationship. It is more likely that stocks within the same industry classification will be driven by the same fundamental factors than two stocks from different industries. For example, consider the stock price reactions of two banks to the news of a removal of import tariffs in the automotive industry, as opposed to the reactions of a bank and a local car manufacturer. It could reasonably be expected that the local car manufacturers' share price would decrease, especially if it was widely known that it was relying on those protectionist policies to remain feasible. It is likely that the share p[rice of the bank would be relatively immune from the removal of any protectionist policies relating to the automotive industry. For this reason we require our trading pairs to be from the same industry.

A cointegrating relationship which can be explained by some theoretical reasoning, such as that described above, is more robust than a cointegrating phenomena without sound justification. If we identify a cointegrating relationship in-sample and know that both stocks are driven by the same set of fundamental factors, then we could reasonably argue that the observed equilibrium relationship will likely persist into the future – i.e. the cointegrating relationship would likely remain significant out-of-sample. However, if we were to randomly identify a cointegrating relationship between, say, a financial stock and a resource stock, then this relationship would be more statistical phenomena as opposed to a fundamental relationship deriving from both stocks being driven by a unique set of factors. Thus, there is no economic justification to suggest this statistical relationship will remain significant out-of-sample. For these reasons we require our trading pairs to belong to the same industry classification.

Pairs trading is best suited to bear markets, characterized by uncertain fundamental values and high volatility – who would want to remain market neutral in a bull market? If one considers the trading history of the financial services sector over the past 18 months it becomes clear why this study has chosen financial stocks to identify trading pairs. Market uncertainty surrounding the valuation of intricate, derivative laden mortgage-backed securities has resulted in the demise of several banks and the pummeling of many more. Combine this with the woes of the credit-lending business and it has created an environment which has made it very difficult to accurately value financial stocks. Consequently, the financial services sector has seen large fluctuations in recent times – a perfect recipe for a profitable pairs trading strategy.

The data was sourced from the ASX.

5. Method

The method used in this study can be divided into two broad sections. The first section is concerned with the process of selecting the trading pairs. A detailed discussion is conducted which is concerned with the idea of cointegration and the various tests we use to identify cointegrating relationships between time series variables. We also introduce our modeling procedure, a Vector Error Correction model (VECM) and illustrate how in the presence of cointegration these are essentially equivalent representations (Granger Representation Theorem). We conclude this section by discussing the method we use to obtain the residual spread which we then model as a mean reverting VECM.

The second component of this study is concerned with the identification and estimation of our model and the how we overcome the problem of *overparameterization* in our VECM. We conclude by estimating impulse response functions and variance decomposition analysis which we hope will shed light on the dynamic behaviour and inter-relationships between our stock pairs.

5.1 The process of selecting trading pairs – an introduction to cointegration and the associated VECM.

The first and arguably most important decision within a pairs trading strategy is which stocks to trade. The fundamental property in any long/short trading strategy is the presence of a statistically significant mean reverting relationship between the assets traded. This study uses cointegration as the decision rule for selecting pairs of stocks. Cointegration was first attributed to the work of Engle and Granger (1987) for which helped earn them the Nobel Prize (2003) for statistics. 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. It is the statistical properties that cointegration offers which make it such an attractive possibility across a range of applications for academics and practitioners alike. Let us now introduce cointegration and illustrate why we employ it in this study.

Consider a set of economic variables in long-run equilibrium when

$$\beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_n X_{nt} = 0$$

For notational simplicity, an identical long-run equilibrium can be represented in matrix form as

$$\tilde{\beta}\tilde{X}_t = \tilde{0}$$

Where:

$\tilde{\beta}$=(β_1, β_2, β_3,…, β_n)

$\tilde{X}_t'$=(X_{1t}, X_{2t}, …, X_{nt})

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

$$\tilde{e}_t = \tilde{\beta}\tilde{X}_t$$

The equilibrium is only meaningful if the residual series ($\tilde{e}_t$) is stationary.

5.1.1 What is stationarity?

A time series $\{y_t\}$ is a stationary series if its mean, variance and autocorrelations are well approximated by sufficiently long time averages based on a single set of realisations. When this study refers to stationarity it can be interpreted as that the time series is *covariance stationary*. Covariance stationary means that for a given time series, its mean, variances and autocovariances are unaffected by a change in time origin. This can be summarised by the following conditions:

$$E(y_t) = E(y_{t-s}) = \mu$$
$$E(y_t - \mu)^2 = E(y_{t-s} - \mu)^2 = \sigma_y^2$$
$$E(y_t - \mu)(y_{t-s} - \mu) = E\big(y_{t-j} - \mu\big)\big(y_{t-j-s} - \mu\big) = \delta_s$$

$$Where\ \mu, \sigma_y^2\ and\ \delta_s\ are\ all\ constants.$$

5.1.2 A definition of cointegration

We say that components of the vector $\tilde{X}_t$ are cointegrated of order d,b, which is denoted by $\tilde{X}_t \sim CI(d,b)$ if:

1. All components of $\tilde{X}_t$ are integrated of order d[9].

[9] A series is integrated of order d if it must be differenced d times in order to become stationary. Many macroeconomic and stock price series are integrated of order 1, or I(1). A stationary series is by definition an I(0) process.

2. There exists a vector $\tilde{\beta}$ such that the linear combination

$$\tilde{\beta}\tilde{X}_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt}$$

is integrated of order (d-b), where b>0 and $\tilde{\beta}$ is the cointegrating vector (CV).

Let us now consider a few important aspects of cointegration

1. Cointegration refers to a linear combination of non-stationary variables. The CV is not unique. For example, if $(\beta_1, \beta_2, \ldots, \beta_n)$ is a CV, then for non-zero λ, $(\lambda\beta_1, \lambda\beta_2, \ldots, \lambda\beta_n)$ is also a CV. Typically the CV is normalised with respect to x_{1t} by selecting $\lambda = 1/\beta_n$.
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 – this condition does not imply that all similarly integrated variables are cointegrated, in fact it is usually not the case.
3. If the vector $\tilde{X}_t$ has n components, there may be as many as (n-1) linearly independent cointegrating vectors. For example, if n=2 then there can be at most one independent CV.

To allow us to prove (3) consider the simple vector auto-regression (VAR) model

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$$

We have limited the lag length to just one for simplicity, but in practice lag length should be set so that the error terms are white noise processes. It is important to note that correlation between these error terms is allowable.

Applying lag operators and rearranging we get

$$(1 - a_{11}L)y_t - a_{21}Lz_t \ = \ \varepsilon_{yt}$$
$$-a_{21}Ly_t \ + \ (1 - a_{22}L)z_t \ = \ \varepsilon_{zt}$$

which can be represented in matrix form as

$$\begin{bmatrix} (1 - a_{11}L) & -a_{12} \\ -a_{21} & (1 - a_{22}L) \end{bmatrix} \begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

Using Cramer's rule or matrix inversion

$$y_t \ = \ \frac{[(1 - a_{22}L)\varepsilon_{yt} \ + \ a_{12}L\varepsilon_{zt}]}{[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L_2]}$$

$$z_t \ = \ \frac{[a_{21}L\varepsilon_{yt} \ + \ (1 - a_{11}L)\varepsilon_{zt}]}{[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L_2]}$$

The two-variable first-order system has been converted into two univariate second order difference equations, where both have the same inverse characteristic equations. That is, setting

$$(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L_2 \ = \ 0 \ and \ \lambda = {}^{1}\!/_{L}$$

Then this implies that

$$\lambda_2 - (a_{11}a_{22})\lambda \ + \ (a_{11}a_{22} - a_{12}a_{21}) \ = \ 0$$

whereby the characteristic roots λ_1 and λ_2 determine the time paths of both variables.

1. If (λ_1, λ_2) lie inside the unit circle then stable solutions for the series {y_t} and {z_t} exist and the variables are stationary. An important implication of this is that they cannot not be cointegrated of order (1,1).
2. If either root lies outside the unit circle then the solutions are explosive:

$$y_t = \frac{[(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]}{[(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L_2]}$$

$$= \frac{[(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]}{(1 - \lambda_1 L)(1 - \lambda_2 L)}$$

Roots of the characteristic equation

$$(1 - \lambda_1 L)(1 - \lambda_2 L) = 0 \; are \; \lambda_1 \; and \; \lambda_2$$

If $|\lambda_i| < 1$, i=1,2 then the solution is stable, however if at least one $|\lambda_i| > 1$, for i=1,2, or if either root lies outside the unit circle then the system is explosive. Neither variable is difference stationary which implies that the *variables cannot be cointegrated of order (1,1).*

3. If $a_{12} = a_{21} = 0$, then

$$y_t = a_{11}y_{t-1} + \varepsilon_{yt}$$
$$z_t = a_{22}z_{t-1} + \varepsilon_{zt}$$

and the solution is trivial. If $a_{11}=a_{22}=1$, then {y_t} and {z_t} are unit root processes. This implies that $\lambda_1 = \lambda_2 = 1$ and

$$y_t = y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$

That is, the two variables evolve without any long-run equilibrium relationship.

4. For $\{y_t\}$ and $\{z_t\}$ to be CI(1,1) it is necessary for one characteristic root to be unity and the other less than unity in absolute value. In this case, each variable will have the same stochastic trend and the first difference of each variable will be stationary. For example, If λ_i=1 then;

$$y_1 = \frac{[(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]}{[(1 - L)(1 - \lambda_2 L)]}$$

$$(1 - \lambda)y_1 = \Delta y_t = \frac{[(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}]}{(1 - \lambda_1 L)}$$

which is stationary if $|\lambda_2| < 1$.

Thus, to ensure variables are CI(1,1), we must set one of the characteristic roots equal to unity and find the other less than unity in absolute value.

For the larger of the two roots to be unity it must be the case that

$$0.5(a_{11} + a_{22}) + 0.5\surd(a_{112} + a_{222} - 2a_{11}a_{22} + 4a_{12}a_{21}) = 1$$

$$\rightarrow a_{11} = \frac{[(1 - a_{22}) - a_{12}a_{21}]}{(1 - a_{22})}$$

Now consider second characteristic root – since a_{12} and/or a_{21} must differ from zero if the variables are cointegrated,

$$|\lambda_2| < \rightarrow a_{22} > -1 \; and \; a_{12}a_{21} + a_{222} < 1$$

Let us see how these coefficient restrictions bear on the nature of the solution. Recall the simple VAR model

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$$

Taking differences

$$y_t - y_{t-1} = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t - z_{t-1} = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}$$

Which can be expressed in matrix form as

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{bmatrix} (a_{11} - 1) & a_{12} \\ a_{21} & (a_{22} - 1) \end{bmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

Now $a_{11}-1 = \frac{-a_{12}a_{21}}{(1-a_{22})}$

therefore

$$\Delta y_t = [-a_{12}a_{21}/(1 - a_{22})]y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$

$$\Delta z_t = a_{21}y_{t-1} - (1 - a_{22})z_{t-1} + \varepsilon_{zt}$$

If $a_{12}\neq 0$ and $a_{21}\neq 0$, we can normalise the cointegrating vector with respect to either variable. Therefore normalising with respect to y_{t-1}

$$\Delta y_t = \alpha_y(y_{t-1} - \beta z_{t-1}) + \varepsilon_{yt}$$
$$\Delta z_t = \alpha_z(y_{t-1} - \beta z_{t-1}) + \varepsilon_{zt}$$

where

$$\alpha_y = \frac{-a_{12}a_{21}}{(1 - a_{22})}$$

$$\alpha_z = a_{21}$$

$$\beta = \frac{(1 - a_{22})}{a_{21}}$$

This is referred to as an Error-Correction model (ECM) and $\{y_t\}$ and $\{z_t\}$ change in response to the previous periods deviation from long-run equilibrium: $y_{t-1} - \beta z_{t-1}$.

If $y_{t-1} = \beta z_{t-1}$ then both $\{y_t\}$ and $\{z_t\}$ change only in response to ε_{yt} and ε_{zt} shocks.
If $\alpha_y<0$ and $\alpha_x>0$, then $\{y_t\}$ decreases and z_t increases in response to a positive deviation from the long run equilibrium.

The conditions $a_{22}>-1$ and $a_{12}a_{21} + a_{222}<1$ ensure that $\beta\neq0$. At least one of the speed of adjustment parameters (i.e. α_y and α_z) must be significantly different from zero.

A special case arises when either a_{12} or a_{21} equals zero. To illustrate, set $a_{12}=0$ so that $\alpha_y=0$. When this occurs the error correction component drops out completely and $\{y_t\}$ changes only in response to shocks to the system (ε_{yt}) since $\Delta y_t= \varepsilon_{yt}$. The sequence $\{z_t\}$ does all the correction to eliminate any deviation from long-run equilibrium.

We will now consider some important implications of this simple model:

1. *The restrictions necessary to ensure that the variables are CI(1,1) guarantee that an error correction model exists.*

The individual series $\{y_t\}$ and $\{z_t\}$ are unit root processes, but the linear combination $y_t - \beta z_t$ is stationary, with the normalised CV given by $[1, \frac{-(1-a_{22})}{a_{21}}]$. The variables have an error-correction representation with speed of adjustment coefficients given by $\alpha_y = (-a_{12}a_{21})/(1-a_{22})$, and $\alpha_z = a_{21}$

We have also shown that an error-correction model for I(1) variables necessarily implies cointegration. This illustrates the *Granger Representation Theorem* which states that for any set of I(1) variables, error-correction and cointegration are equivalent representations.

2. *Cointegration necessitates coefficient restrictions in a VAR model.*

Recall that the simple VAR model

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$

$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$$

can be written in matrix form as:

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{bmatrix} (a_{11} - 1) & a_{12} \\ a_{21} & (a_{22} - 1) \end{bmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

or

$$\Delta\tilde{X}_t = \tilde{\Pi}\tilde{X}_{t-1} + \tilde{\varepsilon}_t \qquad (5.1)$$

It is appropriate to estimate a VAR of cointegrated variables using only first differences. Estimating equation (5.1) without $\tilde{\Pi}\tilde{X}_{t-1}$ eliminates the error-correction component from the model.

Also, the rows of $\tilde{\Pi}$ are **not** linearly independent if the variables are cointegrated – multiplying row 1 by $\frac{-(1-a_{22})}{a_{21}}$ yields the corresponding element in row 2. This illustrates the very important insights of Johansen (1988) and Stock and Watson (1988) which says that the rank of Π can be used to determine whether or not two variables $\{y_t\}$ and $\{z_t\}$ are cointegrated.

3. *It is necessary to reinterpret Granger causality in a cointegrated system.*

In the simple two variable case - $\{y_t\}$ and $\{z_t\}$, Granger Causality determines how much of the current value of y_t can be explained by past values of y_t, and whether lagged values of z_t improve the explanation. To illustrate, take the simple VAR in differences

$$\Delta y_t = b_{11}\Delta y_{t-1} + b_{12}\Delta z_{t-1} + \varepsilon_{yt}^{'}$$

$$\Delta z_t = b_{21}\Delta y_{t-1} + b_{22}\Delta z_{t-1} + \varepsilon_{zt}^{'}$$

New interpretation: $\{z_t\}$ *does not Granger Cause* $\{y_t\}$ if lagged values of Δz_{t-1} do not enter the Δy_t equation, i.e. $b_{12}=0$. Similarly, $\{y_t\}$ *does not Granger cause* $\{z_t\}$ if lagged values of Δy_{t-1} do not enter the Δz_t equation, i.e. $b_{21}=0$.

Now, if $\{y_t\}$ and $\{z_t\}$ are cointegrated, by the Granger Representation Theorem the VAR model becomes the following Vector-Error-Correction (VECM) model.

$$\Delta y_t = \alpha_y(z_{t-1} - \beta y_{t-1}) + c_{11}\Delta y_{t-1} + c_{12}\Delta z_{t-1} + \varepsilon_{yt}$$

$$\Delta z_t = \alpha_z(z_{t-1} - \beta y_{t-1}) + c_{21}\Delta y_{t-1} + c_{22}\Delta z_{t-1} + \varepsilon_{zt}$$

5.1.3 Testing for Cointegration: The Engle-Granger Methodology

Engle and Granger (1987) propose a straightforward test to determine whether two given time series, say $\{y_t\}$ and $\{z_t\}$ and cointegrated of order CI(1,1). Unfortunately, this approach suffers from several drawbacks which will be discussed in more detail in section 4.1.4. As a result, this study utilises the alternative Johansen test (1988) which overcomes these issues. Nevertheless, in order to appreciate the benefits of the Johansen approach it is critical to understand the method employed by Engle and Granger.

1. The first step is to test each series individually for their order of integration. The augmented Dickey-Fuller test can be used to test for the presence of a unit-root

and hence whether a series is stationary. If the individual time series are integrated of different orders then it can be concluded with certainty that they are not cointegrated. A cointegrating relationship may be only present between variables integrated of the same order.

2. The second step is to estimate the long-run equilibrium relationship between the time series. If $\{y_t\}$ and $\{z_t\}$ are both I(1) processes then the long-run relationship takes the form

$$y_t = \beta_0 + \beta_1 z_t + e_t$$

If the variables are in fact cointegrated then OLS regression yields "super-consistent" estimates of the cointegrating parameters β_0 and β_1 (CV). It has been shown by Stock (1987) that OLS coefficient estimates converge faster towards their parameter values in the presence of a cointegrating relationship compared with regressions involving stationary variables. If deviations from the long-run equilibrium $\{e_t\}$ are found to be stationary, I(0), then the $\{y_t\}$ and $\{z_t\}$ sequences are cointegrated of order (1,1). The augmented Dickey-Fuller test can be used to determine the stationarity of the residual series $\{e_t\}$.

The Engle and Granger approach is relatively straight forward and easily implemented in practice. For this reason it remains useful as a secondary source of evidence for the presence of any cointegrating relationships between time series. However this study acknowledges and appreciates the significant drawbacks of the Engle and Granger approach and consequently employs an alternative method to detect the presence of cointegration. The main drawbacks can be summarised as follows:

1. *The Engle and Granger test for cointegration uses residuals from either of the two "equilibrium" equations.*

$$y_t = \beta_{10} + \beta_{11} z_t + e_{1t}$$
$$z_t = \beta_{20} + \beta_{21} y_t + e_{2t}$$

As the sample size increase indefinitely, asymptotically a test for a unit root in $\{e_{1t}\}$ is equivalent to that for $\{e_{2t}\}$. This is not applicable to smaller sample sizes.

2. *The major problem regarding the Engle and Granger procedure is that it relies on a two-step estimator.*

1st step: Generate the residual series from one of the equilibrium equations $\{\hat{e}_t\}$
2nd step: Use generated errors to estimate a regression of the following form;

$$\Delta\hat{e}_t = b_1\Delta\hat{e}_{t-1} + b_2\Delta\hat{e}_{t-2} + \ldots . b_n\Delta\hat{e}_{t-n} + u_t \qquad (ADF\ equation)$$

The coefficient b_1 is obtained by regressing the residuals from another regression on lagged differences of itself. This two-stage test means that any errors introduced in the first step are now carried forward to the second step. Essentially, this approach is subject to twice the estimation error.

5.1.4 An alternative approach: The Johansen test for Cointegration (1988)

This study uses the Johansen test (1988) to identify cointegrating relationships between stock price series. The Johansen and Stock & Watson maximum likelihood estimators circumvent the use of a two-step estimator and in doing so avoid the drawbacks faced by Engle and Granger. Instead, the Johansen (1988) procedure relies heavily on the relationship between the rank of a matrix and its characteristic roots. One intuitive explanation suggests that the Johansen procedure is nothing more than a multivariate generalisation of the Dickey-Fuller test. We will now illustrate the reasoning behind the Johansen approach:

Consider the univariate case:

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

$$or\ \Delta y_t = (a_1 - 1)y_{t-1} + \varepsilon_t$$

If $(a_1-1)=0$ then we conclude that $\{y_t\}$ has a unit root which we know to be a non-stationary process. If $(a_1-1)\neq 0$ then $\{y_t\}$ is a stationary process.

