

A pairs trading strategy based on linear state space models and the Kalman filter

CARLOS EDUARDO DE MOURA[†], ADRIAN PIZZINGA^{*‡} and JORGE ZUBELLI[†]

[†]Associação Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, Brazil

[‡]Institute of Mathematics and Statistics - Fluminense Federal University (UFF), Niterói, Brazil

(Received 28 December 2014; accepted 18 February 2016; published online 25 April 2016)

Among many strategies for financial trading, pairs trading has played an important role in practical and academic frameworks. Loosely speaking, it involves a statistical arbitrage tool for identifying and exploiting the inefficiencies of two long-term, related financial assets. When a significant deviation from this equilibrium is observed, a profit might result. In this paper, we propose a pairs trading strategy entirely based on linear state space models designed for modelling the *spread* formed with a pair of assets. Once an adequate state space model for the spread is estimated, we use the Kalman filter to calculate conditional probabilities that the spread will return to its long-term mean. The strategy is activated upon large values of these conditional probabilities: the spread is bought or sold accordingly. Two applications with real data from the US and Brazilian markets are offered, and even though they probably rely on limited evidence, they already indicate that a very basic portfolio consisting of a sole spread outperforms some of the main market benchmarks.

Keywords: Kalman filter; Mean-reverting conditional probabilities; Pair; Pairs trading; Spread; State space models; Statistical arbitrage

1. Introduction

Pairs trading is a type of statistical arbitrage that was first implemented in the mid-80s by Nunzio Tartaglia and his group at Morgan Stanley (*cf.* Vidyamurthy 2004). Currently, pairs trading is widely used by investment banks and hedge funds. In general terms, a pairs trading strategy aims to identify and exploit market inefficiencies observed with two long-term, related assets, mostly by using statistical methods. The two assets are said to form a *pair*. When a significant deviation of the prices between the two assets is detected, a trading position is taken: the higher priced asset is sold (this is the so-called *short position* by market practitioners) and the lower priced asset is bought (that is, a *long position* is taken), with the hope that mispricing will correct to the long-term equilibrium value (*cf.* Elliott *et al.* 2005, Vidyamurthy 2004).

In this paper, we consider two linear state space models that are appropriate for modelling *spreads* (stationary linear combinations of long-term related assets), with the intent of testing a new quantitative strategy involving pairs trading. The first model is the unobserved component model proposed by Elliott *et al.* (2005). Such a model, which has a Gaussian linear state space form, is a discrete-time version of the linear mean-reverting Ornstein–Uhlenbeck model. The second model is the traditional stationary autoregressive moving average, or

ARMA, model (*cf.* Brockwell and Davis 1991, 2003, Hamilton 1994, Enders 2004), and its particular specifications are also dealt with in this paper under appropriate linear state space forms. In fact, we will prove that this second class of models, even though they lack theoretical finance support, encompasses the proposal made by Elliott *et al.* (2005) as a particular case.

Subsequently, we develop a methodology for calculating conditional probabilities (given the past and actual spread data) that the spread will return to its long-term mean *k*-steps ahead (the frequency can be daily or intra-daily) whenever significant deviations are observed. We propose an alternative augmented state space form for a previously selected model estimated using spread data, and with this enlarged state space form, we apply the Kalman filter *k*-steps-ahead prediction (see, for instance, Harvey 1989, Durbin and Koopman 2001) to obtain conditional mean vectors and covariance matrices of the *k* future spreads. These first- and second-order moments are all that are needed for calculating the conditional probabilities previously mentioned. The quantitative strategy we pursue here is activated according to the following rule: if the spread is found to be considerably below (above) its long-term mean and the conditional probability that the spread will increase above (decrease below) its long-term mean by *k*-steps ahead is reasonably large, buy (sell) the spread.

The contribution made by this paper to the literature on pairs trading is the paradigm related to the trading rule briefly

*Corresponding author. Email: adrianhp@est.uff.br

described above: one takes positions on the assets forming the pair by checking whether the spread is too positive or too negative and also by examining the probability that the spread will not take too long to cross its long-term value (which is the probability that a profit will result soon).

The paper is organized as follows. Section 2 reviews the literature on pairs trading, without claiming exhaustiveness. Section 3 discusses pair trading from the statistical arbitrage standpoint, enumerating some of its main practical features. Section 4 presents the two aforementioned linear state space models, discusses their mathematical properties and embeds each of them into the state space modelling/Kalman filter framework. Section 5 formally discusses how the conditional probabilities that the spread will mean-revert are calculated, addresses the corresponding computational issues and describes step by step how the quantitative strategy is implemented. Section 6 offers two applications to real data from the US and Brazilian markets and compares the performance of the proposed strategy with the main benchmarks and with a former pairs trading strategy already used by market practitioners. For each of the examples, the justification for the pair used is initially addressed using a fundamental analysis of the expected equilibrium between the two corresponding assets (section 6.1); in the sequel, the advocated equilibrium relation is assessed using proper econometric cointegration tests (beginning of section 6.2). An analysis of the computational effort related to estimation and goodness of fit is included as well. Section 7 discusses the main results obtained in the former section, provides some economic arguments in favour of our methodology and lists some comments on the use of the latter in real scenarios. The appendices review the Kalman filter methods used in the paper, provide the proofs of the technical results and explain some of the financial returns calculated in the applications.

