

A New Approach to Modeling and Estimation for Pairs Trading

Binh Do ^{*} Robert Faff [†] Kais Hamza [‡]

May 29, 2006

Abstract

Pairs trading is an speculative investment strategy based on relative mispricing between a pair of stocks. Essentially, the strategy involves choosing a pair of stocks that historically move together. By taking a long-short position on this pair when they diverge, a profit will be made when they next converge to the mean by unwinding the position. Literature on this topic is rare due to its proprietary nature. Where it does exist, the strategies are either adhoc or applicable to special cases only, with little theoretical verification. This paper analyzes these existing methods in detail and proposes a general approach to modeling relative mispricing for pairs trading purposes, with reference to the mainstream asset pricing theory. Several estimation techniques are discussed and tested for state space formulation, with Expectation Maximization producing stable results. Initial empirical evidence shows clear mean reversion behavior in selected pairs' relative pricing.

^{*}PhD Candidate, Department of Accounting and Finance, Monash University

[†]Director of Research, Department of Accounting and Finance, Monash University

[‡]School of Mathematical Sciences, Monash University

1 Introduction

Pairs trading is one of Wall Street's quantitative methods of speculation which dates back to the mid-1980s (Vidyamurthy, 2004). In its most common form, pairs trading involves forming a portfolio of two related stocks whose relative pricing is away from its "equilibrium" state. By going long on the relatively undervalued stock and short on the relatively overvalued stock, a profit may be made by unwinding the position upon convergence of the *spread*, or the measure of relative mispricing. Whilst the strategy appears simple and in fact has been widely implemented by traders and hedge funds, due to the proprietary nature of the area, published research has been largely limited. Most referenced works include Gatev, Goetzmann and Rouwenhorst (1999), Vidyamurthy (2004), and Elliott, van der Hoek and Malcolm (2005). The first paper is an empirical piece of research that, using a simple standard deviation strategy, shows pairs trading after costs can be profitable. The second of these papers details an implementation strategy based on a cointegration based framework, without empirical results. The last paper applies a Kalman filter to estimating a parametric model of the spread. These methods can be shown to be applicable for special cases of the underlying equilibrium relationship between two stocks. A pairs trading strategy forcing an equilibrium relationship between the two stocks with little room for adaptation, may lead to a conclusion of "non-tradeability" at best and non-convergence at worst. This paper attempts to provide a uniform, analytical framework to design and implement pairs trading on any arbitrary pairs although it is acknowledged that pairs trading is best based on a priori expectation of co-movement verified by historical time series. Econometric techniques involved in the implementation phase are discussed and some empirical results are provided.

To define the boundary of this project, it is necessary to identify pairs trading relative to other seemingly related hedge fund strategies. There are as many classification themes in the industry as the number of strategies. After compiling both academic sources and informal, internet-based sources, pairs trading falls under the big umbrella of the long/short investing approach that is based on simultaneous exploitation of overpricing and underpricing, by going long on perceived under-priced assets and short on perceived overpriced ones. Under the long/short investing umbrella (as opposed to say, event driven strategies),

there are market neutral strategies and pairs trading strategies. Originally suggested by Jacobs and Levy (1993), Jacobs, Levy and Starer (1998, 1999), and debated in Michaud (1993), market neutral investing is a portfolio optimization exercise that aims to achieve negligible exposure to systematic risks, whilst “harvesting” two alphas, or active returns, one from the long position on the winners and one from the short position in the losers. There are also market neutral strategies that earn both beta return and two alphas, via the use of derivatives, such as the equitized strategy and hedge strategy (see Jacobs and Levy, 1993). Alternatively, market neutral investing can also achieve alpha return in one presumably less efficient market and beta return in another more efficient market, a practice known as *alpha transport*. Success of market neutral investing is derived from securities selection skills, leverage, and mathematical optimization, the latter is particularly proprietary, sometimes labeled ambiguously as “integrated optimization” (Jacob and Levy, 2005).¹

Pairs trading, on the other hand, exploits short term mispricing (sometimes heuristically called arbitrage), present in a pair of securities. It often takes the form of either *statistical arbitrage* or *risk arbitrage* (Vidyamurthy, 2004). Statistical arbitrage, the object of this study, is an equity trading strategy that employs time series methods to identify relative mispricings between stocks. Risk arbitrage, on the other hand, refers to strategies involving stocks of merging companies. The success of pairs trading, especially statistical arbitrage strategies, depends heavily on the modeling and forecasting of the spread time series although fundamental insights can aid in the pre-selection step. Pairs trading needs not be market neutral although some say it is a particular implementation of market neutral investing (Jacob and Levy, 1993).

This paper contributes to the literature by proposing an asset pricing based approach to parameterize pairs trading with a view to incorporate theoretical considerations into

¹To see the leverage impact, consider a typical market neutral strategy that involves an initial capital of \$100. In the absence of margining requirement, the manager can invest \$100 in the long position, and short up to \$100 worth of securities, with the cash placed with the broker as collateral. Thus the total exposure is \$200, or a leverage of two-for-one, plus cash. The manager then benefits from both the long position, and the short position, in the form of residual, or active return (alpha), plus interest earned from the cash proceeds. Clearly, unconstrained short selling is the key to creating this leverage, something long-only investing cannot compete.

the strategy as opposed to basing it purely on statistical history, as inherent in existing methods. The use of a parametric model enables rigorous testing and forecasting. In addition, the proposed approach removes the restriction of “return parity” often implicitly assumed in existing methods, hence widening the universe of tradeable pairs, and avoiding forcing incorrect state of equilibrium. A technical contribution of this paper lies in the estimation of a Gaussian and linear state model with exogenous inputs in both transition equation and observation.

The remainder of the paper is organized as follows. Section 2 outlines three existing pairs trading methods and their assumptions/limitations. Section 3 proposes what is termed a *stochastic residual spread model* of pairs trading. Section 4 discusses two alternative estimation methods, Maximum Likelihood Estimation and joint filtering, and suggests an integrated approach that combines the two. Simulation is performed to demonstrate comparative performance of the methods. Section 5 presents some preliminary empirical results. Section 7 concludes.

2 Existing Pairs Trading Methods

This section describes three main methods to implement pairs trading, which we label: the distance method, the cointegration method and the stochastic spread method. The distance method is used in Gatev et al (1999) and Nath (2003) for empirical testing whereas the cointegration method is detailed in Vidyamurthy (2004). Both of these are known to be widely adopted by practitioners. The stochastic spread approach is recently proposed in Elliot et al (2004).

2.1 The distance method

Under the distance method, the co-movement in a pair is measured by what is known as the *distance*, or the sum of squared differences between the two normalized price series. Trading is triggered when the distance reaches a certain threshold, as determined during a formation period. In Gatev et al (1999), the pairs are selected by choosing, for each stock,

a matching partner that minimizes the distance. The trading trigger is two historical standard deviations as estimated during the formation period. Nath (2003) keeps a record of distances for each pair in the universe, in an empirical distribution format so that each time an observed distance crosses a trigger of 15 percentile, a trade is entered for that pair. Risk control is instigated by limiting a trading period at the end of which positions have to be closed out regardless of outcomes. Nath (2003) also adopts a stop-loss trigger to close the position whenever the distance widens further to hit the 5 percentile. In overall, the distance approach purely exploits the statistical relationship of a pair, at a price level. As the approach is economic model-free, it has the advantage of not being exposed to model mis-specification and mis-estimation. On the other hand, being non-parametric means that the strategy lacks forecasting ability regarding the convergence time or expected holding period. What is a more fundamental issue is its underlying assumption that the price level distance is static through time, or returns of the two stocks are in parity. Although such assumption may be valid in short periods of time, it is so only 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.