Generalizing to the two variable case:

$$\begin{Bmatrix} y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \\ \\ z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt} \end{Bmatrix} \tag{5.2}$$

$$or\ \Delta\tilde{X}_t = \tilde{\prod}\tilde{X}_{t-1} + \tilde{\varepsilon}_t$$

Where $\Delta\tilde{X}_t = \begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix}$ $\quad X_{t-1} = \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix}$ $\quad \tilde{\prod} = \begin{bmatrix} (a_{11}-1) & a_{12} \\ a_{21} & (1-a_{22}) \end{bmatrix}$

Recall our previous statement that the rows of $\tilde{\prod}$ are not linearly independent if the variables are cointegrated. Let us now explore this statement a little further assuming the $\tilde{\prod}$ matrix has two rows.

I. If the $\tilde{\prod}$ matrix is of full rank, then its rank is equal to the number of linearly independent rows and that should equal the number of rows (i.e. 2).

   Suppose $\tilde{\prod}$ is of full rank. Then the long run solution to (5.2)

$$\begin{bmatrix} y_t = y_{t-1} = \cdots = \tilde{y} \\ z_t = z_{t-1} = \cdots = \tilde{z} \\ \varepsilon_{yt} = \varepsilon_{yt-1} = \cdots = 0 \\ \varepsilon_{zt} = \varepsilon_{zt-1} = \cdots = 0 \end{bmatrix}$$

   is given by two independent equations

$$(1 - a_{11})\tilde{y} + a_{12}\tilde{z} = 0$$
$$a_{21}\tilde{y} + (1 - a_{22})\tilde{z} = 0$$

Or alternatively,

$$\Pi_{11}\tilde{y} + \Pi_{12}\tilde{z} = 0$$
$$\Pi_{21}\tilde{y} + \Pi_{22}\tilde{z} = 0$$

Each of the above equations is an independent restriction on the long-run solution to the variables:

- The two variables in the system face two long-run constraints
- The two variables contained in $\tilde{X}_t$ must be stationary with long-run values given by the above independent equations.

II. If the rank of $\tilde{\prod}$ is zero, then the elements of $\tilde{\prod}$ must be zero. This implies that

1) $\Delta\tilde{X}_t = \tilde{e}_t$
2) Δy_t and Δz_t are both now I(0) processes

(4.1) becomes:

$$y_t = y_{t-1} + \varepsilon_{yt}(a_{11} - 1) = 0 \rightarrow a_{11} = 1$$
$$z_t = z_{t-1} + \varepsilon_{zt}(a_{22} - 1) = 0 \rightarrow a_{22} = 1$$
$$a_{12} = a_{21} = 0$$

and $\{y_t\}$ and $\{z_t\}$ are both unit root processes with no linear combination that is stationary.

III. If the rank of $\tilde{\prod}$ is r where r=1, then the rows of $\tilde{\prod}$ are not linearly independent. There is a single CV given by any row of the matrix $\tilde{\prod}$. Each $\{x_{it}\}$ sequence can be written in error-correction form. For example,

$$\Delta y_t = n_{11}y_{t-1} + n_{12}z_{t-1} + \varepsilon_{yt}$$

Normalising with respect to y_{t-1}, we can set $\alpha_1 = n_{11}$ and $\beta_{12} = \frac{n_{12}}{n_{11}}$ to obtain:

$$\Delta y_t = \alpha_1(y_{t-1} + \beta_{12}z_{t-1}) + \varepsilon_{yt}$$

In the long run, the series $\{x_{it}\}$ will satisfy the relationship

$$y_{t-1} + \beta_{12}z_{t-1} = 0$$

The normalised CV is given by $(1, \beta_{12})$ and the speed of adjustment parameter by α_1.

A slight but important digression

As with the augmented Dickey-Fuller test, the multivariate model given by (5.1) can be generalized to allow for higher-order autoregressive processes. Consider the generalization of (5.1) incorporating three lags

$$y_t = a_{11}y_{t-1} + b_{11}y_{t-2} + c_{11}y_{t-3} + a_{12}z_{t-1} + b_{12}z_{t-2} + c_{12}z_{t-3} + \varepsilon_{yt}$$
$$z_t = a_{21}y_{t-1} + b_{21}y_{t-2} + c_{21}y_{t-3} + a_{22}z_{t-1} + b_{22}z_{t-2} + c_{22}z_{t-3} + \varepsilon_{zt}$$

$$\begin{pmatrix} y_t \\ zt \end{pmatrix} = \begin{bmatrix} a_{11} & b_{11} & c_{11} \\ a_{21} & b_{21} & c_{21} \end{bmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{pmatrix} + \begin{bmatrix} a_{12} & b_{12} & c_{12} \\ a_{22} & b_{22} & c_{22} \end{bmatrix} \begin{pmatrix} z_{t-1} \\ z_{t-2} \\ z_{t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}$$

$$\tilde{X}_t = \tilde{A}_1\tilde{X}_{t-1} + \tilde{A}_2\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} + \tilde{\varepsilon}_t$$
$$\Delta\tilde{X}_t = (\tilde{A}_1 - \tilde{I})\,\tilde{X}_{t-1} + \tilde{A}_2\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} + \tilde{\varepsilon}_t$$

Add and subtract $(\tilde{A}_1 - \tilde{I})\tilde{X}_{t-2}$:

$$\Delta\tilde{X}_t = (\tilde{A}_1 - \tilde{I})\tilde{X}_{t-1} + \tilde{A}_2\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} - (\tilde{A}_1 - \tilde{I})\tilde{X}_{t-2} + (\tilde{A}_1 - \tilde{I})\tilde{X}_{t-2} + \tilde{\varepsilon}_t$$

$$= (\tilde{A}_1 - \tilde{I})\Delta\tilde{X}_{t-1} + (\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} + \tilde{\varepsilon}_t$$

Add and subtract $(\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\tilde{X}_{t-3}$:

$$\Delta\tilde{X}_t = (\tilde{A}_1 - \tilde{I})\Delta\tilde{X}_{t-1} + (\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} - (\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\tilde{X}_{t-3} + (\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\tilde{X}_{t-3} + \tilde{\varepsilon}_t$$

$$\Delta\tilde{X}_t = (\tilde{A}_1 - \tilde{I})\Delta\tilde{X}_{t-1} + (\tilde{A}_2 + \tilde{A}_1 - \tilde{I})\Delta\tilde{X}_{t-2} + \tilde{A}_3\tilde{X}_{t-3} + \tilde{\varepsilon}_t$$

Continuing in this fashion for p=3:

$$\Delta\tilde{X}_t = \sum_{i=1}^{p-1} \tilde{n}_i \, \Delta\tilde{X}_{t-1} + \tilde{n}\tilde{X}_{t-p} + \tilde{\varepsilon}_t$$

$$Where\ \tilde{n}_i = -(\tilde{I} - \sum_{j=1}^{i} \tilde{A}_j)\ and\ \tilde{n} = -(\tilde{I} - \sum_{j=1}^{p} \tilde{A}_i)$$

The number of distinct cointegrating vectors can be obtained by checking the significance of the characteristic roots of $\tilde{\prod}$. The rank of a matrix is equal to the number of characteristic roots that differ from zero. Suppose we order the n characteristic roots of $\tilde{\prod}$ such that $\lambda_1>\lambda_2>\ldots>\lambda_n$. If the variables are not cointegrated, the rank of $\tilde{\prod}$ is zero as are the characteristic roots. Since ln(1)=0, each of $\ln(1-\lambda_i)=0$ if the variables are not cointegrated. Similarly, if the rank of $\tilde{\prod}$ is unity, then $0<\lambda_1<1$ so that the first expression $\ln(1-\lambda_1)$ will be negative and all other $\lambda_i=0$ so that $\ln(1-\lambda_2)=\ln(1-\lambda_3)=\ldots=\ln(1-\lambda_n)=0$.

In practice, we can only obtain estimates of $\tilde{\prod}$ and the characteristic roots. The test for the number of characteristic roots that are insignificantly different from unity can be conducted using the following two test statistics:

$$\lambda trace(r) = -T\sum_{i=r+1}^{n} \ln(1-\hat{\lambda}_i)$$

$$\lambda max(r, r+1) = -T ln(1-\hat{\lambda}_{r+1})$$

Where $\hat{\lambda}_i$=estimator of characteristic roots (eigenvalues from $\tilde{\prod}$ matrix)

T= number of usable observations

$\lambda trace(r)$ tests the null hypothesis that the number of distinct cointegrating vectors is less than or equal to r against a general alternative.

$\lambda trace(r) = 0$ when all λ_i=0

The further the estimated characteristic roots are from zero, the more negative is ln(1-λ_i) and the larger the λtrace statistic.

Johansen and Juselus (1990) provide critical values of the $\lambda trace$ statistic which was obtained from simulation studies.

5.1.5 Obtaining the residual spread

Once we have identified two stock price series which are cointegrated the next step is to obtain the residual series {ε_t} from the cointegrating equation. Let us suppose that y_t and z_t are found to be cointegrated and we know that z_t leads y_t from our Granger Causality testing. This leads us to estimate the following cointegrating equation

$$y_t = \beta_0 + \beta_1 z_t + \varepsilon_t$$

which can be rearranged to give

$$\varepsilon_t = y_t - \beta_0 - \beta_1 z_t$$

where β_0 is an intercept term and β_1 is referred to as the "cointegrating coefficient". Recall that for the cointegrating relationship to be meaningful the residual series $\{\varepsilon_t\}$ must now be a stationary, or an I(0) process. It can also be helpful when interpreting ε_t by realising that it shares identical dynamics to the underlying long/short portfolio which is represented by $\{y_t - \beta_0 - \beta_1 z_t\}$. The mean-reverting long/short portfolio is created by longing 1 unit of y_t for every β_1 units short of z_t. The portfolio fluctuates around its long-run equilibrium value of β_0.
For completeness if not necessitation we then subjected the estimated residual series $\{\varepsilon_t\}$ to the augmented Dickey-Fuller test to clarify its stationarity.

We are now in a position where we can calibrate the VECM to fit the residual series. We hope that this will provide us with a valuable insight into the finer dynamics of the residual series and expose some of the complex inter-relationships between our stock pairs.

5.2 Modeling procedure and variance analysis

5.2.1 Deriving a usable VAR model

When we are not confident that a variable is actually exogenous, a natural extension of transfer function analysis is to treat each variable symmetrically. In our pairs trading strategy which involves two variables, we can let the time path of Δy_t be a function of lagged differences of y_t, combined with current and past realizations of the Δz_t sequence. The dynamics of the Δz_t sequence is simply a mirror image of that described for the Δy_t sequence. This can be represented by the system shown below, where lags are set at unity for simplicity:

$$\Delta y_t = b_{10} - b_{12}\Delta z_t + \gamma_{11}\Delta y_{t-1} + \gamma_{12}\Delta z_{t-1} + \varepsilon_{yt} \tag{5.3a}$$

$$\Delta z_t = b_{20} - b_{21}\Delta y_t + \gamma_{21}\Delta y_{t-1} + \gamma_{22}\Delta z_{t-1} + \varepsilon_{zt} \tag{5.3b}$$

Note here that we choose to model differences rather than levels to avoid spurious regressions. Spurious regressions occur when one attempts to model non-stationary variables. Since our stock price series are known to be integrated of order one, taking first differences yields a stationary sequence which we can then accurately fit to a given model.

This bivariate VAR also assumes that the error terms (ε_{yt} and ε_{zt}) are white noise processes with standard deviations given by σ_y and σ_z respectively. They are assumed to be uncorrelated with each other.

Equations (5.3a) and (5.3b) constitute a *first-order* VAR since the lag length is set at unity. We will proceed to use this simple model to illustrate some of the issues we face when estimating our model. It should be remembered that we actually estimate a VECM in this study which incorporates an error-correction component to account for the presence of a cointegrating relationship. However for simplicity, this simple VAR will suffice to aid an explanation of our ideas.

To begin with, it is important to note that the structure of the system incorporates feedback since Δy_t and Δz_t are allowed to affect each other. For example, $-b_{12}$ is the contemporaneous effect of a unit change in Δz_t on Δy_t and γ_{21} the effect of a unit change in Δy_{t-1} on Δz_t. The error terms ε_{yt} and ε_{zt} are pure innovations (or shocks) in the Δy_t and z_t sequences, respectively. Furthermore, it is evident that when b_{21} is significantly different from zero, ε_{yt} has an indirect contemporaneous effect on Δz_t, and if b_{12} is significantly different from zero then ε_{zt} also has an indirect contemporaneous effect on Δy_t. It could very well be plausible that such a system could be used to describe the stock price dynamics in this study.

Equations (5.3a) and (5.3b) are not reduced-form equations since Δy_t has a contemporaneous effect on Δz_t and Δz_t has a contemporaneous effect on Δy_t. There is a certain simultaneity present which inhibits the direct estimation of the bivariate VAR in its structural form. Fortunately, it is possible to transform the system of equations into a more usable form. The present system can be represented in matrix form as:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

or

$$BX_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t$$

where

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, X_t = \begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Premultiplication by the inverse of B (B^{-1}) allows us to obtain the vector autoregressive (VAR) model in *standard* form:

$$x_t = A_0 + A_1 x_{t-1} + e_t \qquad (5.4)$$

Where

$A_0 = B^{-1}\Gamma_0$

$A_1 = B^{-1}\Gamma_1$

$e_t = B^{-1}\varepsilon_t$

For notational purposes, we can define a_{io} as element i of the vector A_o, a_{ij} as the element in row i and column j of the matrix A_1, and e_{it} as the element i of the vector e_t. Using this new notation, we can rewrite (4.2.3) in expanded format as:

$$\Delta y_t = a_{10} + a_{11}\Delta y_{t-1} + a_{12}\Delta z_{t-1} + e_{1t} \qquad (5.5a)$$

$$\Delta z_t = a_{20} + a_{21}\Delta y_{t-1} + a_{22}\Delta z_{t-1} + e_{2t} \qquad (5.5b)$$

To distinguish between the different representations we will refer to the initial system (5.3a/b) as a *structural* VAR of the *primitive system*, while the new system derived above will be called a VAR in *standard form*. It should be noted that that the error terms (i.e. e_{1t} and e_{2t}) in the standard VAR are composites of the two shocks (ε_{yt} and ε_{zt}) from the primitive system. Since $e_t = B^{-1}\varepsilon_t$, we can compute e_{1t} and e_{2t} as:

$$e_{1t} = (\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12}b_{21}) \qquad (5.6)$$

$$e_{2t} = (\varepsilon_{zt} - b_{21}\varepsilon_{yt})/(1 - b_{12}b_{21}) \quad (5.7)$$

Since the new error terms are simply linear combinations of those error terms from the structural VAR then it will be the case that they share similar statistical properties. Given that the residuals from the primitive system are white noise processes, it follows that both e_{1t} and e_{2t} have zero means, constant variances, and are individually serially uncorrelated[10]. It is critical to acknowledge however that although e_{1t} and e_{2t} are both stationary processes they will also be correlated. The covariance of the two terms is

$$Ee_{1t}e_{2t} = \frac{E\left[(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{zt} - b_{21}\varepsilon_{yt})\right]}{(1 - b_{12}b_{21})^2}$$

$$= \frac{-(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)}{(1 - b_{12}b_{21})^2}$$

In general, it can be expected that this covariance term will not be zero, so that the two shocks will be correlated. In the special case where $b_{12} = b_{21} = 0$ (i.e. if there are no contemporaneous effects of Δy_t on Δz_t and Δz_t on Δy_t), the shocks will be uncorrelated.

5.2.2 The problem of identification

In this section we discuss some of the issues faced when we come to estimating our VAR as well as the statistical techniques that we can use to overcome these problems. For simplicity, we will use the structural form of the bivariate VAR equations (5.3a/b) introduced earlier to aid this discussion. Recall the primitive system is the model we are essentially trying to estimate and is given by:

$$\Delta y_t = b_{10} - b_{12}\Delta z_t + \gamma_{11}\Delta y_{t-1} + \gamma_{12}\Delta z_{t-1} + \varepsilon_{yt} \quad (5.3a)$$

$$\Delta z_t = b_{20} - b_{21}\Delta y_t + \gamma_{21}\Delta y_{t-1} + \gamma_{22}\Delta z_{t-1} + \varepsilon_{zt} \quad (5.3b)$$

[10] See Enders (1995) p. 302 for proofs

Due to the feedback inherent in the system, these equations cannot be estimated directly using OLS. The reason is that Δz_t is correlated with the error term ε_{yt} and Δy_t with the error term ε_{zt}. One of the fundamental assumptions of the Classical Linear Regression Model (CLRM) is that the explanatory variables are independent of the error terms. Note that there is no such problem in estimating the VAR system in the standard form represented by (5.5a) and (5.5b). OLS can provide estimates of the two elements of A_0 and four elements of A_1. Furthermore, by obtaining the residual series [{e_1} and {e_2}] from the regression we can calculate estimates of var(e_1), var(e_2) and cov(e_1e_2). The issue then is whether it is possible to recover all of the information present in the structural form of the VAR from the information estimated from the standard form. Alternatively worded, is the primitive form *identifiable* given the OLS estimates from the standard form of the VAR model represented by equations (5.5a) and (5.5b)?

At face value the answer to this question is unequivocally no! In order to make this system identifiable we must appropriately restrict one of the coefficients in the primitive system. To understand why this is the case let us compare the number of coefficients required by each system. Using OLS to estimate the standard system represented by equations (5.5a) and (5.5b) yields six parameter estimates (a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, and a_{22}) and the calculated values of var(e_{1t}), var(e_{2t}), and cov($e_{1t}e_{2t}$). However, the primitive system represented by equations (5.3a) and (5.3b) contains ten parameter estimates. In addition to the two intercept coefficients (b_{10} and b_{20}), the four autoregressive coefficients (γ_{11}, γ_{12}, γ_{21}, and γ_{22}), and the two contemporaneous feedback coefficients (b_{12} and b_{21}), there are the two standard deviation estimates (σ_y and σ_z). In all, the primitive system requires the specification of ten coefficient estimates whereas the standard VAR estimation yields us only nine parameters. Unless one of the parameters is restricted, it is not possible to identify the primitive system – we say that the primitive form of the VAR is *underidentified*.

To make certain that our system is properly identified we employ the type of *recursive* system proposed by Sims (1980). Essentially we must choose to restrict one of the parameters in the primitive system to equal zero. For example, suppose we set the constraint $b_{21}=0$. Writing (5.3a) and (5.3b) with the constraint imposed gives us:

$$\Delta y_t = b_{10} - b_{12}\Delta y_t + \gamma_{11}\Delta y_{t-1} + \gamma_{12}\Delta z_{t-1} + \varepsilon_{yt} \tag{5.8}$$

$$\Delta z_t = b_{20} + \gamma_{21}\Delta y_{t-1} + \gamma_{22}\Delta z_{t-1} + \varepsilon_{zt} \tag{5.9}$$

Given the restriction it is clear that Δz_t has a contemporaneous effect on Δy_t, but Δy_t only affects Δz_t through a lagged response, that is, there is no contemporaneous effect of Δy_t on Δz_t. Imposing the restriction $b_{21}=0$ means that B^{-1} is given by:

$$B^{-1}=\begin{bmatrix}1 & -b_{12}\\ 0 & 1\end{bmatrix}$$

Now, premultiplication of the primitive system by B^{-1} yields:

$$\begin{bmatrix}y_t\\ z_t\end{bmatrix}=\begin{bmatrix}1 & -b_{12}\\ 0 & 1\end{bmatrix}\begin{bmatrix}b_{10}\\ b_{20}\end{bmatrix}+\begin{bmatrix}1 & -b_{12}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}y_{t-1}\\ z_{t-1}\end{bmatrix}+\begin{bmatrix}1 & -b_{12}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\varepsilon_{yt}\\ \varepsilon_{zt}\end{bmatrix}$$

or

$$\begin{bmatrix}y_t\\ z_t\end{bmatrix}=\begin{bmatrix}b_{10}-b_{12}b_{20}\\ b_{20}\end{bmatrix}+\begin{bmatrix}\gamma_{11}-b_{12}\gamma_{21} & \gamma_{12}-b_{12}\gamma_{22}\\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}y_{t-1}\\ z_{t-1}\end{bmatrix}+\begin{bmatrix}\varepsilon_{yt}-b_{12}\varepsilon_{zt}\\ \varepsilon_{zt}\end{bmatrix} \tag{5.10}$$

Estimating the system using OLS yields the theoretical parameter estimates:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$
$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

$$where$$
$$a_{10} = b_{10} - b_{12}b_{20}$$
$$a_{11} = \gamma 11 - b_{12}\gamma_{21}$$
$$a_{12} = \gamma 12 - b_{12}\gamma_{22}$$
$$a_{20} = b_{20}$$
$$a_{21} = \gamma_{21}$$

$$a_{22} = \gamma_{22}$$

Since $e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt}$ and $e_{2t} = \varepsilon_{zt}$, we can calculate the parameters of the variance/covariance matrix as

$$Var(e_1) = \sigma_y^2 + b_{12}^2\sigma_z^2 \tag{5.11a}$$

$$Var(e_2) = \sigma_z^2 \tag{5.11b}$$

$$Cov(e_1, e_2) = -b_{12}\sigma_z^2 \tag{5.11c}$$

Thus, we have nine parameter estimates a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22}, var(e_1), var(e_2), and cov(e_1e_2) that can be substituted into the nine equations above in order to simultaneously solve for b_{10}, b_{12}, γ_{11}, γ_{12}, b_{20}, γ_{21}, γ_{22}, σ_y^2, and σ_z^2.