2. Pairs trading: a glimpse at the literature

This section discusses earlier studies on pairs trading strategies, focusing mainly on spread modelling. A feature common to most of these models reviewed in section 2.1 consists of recognizing the spread associated with a pair of stocks (*cf.* the naive definition of ‘pair’ already given and used in section 1) as some kind of mean-reverting stochastic process, the parameters of which are estimated using financial market data. In section 2.2, we explain how this paper fits within the literature.

2.1. The review

The first reference that we discuss here is Vidyamurthy (2004). In this book, a good background is provided on the pairs trading universe as well as several techniques for choosing pairs trading, with a focus on cointegration tests. Moreover, the author explains how pairs trading works and surveys some methods for addressing the problem in real settings—for instance, common trends/cointegration models, arbitrage pricing theory (APT), distance measures and state space models/Kalman filter.

Gatev *et al.* (2006) studied pairs trading strategies in the US equity market with daily data over the period 1962–2002.

In their study, stocks from companies that had at least one day out of business were discarded. A pair formation for each stock was found by minimizing the squared deviations between the two normalized daily price series, where the dividends were reinvested. The basic strategy consists of opening a position in a pair when prices diverge by more than two historical standard deviations and unwinding the position whenever the prices cross each other. Should prices not cross after the end of the trading interval, gains and losses are calculated at the end of the last trading day. The performance of this strategy used by Gatev *et al.* (2006) was addressed for a Brazilian stock market case by Perlin (2007). The latter investigated the period from 2000 until 2006 and tested different conditions of long and short, ranging between 1.5 and 3 standard deviations. For the data set used, the best options were those between 1.5 and 2 standard deviations.

A very seminal paper entirely dedicated to state space modelling for spreads was authored by Elliott *et al.* (2005), where a Gaussian linear state space model for the mean-reversion behaviour of the spread between paired stocks was developed under a continuous time setting. It is assumed that the ‘observed’ spread S_t is a noisy observation of some mean-reverting ‘unobserved’ spread x_t . The set-up for parameter estimation is based on a version of the expectation–maximization algorithm previously developed in Elliott and Krishnamurthy (1999). The pairs trading strategy proposed is the following: if S_t is larger/smaller than the one-step-ahead estimate $\hat{x}_{t|t-1}$, then the spread is regarded as too large/small, and thus, the trader could take a *short/long* position in the spread portfolio. Therefore, a profit is expected whenever a price correction occurs.

Another paper on state space models for spread data is that by Triantafyllopoulos and Montana (2009), where the modelling framework proposed in Elliott *et al.* (2005) is extended in several ways. First, they introduce time-varying autoregressive (or mean-reverting) parameters, which potentially allows the model to adapt itself to sudden changes in the data. Second, they develop and implement a Bayesian approach for estimating the parameters and provide an on-line estimation scheme. Finally, they advocate a procedure known as flexible least squares to estimate the cointegration coefficient recursively, unveiling a possible time-varying cointegration relationship between the two asset prices.

We now discuss two other works published in 2009. The first is a paper by Bertram (2009), where the theory of Itô diffusion processes comes into play for determining optimal trading strategies that also take into account transaction costs. The empirical content of the paper makes use of the Ornstein–Uhlenbeck modelling of the spread of a security traded simultaneously in both the Australian and New Zealand stock markets. The second paper, by Huck (2009), offers a data-driven and multi-criteria decision method for selecting pairs and implements the latter using weekly returns of S&P100 stocks.

In 2010, at least five papers addressing pairs trading techniques were published. Bertram (2010) complements his work published in 2009 by deriving analytical solutions for the expected return, the variance of the return and the expected trade length of his continuous time trading strategy—these parameters are used for constructing optimal strategies. Similarly,

Huck (2010) authored a continuation paper (of the one just surveyed in the last paragraph), where the multi-criteria decision method is enhanced by adding neural network forecasting techniques. The paper by Avellaneda and Lee (2010) employed principal component analysis with sectors Exchange Trade Funds for extracting risk factors. Baronyan *et al.* (2010) investigated 14 market-neutral trading strategies combined with different trading methods and pairs selection methods. From empirical evidence of weekly data on stocks that comprise the Dow Jones 30 index, they find that the performance of market-neutral equity trading is superior in the complicated year of 2008, the first one of the global financial crisis. Finally, Wissner-Gross and Freer (2010) proposed an econophysical perspective to generalize statistical arbitrage trading strategies for space-like separated world trading locations: one of their findings is that optimal intermediate locations exist between trading centres.