2.2 The cointegration method

The cointegration approach outlined in Vidyamurthy (2004) is an attempt to parameterize pairs trading, by exploring the possibility of cointegration (Engle and Granger, 1987). Cointegration is the phenomenon that two time series that are both integrated of order d , can be linearly combined to produce a single time series that is integrated of order $d - b$, $b > 0$, the most simple case of which is when $d = b = 1$. As the combined time series is stationary, this is desirable from the forecasting perspective. Co-integrated time series can also be represented in an Error Correction Model (ECM) in which the dynamics of one time series at the current time is a correction of last period's deviation from the equilibrium (called the error correction component) and possibly some lag dynamics (and noises). The significance of this is that forecast can be done based on the past information. Vidyamurthy (2004) observes that as the logarithm of two stock prices are often assumed

to follow a random walk, or be non-stationary, there is a good chance that they will be co-integrated. If that is the case, cointegration results can be used to determine how far the spread is away from its equilibrium so that long/short positions can be entered to profit from the mispricing. To test for co-integration, Vidyamurthy (2004) adopts Engle and Granger’s 2-step approach (Engle and Granger, 1987) in which log price of stock A is first regressed against log price of stock B in what is called *cointegrating regression*:

$$\log(p_t^A) - \gamma \log(p_t^B) = \mu + \epsilon_t \quad (1)$$

where γ represents the cointegration coefficient and the constant term μ captures some sense of “premium” in stock A versus stock B. The estimated residuals are then tested for stationarity, hence cointegration, using the Augmented Dickey-Fuller test. Under this procedure, results are sensitive to the ordering of the variables, i.e if instead $\log(p_t^B)$ is regressed against $\log(p_t^A)$, a different set of standard errors will be found from the same sample. This issue can be resolved by using the t-statistics from Engle and Yoo (1987).

However, Vidyamurthy’s procedure is not necessarily premised on the cointegration condition, instead it looks for evidence of mean reversion in the spread time series, defined as $y_t = \log(p_t^A) - \gamma \log(p_t^B)$, heuristically interpreted as the return on a portfolio consisting of long 1 unit of A and short γ units of B. Cointegration means that the spread has a long run mean of μ , such that any deviation from it suggests disequilibrium. Vidyamurthy then analyzes the residuals for mean reversion, based on which trading rules are formed. Two general approaches are suggested for this analysis. One approach models the residuals as a mean reverting process, such as an ARMA process. The other approach manually constructs an empirical distribution of zero crossings from the data sample. A high rate of zero crossings is used as evidence of mean reversion, although it is not clear how to define the trigger point. The latter “model-free” approach appears to be favored by Vidyamurthy due to its simplicity and avoidance of model mis-specification.

Apart from being rather adhoc, Vidyamurthy’s approach may be exposed to errors arising from the econometric techniques employed. For one thing, the 2-step cointegration procedure renders results sensitive to the ordering of the variables, therefore the residuals may have different sets of statistical properties. For another, if the bivariate series are not cointegrated, the “cointegrating regression” leads to spurious estimators (Lim and

Martin, 1995), making the mean reversion analysis on residuals unreliable. So what can be done to improve this simple but intuitive approach? One way is to perform more rigorous testing of cointegration, including using Johansen's testing approach based on a Vector Error Correction Model (VECM) and comparing the outcome to the Engle-Granger results. But more importantly, if the cointegration test fails, one should refrain from trading based on residuals whose properties are unknown.

One major issue with this cointegration approach is the difficulty in associating it with theories on asset pricing. Although pairs trading has been originally premised on pure statistical results, economic theory considerations are necessary in verifying the strategy as the trader should not lose sight of fundamentals driving the values of the assets. In this regard, how do we interpret γ as the cointegration coefficient? Vidyamurthy attempts to relate the cointegration model to the Arbitrage Pricing Theory (APT)(Ross, 1976), and suggests that γ may have the meaning of a constant risk exposure proportionality. That is, if in the APT framework, for 1 unit exposure by stock B to all risk factors, stock A is exposed to γ units, then A and B satisfy the condition of cointegration. The argument makes use of the common trend representation of cointegrated series in which individual time series are driven by some common trends, identical up to a scalar, and a specific component:

$$\begin{aligned} \log(p_t^A) &= n_t^A + u_t^A \\ \log(p_t^B) &= n_t^B + u_t^B \end{aligned}$$

Therefore the return time series are:

$$\begin{aligned} R_t^A &= R_t^{c,A} + R_t^{s,A} \\ R_t^B &= R_t^{c,B} + R_t^{s,B} \end{aligned}$$

where R^c and R^s have the meaning of return component due to the trend component, and the specific, stationary component, respectively. A result of co-integration is that the common return components of both return time series should be identical up to a scalar, or $R_t^{c,A} = \gamma R_t^{c,B}$, as such

$$R_t^A = \gamma R_t^{c,B} + R_t^{s,A} \quad (2)$$

From this result, Vidyamurthy then asserts that if the APT theory holds true for every time step, then we have a cointegrated system if the factor exposure vectors of the two stocks are identical up to a scalar:

$$\begin{aligned} R_t^A &= \gamma(r_{1,t}b_1 + r_{2,t}b_2 + \dots + r_{n,t}b_n) + R_t^{s,A} \\ R_t^B &= (r_{1,t}b_1 + r_{2,t}b_2 + \dots + r_{n,t}b_n) + R_t^{s,B} \end{aligned}$$

where r_1, r_2, \dots are excess returns from exposure to risk factors, and b_1, b_2, \dots degrees of exposure, or beta in the factor models' language. However, an inspection of the equations reveals some fundamental error in the argument. Recall that under the APT theory, the return due to exposure to risk factors is on top of the risk free return, or:

$$\begin{aligned} R_t^A &= R_{f,t} + \gamma(r_{1,t}b_1 + r_{2,t}b_2 + \dots + r_{n,t}b_n) + R_t^{s,A} \\ R_t^B &= R_{f,t} + (r_{1,t}b_1 + r_{2,t}b_2 + \dots + r_{n,t}b_n) + R_t^{s,B} \end{aligned}$$

This suggests that when the risk exposure profiles of A and B are identical up to a scalar, it is generally not true that the return on 1 unit of A is identical to the return on γ units of B plus some Gaussian noise, as projected by (2). In other words, the cointegration model (1) does not reconcile well with the “mainstream” asset pricing models. It will be interesting to see how this statistical model fares in the empirical test.

2.3 The stochastic spread method

Elliott et al (2005) explicitly model the mean reversion behavior of the spread between the paired stocks in a continuous time setting, where the spread is defined as the difference between the two prices. The spread is driven by a latent state variable x , assumed to follow a Vasicek process:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t \quad (3)$$

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 speed κ . By making the spread equal to the state variable plus a Gaussian noise, or:

$$y_t = x_t + H\omega_t \quad (4)$$

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

The above model offers three major advantages from the empirical perspective. First, it captures mean reversion which underlies pairs trading. The fact that x can be negative is not a problem because the spread so defined can take on negative values. However, although it is not clear from Elliott et al (2005), it should be stressed here that strictly speaking, the spread should be defined as the difference in logarithms of the prices: $\log(p_t^A) - \log(p_t^B)$. Generally, the long term mean of the level difference in two stocks should not be constant, but widens as they go up and narrows as they go down. The exception is when the stocks trade at similar price points. By using the spread as log differences, this is no longer a problem.²

Second, being a continuous time model, it is convenient for forecasting purposes. As will be shown in a later section, 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 $E[\tau|x_t]$ where τ 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 (3) in a discrete time transition equation, motivated by the fact that the solution to (3) is

²To see this, assume stock A and B both return r in 1 unit of time so that $p_{t+1}^A = p_t^A e^r$ and $p_{t+1}^B = p_t^B e^r$. The log difference is

$$\begin{aligned} \log(p_{t+1}^A) - \log(p_{t+1}^B) &= (\log(p_t^A) + r) - (\log(p_t^B) + r) \\ &= \log(p_t^A) - \log(p_t^B) \end{aligned}$$

Markovian:

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

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

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

where Δ denotes the time interval (in years) between two observations, and the variance of the random process ϵ happens to be a constant $v = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta})$. 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 $\Psi = \{\theta, \kappa, \sigma, h\}$.³

Despite the several advantages, this approach does have a fundamental issue is that the model restricts the long run relationship between the two stocks to one of return parity, i.e in the long run, the two stocks chosen must provide the same return such that any departure from it will be expected to be corrected in the future. (See the previous footnote for the proof). This is a huge restriction as in practice it is rare to find two stocks with identical returns. Although one can invoke the factor models to argue that stocks with same risk factor exposures should have the same expected returns, in reality it is not necessarily the case because there are also firm specific returns that make the two total returns different. Note also that the notion of diversification that cancels unsystematic returns does not apply here because a pairs portfolio is not a diversified portfolio. When then can the Elliot et al's formulation be applicable? One possible case is companies that adopt a dual listed company (DLC) structure, effectively a merger between two companies domiciled in two different companies with separate shareholder registries and identities. Globally, there are only a small number of dual listed companies, with notable example including Unilever NV/PLC, Royal Dutch Petroleum/Shell (which dropped its structure in

³For introduction to the state space model and Kalman filter, see Durbin and Koopman (2001).