Note also that the estimates of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences can be recovered. The residuals from the second equation (i.e. the $\{e_{2t}\}$ sequence) are estimates of the $\{\varepsilon_{zt}\}$ sequence. Combining these estimates along with the solution for b_{12} allows us to calculate an estimate of the $\{\varepsilon_{yt}\}$ sequence using the relationship $e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt}$.

Let us now examine some of the practical implications this restriction poses for our interpretation of the dynamics of this model. In (5.9) the constraint b_{21}=0 means that Δy_t does not have a contemporaneous effect on Δz_t. In (5.10), the restriction manifests itself such that both ε_{yt} and ε_{zt} shocks affect the contemporaneous value of Δy_t, but only ε_{zt} shocks affect the contemporaneous value of Δz_t. The observed values of e_{2t} are completely attributed to pure shocks to the $\{\Delta z_t\}$ sequence. Decomposing the residuals in this triangular fashion in referred to as a **Choleski decomposition.**

An obvious question is how do we decide which parameter within the primitive system to constrain? Our decision is based on our results from the Ganger Causality tests which we conducted earlier in our method. These tests tell us which stock price sequence in the pair is leading the other stock price sequence. The feedback coefficient in the primitive system which is not responsible for any explanatory power in the model is then constrained to equal zero. For example, if it was determined that Δz_t could significantly help to explain movements in Δy_t, but values of Δy_t could not help explain values of Δz_t

then we conclude that Δz_t leads Δy_t and that the relationship is uni-directional. In the primitive system we would represent this by constraining b_{21} to equal zero.

5.2.3 Choosing an appropriate lag length: Akaike Information Criterion

The Akaike Information Criterion (AIC) is the method used to determine the lag length in many of the models estimated in this study. It can be interpreted as a selection criterion between competing models. When choosing a certain model specification the aim is usually to maximise the goodness-of-fit in-sample (R^2) which is done by minimizing the sum of squared residuals (RSS). The AIC says that in addition to being a good fit, it is also beneficial if that model is parsimonious. Thus, instead of simply trying to minimise RSS the AIC imposes a penalty for including regressors in the model which do not significantly improve the explanatory power of the model. It is the aim of AIC to minimise the following AIC statistic:

$$AIC = e^{2k/n}\frac{RSS}{n}$$

where k is the number of regressors (including the intercept term) and n is the number of observations.

5.2.4 Impulse Response functions

The impulse response function (given above) is critical for analyzing the inter-relationships between price series represented in a VAR. The impulse response function is essentially the vector moving average representation of the VAR (5.12) in that the variables (Δy_t and Δz_t) are expressed in terms of the current and past values of the two types of shocks (i.e. e_{1t} and e_{2t}).

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} \tag{5.12}$$

$$where\ \mu = [\bar{y}\ \bar{z}]'$$

$$and\ \bar{y} = \frac{[a_{10}(1-a_{22}) + a_{12}a_{20}]}{\Delta}, \bar{z} = \frac{[a_{20}(1-a_{11}) + a_{21}a_{10}]}{\Delta}$$

$$\Delta = (1 - a_{11})(1 - a_{22}) - a_{12}a_{21}$$

The VMA representation allows you to trace out the time path of the various shocks to the variables contained in the VAR system. From a pairs trading perspective it would be interesting to know how shocks to one of the variables are filtered through each of the individual price series and how long it would take for equilibrium to be restored within the system.

Let us now recall the simple VAR model we have been using up to this point to show how we estimate the impulse response functions. Writing the standard form of the VAR in matrix notation gives us:

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{5.13}$$

Or, using (5.12), we obtain

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \Delta \bar{y} \\ \Delta \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \tag{5.14}$$

Equation (5.14) expresses Δy_t and Δz_t in terms of the $\{e_{1t}\}$ and $\{e_{2t}\}$ sequences. However, it is insightful to rewrite (5.14) in terms of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences. From (5.6) and (5.7), the vector of errors can be written as

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = [1/(1 - b_{12}b_{21}) \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \tag{5.15}$$

So that (5.14) and (5.15) can be combined to form

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \Delta \bar{y} \\ \Delta \bar{z} \end{bmatrix} + [\frac{1}{1 - b_{12}b_{21}}] \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

To simplify, let us define 2 x 2 matrix Φ_i, with elements $\Phi_{jk}(i)$:

$$\Phi_i = [\frac{A_1^i}{1 - b_{12}b_{21}}] \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$$

Hence, the MVA representation of (5.14) and (5.15) can be written in terms of the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences:

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \Delta \bar{y} \\ \Delta \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \Phi_{11}(i) & \Phi_{12}(i) \\ \Phi_{21}(i) & \Phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

Or more compactly

$$x_t = \mu + \textstyle\sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i} \qquad (5.16)$$

The impulse response function (VMA representation) is an especially useful tool to examine the interaction between the $\{\Delta y_t\}$ and $\{\Delta z_t\}$ sequences. The coefficients of Φ_i can be used to generate the effects of ε_{yt} and ε_{zt} shocks on the entire time paths of the $\{\Delta y_t\}$ and $\{\Delta z_t\}$ sequences. It can be seen from the notation that the four elements $\Phi_{jk}(0)$ are **impact multipliers**. For example, the coefficient $\Phi_{12}(0)$ is the instantaneous impact of a one-unit change in ε_{zt} on Δy_t. In the same way, the elements of $\Phi_{11}(1)$ and $\Phi_{12}(1)$ are the one period responses of unit changes in ε_{yt-1} and ε_{zt-1} on Δy_t, respectively. Updating by one period indicates that $\Phi_{11}(1)$ and $\Phi_{12}(1)$ also represent the effects of unit changes in ε_{yt} and ε_{zt} on Δy_{t+1}.

The accumulated effects of unit impulses in ε_{yt} and ε_{zt} can be obtained by the appropriate summation of the coefficients of the impulse response functions. For example, note that after *n* periods the effect of ε_{zt} on the value of Δy_{t+n} is $\Phi_{12}(n)$. Thus, after *n* periods, the cumulated sum of the effects of ε_{zt} on the $\{\Delta y_t\}$ sequence is

$$\sum_{i=0}^{n} \Phi_{12}(i)$$

Letting n approach infinity yields the **long-run multiplier**. Since the $\{\Delta y_t\}$ and $\{\Delta z_t\}$ sequences are assumed to be stationary, it must be the case that for all j and k,

$$\sum_{i=0}^{\infty} \Phi_{jk}^2\,(i)\ is\ finite.$$

The four sets of coefficients $\Phi_{11}(i)$, $\Phi_{12}(i)$, $\Phi_{21}(i)$, and $\Phi_{22}(i)$ are called the **impulse response functions.** Plotting the impulse response functions (i.e. plotting the coefficients of $\Phi_{jk}(i)$ against i) is a practical way to visually represent the behaviour of the $\{\Delta y_t\}$ and $\{\Delta z_t\}$ series in response to the various shocks.

5.2.4 Variance Decomposition

Since unrestricted VARs are overparameterized, they are not particularly useful for short-term forecasts. However, understanding the properties of the forecast errors is exceedingly helpful in uncovering the intricate inter-relationships between the variables in the system. Suppose that we knew the coefficients A_0 and A_1 and wanted to forecast the various values of x_{t+i} conditional on the observed value of x_t. Updating (5.4) one period (i.e., $x_{t+1}=A_0+A_1x_t+e_{t+1}$) and taking the conditional expectation of x_{t+1}, we obtain

$$E_t x_{t+1} = A_0 + A_1 x_t$$

Note that the one step ahead forecast error is $x_{t+1} - E_t x_{t+1} = e_{t+1}$. Similarly, updating two periods we get

$$\begin{aligned} x_{t+2} &= A_0 + A_1 x_{t+1} + e_{t+2} \\ &= A_0 + A_1(A_0 + A_1 x_t + e_{t+1}) + e_{t+2} \end{aligned}$$

If we take conditional expectations, the two step ahead forecast of x_{t+2} is

$$E_t x_{t+2} = (I + A_1)A_0 + A_1^2 x_t$$

The two-step ahead forecast error (i.e. the difference between the realization of x_{t+2} and the forecast) is $e_{t+2} + A_1 e_{t+1}$. More generally, it is easily verified that the n-step ahead forecast is

$$E_t x_{t+n} = (I + A_1 + A_1^2 + \cdots + A_1^{n-1})A_0 + A_1^n x_t$$

And the associated forecast error is

$$e_{t+n} + A_1 e_{t+n-1} + A_1^2 e_{t+n-2} + \cdots + A_1^{n-1} e_{t+1} \qquad (5.17)$$

We can also consider these forecast errors in terms of (5.16) (i.e. the VMA form of the model). Of course, the VMA and VAR models contain exactly the same information but it is helpful to describe the properties of the forecast errors in terms of the $\{\varepsilon_t\}$ sequence. If we use (5.16) to conditionally forecast x_{t+1}, the one-step ahead forecast error is $\Phi_0 \varepsilon_{t+1}$. In general,

$$x_{t+n} = \mu + \sum_{i=0}^{n-1} \Phi_i \varepsilon_{t+n-i}$$

So that the n period forecast error $x_{t+n} - E_t x_{t+n}$ is

$$x_{t+n} - E_t x_{t+n} = \mu + \sum_{i=0}^{n-1} \Phi_i \varepsilon_{t+n-i}$$

Focusing solely on the $\{\Delta y_t\}$ sequence, we see that the n-step ahead forecast error is

$$\Delta y_{t+n} - E_t \Delta y_{t+n} = \Phi_{11}(0)\varepsilon_{yt+n} + \Phi_{11}(1)\varepsilon_{yt+n-1} + \cdots + \Phi_{11}(n-1)\varepsilon_{yt+1} + \Phi_{12}(0)\varepsilon_{zt+n} + \Phi_{12}(1)\varepsilon_{zt+n-1} + \cdots + \Phi_{12}(n-1)\varepsilon_{zt+1}$$

Denote the variance of the n-step ahead forecast error variance of Δy_{t+n} as $\sigma_y(n)^2$

$$\sigma_y(n)^2 = \sigma_y^2[\Phi_{11}(0)^2 + \Phi_{11}(1)^2 + \cdots + \Phi_{11}(n-1)^2] + \sigma_z^2[\Phi_{12}(0)^2 + \Phi_{12}(1)^2 + \cdots + \Phi_{12}(n-1)^2]$$

Since all values of $\Phi_{jk}(i)^2$ are necessarily non-negative, the variance of the forecast error increases as the forecast horizon n increases. Note that it is possible to decompose the n-step ahead forecast error variance due to each one of the shocks. Respectively, the proportions of $\sigma_y(n)^2$ due to shocks in the $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ sequences are:

$$\frac{\sigma_y^2[\Phi_{11}(0)^2 + \Phi_{11}(1)^2 + \cdots + \Phi_{11}(n-1)^2]}{\sigma_y(n)^2}$$

and

$$\frac{\sigma_z^2[\Phi_{12}(0)^2 + \Phi_{12}(1)^2 + \cdots + \Phi_{12}(n-1)^2]}{\sigma_y(n)^2}$$

The **forecast error variance decomposition** tells us the proportion of the movements in a sequence due to its "own" shocks versus shocks to other variable. If ε_{zt} shocks explain none of the forecast error variance of $\{\Delta y_t\}$ at all forecast horizons, we can say that the $\{\Delta y_t\}$ sequence is exogenous. In such a circumstance, the $\{\Delta y_t\}$ sequence would evolve independently of the ε_{zt} shocks and $\{\Delta z_t\}$ sequence. At the other extreme, ε_{zt} shocks could explain all the forecast error variance in the $\{\Delta y_t\}$ sequence at all forecast horizons, so that $\{\Delta y_t\}$ would be entirely endogenous. In applied research, it is typical for a variable to explain almost all its forecast error variance at short horizons and smaller proportions at longer horizons. We would expect this pattern if ε_{zt} shocks had little contemporaneous effect on Δy_t, but acted to affect the $\{\Delta y_t\}$ sequence with a lag.

6. Results

In this section of the study we present and interpret our results. We first present our findings relating to the identification of the trading pairs, including output from the cointegration tests and Granger Causality tests. This then allows us to obtain the residual spread. For completeness, we then present the findings of the augmented Dickey-Fuller tests run on the residual series. If the cointegrating relationship is meaningful we expect it to show the residual series is stationary. We then present and interpret the findings of our estimated VECM and seek to determine whether the speed of adjustment coefficients conform to expectations regarding sign and statistical significance. Finally, we present the findings of the impulse response functions and variance decomposition analysis in an attempt to understand what these tools tell us regarding the short-term dynamic behaviour of the individual price series in response to shocks.

6.1 Identifying Trading Pairs

6.1.1 Cointegration Test Output

The first step in this study was to test each potential pair in our sample of 17 financial stocks listed in the ASX for the presence of a cointegrating relationship. Implementing the Johansen test meant that for 17 individual stocks there were $(17^2-17)/2=136$ potential trading pairs. The Johansen test was initially implemented using daily data from the five years of data with an acceptable occurrence of type one errors set at one per cent. Although this level of significance is relatively unforgiving, it was deemed necessary when one considers what is potentially at stake. The single most important feature of a successful pairs trading strategy is the presence of a mean-reverting equilibrium relationship between the pairs. Cointegration provides us with this necessary condition, and, as such, we have decided to set α=1% so that we can be as sure (as possible) that any detected cointegrating relationship is robust.

In a test of robustness, the identified pairs were then re-tested using weekly data. The data sample was filtered to include only prices from the Wednesday of each week over the five years. The choice on which day to base this test is not critical. However, this study uses Wednesday because it was thought that this day was most insulated from certain market irregularities such as thin-trading earlier and later in the week which could potentially bias pricing patterns.

Trading pairs were retained if their p-values, which were calculated from the cointegration tests using weekly data, were less than 10%. This means that on average the likelihood of rejecting the null hypothesis of "no cointegrating relationship" when in fact it is true will be less than 10 in every 100. This secondary test for cointegration using the weekly data was deemed important in separating those pairs which were truly cointegrated, from those which were only mildly or weakly cointegrated. It should be the aim of every pairs trading strategy to only base trades on pairs of stocks which exhibit an extremely strong and robust cointegrating relationship.

Diagram 1: Estimation output from cointegration test for WBC and BOQ.

Date: 09/03/08 Time: 13:11
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: WBC BOQ
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.017693	23.14649	15.49471	0.0029
At most 1	0.000232	0.296424	3.841466	0.5861

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Diagram 1 clearly illustrates the presence of a statistically significant cointegrating relationship between the stocks WBC and BOQ. When we consider the p-value of 0.0029 associated with the trace statistic for the null hypothesis of "no cointegrating relationships", it is clear that this can be rejected at the 1% level. Thus, we can conclude that there exists one statistically significant cointegrating equation.

Alternatively, consider the output from the cointegration test for the stocks BOQ and SUN in diagram 2 below.

Diagram 2: Cointegration test output for stocks BOQ and SUN

Date: 10/14/08 Time: 13:00
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: BOQ SUN
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.002637	3.462530	15.49471	0.9420
At most 1	6.41E-05	0.082109	3.841466	0.7744

Trace test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

It can be seen that the reported p-value for the trace statistic is only 0.9420. This is obviously larger than 0.01 and so we accept the null hypothesis that there is no cointegrating relationship present.

This study identifies 14 potential trading pairs which conform to the filters outlined above. These pairs are summarized, along with their p-values from both tests in table 2 below.

Table 2: Summary of trading pairs and associated p-values from both cointegration tests.

Trading Pairs	P-value: daily sample	P-value: weekly sample
ANZ/AMP	>1%	>5%
BOQ/WBC	>1%	>10%
SGB/WBC	>1%	>1%
FKP/WBC	>1%	>10%
LLC/ASX	>1%	>5%
ASX/CBA	>1%	>1%
ASX/SGB	>1%	>10%
BEN/AMP	>1%	>1%
ASX/BEN	>1%	>1%
AXA/AMP	>1%	>5%
ASX/AMP	>1%	>1%
PPT/AMP	>1%	>5%
QBE/AMP	>1%	>1%
CBA/AMP	>1%	>1%

The results of the tests for cointegration for our 14 pairs identified in table 1 are included in Appendix A.

6.1.2 Testing for Granger Causality

Granger Causality plays an important role in this study and in some sense links the trading pairs identification problem with the modeling procedure. For each of the 14 potential trading pairs which have now just been identified, we ran Granger Causality tests on each pair to determine which stock price series *informationally* led the other. This insight into the dynamics of the cointegrating relationship for a given pair of stocks was important for two reasons. Firstly, it tells us which stock is the dependent variable in the cointegrating equation, and which stock is independent, or responsible for driving price changes in the dependent variable. Secondly, it provides us with a theoretical model for justifying which coefficient in the primitive form of the VAR to constrain to zero. This is known as the problem of identification when estimating a VAR and was introduced in the method section.

Following on from our example, consider the Granger Causality output presented in diagram 3 (below) for the trading pair WBC/BOQ.

Diagram 3: Output from the Granger Causality test for WBC and BOQ

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:21
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
BOQ does not Granger Cause WBC	1283	11.8526	7.9E-06
WBC does not Granger Cause BOQ		1.69379	0.18423

With a p-value (7.9E-06) of effectively zero, we can reject the null hypothesis that BOQ does not Granger Cause WBC and instead conclude that it does. We conclude that explaining the contemporaneous value of WBC can be significantly improved by incorporating past values of BOQ, in addition to just past values of itself. The results presented here show an example of uni-directionality. If both p-values were less than 0.05 then we would conclude that both stock price sequences Granger Cause each other. This is not a helpful result for this study because we wish to constrain the contemporaneous effects of one series equal to zero so that the VAR system becomes identifiable.

Granger Causality tests were conducted for each of the 14 pairs and the output from these tests are reported in Appendix B. For those pairs where it was found both stocks led each other, these pairs were removed from the study.