July 2005), BHP Billiton Limited/PLC and Rio Tinto Limited/PLC. In a DLC structure, both groups of shareholders are entitled to the same cash flow, although shares are traded on two different exchanges, and often attract different valuations. The fact that the shares can not be exchanged for each other precludes riskless arbitrage, although they present a clear opportunity for pairs traders, as has been widely exploited by hedge funds. Another candidate for pairs trading assuming return 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 Depositary Receipts (ADRs). Companies may also cross list in different exchanges within a country, such as the NASDAQ and NYSE in America.⁴

The next section proposes a new parametric approach to pairs trading, called a *stochastic residual spread* method that addresses issues encountered in the existing methods.

3 A New Pairs Trading Method: The Stochastic Residual Spread

Pairs trading is essentially predicated on the existence of mean reversion in relative mispricing between two assets. A pairs trading strategy ideally must be able to quantify the level of mispricing and the strength of the mean reversion in some way, based on which to determine tradeability, and subsequently, entry and exit rules. The existing methods address these issues purely on a statistical basis, leading to adhoc trading rules. It is therefore motivating to explore other approaches that incorporate some theoretical flavour and evaluate how they fare against those statistical rules.

The method of stochastic residual spread proposed herein starts with an 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* $G(R_t^A, R_t^B, U_t)$ where U denotes some exogenous vector potentially present in formulating the equilibrium. The term “residual spread”

⁴See Badi and Tennant, 2002 for more information on DLCs and cross listing

emphasizes that the function captures any excess over and above some long term spread, and may take non-zero values, depending on formulation of the spread. By the market forces, the relative valuation should mean revert to equilibrium in the long run. When the disequilibrium is sufficiently large and the expected correction timing is sufficiently short, a pairs trading transaction can be executed to make a profit. The proposed method then adopts the same modeling framework as in Elliot et al (2005) to implement this idea, that is to use a one factor stochastic model to describe the state of mispricing or disequilibrium, and let some noise contaminate its actual observation being measured by the above specified function G . In particular, let x be the state of mispricing or residual spread, with respect to a given equilibrium relationship, whose dynamic is governed by an Vasicek process:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t \quad (5)$$

The observed mispricing is :

$$y_t = G_t = x_t + \omega_t \quad (6)$$

These two equations constitute a state space model of relative mispricing, defined with respect to some equilibrium relationship between two assets. Note that with this model, the state of mispricing is not fully observed, rather it is observed up to some measurement noise. How is such a measurement noise justified in this problem? Dynamic asset pricing studies often use measurement noises to allow for pricing errors existing across a cross section of assets. Yet, in this problem, there is only one single observation of the residual spread, such that there is no cross section consistency issue here to be resolved by measurement errors. It is also not due to the presence of bid-ask spreads or human errors in data handling that gives rise to measurement errors, because such noises would have negligible impact on the residual spread observed. Instead, the measurement noise is set to capture the uncertainty in the so-called equilibrium relationship, embedded in the residual spread function G_t , which is currently used as the observation in Equation 5. More specifically, the equilibrium relationship is not known and needs to be estimated, giving rise to uncertainty, or noise. This consequently implies that the observation in the above state space model is in fact, not fully observed. This issue will be resolved shortly.

Let us now focus on the main aspect of this method, which is to specify the equilibrium

relationship, or alternatively, the residual spread function G . The concept of relative pricing between two assets is, unfortunately, not well explored within the mainstream asset pricing literature, which mainly operates on a portfolio basis. It is also outside the scope of this paper to propose a theoretical framework on relative asset pricing. Instead, in addressing this issue, we are motivated by the factor models in asset pricing, in particular, the APT (Ross, 1976), which asserts that the return on a risky asset, over and above a risk free rate, should be the sum of risk premiums times the exposure, where specification of the risk factors is flexible, and may, for instance, take the form of Fama-French 3-factor model:

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

where $\beta = [\beta_1^i \beta_2^i \dots \beta_n^i]$ and $r^m = [(R^1 - r_f)(R^2 - r_f) \dots (R^n - r_f)]^T$, R^i denotes the raw return on the i th factor. The residual, η has expected value of zero, reflecting that the APT works on a diversified portfolio such that unsystematic or company specific risks are unrewarded, although its actual value may be non-zero. A “relative” APT on two stocks A and B can be written as:

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

where $\Gamma = [(\beta_1^A - \beta_1^B) \quad (\beta_2^A - \beta_2^B) \dots (\beta_n^A - \beta_n^B)]$, a vector of exposure differentials and e is a residual noise term. In addition, we assume that the above relationship holds true in all time periods, such that we can write:

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

If we are prepared to embrace the above equilibrium model, we can specify the residual spread function as follows:

$$G_t = G(p_t^A, p_t^B, U_t) = R_t^A - R_t^B - \Gamma r_t^m \quad (7)$$

If the values of Γ is known (and r_t^m specified), G_t is completely observable and we have a completely tractable model of mean reverting relative pricing for two stocks A and B , ready to be used for pairs trading. Below is reproduction of the model, in a state space form:

The transition equation:

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t$$

The measurement equation:

$$y_t = G_t = x_t + \omega_t$$

where G_t is specified in (7).

In a discrete time format, we have:

The transition equation:

$$x_k = \theta(1 - e^{-\kappa\Delta}) + e^{-\kappa\Delta}x_{k-1} + \epsilon_k \quad (8)$$

The measurement equation:

$$y_k = x_k + H\omega_k \quad (9)$$

Note that this model nests Elliot et al's model when Γ is a zero vector.

This state space model remains problematic with the observation 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 regressor 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. Another solution that is adopted in this paper is to redefine the observation $y = R^A - R^B$ such that the measurement equation is rewritten as:

$$y_k = x_k + \Gamma r_k^m + H\omega_k \quad (10)$$

This formulation allows the mispricing dynamic and the 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 8 and 10 constitute a model of stochastic residual spread for a pairs trading implementation. This is a linear and Gaussian state space model, which can be estimated by Maximum Likelihood Estimation or some form of filtering, to be discussed in the next section.

To summarize, what has been done so far is formulation of a continuous time model of mean reversion in the relative pricing between two assets, with the relative pricing model being adapted from the APT model of single asset pricing. An econometric framework has been also formulated to aid in the estimation process. At this juncture, one may question the validity of this approach, on the basis of its reliance of the APT model. In fact, the proposed method does not at all make any assumption on the validity of the APT model. Rather it adapts the factor structure of the APT to derive a relative pricing framework, without requiring the APT to strictly valid to the fullest sense. Therefore, whereas a strict application of the APT may mean the long run level of mispricing, or θ , should need to be close to zero, a non-zero estimate should not serve to invalidate the APT or the pairs trading model as a whole. Rather, it may mean that there is a firm specific premium commanded by one company versus the other, to reflect such things as management superiority. On this note, one may redefine the function G to reflect this premium by adding a constant term, for example, $G_t = R_t^A - R_t^B - \Gamma r_t^m - \mu$. However, this formulation will only add further complication to the estimation by increasing the number of parameters to estimate, whereas it can be “absorbed” in parameter θ . Another reason for using the APT model is to adapt the flexibility structure of the APT in allowing the method’s implementers to factor in their prior belief in designing an appropriate trading rule. In other words, in computing the spread, the traders may want to incorporate risk factors deemed to be relevant to the pair, in a linear factor format as in the APT. The most straightforward design is to use one single risk factor, the market premium, in which case relative pricing is now based on the CAPM. In fact, the simulation and empirical testing following this section adopts CAPM as the asset pricing model.