Table 3: A summary of the results from the Granger Causality tests

Trading Pair	Direction of Causality (p-value)	Direction of Causality (p-value)
ANZ/AMP	ANZ does not lead AMP (0.00016)*	AMP does not lead ANZ (0.12127)
BOQ/WBC	BOQ does not lead WBC (7.9E-06)*	WBC does not lead BOQ (0.18423)
SGB/WBC	SGB does not lead WBC (0.0073)*	WBC does not lead SGB (0.08434)
FKP/WBC	FKP does not lead WBC (0.00032)*	WBC does not lead FKP (0.19337)
LLC/ASX	LLC does not lead ASX (0.00879)*	ASX does not lead LLC (0.07776)
ASX/CBA	ASX does not lead CBA (1.3E-05)*	CBA does not lead ASX (0.88194)
ASX/SGB	ASX does not lead SGB (0.02394)*	SGB does not lead ASX (0.02662)*
BEN/AMP	BEN does not lead AMP (0.00014)*	AMP does not lead BEN (0.03938)*
ASX/BEN	ASX does not lead BEN (6.0E-05)*	BEN does not lead ASX (7.1E-05)*
AXA/AMP	AXA does not lead AMP (0.0002)*	AMP does not lead AXA (0.51577)
ASX/AMP	ASX does not lead AMP (0.00028)*	AMP does not lead ASX (0.41104)
PPT/AMP	PPT does not lead AMP (6.3E-05)*	AMP does not lead PPT (0.37082)
QBE/AMP	QBE does not lead AMP (0.00075)*	AMP does not lead QBE (0.56121)
CBA/AMP	CBA does not lead AMP (3.2E-05)*	AMP does not lead CBA (0.01437)*

* indicates rejection at 5% level

Once we removed the trading pairs for which a bi-directional Granger Causality relationship exists, we are then left with ten trading pairs. These trading pairs together with their direction of Granger Causality are summarized in table 4.

Table 4: Summary of trading pairs and direction of Granger Causality

Trading Pair (→ = "Granger Causes")
ANZ→AMP
BOQ→WBC
SGB→WBC
FKP→WBC
LLC→ASX
ASX→CBA
AXA→AMP
ASX→AMP
PPT→AMP
QBE→AMP

6.1.3 Cointegrating Equation and Residual Spread

Once we know the direction of causality we are now in a position to estimate the long-run equilibrium relationship. For two stocks y_t and z_t, the equilibrium relationship takes the following form

$$y_t = \beta_0 + \beta_1 z_t + \varepsilon_t$$

where ε_t is a stationary process if the cointegrating relationship is a meaningful one. Lets now consider the estimation output from our BOQ/WBC example and interpret the coefficients.

Diagram 4: Estimation output from the cointegrating equation for BOQ/WBC

Dependent Variable: WBC
Method: Least Squares
Date: 09/03/08   Time: 13:12
Sample: 1 1285
Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BOQ	1.123173	0.007511	149.5391	0.0000
C	6.442180	0.088783	72.56082	0.0000

R-squared	0.945739	Mean dependent var	19.22138
Adjusted R-squared	0.945697	S.D. dependent var	3.703152
S.E. of regression	0.862948	Akaike info criterion	2.544632
Sum squared resid	955.4245	Schwarz criterion	2.552660
Log likelihood	-1632.926	F-statistic	22361.95
Durbin-Watson stat	0.069961	Prob(F-statistic)	0.000000

The cointegrating coefficient (1.123173) is practically interpreted to as the number of units of BOQ held short, for every single unit of WBC held long so that the resulting portfolio is mean reverting. The value of the portfolio, which can essentially be represented by [C + ε_t] has an equilibrium value of 6.442180 and fluctuates around this value with dynamics governed by those of ε_t. Thus, an insight into the dynamic behaviour of ε_t provides us with an insight into the dynamic behaviour of the total portfolio.

If the cointegrating relationship is meaningful, then ε_t is stationary. If ε_t is indeed stationary then it will have constant (in this case equal to zero) mean, constant variance and constant autocorrelations, and its dynamic behaviour should be well described as exhibiting a strong level of mean-reversion. It should be recognized that these properties of ε_t are all extremely important for any trading rule based on this pairs trading strategy.

Diagram 5: Plot of WBC/BOQ residual series against time

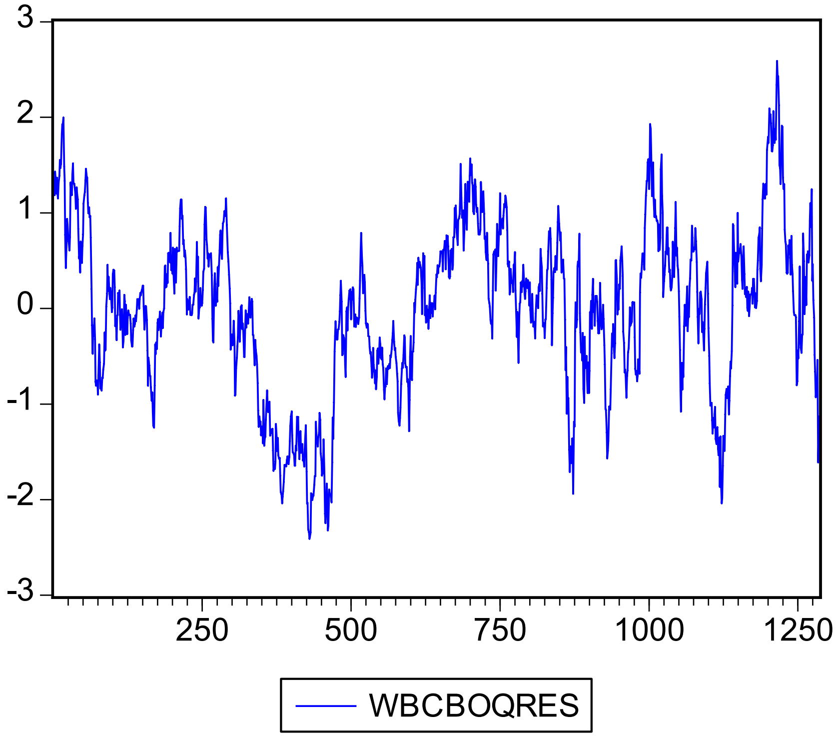


Diagram 5 illustrates a great example of the types of dynamic behaviour we want to see present in our residual series. The key features are the lack of any trend, high levels of volatility and mean reversion around an apparent equilibrium value of zero. This series certainly appears stationary. For completeness, we can subject the residual series from each of our cointegrating equations to the augmented Dickey-Fuller (ADF) test to determine whether the residual series has a unit-root. A unit-root is a key feature of a random walk model which we know to be non-stationary. We want to be able to reject the null hypothesis that there exists a unit root so that we may instead conclude that the residual series is stationary.

Let us now interpret the output from the ADF test on the WBC/BOQ residuals.

Diagram 6: Output for the ADF test on the WBC/BOQ residual series.

Null Hypothesis: BOQWBCRES has a unit root			
Exogenous: Constant, Linear Trend			
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.857536	0.0004
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	
*MacKinnon (1996) one-sided p-values.			

The p-value corresponding to the t-statistic of -4.858 is 0.0004 which is certainly less than 0.05. Thus we are able to reject the null hypothesis that there exists a unit root and we conclude that the WBC/BOQ residual series is in fact a stationary series. This is consistent with our expectations.

Estimation output, plots of the residual series and ADF test output for each pair are reported in Appendix C. Upon visual and statistical inspection of the residual series, those clearly lacking mean reverting behaviour, or failing the ADF test at 5% are omitted from the study.

One of the trading pairs which failed both the visual inspection and the ADF test was PPT and AMP. Consider the residual plot and the ADF test output below.

Diagram 7: Plot of the residual series for PPT/AMP

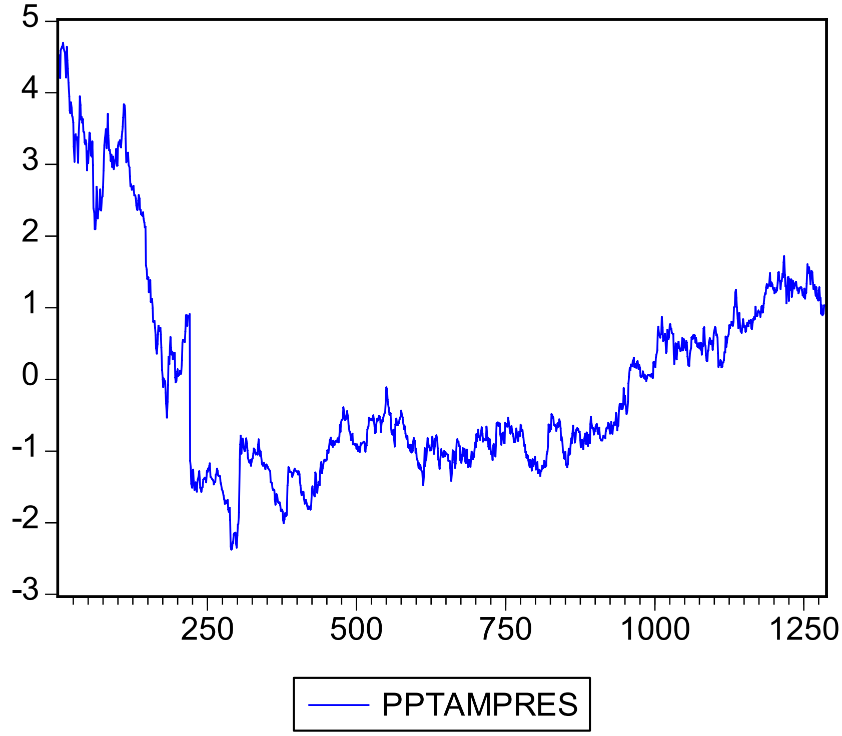


Diagram 8: ADF test output for PPT/AMP residual series

Null Hypothesis: PPTAMPRES has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.803079	0.1963
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

A fundamental question that must be asked when considering diagram 7 is whether it exhibits sufficient levels of mean reversion so that the portfolio could be traded regularly enough to make money. If the first 250 observations were removed it appears as though there is a strong positive trend in the data. This is strong evidence to suggest that the residual series is (1) not stationary, and (2) not a good candidate for pairs trading. The statistical output for the ADF test supports this observation with a p-value (0.1963) far greater than 0.05. Thus we are not able to reject the null hypothesis and instead conclude that the residual series does possess a unit-root and is not stationary.

A summary of the final trading pairs is presented in table 5 below.

Table 5: Summary of final trading pairs

Final Trading Pairs
BOQ→WBC
SGB→WBC
FKP→WBC
LLC→ASX
ASX→CBA

6.2 Calibrating a Vector Error-Correction model (VECM)

In this section we present the estimation output from our modeling procedure. We have chosen to model the residual series as a vector-error-correction model (VECM) which is essentially a vector-autoregression model with an *error-correction* component. The VECM is the correct specification of the VAR model when the component series are cointegrated (Granger Representation Theorem). Let us now consider the VECM estimated for the residual series belonging to the WBC/BOQ example.

Diagram 9: VEMC output for WBC/BOQ

Vector Autoregression Estimates
Date: 10/02/08 Time: 11:49
Sample (adjusted): 4 1285
Included observations: 1282 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	∆BOQ	∆WBC
∆BOQ (-1)	-0.163967	-0.086787
	(0.02877)	(0.03498)
	[-5.69827]	[-2.48121]
∆BOQ (-2)	-0.028951	-0.008412
	(0.02928)	(0.03560)
	[-0.98863]	[-0.23632]
∆WBC(-1)	0.023823	0.015252
	(0.02377)	(0.02889)
	[1.00240]	[0.52794]
∆WBC(-2)	0.059918	-0.031605
	(0.02373)	(0.02885)
	[2.52462]	[-1.09551]
C	0.009796	0.008912
	(0.00441)	(0.00536)
	[2.22277]	[1.66374]
BOQWBCRESID	-0.020684	0.014732
	(0.00518)	(0.00629)
	[-3.99538]	[2.34114]
R-squared	0.037150	0.011005
Adj. R-squared	0.033377	0.007130
S.E. equation	0.157135	0.191007
F-statistic	9.846461	2.839783
Log likelihood	556.4598	306.2106
Akaike AIC	-0.858752	-0.468347

To begin with, it is interesting to notice that lags of ΔWBC offer no significant explanatory power in predicting the contemporaneous value of itself. The only independent variable which has significant explanatory power in predicting ΔWBC is the first lag of ΔBOQ (ΔBOQ -1). This finding is consistent with the results for Granger Causality which indicates that ΔBOQ informationally leads ΔWBC. Although the model does not explain much of the variation in ΔWBC, in fact it explains little more than 1%, the F-statistic (2.84) tells us that the probability of each coefficient being (jointly) insignificantly different from zero is very close to zero. Furthermore, the goodness of fit (R^2) for each equation is not of concern since we are dealing with a system of equations and are more concerned with the estimated sign and significance of certain key coefficients.

Unlike ΔWBC, lags of ΔBOQ do offer significant explanatory power in predicting the contemporaneous value of itself, especially the first lag with a t-statistic of -5.7. It is also apparent that lags of ΔWBC do not offer much in terms of helping to explain variation in ΔBOQ. This is also consistent with the results of Granger Causality which suggest uni-directional causality from ΔBOQ → ΔWBC.

The most important finding from these results is that the sign and significance of the speed of adjustment coefficients (BOQWBCRESID in diagram 9) conform to those hypothesized by cointegration. Cointegration says that in a VECM one or both of these parameters must be significantly different from zero. These coefficients tell us which stock is responsible for returning the system to its long-run equilibrium relationship. If both of these terms were no different from zero then we have actually just estimated a bivariate VAR as there is no term in the model which would cause the stock prices to return to some long-run equilibrium relationship.
In understanding the dynamics of the relationship between ΔBOQ and ΔWBC it is interesting to note that both sequences actively move in opposite directions to restore equilibrium following a shock to the system. An alternative might be the case where both sequences move in the same direction, only one at a faster rate than the other to restore equilibrium.

Let us now consider the VECM output for another of our identified pairs where we hope the estimated coefficients of the VECM better reflect our expectations based on the Granger Causality testing. The pair under consideration here is FKP/WBC where FKP informationally leads WBC and the causality is uni-directional. Thus, we should expect that the lags of ΔFKP have some significant explanatory power for ΔWBC, but lags of ΔWBC should not be significantly different from zero when explaining ΔFKP.

Diagram 10: VECM output for FKP/WBC

Vector Autoregression Estimates		
Date: 10/02/08 Time: 12:04		
Sample (adjusted): 4 1285		
Included observations: 1282 after adjustments		
Standard errors in ( ) & t-statistics in [ ]		
	ΔFKP	ΔWBC
ΔFKP(-1)	-0.147828	0.069899
	(0.02839)	(0.06812)
	[-5.20716]	[1.02615]
ΔFKP(-2)	-0.082727	-0.153082
	(0.02843)	(0.06822)
	[-2.90959]	[-2.24388]
ΔWBC(-1)	0.002461	-0.005100
	(0.01184)	(0.02842)
	[0.20782]	[-0.17947]
ΔWBC(-2)	0.008626	-0.021119
	(0.01181)	(0.02834)
	[0.73026]	[-0.74511]
C	0.005837	0.008553
	(0.00223)	(0.00536)
	[2.61351]	[1.59602]
FKPWBCRESID	-0.008607	0.014394
	(0.00243)	(0.00584)
	[-3.53750]	[2.46542]
R-squared	0.031899	0.011307
Adj. R-squared	0.028106	0.007433
S.E. equation	0.079593	0.190978
F-statistic	8.408992	2.918471
Log likelihood	1428.454	306.4060
Akaike AIC	-2.219117	-0.468652

Consistent with our expectations the lags of ΔFKP (ΔFKP-1 and ΔFKP-2) are individually statistically significant in helping explain contemporaneous variation in ΔWBC. Also aligned with our a-priori expectations the lags of ΔWBC (ΔWBC-1 and ΔWBC-2) do not individually contribute to the models ability to explain ΔFKP. Together these observations conform to the results of the Granger Causality test (see Appendix B) where we identified a uni-directional lead from ΔFKP to ΔWBC.

The speed of adjustment coefficients (FKPWBCRESID in diagram 10) also conform to the constraints of cointegration. Importantly both coefficients are statistically significant and are of opposite signs so that they will move in opposite directions to restore equilibrium following a shock to the system.

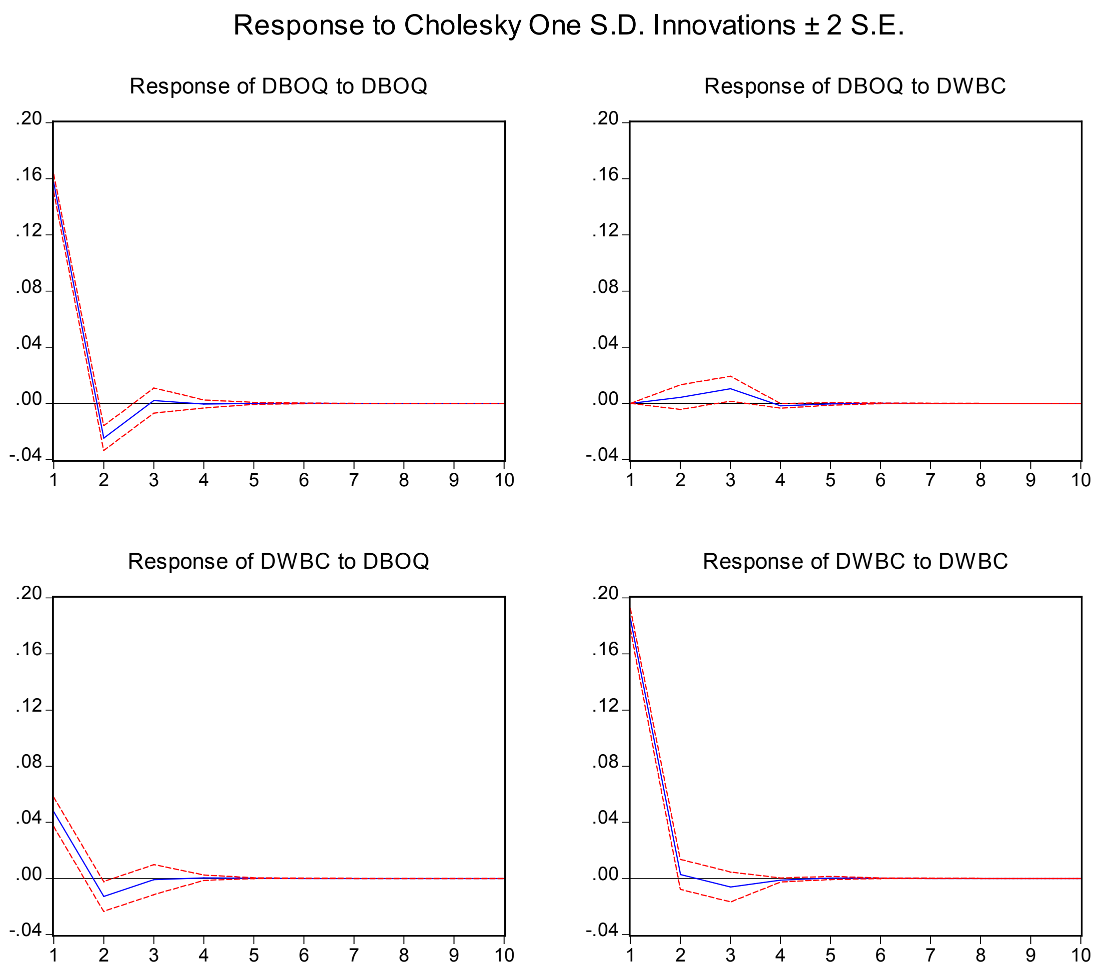
VECM output for the 4 remaining final trading pairs are reported in Appendix D.

6.3 Results of variance analysis

6.3.1 Impulse response functions

In the previous section we estimated a VECM which amongst other things provided estimates of the speed of adjustment coefficients. These parameters describe the longer-term dynamic behaviour of the individual sequences relative to divergences from the long-run equilibrium relationship. A professional investor implementing a pairs trading strategy would also be interested in the short-term dynamics of the system. The impulse response functions maps out the short-term behaviour of each sequence in response to shocks to the system. These shocks are applied to each sequence individually and its impact on both itself and the other sequence are then plotted against time. It is important to remember that since both sequences are assumed to be stationary processes, the cumulative impact of the shock must be finite i.e. the marginal impact of the shock in each subsequent time period must progressively reach zero.

Diagram 11: Impulse response functions for ΔBOQ/ΔWBC



Let us now proceed to interpret the impulse response functions for ΔBOQ/ΔWBC in diagram 11. The top left (bottom right) panel plots the dynamic reaction of ΔBOQ (ΔWBC) in response to a one standard deviation shock to ΔBOQ (ΔWBC). The blue line represents the estimated impulse response function while the surrounding red lines are the 95% confidence intervals. In the top left (TL) panel we can see that, on average, ΔBOQ increases after one period by 0.157 units and decreases by 0.025 units following one standard deviation shock to ΔBOQ, both of which are statistically significant. However any movement in response to the shock after two periods is not significant and so we say that the dynamic effects of the shocks last two periods and after that it dies out. We have used diagram 12 to verify the exact values in the TL panel in diagram 11.

Diagram 12: Response of ΔBOQ in period d

	ΔBOQ	ΔWBC
1	0.157135	0.000000
	(0.00310)	(0.00000)
2	-0.024633	0.004407
	(0.00441)	(0.00440)
3	0.002029	0.010430
	(0.00447)	(0.00445)
4	-0.000414	-0.001816
	(0.00142)	(0.00081)
5	-3.36E-05	-0.000402
	(0.00036)	(0.00044)
6	4.43E-05	5.69E-05
	(0.00011)	(0.00012)
7	-3.25E-06	1.93E-05
	(4.7E-06)	(3.3E-05)
8	-1.25E-06	3.59E-07
	(3.6E-06)	(9.0E-06)
9	-2.02E-08	-1.22E-06
	(6.5E-07)	(2.2E-06)
10	4.83E-08	-1.16E-07
	(1.4E-07)	(5.7E-07)

Similarly, the bottom right (BR) panel in diagram 11 plots the dynamic behaviour of ΔWBC following a one standard deviation shock to itself. Diagram 13 tells us that ΔWBC increases by 0.185 units following this shock which is statistically significant but all subsequent impacts are not significantly different from zero. Thus, on average, a one standard deviation shock to ΔWBC impacts on the dynamic behaviour of ΔWBC for only a single time period following the shock before wilting away to insignificance.

Diagram 13: Response of ΔWBC in period d

	ΔWBC	ΔBOQ
1	0.047510	0.185004
	(0.00525)	(0.00365)
2	-0.012913	0.002822
	(0.00533)	(0.00534)
3	-0.000883	-0.006186
	(0.00536)	(0.00535)
4	0.000426	-0.001126
	(0.00097)	(0.00070)
5	5.33E-05	0.000248
	(0.00014)	(0.00057)
6	-6.25E-06	8.95E-05
	(2.9E-05)	(0.00010)
7	-5.34E-06	-8.04E-06
	(1.0E-05)	(3.8E-05)
8	2.52E-08	-5.11E-06
	(3.2E-06)	(8.0E-06)
9	3.05E-07	-1.73E-08
	(7.3E-07)	(2.5E-06)
10	1.61E-08	2.64E-07
	(1.9E-07)	(6.1E-07)

The truly interesting feature of these functions is how each sequence responds to shocks to the other variable. Let us now recall the primitive form (5.3a/b) of the VAR introduced in the method section which will hopefully help us to illustrate some of the non-trivial inter-relationships between these sequences.

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} \qquad (5.3a)$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} \qquad (5.3b)$$

Remember that the primitive system cannot be estimated directly due to the feedback inherent the system. Because z_t is correlated with ε_{yt} and y_t is correlated with ε_{zt}, these equations need to be transformed into a more usable form. Rearranging the primitive system yields the standard form of the VAR:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \tag{5.5a}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \tag{5.5b}$$

The important point to realize is that there remain no feedback issues with this new form of the VAR and OLS can be directly applied to estimate both of these equations. However, as described in the method section we must constrain one of the variables in the primitive system in order to make it identifiable. This study uses the Granger Causality test to determine which one of the contemporaneous coefficients (b_{12} or b_{21}) to constrain to zero in the primitive system. We then estimate the coefficients from the standard form of the VAR and "back-out" the parameter values of the primitive system. This is called a Choleski decomposition**.** To illustrate, let us suppose that we set the constraint $b_{21}=0$ in the primitive system so that the contemporaneous value of y_t does not have a contemporaneous effect on z_t. In terms of equation (5.15), the error terms can be decomposed as follows:

$$e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt} \tag{6.1}$$

$$e_{2t} = \varepsilon_{zt} \tag{6.2}$$

It can be seen as a result of the Choleski decomposition that although a ε_{yt} shock has no direct impact on z_t, there is an indirect effect in that lagged values of y_t affect the contemporaneous value of z_t. The key point is that the decomposition forces a potentially important asymmetry on the system since an ε_{zt} shock has contemporaneous effects on both y_t and z_t. For this reason (6.1) and (6.2) are said to imply an ordering of the variables. A ε_{zt} shock directly affects e_{1t} and e_{2t} but an ε_{yt} shock does not affect e_{2t}. Hence, z_t is "prior" to y_t.

Let us now replace the variables "y_t" and "z_t" in our theoretical equation with those stocks ΔBOQ and ΔWBC respectively. Since we have constrained the system so that ΔWBC does not have a contemporaneous effect on ΔBOQ, the same way y_t does not contemporaneously effect z_t, can we identify patterns in the impulse response functions which comply with how we expect them to behave.

Consider the TR panel of diagram 11 which plots the impact a one standard deviation shock to ΔWBC has on ΔBOQ. We can see that a shock to ΔWBC has no contemporaneous affect on ΔBOQ which is consistent with our expectations. This is because we know that BOQ Granger Causes WBC so that we constrain the primitive form of the VAR so that the contemporaneous value of ΔWBC has no impact on the contemporaneous value of ΔBOQ. The shock has no significant effect on ΔBOQ in the first or second periods after the time of the shock, but does show a statistically significant positive effect in the third period. The effects have totally died out by the fourth period. We expect this since for stationary variables the total effect of a shock must be finite. This is also consistent with our expectations since lagged values of ΔWBC are allowed to have an impact on the contemporaneous value of ΔBOQ. It is only the coefficient relating the contemporaneous value of ΔWBC in the primitive system that we constrain to zero, we allow the lagged coefficients to be determined by the model.

Now consider the BL panel in diagram 11 which plots the impact a one standard deviation shock to ΔBOQ has on ΔWBC. Consistent with our expectations the effects are statistically significant contemporaneously and also in both of the first lagged time periods. By the third time period after the shock the effects to ΔWBC have died out. This is also consistent with the stationarity of our variables. The contemporaneous effect of the shock is derived from two sources. Firstly, since we know that ΔBOQ informationally drives ΔWBC we allow the contemporaneous value of ΔBOQ to have a significant effect on the contemporaneous value of ΔWBC. The second source of the impact can be best described with reference to equation (6.1). A shock to ΔBOQ, through positive correlation, also results in a shock to ΔWBC. Thus, shocking ΔBOQ is going to contemporaneously impact the value of ΔWBC. Following a shock to the system, ε_{it+1} returns to zero, but the autoregressive nature of the system ensures the individual series do not immediately return to zero. The subsequent values of the

{ΔBOQ} and {ΔWBC} sequences converge to their long-run levels. This convergence is assured by the stability of the system.

6.3.2 Variance decomposition

Diagram 14 tells us the proportion of the movements in the ΔBOQ sequence due to its "own" shocks versus the shocks to ΔWBC. It is quite clear that shocks to ΔWBC explain effectively none of the forecast error variance of ΔBOQ at any of the forecast horizons reported in diagram 14. It is accurate to conclude that ΔBOQ is effectively exogenous of ΔWBC and the ΔBOQ sequence evolves independently of ΔWBC and those shocks to ΔWBC.

Diagram 14: Variance decomposition of ΔBOQ

Time	S.E.	ΔBOQ	ΔWBC
1	0.157135	100.0000	0.000000
2	0.159115	99.92327	0.076727
3	0.159470	99.49588	0.504124
4	0.159481	99.48298	0.517022
5	0.159481	99.48235	0.517653
6	0.159481	99.48233	0.517665
7	0.159481	99.48233	0.517667
8	0.159481	99.48233	0.517667
9	0.159481	99.48233	0.517667
10	0.159481	99.48233	0.517667

Diagram 15 tells us the proportion of the movements in the ΔWBC sequence due to its "own" shocks versus the shocks to ΔBOQ. We can see that shocks to the sequence ΔBOQ are partly responsible for explaining the forecast error variance of ΔWBC. We conclude that ΔWBC evolves endogenously with ΔBOQ.

Diagram 15: Variance Decomposition of ΔWBC

Time	S.E.	ΔBOQ	ΔWBC
1	0.191007	6.186837	93.81316
2	0.191463	6.612191	93.38781
3	0.191565	6.607277	93.39272
4	0.191569	6.607510	93.39249
5	0.191569	6.607506	93.39249
6	0.191569	6.607505	93.39250
7	0.191569	6.607505	93.39250
8	0.191569	6.607505	93.39250
9	0.191569	6.607505	93.39250
10	0.191569	6.607505	93.39250

At this stage a reasonable question is whether these findings are consistent with what we expect given our previous results? The results of the variance decomposition analysis are consistent with our previous results. The Granger Causality tests tell us that ΔBOQ does informationally lead ΔBOQ, and so we would expect to find that shocks to ΔBOQ are responsible for explaining some of the forecast error variance of ΔWBC. Similarly, we know that ΔWBC does not Granger Cause ΔBOQ and thus, we are not skeptical as to why ΔWBC evolved exogenously of ΔBOQ.

6.4 Is the assumption of error term normality critical?

6.4.1 The issue of normality

The results provided in this study are useful to gauge an insight into the co-movement between our trading pairs. By modeling the residual spread we gain some understanding into the dynamic behaviour of the mean-reverting portfolio. It is important to note, however, that what we are attempting to calibrate to the VECM, a residual series which has been derived from an OLS regression. Gujarati (2003) notes that when performing OLS the researcher must assume that the error term is normally distributed, has an expected value of zero, constant variance, and constant covariances. Gujarati (2003) continues to say that if autocorrelation persists within the residual series, then this will lead to bias in the estimator. If we have biased estimators then any conclusions that have been drawn from those estimated coefficients are incorrect. Thus, a logical question to ask is what do we do if it is found that autocorrelation persists in the residual series?

6.4.2 Common trends model and APT

Common trends cointegration model

In order to avoid the issues relating to biased estimators, we will now attempt to analyse the cointegrating relationships via the use of the Common Trends Model (CTM) (Stock and Watson,1991) and reconcile it to the Arbitrage Pricing Theory (APT) (Ross, 1976). The common trends model says that given two series y_t and z_t, these stock price series can be decomposed as follows

$$\begin{aligned} y_t &= n_{yt} + \varepsilon_{yt} \\ z_t &= n_{zt} + \varepsilon_{zt} \end{aligned} \tag{6.3}$$

where n_{yt} and n_{zt} represent the so-called common trends or random walk components of the two time series; ε_{yt}, ε_{zt} are the stationary and firm-specific components of the time series. If the two time series are cointegrated, then their common trends must be identical up to a scalar,

$$n_{yt} = \gamma n_{zt} \tag{6.4}$$

where γ is the cointegrating coefficient.

In a cointegrated system with two individual time series, the innovations sequences derived from the common trend components must be perfectly correlated. This can be practically interpreted as the correlation coefficient (ρ) being equal to positive or negative unity. We will denote the innovation sequences derived from the common trends of the two series as r_{yt} and r_{zt}. The innovation sequence for a random walk is obtained by simply taking first differences. In equation form, we have:

$$n_{yt+1} - n_{yt} = r_{yt+1} \tag{6.5}$$
$$n_{zt+1} - n_{zt} = r_{zt+1}$$

According to the common trends model, cointegration requires that the common trend component of each stock price must be identical up to a scalar:

$$n_{yt} = \gamma n_{zt} \tag{6.6}$$

Moving forward a single period this relationship should still hold so that

$$n_{yt+1} = \gamma n_{zt+1} \tag{6.7}$$

and now it should be clear that from (6.6) and (6.7) that

$$r_{yt+1} = \gamma r_{zt+1} \tag{6.8}$$

This means that if two time series are cointegrated, then according to the common trends model, their innovations must also be identical up to a scalar. Now, if two variables are identical up to a scalar (in this case the cointegrating coefficient (γ)), they must be perfectly correlated.
Thus, in a cointegrated system the innovation sequences derived from the common trends must also be perfectly correlated.

Based on our discussion so far, we have established a method for determining the cointegrating coefficient (γ). The cointegrating coefficient may be obtained from regressing the innovation sequences of the common trends against each other. Up to this point, we have established that we have a linear relationship between the innovation sequences, given as

$$r_{yt} = \gamma r_{zt} \tag{6.9}$$

Simply regressing one innovation sequence against the other yields the cointegrating coefficient

$$\gamma = \frac{cov(r_y, r_z)}{var(r_z)} \tag{6.10}$$

In summary, there are two key conditions which must be satisfied for the existence of a cointegrating relationship in a common trends model. Firstly, the innovation sequences derived from the common trends of the two series must be identical up to a scalar. Secondly, the firm-specific components of the individual stock price series must be stationary. The common trends component may be stationary, or non-stationary in the presence of cointegration.

Common Trends Model and APT

We start this analysis by realising that the log of stock prices can also be decomposed into a random walk component (non-stationary) and a stationary component:

$$\log(p_t) = n_t + \varepsilon_t \tag{6.11}$$

where n_t is the random walk, and ε_t is the stationary component. Differencing the log of stock price yields the sequence of returns. Therefore, based on equation (6.11), the return r_t, at time t may also separated into two parts

$$\log(p_t) - \log(p_{t-1}) = n_t - n_{t-1} + (\varepsilon_t - \varepsilon_{t-1}) \tag{6.12}$$

$$r_t = r_t^c + r_t^s \tag{6.13}$$

where r_t^c is the return due to the non-stationary trend component, and r_t^s is the return due to the stationary component.

It is interesting to note that the return due to the trend component (r_t^c) is identical to the innovation derived from the trend component. Therefore, the cointegration requirements pertaining to the innovations of the common trend may be rephrased as follows: If two stocks are cointegrated, then the returns from their common trends must be identical up to a scalar.

At this stage, one may naturally query why we would ever expect two individual stocks to have the same return? The Arbitrage Pricing Theory (APT) model (Ross, 1976) can be used to explain why certain stocks can be expected to generate identical return payoffs. The APT model says that stock returns can be separated into common factor returns (returns based on the exposure of stocks to different risk factors)[11]. If two stocks share the same risk factor exposure profile, then the common factor returns for both the stocks must be the same. This provides us with an economic rationale for the circumstances under which we might expect stocks to share a common return component.

We are now in a position where we can reconcile the APT model with the common trends model. According to APT, stock returns for a single time period can be decomposed into two types: common factor returns and firm-specific returns. Let these correspond to the common trend innovation and the first difference of the specific component in the common trends model (6.12). For the correspondence to be valid, the integration of the specific returns must be a stationary process.

Alternatively, as we saw earlier, the specific returns (r_t^s) must not be white noise. Since the APT is a static model and cannot provide us with any guarantee pertaining to the dynamic behaviour of the time series of specific returns. As a result, we must make this assumption that the specific returns are not white noise. It is reassuring, to a certain extent, to know that the validity of this assumption is tested when running the cointegration tests and pairs where the specific component is non-stationary are eliminated. We can now interpret the inferences from the common trends model in APT terms. The correlation of the innovation sequence is the common factor correlation.

[11] Factors to use in the APT are extensive and varied, however, should be set so that the unexplained component, the firm-specific return is relatively small. See Fama and French (1993) 3-factor model for a widely used combination of factors used in the literature.

It can now be said that a pair of stocks with the same risk factor exposure profiles, satisfies the necessary conditions for cointegration. Consider two stocks A and B with risk factor exposure vectors γβ and β, respectively. The factor exposure vectors in this case are identical up to a scalar. We denote the factor exposures as

Stock A: $\gamma\beta = (\gamma\beta_1, \gamma\beta_2, \gamma\beta_3, \dots, \gamma\beta_n)$

Stock B: $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_n)$

Geometrically, it may be interpreted that the factor exposure vectors of the two stocks point towards the same direction; that is, the angle between them is zero.

If $x=(x_1,x_2,x_3,\dots x_n)$ is the factor returns vector, and r_A^{spec} and r_B^{spec} are the specific returns for stocks A and B, then the returns for the stocks r_A and r_B are given as

$$r_A = \gamma(\beta_1 x_1 + \beta_2 x_2 +, \dots, +\beta_n x_n) + r_A^{spec}$$
$$r_B = (\beta_1 x_1 + \beta_2 x_2 +, \dots, +\beta_n x_n) + r_B^{spec}$$

The common factor returns for the stocks are therefore

$$r_A^{cf} = \gamma(\beta_1 x_1 + \beta_2 x_2 +, \dots, +\beta_n x_n)$$
$$r_B^{cf} = (\beta_1 x_1 + \beta_2 x_2 +, \dots, +\beta_n x_n)$$

Thus, $r_A^{cf} = \gamma r_B^{cf}$. The innovation sequences of the common trend are identical up to a scalar. This satisfies the first condition for cointegration. Additionally, the spread series must be stationary for cointegration (r_A-γr_B).

As we established previously, a necessary condition of cointegration is that the two stocks share an identical risk factor vector. Geometrically, this can be interpreted as the requirement that the vectors point towards the same direction. Thus, a key area to look at for a measure of cointegration is the risk factor exposure profiles of the individual stocks and how closely aligned they are.

The Distance Measure

Recall from the discussion on the common trends model that the necessary condition for cointegration is that the innovation sequences derived from the common trends must be perfectly correlated. We also established that the common factor return of the APT model might be interpreted as the innovations derived from the common trends. The correlation between the innovation sequence sequences is therefore the correlation between the common factor returns. The closer the absolute value of this measure is to unity, the greater will be the degree of co-movement. The distance measure proposed in Vidyamurthy (2004) is exactly that: the absolute value of the correlation of the common factor returns. The formula for the distance measure is therefore given as

$$|\rho| = \left|\frac{x_A F x_B}{(x_A F x_B)(x_B F x_B)}\right| \tag{6.14}$$

where x_A and x_B are the factor exposure vectors of the two stocks A and B, and F is the covariance matrix.

Interpreting the Distance Measure

In subsequent sections it was hinted that perfect alignment of the factor exposure vectors, that is, a zero angle between them, is indicative of cointegration. Vidyamurthy (2004) transforms the factor exposure vectors from the space of factor exposures to the space of returns and then measures the angle between the transformed vectors. This is necessary since all the factors in the multi-factor model are not created equal. Returns are more sensitive to changes in some factors versus others. Vidyamurthy (2004) shows that

$$cos\theta = \frac{cov(r_A, r_B)}{\sqrt{var(r_A) var(r_B)}} = \rho \tag{6.15}$$

Thus, it is clear that when $\theta=0$, then $\rho=1$ and we have a cointegrating relationship between the two stocks.

7. Conclusion

The purpose of this study was to develop a method for selecting trading pairs which could then be used in a pairs trading strategy. Upon the identification of the pairs, an analysis of the co-movement between the stocks was conducted with the aim of providing insight into the dynamic behaviour of the stocks, which would be valuable to an institutional investor.

This study employed the Johansen test (1988) to identify cointegrated stock pairs which share a long-run equilibrium pricing relationship. The search for trading pairs was limited to 17 financial stocks trading on the ASX200. We restricted our trading pairs to the most liquid segment of the ASX to reduce the adverse effects associated with "moving" illiquid shares, such as increased transactions costs and market impact costs. By constraining the search to include only stocks from within the same industry group, it was likely that their prices were driven by a common set of fundamental factors. As a result, it is more likely that any cointegrating relationships identified in-sample, will also remain significant out-of-sample.

The study identified 5 (from a possible 136) trading pairs which were cointegrated using both daily prices, and weekly prices. We then estimated the long-run equilibrium relationship between each of the pairs and obtained the residual series. These residual series were all tested for stationarity using the ADF test. For each of these tests, the null hypothesis that the residual series had a unit root (which implies non-stationarity) was rejected, and thus, we concluded that each of the residual series was a stationary process. This finding was consistent with our hypothesis since a cointegrating relationship between two I(1) processes is only meaningful if the resulting residual spread is stationary.