What remains to be examined is development of an optimal estimation strategy, and formulation of trading rules. The former warrants a in-depth analysis due to the peculiar structure of the state space model (8) and (10), hence delayed until the next section. Yet, trading rules based on this modeling strategy is by no means trivial. Unlike existing pairs trading strategy which are predicated on mispricing at the *price* level, the strategy proposed is based on mispricing at the *return* level. Therefore, the existing methods open positions when the prices drift sufficiently apart and unwind when they converge. In contrast, the proposed strategy opens positions when the *accumulated* residual spread in

the returns is sufficient large, and unwind when the *accumulated* spread is equal the long run level of the spread. In other words, correction in the context of this strategy does not occur when the spread is at its long run level, rather it may be at the other side of the long run level, for the accumulated spread to be “neutralised”. To illustrate this point, consider two stocks A and B with, for simplicity, identical risk-return profiles, such that their returns should be identical, and have sustained that behavior for a period of time. Assume now that the last observed period sees A return 5% and B 3%, or a residual spread of 2%. For correction to happen in the next period, the residual spread needs to be around -2%, regardless of the individual direction of the stocks, hence a zero accumulated residual spread.

Therefore, a trading rule for this strategy is to take a long-short position whenever the accumulated spread $\delta_k = \sum_{i=k-l}^k E[x_i|Y_i]$, with l less than or equal to the current time k , exceeds θ by a certain threshold. The trader will have to fix a base from which to determine point l where $\delta_l = 0$. One may also wish to compute the expected convergence time, that is the expectation of $T > k$ such that δ_T first crosses 0, given $\delta_k = c$. We are investigating analytical results of this first passage time question. Meanwhile, one can always use Monte Carlo simulation to compute the expectation. This quantity will determine the expected holding period, hence expected return. Clearly, formulation of trading rules based on the residual spread approach is interesting and requires further investigation.

4 Estimation Methodologies

This section looks at the econometrics of the state space model represented by (8) and (10). Because it is linear and Gaussian (LGSS), the conventional estimation approach is to perform MLE where the likelihood function is of a *error prediction decomposition* form (see Durbin and Koopman, 2001):

$$\begin{aligned} \log L(y) &= \sum_{i=1}^N \log p(y_i|Y_{i-1}) \\ &= -\frac{N}{2} \log 2\pi - \frac{1}{2} \sum_{i=1}^N \left(\log |F_i| + e_i' F_i^{-1} e_i \right) \end{aligned}$$

where $Y_i = \{y_1, y_2, \dots, y_i\}$, $F_i = \text{Var}[y_i|Y_{i-1}]$, $e_i = y_i - E[y_i|Y_{i-1}]$, and N being the length of the time series. The quantities F_i and e_i are routinely computed by the Kalman filter, a celebrated algorithm that produces minimum mean squared error estimates of $E[x_i|Y_i]$ (refer to Haykin, 2001).

The loglikelihood function is then maximized numerically to obtain MLE estimates of the parameters, in this case, $\Psi = [\theta, \kappa, \sigma, \Gamma, H]$. The attractiveness of MLE is that its estimates are known to be efficient and asymptotically normal. Potential issues with this method, as well as the frequentist methods as a whole are their finite sample performance and numerical issues arising from the numerical optimisation step. Shumway and Stoffer (1982) propose an Expectation Maximization (EM) algorithm to compute the MLE estimates without the need for numerical maximisation of the loglikelihood function. It involves treating the latent state variable as missing data, such that parameters are estimated by recursively finding values that maximises the expectation of the complete data loglikelihood function (i.e. $\log p(x, y)$, where the expectation is taken with respect to the posterior density $p(X_N|Y_N)$). Beside avoiding numerical optimization, this strategy ensures increased likelihood and also produces smoothed estimates $x_k|Y_N$ as a by-product, using the Rauch-Tung-Streusel smoother version of the Kalman filter (see Chapter 1, Haykin, 2001). However, derivations of the EM algorithm available in the literature, to the authors' best knowledge, are based on a special case of LGSS:

$$\begin{aligned}x_k &= Ax_{k-1} + Gv_{k-1} \\y_k &= Cx_k + H\omega_k\end{aligned}$$

In contrast, our model represented by (8) and (10) is of a more general form:

$$\begin{aligned}x_k &= Ax_{k-1} + B + Gv_{k-1} \\y_k &= x_k + DU_k + H\omega_k\end{aligned}$$

where U_k is an exogenous input in the output equation, which is not common in state space modeling. Elliott et al (2005) provide a derivation based on the above setup with an exception that there is no U_k in the measurement equation, due to their restrictive model of pairs trading. The addition of U_k is non-trivial in this case because it is time varying and has unknown coefficient D . A derivation of the EM for this general setup, which is believed

to be equally nontrivial, is enclosed in the appendix for interested readers. Naturally, this study also investigates the performance of EM algorithm in this application, in comparison against the optimization based MLE.

Alternatively, due to the finite sample issues inherent in the MLE approach, one may consider Bayesian estimation schemes. For example, one can employ a Markov Chain Monte Carlo (MCMC) simulation to draw posterior distributions of the parameters and the state, conditional on the observed return differentials y and the market excess return. MCMC has been increasingly popular as an estimation method for time series models, where notable applications to state space models include Jacquier, Polson and Rossi (1994) and more recently Eraker (2004). Whilst this MCMC approach is theoretically sound, based on convergence results of Markov chains, it often entails high computational costs in terms of both complexity and intensity. Another approach within the Bayesian framework is joint filtering, sometimes known as the self-organised state space estimation method (Kitagawa, 1998), or joint estimation (Haykin, 2001). The method suggests unknown parameters be treated as random processes with very small variances, and concatenated with the state process to form an augmented state space system. The estimation of the parameters and the state then reduces to one of filtering a now nonlinear state space system. This method has the advantage of being highly model independent and computationally efficient, unlike the MCMC method. On the other hand, its treatment of fixed parameters as random processes (as opposed to random variables under a pure Bayesian framework), is likely to introduce instability (see Liu and West, 2001). However, for applications where parameters are time varying, such as this pairs trading model where the beta differential may be time varying, the joint filtering approach may be able to better capture the variation. Clearly, the performance of joint filtering will be substantially dependent upon the choice of nonlinear filters. Nonlinear filters based on Kalman filtering such as the Extended Kalman filter and Unscented Kalman Filter (see Haykin, 2001) are unsuitable for the (augmented) model in question, due to the presence of unknown noise variances. A recent generation of nonlinear filters is particle filtering (Doucet, de Freitas and Gordon, 2001), an established variant of which is the Auxiliary Particle Filter (Pitt and Shephard, 1999). The essence of particle filtering is to efficiently approximate the posterior distribution $x_k|Y_k$ using sequential importance sampling and resampling. Unlike

Kalman filtering, particle filtering is not restricted to linearity nor Gaussianity. However, an important caveat that is often unclear from technical references on particle filtering is that it operates on the basis that the distributional form of the system noises is correctly specified. This means that particle filtering may be less robust to model misspecification than Kalman filter based MLE.

Finally, one can integrate these methods into one single procedure to obtain optimal results. For example, it has been suggested before (for example, Durbin and Koopman, 2001) that the EM algorithm can be employed in the early state of an optimisation scheme if the former proves to be slow in convergence when it moves closer to the optimum. Similarly, these two (EM and optimisation based MLE) can also be used to initialize particles for a particle filtering procedure. For example, one can assume a normal distribution for the initializing particles with mean and variance taken from the MLE estimates and their easily computable standard errors.