Each of these residual series was modeled as a Vector-Error-Correction model (VECM) and variance analysis conducted for each. It can be concluded that in the presence of a cointegrating relationship between two time series, at least one of the speed of adjustments coefficients must be significantly different from zero. This ensured that the system mean-reverted following a deviation from its long-run equilibrium. It can also be concluded that shocks to one of the time series can also have an effect on the other time series. This impact is generally only statistically significant when shocking the leading variable. For example, suppose x_t leads y_t, then shocking x_t obviously effects its own future time path, but also the time path of y_t since x_t is a key factor explaining the dynamics of y_t. However, we generally find that the impact on x_t from shocking y_t is little different from zero. The contemporaneous impact on x_t from a shock to y_t is always zero as this is a straight forward result of the Choleski decomposition.

There are two main directions for future research which could extend this study. The purpose of any trading strategy should be to maximise return, for a given risk tolerance. The purpose of this study was not to access the profitability of this pairs trading approach. Nevertheless, this remains its ultimate goal. It would be interesting to discuss trading rules and risk management devices to determine if this approach to pairs trading could be used by professional money managers to profit. Secondly, a key assumption of the analysis conducted in this study is that the error terms from the cointegrating equation conformed to the assumptions of OLS. These assumptions can be summarised by stating that the error term is required to be normally distributed. This study introduced the idea of a Common Trends Model (CTM) and attempted to reconcile it to the Arbitrage Pricing Theory (APT). Future research could look at the ability of the CTM and APT to overcome the issue of autocorrelation in the residual series and the problems it provides for coefficient interpretation.

Bibliography

Alexander, C. and A. Dimitriu, 2002, The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies, *working paper* (ISMA Center of the University of Reading, UK)

Alexander, C. and A. Dimitriu, 2005, Index and Statistical Arbitrage: Tracking error or Cointegration? *working paper* (SMA Center of the University of Reading, UK)

Alexander, C., I. Giblin and W. Weddington III, 2001, Cointegration and asset management: A new active hedge fund strategy, *working paper* (SMA Center of the University of Reading, UK)

Bedi, J. and P. Tennant, 2002, Dual-listed Companies, *Reserve Bank of Australian Bulletin,* October

Beveridge, S. and C.R. Nelson, 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, pp. 151-74

Black, F., 1972, Capital Market Equilibrium and Restricted Borrowing, *Journal of Business* 45, pp.444-454

Brooks, C. and H. Kat, 2001, The Statistical Properties of Hedge Fund Index Returns, ISMA Centre Discussion Papers in Finance 2001-09, forthcoming *Journal of Alternative Investments,* 2002

Chiarella, C., S. Gao and M. Stevenson, 2008, Resolving the Price-Earnings Puzzle and Related Econometric Issues, *working paper* (Faculty of Economics and Business, University of Sydney)

Do, B., R. Faff and K. Hamza, 2006, A new approach to modeling and estimation for pairs trading", *working paper* (Monash University)

Durbin, J. and S. Koopman, 2001, *Time Series Analysis by State Space Models*, Oxford University Press

Elliott, R., J. Van der Hoe and W. Malcolm, 2005, Pairs Trading, *Quantitative Finance* 5, (3), pp. 271-276.

Enders, W.,1995, *Applied Econometric Time Series*, John Wiley & Sons, New York

Engle, R.F and C.W.J. Granger, 1987, Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica 55,* (2), pp.251-76

Engle, R. and C. Granger, 1987, Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica 55,* (2), pp. 251-276

Engle, R. and B. Yoo, 1987, Forecasting and Testing in Co-integrated Systems, *Journal of Econometrics* 35, pp. 143-159

Fama, E.F. and K. R. Fama, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, pp. 3-56

Fund, W. and D. Hsieh, 1999, A primer of hedge funds, *Journal of Empirical Finance* 6, pp.309-331

Gatev, E., G., W. Goetzmann and K. Rouwenhorst, 1999, Pairs Trading: Performance of a Relative Value Arbitrage Rule, *working paper* (Yale School of Management)

Gujarati, D., 2003, *Basic Econometrics (4th ed.),* McGraw-Hill/Irwin, New York

Jacobs, B. and K. Levy, 1993, Long/Short Equity Investing, *Journal of Portfolio Management* 20 (1), pp. 52-64

Jacobs, B., K. Levy and D. Starer, 1998, Long-Short Portfolio Management: An Integrated Approach, *Journal of Portfolio Management,* Winter, pp. 23-32.

Jacobs, B., K. Levy and D. Starer, 1998, On the Optimality of Long-Short Strategies", Financial Analysts Journal, Vol. 54(2), pp. 40-50

Jensen, M., 1969, The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance* 23, pp.389-416.

Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control* 12, pp. 231-254

Johansen, S. and K. Juselius, 1990, Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics* 52, pp. 169-210

Kalman, R.E., 1960, A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering* 82, pp.35-45.

Lim, G and V. Martin, 1995, Regression-based Cointegration Estimators, *Journal of Economic Studies* 22 (1), pp. 3-22

Lin, Y., M. McRae and C. Gulati, 2006, Loss Protection in Pairs Trading through Minimum Profit Bounds: A Cointegration Approach, *Journal of Applied Mathematics and Decision Sciences 6*, pp.1-14

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, pp.13-37

Markowitz, H., 1959, *Portfolio Selection: Efficient Diversification of Investments,* John Wiley & Sons, New York

Nath, P., 2003, High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds, *working paper* (London Business School)

Rice, S.O., 1945, Mathematical Analysis of Random Noise, *Bell System Technical Journal* 24, pp.146-156

Ross, S., 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory* 13, pp. 341-360

Sharpe, W., 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance* 19, pp. 425-442

Sims, C.A., 1980, Macroeconomics and Reality, *Econometrica* 48, pp.1-48

Stock, J.H. and M.W. Watson, 1991, Variable Trends in Economic Time Series in R.F. Engle and C.W.J. Granger, *Long-run Economic Relationships*, Oxford University Press, pp. 17-50

Stock, J.H. and M.W. Watson, 1988, Testing for Common Trends, *Journal of the American Statistical Association* 83 (404) pp. 1097-1107

Vidyamurthy, G., 2004, Pairs Trading, *Quantitative Methods and Analysis*, John Wiley & Sons, Canada

8. Appendix

Appendix A – Results for Johansen Tests

Date: 09/03/08 Time: 13:07
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: AMP ANZ
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.015383	19.97145	15.49471	0.0099
At most 1	0.000100	0.128265	3.841466	0.7202

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 13:15
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: SGB WBC
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.018584	24.28758	15.49471	0.0018
At most 1	0.000216	0.276517	3.841466	0.5990

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 13:34
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: WBC FKP
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.015380	19.98666	15.49471	0.0098
At most 1	0.000116	0.147868	3.841466	0.7006

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 13:38
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: ASX LLC
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.014721	20.76646	15.49471	0.0073
At most 1	0.001393	1.784160	3.841466	0.1816

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 11:43

Sample (adjusted): 3 1285

Included observations: 1283 after adjustments

Trend assumption: Linear deterministic trend

Series: CBA ASX

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.018280	26.56725	15.49471	0.0007
At most 1	0.002255	2.897072	3.841466	0.0887

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 13:48

Sample (adjusted): 6 1285

Included observations: 1280 after adjustments

Trend assumption: Linear deterministic trend

Series: SGB ASX

Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.013258	20.58040	15.49471	0.0078
At most 1	0.002728	3.496438	3.841466	0.0615

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 14:12

Sample (adjusted): 3 1285

Included observations: 1283 after adjustments

Trend assumption: Linear deterministic trend

Series: BEN AMP

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.016535	21.39979	15.49471	0.0057
At most 1	6.37E-06	0.008178	3.841466	0.9275

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 14:28

Sample (adjusted): 6 1285

Included observations: 1280 after adjustments

Trend assumption: Linear deterministic trend

Series: AMP ASX

Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.014006	20.31862	15.49471	0.0087
At most 1	0.001767	2.263994	3.841466	0.1324

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 10:39
Sample (adjusted): 3 1285
Included observations: 1283 after adjustments
Trend assumption: Linear deterministic trend
Series: AMP PPT
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.014517	19.32169	15.49471	0.0126
At most 1	0.000437	0.560491	3.841466	0.4541

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 11:22
Sample (adjusted): 6 1285
Included observations: 1280 after adjustments
Trend assumption: Linear deterministic trend
Series: AMP QBE
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.015424	20.93675	15.49471	0.0068
At most 1	0.000812	1.039983	3.841466	0.3078

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Date: 09/03/08 Time: 11:38

Sample (adjusted): 3 1285

Included observations: 1283 after adjustments

Trend assumption: Linear deterministic trend

Series: AMP CBA

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.015126	19.79985	15.49471	0.0105
At most 1	0.000191	0.245404	3.841466	0.6203

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Appendix B – Estimation output for Granger Causality testing

Pairwise Granger Causality Tests

Date: 10/21/08 Time: 10:52

Sample: 1 1285

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
AMP does not Granger Cause ANZ	1283	2.11324	0.12127
ANZ does not Granger Cause AMP		8.79338	0.00016

Pairwise Granger Causality Tests
Date: 10/21/08 Time: 10:45
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
WBC does not Granger Cause SGB	1283	2.47773	0.08434
SGB does not Granger Cause WBC		4.93901	0.00730

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:23
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
FKP does not Granger Cause WBC	1283	8.10548	0.00032
WBC does not Granger Cause FKP		1.64525	0.19337

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:24
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
ASX does not Granger Cause LLC	1283	2.55930	0.07776
LLC does not Granger Cause ASX		4.75226	0.00879

Pairwise Granger Causality Tests

Date: 09/03/08 Time: 11:42

Sample: 1 1285

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
ASX does not Granger Cause CBA	1283	11.3852	1.3E-05
CBA does not Granger Cause ASX		0.12564	0.88194

Pairwise Granger Causality Tests

Date: 09/03/08 Time: 12:28

Sample: 1 1285

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
ASX does not Granger Cause SGB	1283	3.74302	0.02394
SGB does not Granger Cause ASX		3.63622	0.02662

Pairwise Granger Causality Tests

Date: 09/03/08 Time: 12:29

Sample: 1 1285

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
AMP does not Granger Cause BEN	1283	3.24258	0.03938
BEN does not Granger Cause AMP		8.96944	0.00014

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:30
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
ASX does not Granger Cause BEN	1283	9.78929	6.0E-05
BEN does not Granger Cause ASX		9.62176	7.1E-05

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:31
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
AMP does not Granger Cause AXA	1283	0.66245	0.51577
AXA does not Granger Cause AMP		8.59394	0.00020

Pairwise Granger Causality Tests
Date: 09/03/08 Time: 12:32
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
ASX does not Granger Cause AMP	1283	8.24753	0.00028
AMP does not Granger Cause ASX		0.88969	0.41104

Pairwise Granger Causality Tests
Date: 09/03/08   Time: 10:38
Sample: 1 1285
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
PPT does not Granger Cause AMP	1283	9.73983	6.3E-05
AMP does not Granger Cause PPT		0.99281	0.37082

Pairwise Granger Causality Tests
Date: 09/03/08   Time: 11:23
Sample: 1 1285
Lags: 5

Null Hypothesis:	Obs	F-Statistic	Probability
QBE does not Granger Cause AMP	1280	4.26516	0.00075
AMP does not Granger Cause QBE		0.78398	0.56121

Pairwise Granger Causality Tests
Date: 09/03/08   Time: 11:38
Sample: 1 1285
Lags: 2

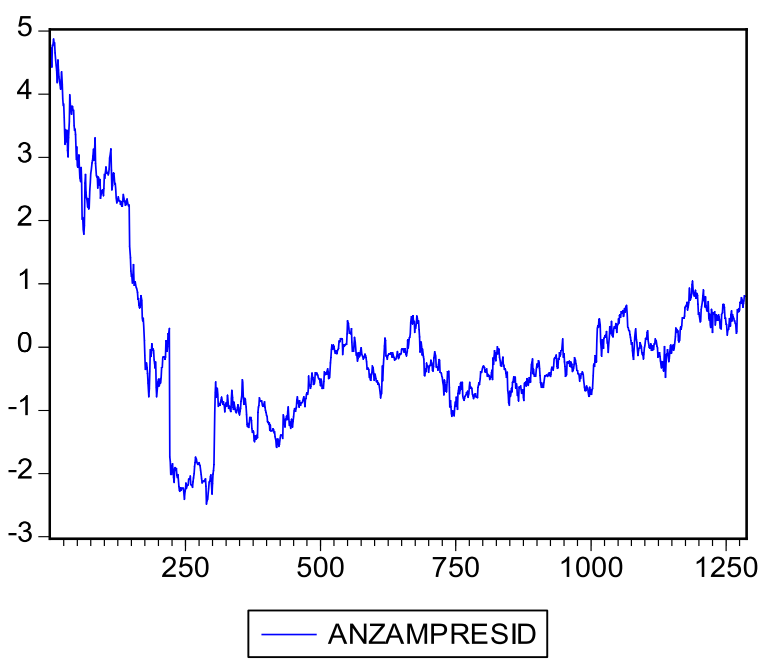
Null Hypothesis:	Obs	F-Statistic	Probability
CBA does not Granger Cause AMP	1283	10.4315	3.2E-05
AMP does not Granger Cause CBA		4.25640	0.01437

Appendix C – Output from cointegrating equation, plot of residuals and ADF test

Dependent Variable: AMP
Method: Least Squares
Date: 10/20/08 Time: 15:11
Sample: 1 1285
Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ANZ	0.359272	0.008397	42.78823	0.0000
C	-1.062734	0.184118	-5.772041	0.0000

R-squared	0.587968	Mean dependent var	6.660409
Adjusted R-squared	0.587646	S.D. dependent var	2.028269
S.E. of regression	1.302448	Akaike info criterion	3.367924
Sum squared resid	2176.444	Schwarz criterion	3.375952
Log likelihood	-2161.891	F-statistic	1830.832
Durbin-Watson stat	0.009208	Prob(F-statistic)	0.000000



Null Hypothesis: ANZAMPRES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.224803	0.0800
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: WBC

Method: Least Squares

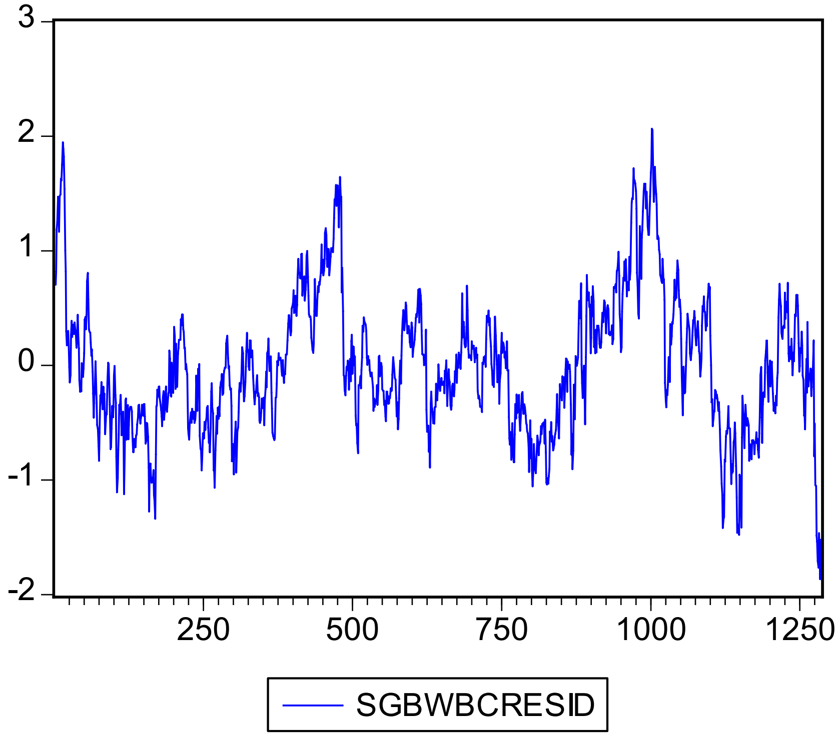
Date: 10/20/08 Time: 15:20

Sample: 1 1285

Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.037958	0.080508	25.31372	0.0000
SGB	0.693041	0.003176	218.2066	0.0000

R-squared	0.973761	Mean dependent var	19.22138
Adjusted R-squared	0.973741	S.D. dependent var	3.703152
S.E. of regression	0.600085	Akaike info criterion	1.818063
Sum squared resid	462.0101	Schwarz criterion	1.826092
Log likelihood	-1166.105	F-statistic	47614.12
Durbin-Watson stat	0.090896	Prob(F-statistic)	0.000000



Null Hypothesis: SGBWBCRES has a unit root

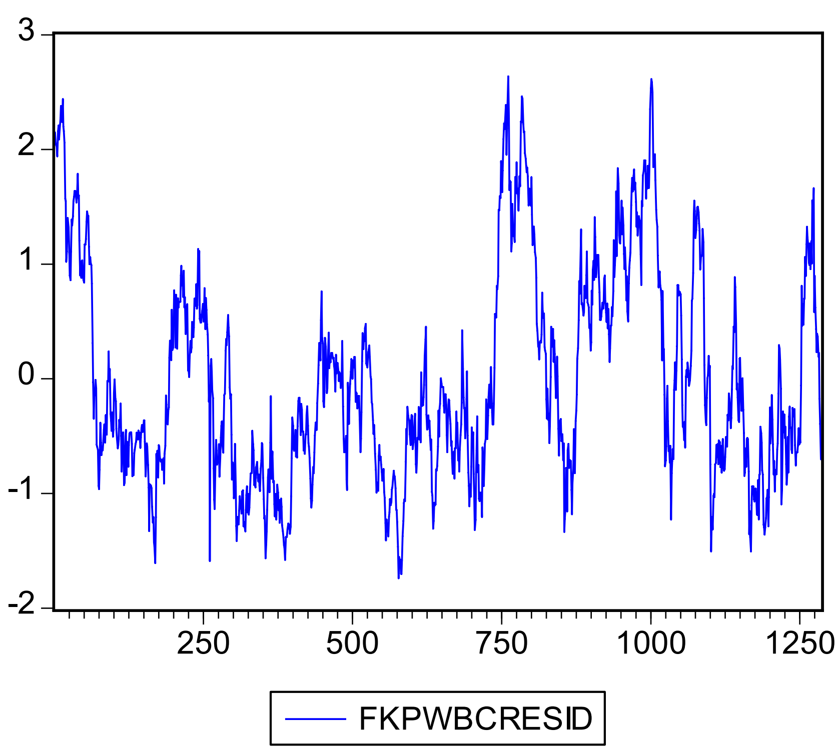
Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.236605	0.0001
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: WBC				
Method: Least Squares				
Date: 10/20/08 Time: 15:21				
Sample: 1 1285				
Included observations: 1285				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.66849	0.060435	193.0760	0.0000
FKP	2.056036	0.014863	138.3292	0.0000
R-squared	0.937163	Mean dependent var		19.22138
Adjusted R-squared	0.937114	S.D. dependent var		3.703152
S.E. of regression	0.928641	Akaike info criterion		2.691366
Sum squared resid	1106.425	Schwarz criterion		2.699394
Log likelihood	-1727.202	F-statistic		19134.97
Durbin-Watson stat	0.061816	Prob(F-statistic)		0.000000



Null Hypothesis: FKPWBCRES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.815558	0.0004
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: ASX