We have implemented these alternative approaches on the model (8) and (10) in a simulation setting. In particular, for each simulation run, a time series of x is simulated based on true parameters, and equation (8), which is the exact discrete time representation of the Vasicek process. The market excess return is generated assuming a geometric Brownian motion. A time series of y is then simulated according to equation (10) and simulated value of x and market excess return. Alternative estimation procedures are then applied to y to estimate x and the parameters. Outcomes are then aggregated across simulations to obtain sample averages. Final results show little variation amongst the methods in terms of estimation errors, hence not reported here. For a sample of size 100, the EM algorithm quickly converges, such that the subsequent deployment of numerical optimization and/or particle filtering does not add significant value. Figures 1, 2, 3 and 4 display the EM results based on the following parameter values: $A = 0.9$, $B = 0.005$, $D = 0.1$, $G = 0.0529$ and $H = 0.05$, which correspond to $\theta = 0.05$, $\kappa = 5$, $\sigma = 0.4$, $\Gamma = 0.1$, $H = 0.05$. The procedure is initialized with $x_0 = 0$, $P_0 = 0.1$, $A_0 = 0.1$, $B_0 = 0.1$, $D_0 = -0.2$, $G_0 = 0.1$ and $H_0 = 0.1$. To ensure strict positivity of the estimate of A , estimation is performed on $\log(A)$.