Method: Least Squares

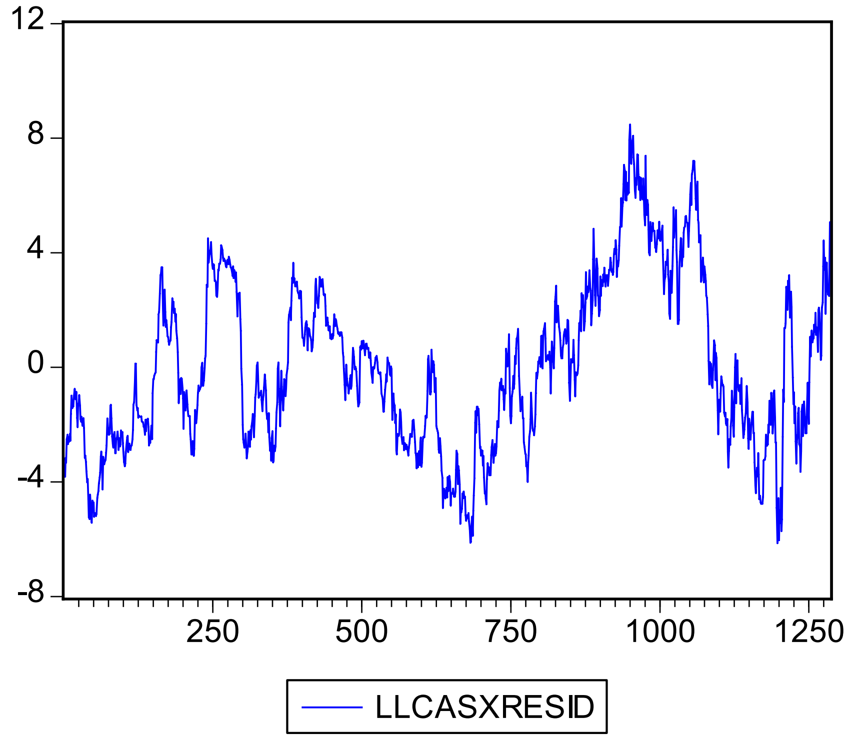
Date: 10/20/08 Time: 15:23

Sample: 1 1285

Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LLC	3.238890	0.026917	120.3284	0.0000
C	-18.35087	0.348484	-52.65920	0.0000

R-squared	0.918601	Mean dependent var	22.41877
Adjusted R-squared	0.918538	S.D. dependent var	10.23579
S.E. of regression	2.921457	Akaike info criterion	4.983597
Sum squared resid	10950.29	Schwarz criterion	4.991626
Log likelihood	-3199.961	F-statistic	14478.92
Durbin-Watson stat	0.039223	Prob(F-statistic)	0.000000



Null Hypothesis: LLCASXRESID has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

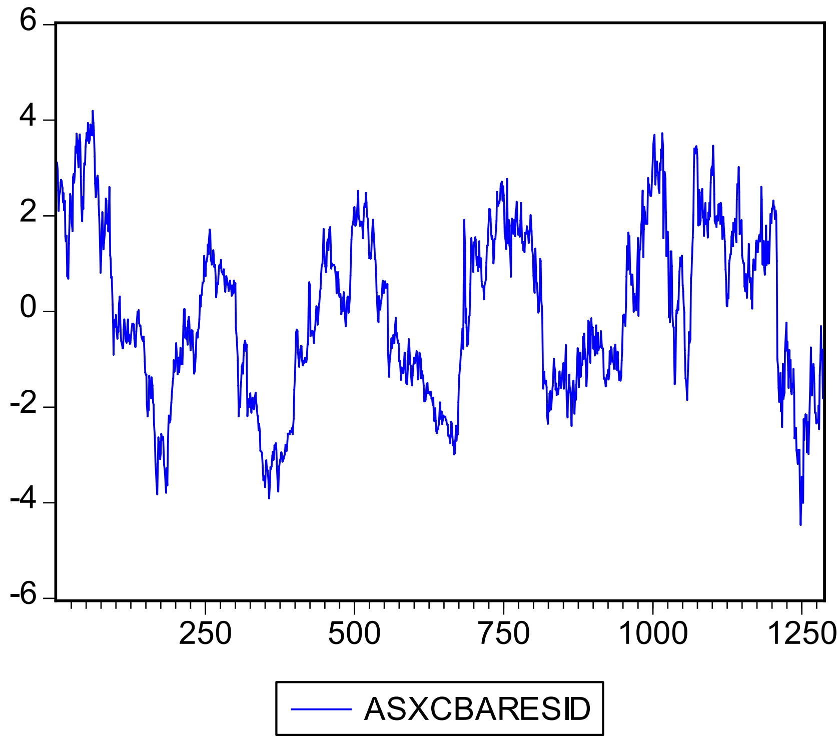
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.459715	0.0093
Test critical values:	1% level	-3.435231	
	5% level	-2.863583	
	10% level	-2.567907	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: CBA
Method: Least Squares
Date: 09/03/08 Time: 11:43
Sample: 1 1285
Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ASX	0.761352	0.004831	157.6086	0.0000
C	19.19414	0.119043	161.2370	0.0000

R-squared	0.950887	Mean dependent var	36.26271
Adjusted R-squared	0.950849	S.D. dependent var	7.991758
S.E. of regression	1.771778	Akaike info criterion	3.983400
Sum squared resid	4027.591	Schwarz criterion	3.991428
Log likelihood	-2557.334	F-statistic	24840.48
Durbin-Watson stat	0.043269	Prob(F-statistic)	0.000000



Null Hypothesis: ASXCBARES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.894820	0.0125
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

Dependent Variable: AMP

Method: Least Squares

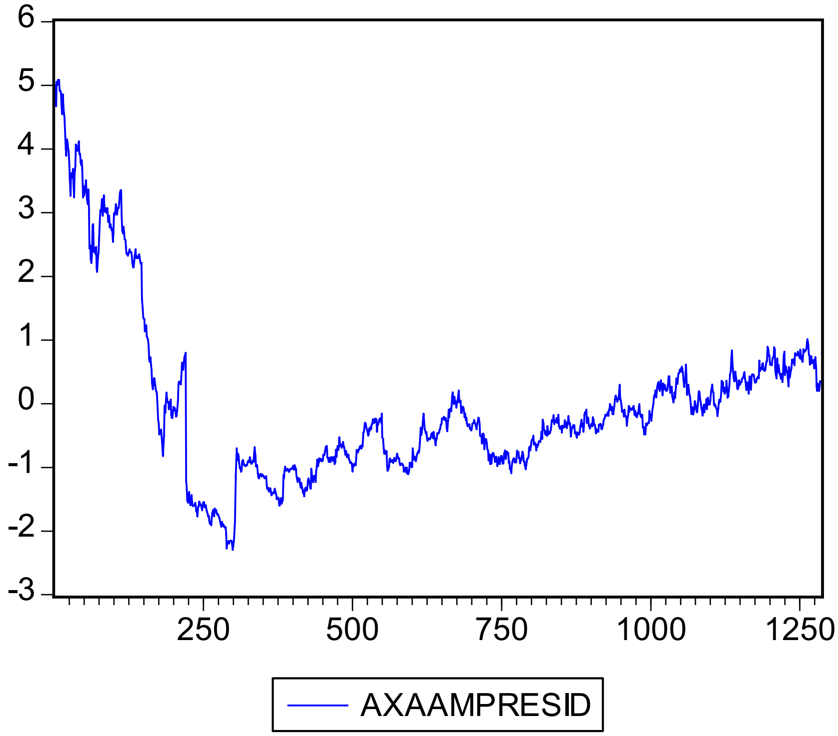
Date: 10/20/08 Time: 18:46

Sample: 1 1285

Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AXA	0.895630	0.022182	40.37627	0.0000
C	2.861072	0.101319	28.23829	0.0000

R-squared	0.559597	Mean dependent var	6.660409
Adjusted R-squared	0.559254	S.D. dependent var	2.028269
S.E. of regression	1.346541	Akaike info criterion	3.434511
Sum squared resid	2326.302	Schwarz criterion	3.442540
Log likelihood	-2204.673	F-statistic	1630.243
Durbin-Watson stat	0.007529	Prob(F-statistic)	0.000000



Null Hypothesis: AXAAMPRES has a unit root

Exogenous: Constant, Linear Trend

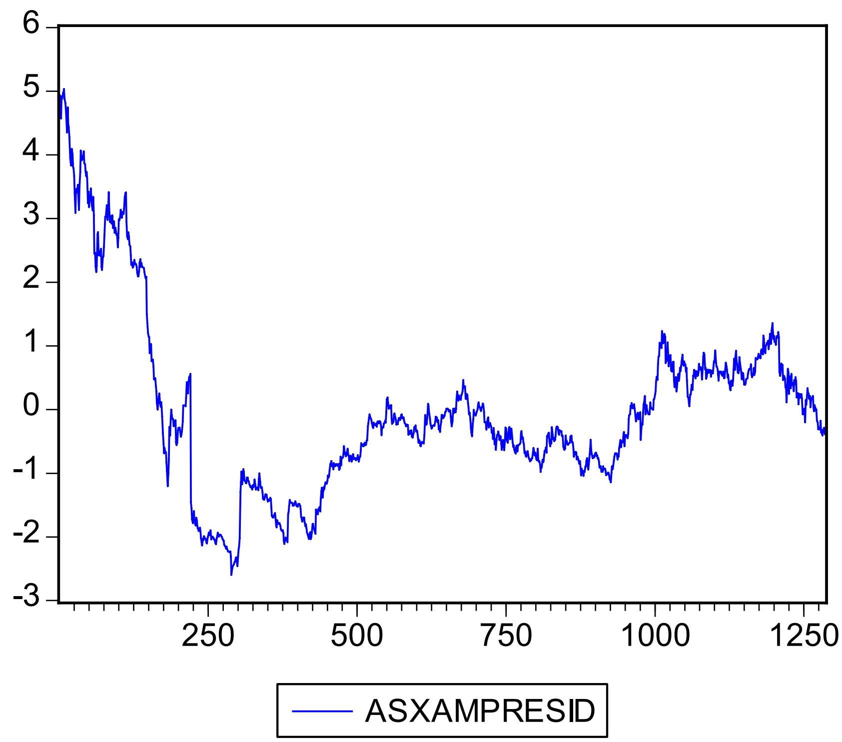
Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.398617	0.0519
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

Dependent Variable: AMP				
Method: Least Squares				
Date: 10/20/08 Time: 18:48				
Sample: 1 1285				
Included observations: 1285				

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ASX	0.142731	0.003837	37.19484	0.0000
C	3.460546	0.094566	36.59395	0.0000

R-squared	0.518837	Mean dependent var	6.660409
Adjusted R-squared	0.518462	S.D. dependent var	2.028269
S.E. of regression	1.407475	Akaike info criterion	3.523027
Sum squared resid	2541.606	Schwarz criterion	3.531056
Log likelihood	-2261.545	F-statistic	1383.456
Durbin-Watson stat	0.007795	Prob(F-statistic)	0.000000



Null Hypothesis: ASXAMPRES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.321495	0.0632
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: AMP

Method: Least Squares

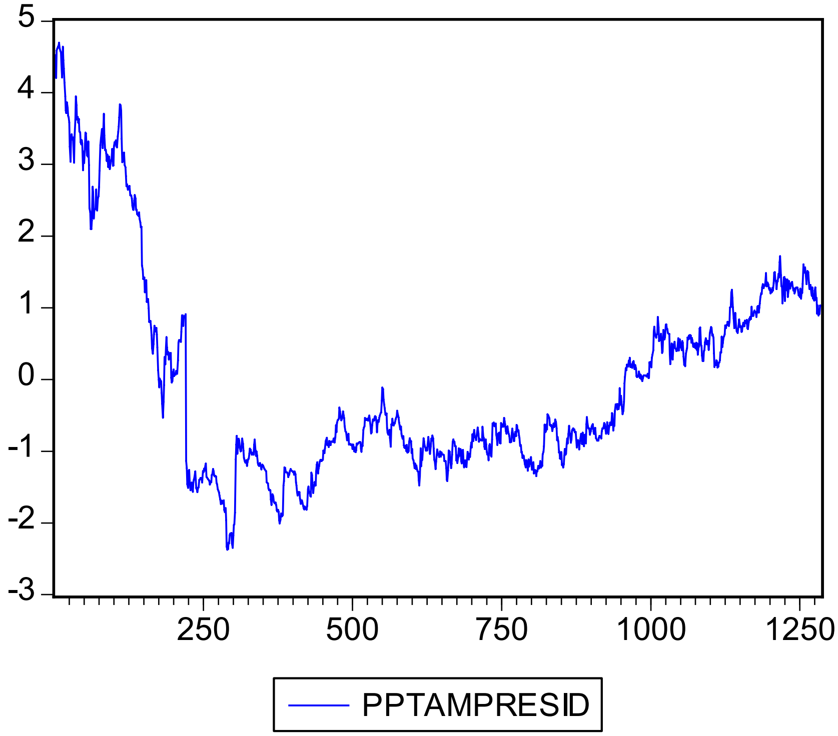
Date: 10/20/08   Time: 18:49

Sample: 1 1285

Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PPT	0.087008	0.002509	34.67915	0.0000
C	1.975508	0.141081	14.00267	0.0000

R-squared	0.483836	Mean dependent var	6.660409
Adjusted R-squared	0.483434	S.D. dependent var	2.028269
S.E. of regression	1.457769	Akaike info criterion	3.593246
Sum squared resid	2726.490	Schwarz criterion	3.601275
Log likelihood	-2306.661	F-statistic	1202.644
Durbin-Watson stat	0.007226	Prob(F-statistic)	0.000000



Null Hypothesis: PPTAMPRES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

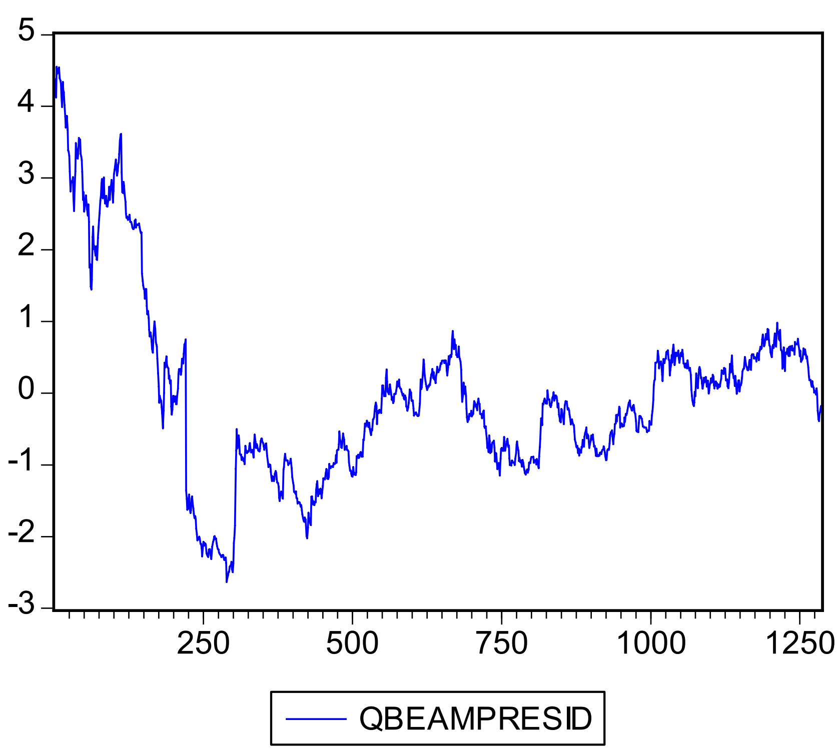
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.803079	0.1963
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Dependent Variable: AMP
Method: Least Squares
Date: 10/20/08 Time: 18:51
Sample: 1 1285
Included observations: 1285

Variable	Coefficient	Std. Error	t-Statistic	Prob.
QBE	0.194123	0.004564	42.53079	0.0000
C	-0.379003	0.169482	-2.236241	0.0255

R-squared	0.585040	Mean dependent var	6.660409
Adjusted R-squared	0.584717	S.D. dependent var	2.028269
S.E. of regression	1.307066	Akaike info criterion	3.375002
Sum squared resid	2191.905	Schwarz criterion	3.383031
Log likelihood	-2166.439	F-statistic	1808.868
Durbin-Watson stat	0.008541	Prob(F-statistic)	0.000000



Null Hypothesis: QBEAMPRES has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic based on SIC, MAXLAG=22)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.572623	0.0326
Test critical values:	1% level	-3.965209	
	5% level	-3.413315	
	10% level	-3.128686	

*MacKinnon (1996) one-sided p-values.

Appendix D – Estimation output from VECM

Vector Autoregression Estimates

Date: 10/21/08 Time: 11:24

Sample (adjusted): 4 1285

Included observations: 1282 after adjustments

Standard errors in ( ) & t-statistics in [ ]

	DSGB	DWBC
DSGB(-1)	-0.077706	-0.007156
	(0.03228)	(0.02623)
	[-2.40702]	[-0.27279]
DSGB(-2)	-0.037789	0.009676
	(0.03230)	(0.02624)
	[-1.17001]	[0.36870]
DWBC(-1)	0.030074	-0.005002
	(0.03975)	(0.03230)
	[0.75661]	[-0.15487]
DWBC(-2)	0.026602	-0.045826
	(0.03968)	(0.03224)
	[0.67043]	[-1.42125]
C	0.015172	0.008326
	(0.00658)	(0.00535)
	[2.30445]	[1.55625]
SGBWBCRESID	-0.023285	0.035614
	(0.01121)	(0.00911)
	[-2.07727]	[3.90978]
R-squared	0.007976	0.013049
Adj. R-squared	0.004088	0.009182
Sum sq. resids	70.35680	46.45681
S.E. equation	0.234816	0.190809
F-statistic	2.051757	3.374213
Log likelihood	41.48568	307.5368

Vector Autoregression Estimates
Date: 10/21/08 Time: 11:30
Sample (adjusted): 4 1285
Included observations: 1282 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	DFKP	DWBC
DFKP(-1)	-0.147828	0.069899
	(0.02839)	(0.06812)
	[-5.20716]	[1.02615]
DFKP(-2)	-0.082727	-0.153082
	(0.02843)	(0.06822)
	[-2.90959]	[-2.24388]
DWBC(-1)	0.002461	-0.005100
	(0.01184)	(0.02842)
	[0.20782]	[-0.17947]
DWBC(-2)	0.008626	-0.021119
	(0.01181)	(0.02834)
	[0.73026]	[-0.74511]
C	0.005837	0.008553
	(0.00223)	(0.00536)
	[2.61351]	[1.59602]
FKPWBCRESID	-0.008607	0.014394
	(0.00243)	(0.00584)
	[-3.53750]	[2.46542]
R-squared	0.031899	0.011307
Adj. R-squared	0.028106	0.007433
Sum sq. resids	8.083477	46.53884
S.E. equation	0.079593	0.190978
F-statistic	8.408992	2.918471
Log likelihood	1428.454	306.4060