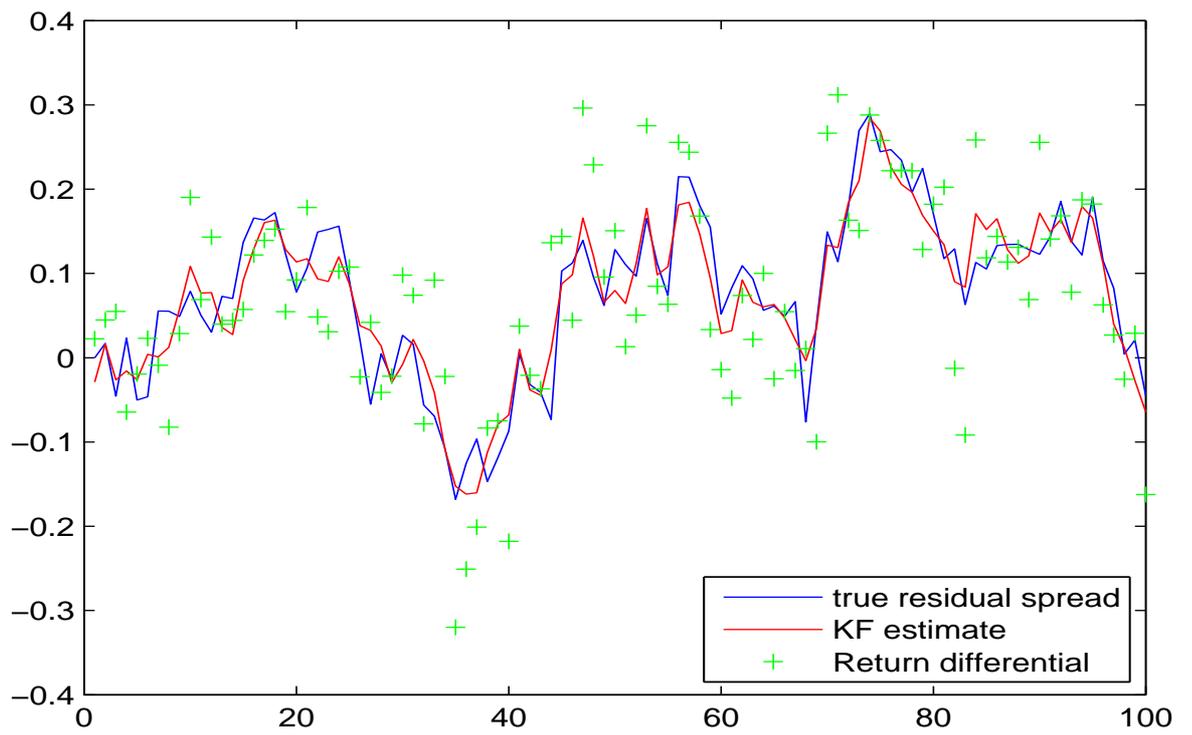


Figure 1: Kalman Smoother Estimate of Residual Spread Given Observed Returns

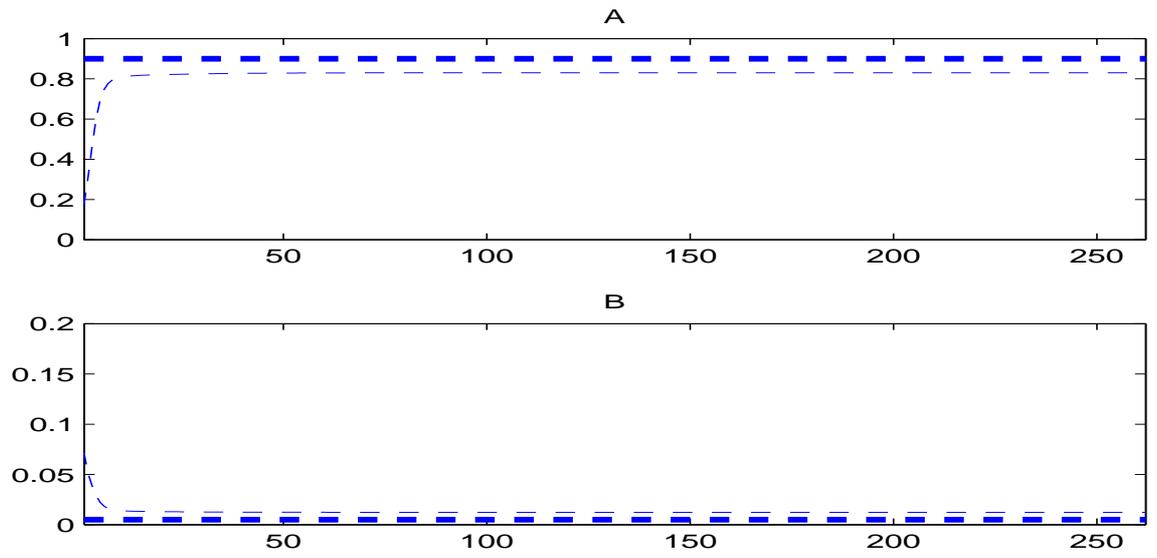


Figure 2: Estimation of A and B

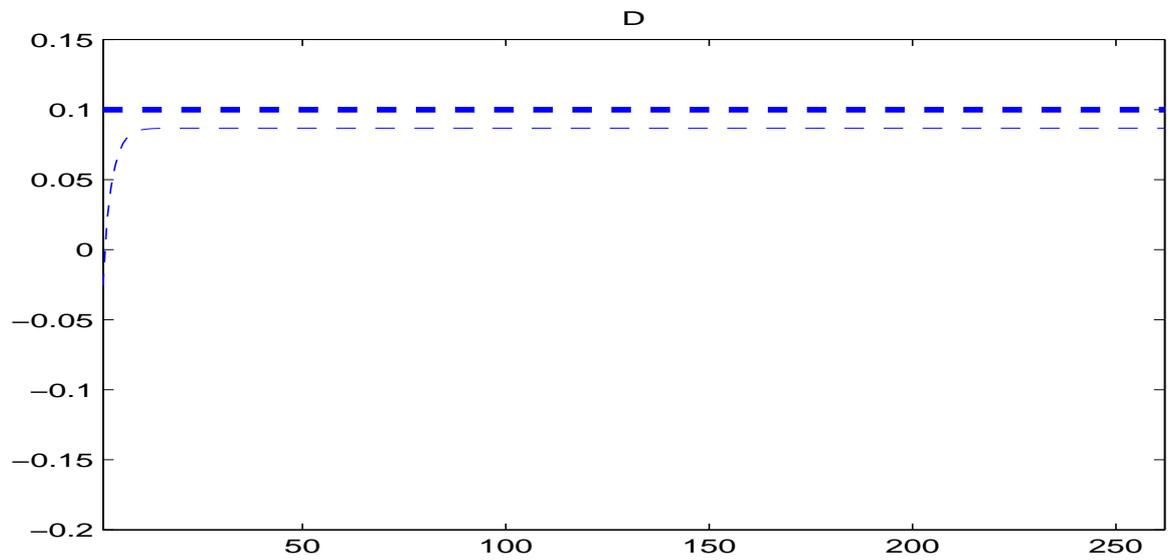


Figure 3: Estimation of D

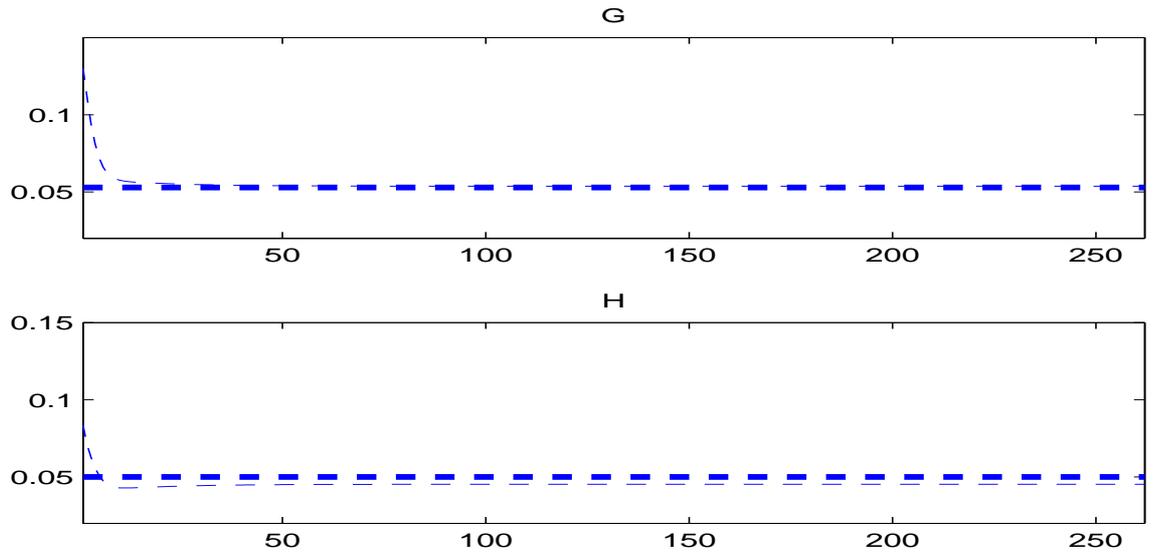


Figure 4: Estimation of G and H

5 Some Empirical Results

This section estimates mean reversion behaviour in three pairs of stocks: BHP-Rio Tinto, Target and Wal-mart and Shell and BP. These pairs are chosen on the basis of industry similarity, with the first pairs being the top two miners in Australia (and the world), the second top retailers in the U.S and the last being the largest energy companies in the UK. For each of the three pairs, an estimation of the model (8) and (10) is performed using EM, on two years of weekly returns. For the Australian pair, the *S&P* 200 index is chosen as the market portfolio. For the US pair, it is the *S&P* 500. The FTSE All Shares index is chosen as the market proxy for the UK pairs. Treasury bond yields in respective countries are used as the risk free rate. Figure 5, 6, and 7 plot the estimated residual spread as implied from the observed return differential. Table 1 reports estimation results.