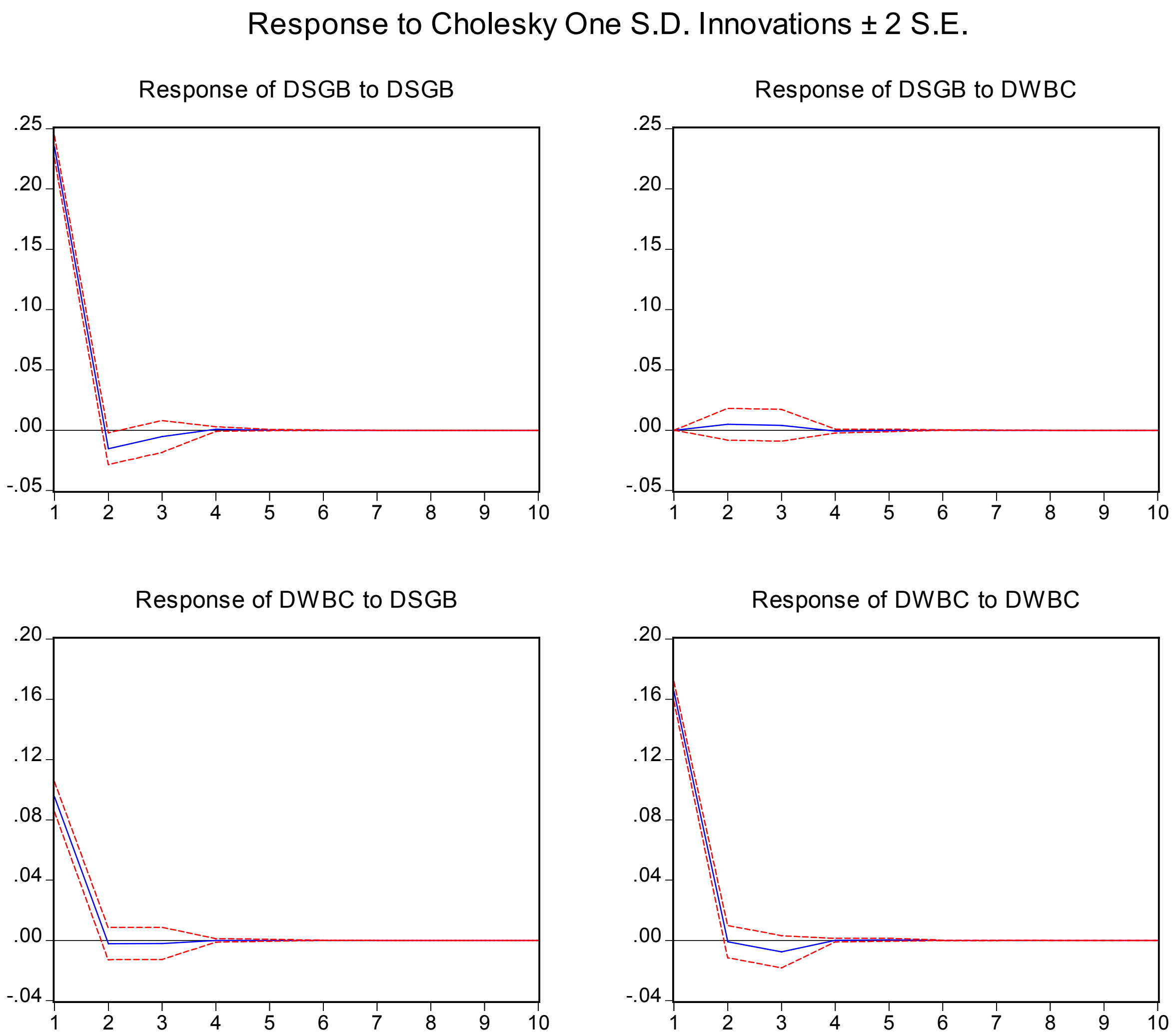
Vector Autoregression Estimates
Date: 10/21/08 Time: 11:33
Sample (adjusted): 4 1285
Included observations: 1282 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	DLLC	DASX
DLLC(-1)	0.007376	0.027250
	(0.02885)	(0.06126)
	[0.25569]	[0.44480]
DLLC(-2)	-0.085669	-0.022473
	(0.02881)	(0.06119)
	[-2.97331]	[-0.36726]
DASX(-1)	-0.006931	-0.018581
	(0.01360)	(0.02888)
	[-0.50963]	[-0.64329]
DASX(-2)	0.029098	-0.070275
	(0.01360)	(0.02888)
	[2.14005]	[-2.43368]
C	0.006802	0.030894
	(0.00472)	(0.01003)
	[1.44033]	[3.08028]
LLCASXRESID	-0.006261	0.001263
	(0.00162)	(0.00345)
	[-3.85693]	[0.36625]
R-squared	0.018990	0.005878
Adj. R-squared	0.015146	0.001982
Sum sq. resids	35.99761	162.3596
S.E. equation	0.167962	0.356709
F-statistic	4.940188	1.508860
Log likelihood	471.0371	-494.5404

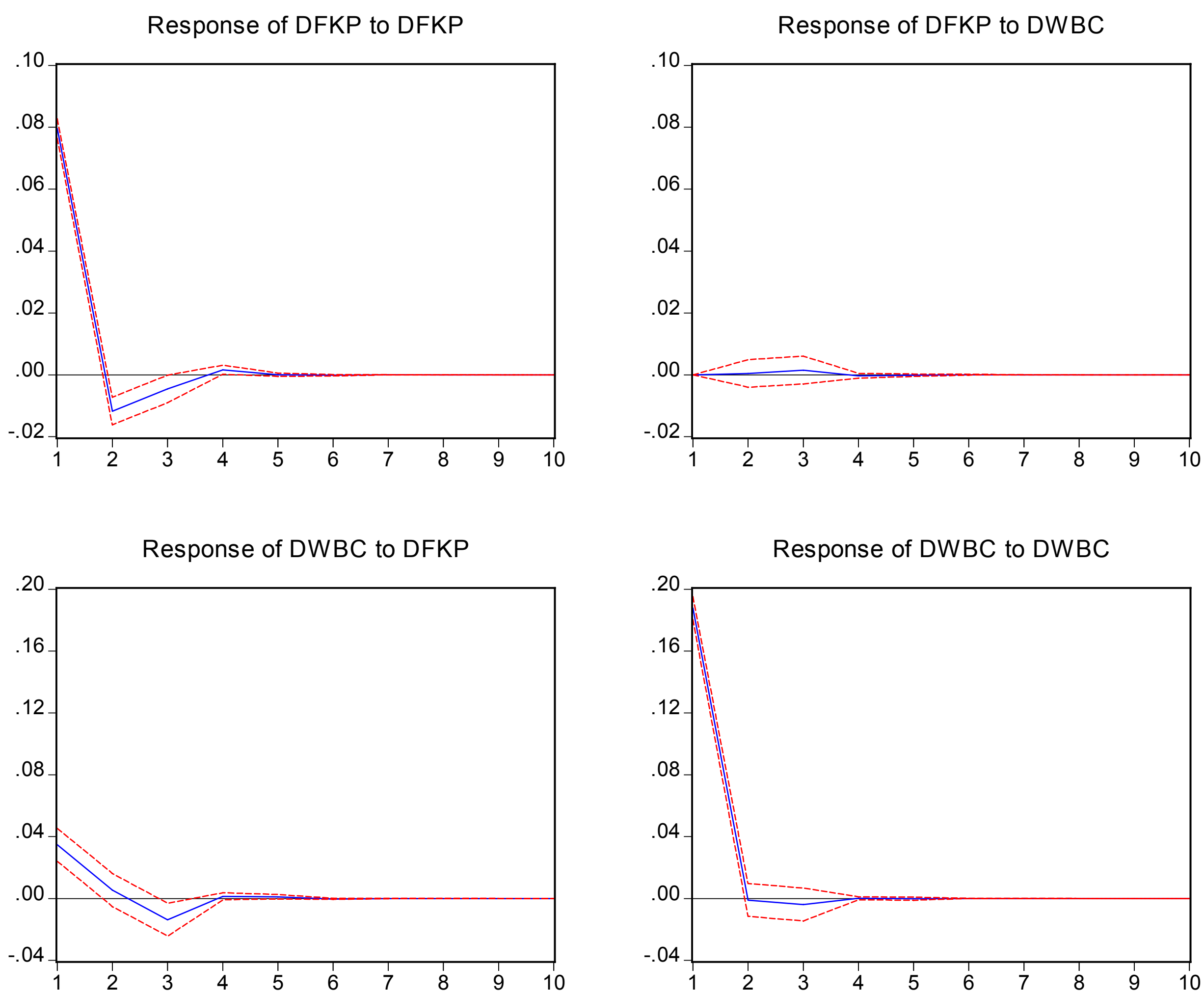
Vector Autoregression Estimates
Date: 10/21/08 Time: 11:36
Sample (adjusted): 4 1285
Included observations: 1282 after adjustments
Standard errors in ( ) & t-statistics in [ ]

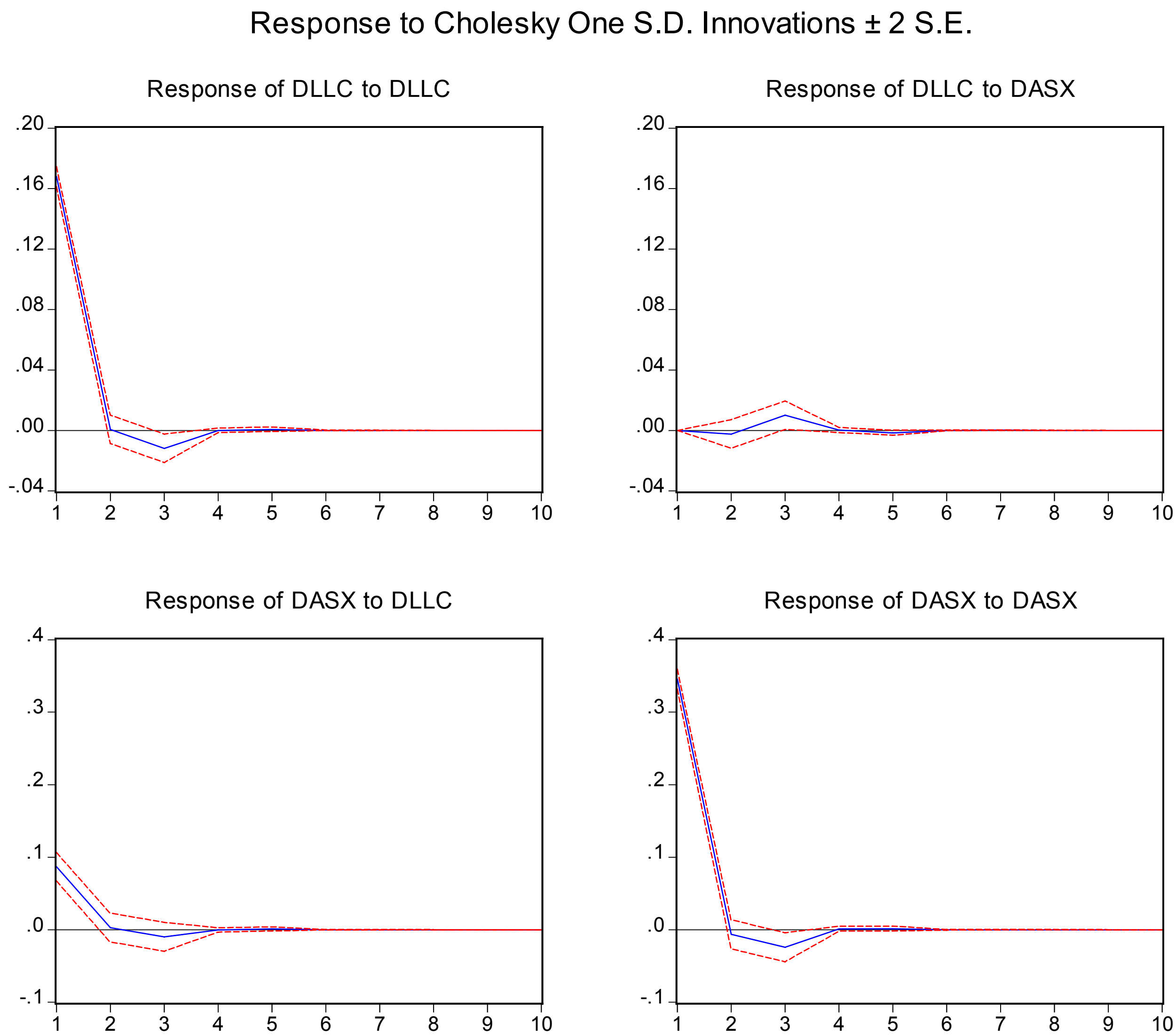
	DASX	DCBA
DASX(-1)	-0.033234	-0.008680
	(0.02912)	(0.02742)
	[-1.14124]	[-0.31650]
DASX(-2)	-0.080606	0.057055
	(0.02906)	(0.02736)
	[-2.77404]	[2.08507]
DCBA(-1)	0.024120	0.000477
	(0.03090)	(0.02910)
	[0.78063]	[0.01638]
DCBA(-2)	-0.006045	-0.025604
	(0.03091)	(0.02911)
	[-0.19559]	[-0.87966]
C	0.031146	0.017437
	(0.00997)	(0.00939)
	[3.12337]	[1.85680]
ASXCBARESID	-0.022865	0.006386
	(0.00567)	(0.00534)
	[-4.03439]	[1.19654]
R-squared	0.018417	0.004435
Adj. R-squared	0.014570	0.000534
Sum sq. resids	160.3117	142.1707
S.E. equation	0.354452	0.333795
F-statistic	4.788144	1.136897
Log likelihood	-486.4039	-409.4250

Appendix E – Impulse Response Functions

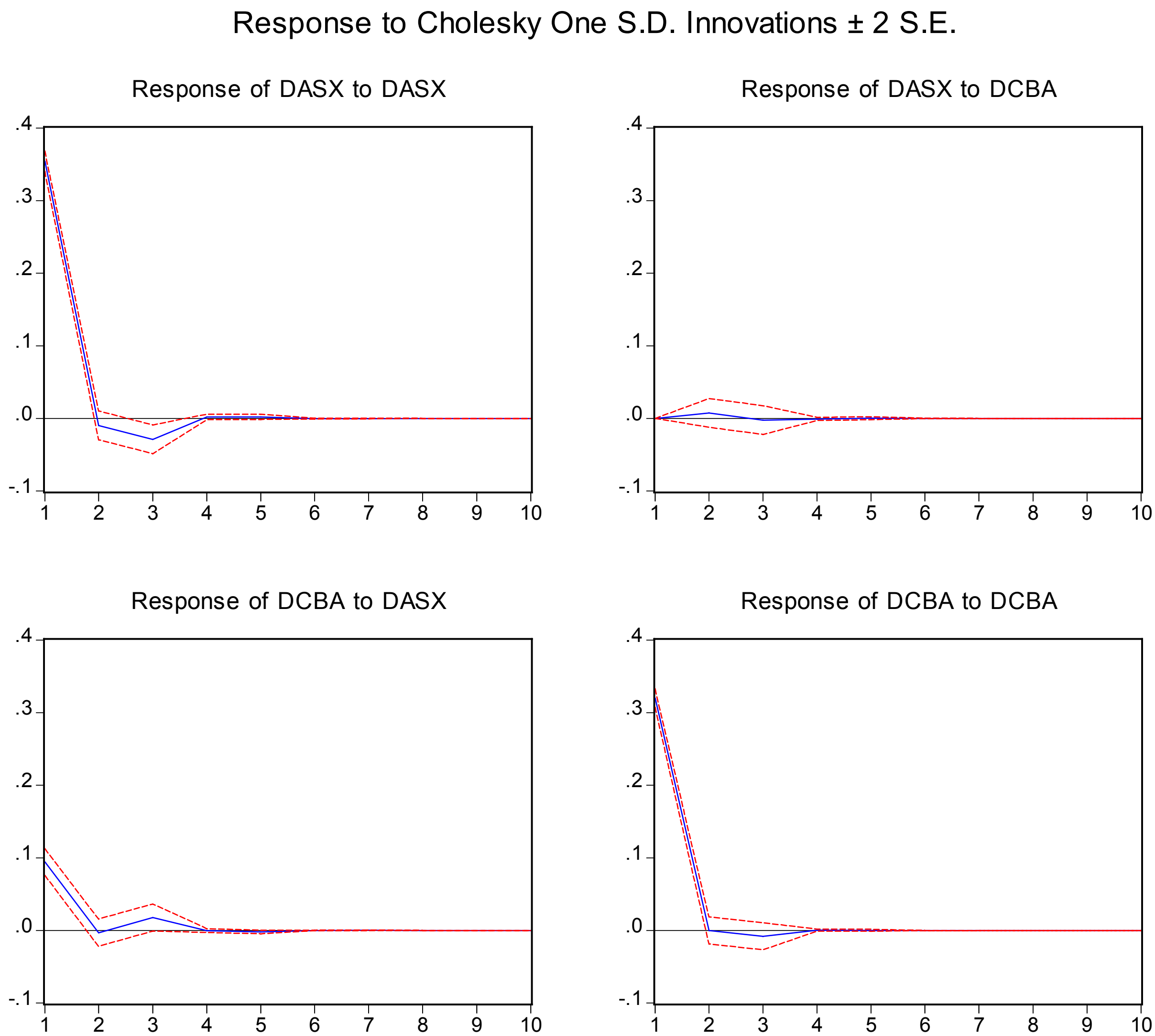


Response to Cholesky One S.D. Innovations ± 2 S.E.





Response to Cholesky One S.D. Innovations ± 2 S.E.
Response of DLLC to DLLC
Response of DLLC to DASX
Response of DASX to DLLC
Response of DASX to DASX
.20
.16
.12
.08
.04
.00
-.04
.4
.3
.2
.1
.0
-.1
1 2 3 4 5 6 7 8 9 10



Response to Cholesky One S.D. Innovations ± 2 S.E.
Response of DASX to DASX
Response of DASX to DCBA
Response of DCBA to DASX
Response of DCBA to DCBA
.4
.3
.2
.1
.0
-.1
1 2 3 4 5 6 7 8 9 10

Appendix F – Variance decomposition

Variance decomposition of DSGB

Period	S.E.	DSGB	DWBC
1	0.234816	100.0000	0.000000
2	0.235372	99.95538	0.044621
3	0.235463	99.92675	0.073254
4	0.235466	99.92574	0.074263
5	0.235466	99.92558	0.074416
6	0.235466	99.92558	0.074424
7	0.235466	99.92558	0.074425
8	0.235466	99.92558	0.074425
9	0.235466	99.92558	0.074425
10	0.235466	99.92558	0.074425

Variance decomposition of DWBC

Period	S.E.	DSGB	DWBC
1	0.190809	24.92960	75.07040
2	0.190823	24.93872	75.06128
3	0.190985	24.90716	75.09284
4	0.190985	24.90715	75.09285
5	0.190985	24.90705	75.09295
6	0.190985	24.90705	75.09295
7	0.190985	24.90705	75.09295
8	0.190985	24.90705	75.09295
9	0.190985	24.90705	75.09295
10	0.190985	24.90705	75.09295

Variance decomposition of DFKP

Period	S.E.	DFKP	DWBC
1	0.079593	100.0000	0.000000
2	0.080447	99.99670	0.003302
3	0.080590	99.95975	0.040247
4	0.080607	99.95852	0.041482
5	0.080607	99.95830	0.041702
6	0.080607	99.95827	0.041728
7	0.080607	99.95827	0.041728
8	0.080607	99.95827	0.041729
9	0.080607	99.95827	0.041729
10	0.080607	99.95827	0.041729

Variance decomposition of DWBC

Period	S.E.	DFKP	DWBC
1	0.190978	3.301606	96.69839
2	0.191056	3.378386	96.62161
3	0.191591	3.875416	96.12458
4	0.191596	3.880746	96.11925
5	0.191600	3.883879	96.11612
6	0.191600	3.884095	96.11590
7	0.191600	3.884098	96.11590
8	0.191600	3.884100	96.11590
9	0.191600	3.884100	96.11590
10	0.191600	3.884100	96.11590

Variance decomposition of DLLC

Period	S.E.	DLLC	DASX
1	0.167962	100.0000	0.000000
2	0.167980	99.97962	0.020377
3	0.168702	99.62186	0.378144
4	0.168702	99.62162	0.378382
5	0.168711	99.61292	0.387081
6	0.168711	99.61292	0.387081
7	0.168711	99.61281	0.387193
8	0.168711	99.61281	0.387193
9	0.168711	99.61281	0.387194
10	0.168711	99.61281	0.387194

Variance decomposition of DASX

Period	S.E.	DLLC	DASX
1	0.356709	5.941635	94.05837
2	0.356779	5.946186	94.05381
3	0.357740	5.991207	94.00879
4	0.357742	5.991232	94.00877
5	0.357747	5.991825	94.00818
6	0.357747	5.991824	94.00818
7	0.357747	5.991829	94.00817
8	0.357747	5.991829	94.00817
9	0.357747	5.991829	94.00817
10	0.357747	5.991829	94.00817

Variance decomposition of DASX

Period	S.E.	DASX	DCBA
1	0.354452	100.0000	0.000000
2	0.354663	99.95259	0.047408
3	0.355845	99.94912	0.050876
4	0.355853	99.94868	0.051318
5	0.355859	99.94863	0.051370
6	0.355860	99.94863	0.051372
7	0.355860	99.94863	0.051373
8	0.355860	99.94863	0.051373
9	0.355860	99.94863	0.051373
10	0.355860	99.94863	0.051373

Variance decomposition of DCBA

Period	S.E.	DASX	DCBA
1	0.333795	8.004691	91.99531
2	0.333809	8.012277	91.98772
3	0.334390	8.270573	91.72943
4	0.334390	8.270592	91.72941
5	0.334397	8.274299	91.72570
6	0.334397	8.274308	91.72569
7	0.334397	8.274335	91.72567
8	0.334397	8.274335	91.72567
9	0.334397	8.274335	91.72566
10	0.334397	8.274335	91.72566