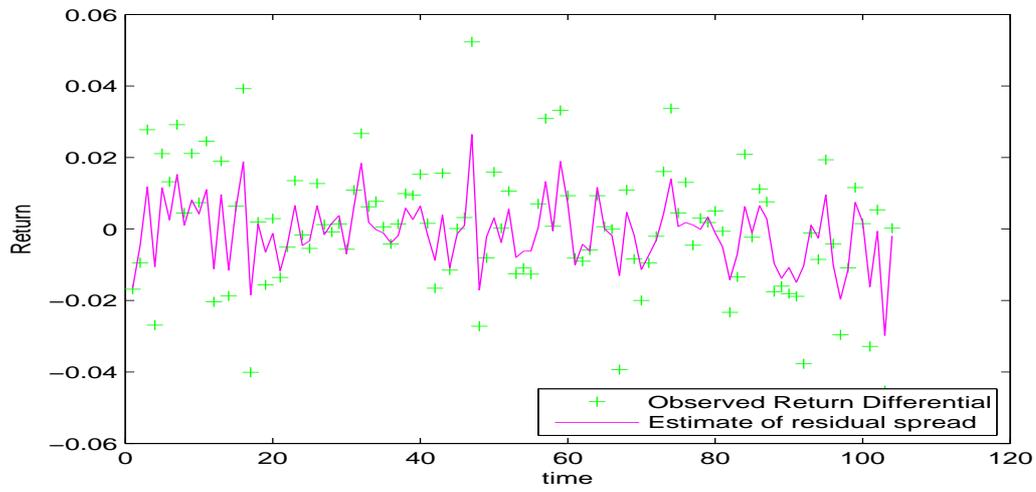


Figure 5: Estimation of BHP-RIO's Residual Spread

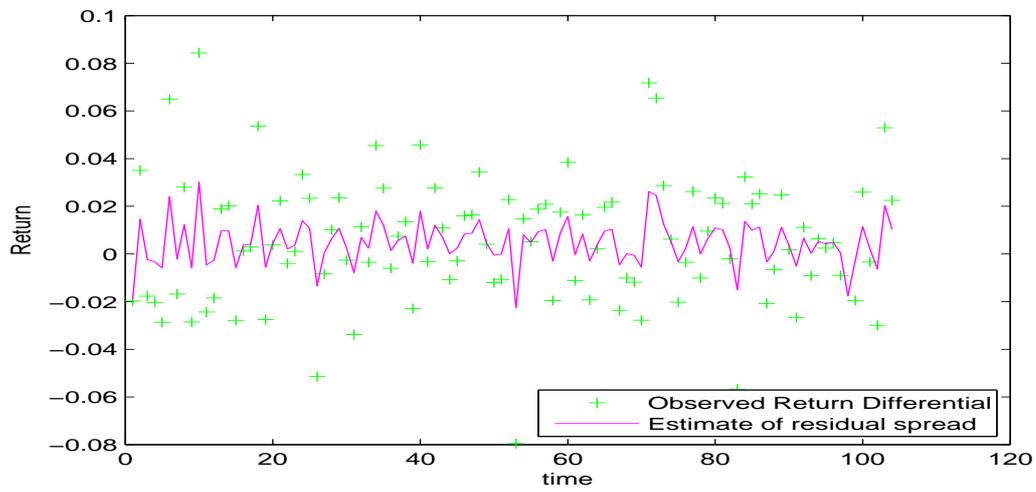


Figure 6: Estimation of WalMart-Target's Residual Spread

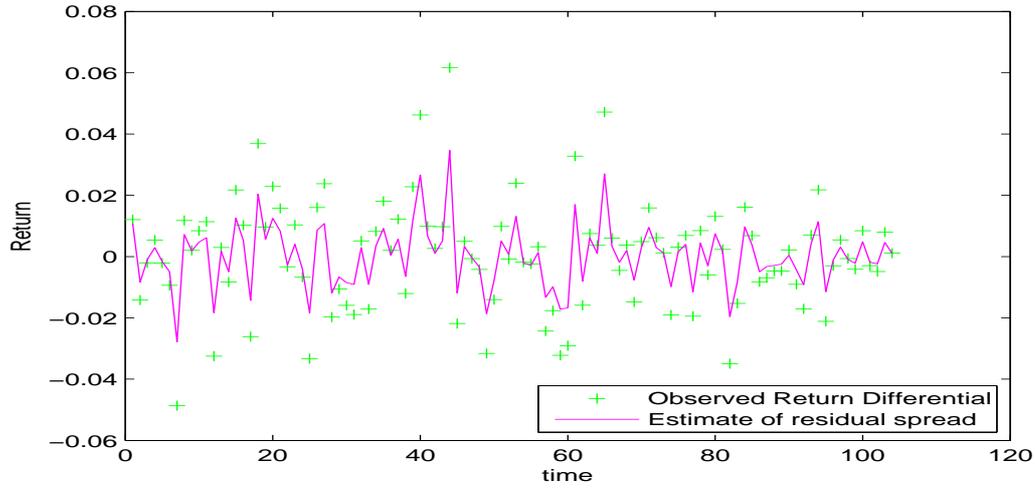


Figure 7: Estimation of BP-Shell's Residual Spread

Table 1: Estimation Results			
	BHP vs RIO	Target vs Walmart	Shell vs BP
log(A)	-6.3665 (-3.5345)	-6.1471 (-6.8993)	-6.5089 (-6.0677)
B	-0.0009 (-5.4785)	0.0048 (18.2916)	-0.0000 (-0.0131)
D	0.4420 (34.1431)	-0.0251 (-1.3306)	-0.0518 (-7.1158)
G	0.0121 (16.3781)	0.0155 (7.5766)	0.0132 (15.7919)
H	0.0110 (13.4408)	0.0226 (15.9084)	0.0117 (12.5962)
θ	-0.0009	0.0048	-0.0000
κ	6.3665	6.1471	6.5089
σ	0.0433	0.0545	0.0475
Note: the numbers in parenthesis are z-statistics			

Despite the limited sample examined, a number of interesting observations can be

drawn from the figures 5-7 and Table 1. First, the estimated coefficients are significant across the three pairs, supporting the Vasicek model of mean reversion in the residual spreads. Second, the level of mean reversion across 3 pairs is strong, reflected by large values of κ , which incidentally are all around 6-6.5. These values are also captured visually in the graphs where the estimated state is shown to quickly revert to its mean. The implication is twofold. On one hand, mean reversion is ample, hence the non-convergence risk is mitigated. On the other, it may be too strong, such that profit opportunities are quick to vanish for those selected pairs. Third, the estimates of θ are not zero, albeit close to zero. This suggests there remains some residual risk over and above the beta risk, that is still priced by the market in a relative sense. In the case of BHP-RIO, on an annualized basis, the residual spread is around 5% in favor of Rio Tinto. This could be attributed to superior management in Rio Tinto, or better asset quality. For Target and Walmart, the spread is nearly 25% p.a, something that cannot be sensibly attributed to nonsystematic risks. An examination of the two stocks' price performance over the two year period in question shows that the long term trend was slightly up for Target and slightly down for Walmart. This is an excellent example of pairs to be avoided: the pairs move together in the short term and the trends diverge in long term, making pairs trading very risky. The long term residual spread in BP and Shell is very negligible. Finally, the beta differentials estimated from the state space models are found to be very close the those obtained from individual market model regression. For example, the regression estimated beta for the sample period is 1.7827 for BHP and 1.3377 for RIO, which is consistent with the fact that the former is exposed to the oil factor whereas the latter is not. The difference is 0.445 which is closed to an estimated D of 0.442.

6 Conclusion

We have proposed a general approach to model relative mispricing for pairs trading purposes, in a continuous time setting. The novelty in this approach lies in its quantification of mean reversion behavior, taking into account theoretical asset pricing relationships. This is in contrast with existing approaches which are purely based on statistical consideration leading to adhoc trading rules. Estimation methods are also extensively discussed,

with an EM algorithm provided and tested for the model in hand. Initial empirical results show evidence of mean reversion in line with priori expectation for the pairs chosen. A natural extension is to investigate the profitability of the strategy on a cross section of pairs, the objective of our next project. Such research will, amongst other things, investigate optimal trading rules, having taken into consideration transaction costs and any regulatory issues concerning short selling.

References

- [1] Bedi, J. and Tennant, P. (2002) “Dual-Listed Companies”, *Reserve Bank of Australian Bulletin*, October.
- [2] Chen, R.-R. and Scott, L. (2003) “Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model”, *Journal of Real Estate Finance and Economics*, Vol. 27(2), pp.143-172.
- [3] Cox, J., Ingersoll, J. and Ross, S. (1985) “A Theory of the Term Structure of Interest Rates”, *Econometrica*, Vol. 53(2), pp.385-408.
- [4] De Rossi, G. (2004a) “Maximum Likelihood Estimation of the Cox-Ingersoll-Ross Model Using Particle Filters”, Working Paper, Cambridge University.
- [5] De Rossi, G. (2004b) “The Two-Factor Cox-Ingersoll-Ross Model as a Self-Organizing State Space”, Working Paper, Cambridge University.
- [6] Doucet, A., de Freitas, N. and Gordon, N. (2001) *Sequential Monte Carlo Methods in Practice*, Springer, New York.
- [7] Duan, J.-C., and Simonato, J.-G. (1999) “Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter”, *Review of Quantitative Finance and Accounting*, Vol. 13, pp.111-135.
- [8] Durbin, J. and Koopman, S. (2001) “Time Series Analysis by State Space Models”, Oxford University Press.
- [9] Elliott, R., van der Hoek, J. and Malcolm, W. (2005) “Pairs Trading”, *Quantitative Finance*, Vol. 5(3), pp. 271-276.
- [10] Engle, R. and Granger, C. (1987) “Co-integration and Error Correction: Representation, Estimation, and Testing”, *Econometrica*, Vol. 55(2), pp. 251-276.
- [11] Engle, R. and Yoo, B. (1987) “Forecasting and Testing in Co-integrated Systems”, *Journal of Econometrics*, Vol. 35, pp. 143-159.

- [12] Gatev, E., G., Goetzmann, W. and Rouwenhorst, K. (1999) “Pairs Trading: Performance of a Relative Value Arbitrage Rule”, Unpublished Working Paper, Yale School of Management.
- [13] Geweke, J. (1989) “Bayesian Inference in Econometric Models Using Monte Carlo Integration”, *Econometrica*, Vol. 57(6), pp. 1317-1339.
- [14] “Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993) Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation”, *IEE Proceedings-F*, Vol. 140(2), pp. 107-113.
- [15] Jacobs, B. and Levy, K. (1993) “Long/Short Equity Investing”, *Journal of Portfolio Management*, Vol. 20(1), pp. 52-64.
- [16] Jacobs, B. and Levy, K. (2005) “Market Neutral Strategies”, John Wiley & Sons, New Jersey.
- [17] Jacobs, B., Levy, K. and Starer, D. (1998) “Long-Short Portfolio Management: An Integrated Approach”, *Journal of Portfolio Management*, Winter, pp. 23-32.
- [18] Jacobs, B., Levy, K. and Starer, D. (1998) “On the Optimality of Long-Short Strategies”, *Financial Analysts Journal*, Vol. 54(2), pp. 40-50.
- [19] Javaheri, A. (2005) “Inside Volatility Arbitrage”, John Wiley & Sons, New Jersey.
- [20] Jazwinski, A. (1970) “Stochastic Processes and Filtering Theory”, Academic Press, New York.
- [21] Kalman, R.E. (1960) “A New Approach to Linear Filtering and Prediction Problems”, *Journal of Basic Engineering*, Vol. 82, pp.35-45.
- [22] Kitagawa, G. (1998) “A Self-Organizing State-Space Model”, *Journal of the American Statistical Association*, Vol. 93, pp. 1203-1215.
- [23] Kitagawa, G. and Sato, S. (2001) “Monte Carlo Smoothing and Self-Organizing State-Space Model”, in: A. Doucer, N. de Freitas and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice* (Springer, New York), pp. 177-196.

- [24] Lamoureux, C.G. and Witte, H.D. (2002) “Empirical Analysis of the Yield Curve: the Information in the Data Viewed Through the Window of Cox, Ingersoll and Ross”, *Journal of Finance*, Vol. 57, pp. 1479-1520.
- [25] Lim, G and Martin, V. (1995) “Regression-based Cointegration Estimators”, *Journal of Economic Studies*, Vol. 22(1), pp. 3-22.
- [26] Liu, J. and West, M. (2001) “Combined Parameter and State Estimation in Simulation-Based Filtering”, in: A. Doucet, N. de Freitas and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice* (Springer, New York), pp. 197-223.
- [27] Michaud, R. (1993) “Are Long-Short Equity Strategies Superior?”, *Financial Analysts Journal*, Vol. 49(6), pp. 44-49.
- [28] Nath, P. (2003) “High Frequency Pairs Trading with U.S Treasury Securities: Risks and Rewards for Hedge Funds”, Working Paper, London Business School.
- [29] Pitt, M. and Shephard, N. (1999) “Filtering via Simulation: Auxiliary Particle Filter”, *Journal of the American Statistical Association*, Vol. 94, pp. 590-599.
- [30] Ross, S. (1976) “The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory*, Vol. 13, pp. 341-360.
- [31] Shumway, R. and Stoffer, D. (1982) “An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm”, *Journal of Time Series Analysis*, Vol. 3(4). pp. 253-264.
- [32] Takahashi, A. and Sato, S. (2001) “Monte Carlo Filtering Approach for Estimating the Term Structure of Interest Rates, *Annals of the Institute of Statistical Mathematics*, Vol. 53, pp. 50-62.
- [33] Vidyamurthy, G. (2004) *Pairs Trading, Quantitative Methods and Analysis*, John Wiley & Sons, Canada.

A Appendix - EM Algorithm For Generalized LGSS Models

Below is a derivation of the EM algorithm for a generalized Linear Gaussian state space model of the form:

$$\begin{aligned}x_k &= Ax_{k-1} + BU_{1,k} + Gv_{k-1} \\y_k &= Cx_k + DU_{2,k} + H\omega_k\end{aligned}$$

where U_1 and U_2 are both exogenous inputs. The following notation applies:

$$\begin{aligned}\mu_0 &= E[x_0] && \text{Mean of the initial value } x_0 \\P_0 &= \text{Var}[X_0] && \text{Variance of the initial value } x_0 \\Y_k^N &= [y_1, y_2, \dots, y_N] && \text{Complete observations on } y \\\hat{x}_k^N &= E[x_k|Y_k^N] && \text{Smoothed estimate of } x_k \\Px_k^N &= \text{Var}[x_k|Y_k^N] && \text{Smoothed covariance matrix of } x_k \\\hat{\Psi}_j &= && \text{Value at the } j\text{-th iteration of the parameter vector}\end{aligned}$$

The **E step** computes the expectation:

$$\begin{aligned}Q(\Psi|\hat{\Psi}_{j-1}) &= \log |P_0| + \text{tr} \left\{ P_0^{-1} \left[Px_0^N + (\hat{x}_0^N - \mu_0)(\hat{x}_0^N - \mu_0)^T \right] \right\} + \\&N \log |G| + \text{tr} \left\{ G^{-1} \left[P_{11} + AP_{00}A^T + \sum_{i=1}^N Bu_{1,i}u_{1,i}^T B^T - \right. \right. \\&P_{10}A^T - AP_{10}^T - \sum_{i=1}^N Bu_{1,i}(\hat{x}_i^N)^T - \sum_{i=1}^N \hat{x}_i^N u_{1,i}^T B^T + \\&\left. \left. \sum_{i=1}^N A\hat{x}_{i-1}^N u_{1,i}^T B^T + \sum_{i=1}^N Bu_{1,i}(\hat{x}_{i-1}^N)^T A^T \right] \right\} + \\&N \log |H| + \text{tr} \left\{ H^{-1} \left[\sum_{i=1}^N CPx_i^N C^T + \right. \right. \\&\left. \left. \sum_{i=1}^N (y_i - C\hat{x}_i^N - Du_{2,i})(y_i - C\hat{x}_i^N - Du_{2,i})^T \right] \right\} \tag{11}\end{aligned}$$

where

$$P_{11} = \sum_{i=1}^N P x_i^N + \hat{x}_i^N (\hat{x}_i^N)^T \quad (12)$$

$$P_{00} = \sum_{i=1}^N P x_{i-1}^N + \hat{x}_{i-1}^N (\hat{x}_{i-1}^N)^T \quad (13)$$

$$P_{10} = \sum_{i=1}^N P x_{i,i-1}^N + \hat{x}_i^N (\hat{x}_{i-1}^N)^T \quad (14)$$

The **M step** is to minimize (13) with respect to each of matrices in Ψ . At the j -th iteration:

$$\begin{aligned} \hat{A}_j &= [P_{10} - \sum_{i=1}^N \hat{x}_i^N u_{1,i}^T (\sum_{i=1}^N u_{1,i} u_{1,i}^T)^{-1} (\sum_{i=1}^N u_{1,i} (\hat{x}_{i-1}^N)^T)] \\ &\quad [P_{00} - \sum_{i=1}^N \hat{x}_{i-1}^N u_{1,i}^T (\sum_{i=1}^N u_{1,i} u_{1,i}^T)^{-1} (\sum_{i=1}^N u_{1,i} (\hat{x}_{i-1}^N)^T)]^{-1} \end{aligned} \quad (15)$$

$$\hat{B}_j = [\sum_{i=1}^N \hat{x}_i^N u_{1,i}^T - \hat{A}_j \sum_{i=1}^N \hat{x}_{i-1}^N u_{1,i}^T] (\sum_{i=1}^N u_{1,i} u_{1,i}^T)^{-1} \quad (16)$$

$$\begin{aligned} \hat{C}_j &= [(\sum_{i=1}^N y_i (\hat{x}_i^N)^T) - (\sum_{i=1}^N y_i u_{2,i}^T) (\sum_{i=1}^N u_{2,i} u_{2,i}^T)^{-1} (\sum_{i=1}^N u_{2,i} (\hat{x}_i^N)^T)] \\ &\quad [P_{11} - (\sum_{i=1}^N \hat{x}_i^N u_{2,i}^T) (\sum_{i=1}^N u_{2,i} u_{2,i}^T)^{-1} (\sum_{i=1}^N u_{2,i} (\hat{x}_i^N)^T)]^{-1} \end{aligned} \quad (17)$$

$$\hat{D}_j = [\sum_{i=1}^N (y_i - \hat{C}_j \hat{x}_i^N) (u_{2,i})^T] (\sum_{i=1}^N u_{2,i} u_{2,i}^T)^{-1} \quad (18)$$

$$\begin{aligned} \hat{G}_j^2 &= \frac{1}{N} \left[P_{11} + \hat{A}_j P_{00} \hat{A}_j^T + \sum_{i=1}^N \hat{B}_j u_{1,i} u_{1,i}^T \hat{B}_j^T - \right. \\ &\quad \left. - P_{10} \hat{A}_j^T - \hat{A}_j P_{10}^T - \sum_{i=1}^N \hat{B}_j u_{1,i} (\hat{x}_i^N)^T - \sum_{i=1}^N \hat{x}_i^N u_{1,i}^T \hat{B}_j^T + \right. \\ &\quad \left. + \sum_{i=1}^N \hat{A}_j \hat{x}_{i-1}^N u_{1,i}^T \hat{B}_j^T + \sum_{i=1}^N \hat{B}_j u_{1,i} (\hat{x}_{i-1}^N)^T \hat{A}_j^T \right] \end{aligned} \quad (19)$$

$$\begin{aligned} \hat{H}_j^2 &= \frac{1}{N} \left[\sum_{i=1}^N \hat{C}_j P x_i^N \hat{C}_j^T + \right. \\ &\quad \left. + \sum_{i=1}^N (y_i - \hat{C}_j \hat{x}_i^N - \hat{D}_j u_{2,i}) (y_i - \hat{C}_j \hat{x}_i^N - \hat{D}_j u_{2,i})^T \right] \end{aligned} \quad (20)$$