Continuing with the literature review, we now mention the paper by Mori and Ziobrowski (2011). In this mostly empirical work, the effectiveness of pairs trading in the US Real Estate Investment Trust market is compared with that in the US general stock market over the period 1987–2008. The authors conclude that the former market was more profitable than the latter between 1993 and 2000, after which pairs trading showed similar performances in both markets.

To conclude, we review three recent works on pairs trading theory and methods. The first two are works by Fasn (2013a, 2013b), who essentially proposes least squares estimators for the parameters of several versions of the Ornstein–Uhlenbeck model and fully investigates their statistical properties, such as consistency and asymptotic distributions. The usual *t*-ratio and Wald tests are also investigated in terms of their asymptotic behaviour. In the third paper by Tourin and Yan (2013), a dynamic model for pairs trading based on the theory of optimal stochastic control is proposed and illustrated using minute-by-minute historical data on two stocks traded on the New York Stock Exchange.

2.2. This paper's contribution

Given the articles and books reviewed here on pairs trading, in this paper, we intend to complement the findings of Elliott *et al.* (2005) along three directions:

- A more general class of possible probabilistic descriptions—the ARMA models—for a given spread time series is proposed. As we demonstrate, such a class encapsulates the mean-reverting model by Elliott *et al.* as a particular case.
- We create a new quantitative pairs trading strategy based upon outputs (specifically: some conditional probabilities) of the stochastic model selected and estimated using spread data.
- The whole procedure (model estimation & trading strategy) is implemented using real financial time series. The results are compared with those from other investment alternatives, including the simple pairs trading strategy proposed by Gatev *et al.* (2006) and re-considered by Perlin (2007).

Finally, we mention that our statistical framework is quite different from that of Triantafyllopoulos and Montana (2009), who work with a model that, by its very definition, has to be recognized under the conditionally Gaussian state space approach (see appendix 1). Moreover, Triantafyllopoulos and Montana make use of the Bayesian perspective for estimating the model parameters. On the other hand, we accomplish such tasks in our model using the maximum likelihood method.

3. Statistical Arbitrage Strategies

Quoting Kaufman (2005), ‘when the two legs of a spread are highly correlated and therefore the opportunity for profit from price divergence is of short duration, the trade is called an arbitrage. True arbitrage has, theoretically, no trading risk, however it is offset by small profits and limited opportunity for volume’.

Statistical arbitrage is a class of strategies widely used by hedge funds and proprietary traders. The distinctive feature of such strategies is that profits can be made by exploiting the statistical *mispricing* of one or more assets, based on their regular behaviour. Despite the use of the term ‘*arbitrage*’, such a class is not riskless. One simple and very popular strategy that fits in with the definition of statistical arbitrage is pairs trading (*cf.* Elliott *et al.* 2005). Other types of statistical arbitrage are discussed in Vidyamurthy (2004) and Pole (2007).

Following Vidyamurthy (2004), the first use of a pairs trading strategy is attributed to the Wall Street ‘quant’ Nunzio Tartaglia, who was at Morgan Stanley in the Mid-1980s. Pairs trading is based on APT (*cf.* Ross 1976). Informally speaking, if two stocks have similar characteristics, the prices of both assets must be more or less the same; that is, they maintain some degree of equilibrium. If prices diverge, then it is likely that one of the assets is overpriced and/or the other is underpriced. Basically, pairs trading schemes involve selling the higher priced asset and buying the lower priced asset with the hope that *mispricing* will be ultimately corrected by the long-term equilibrium value. The difference between the two observed prices is termed *spread*. Therefore, the idea behind a given pairs trading strategy is to trade on the oscillations around the equilibrium value of the spread. The oscillations of the spread occur because the latter is allegedly mean-reverting. One can put on a trade when the spread deviates substantially from its equilibrium value and unwind the trade when the equilibrium is restored (*cf.* Elliott *et al.* 2005). For the trade to be profitable, the deviation must be reasonably larger than trading costs.

Pairs trading is a market-neutral trading strategy. Hence, this strategy strives to provide positive returns in both *bull* and *bear* markets by selecting a large number of long and short positions with no net exposure to the market (*cf.* Nicholas 2000, Jacobs and Levy 2005). The main risks involved in a pairs trading are the following: (1) the *divergence risk*: the long-term equilibrium relation between the assets may change or even vanish; (2) the *horizon risk*: the spread does not converge in a given horizon of time, hence forcing the traders to close the position before the convergence, due to worsened mispricing or margin calls (*cf.* Engelberg *et al.* 2009).

4. Proposed models

4.1. What is a pair?

The idea behind a pair (of stocks, bonds, foreign exchanges, commodities, etc.) is closely linked to the econometric concept of *cointegration*. More precisely, two time series $Y_t \sim I(1)$ and $X_t \sim I(1)$ are said to be *cointegrated* iff $aY_t + bX_t \sim I(0)$ for some $a \neq 0$ and $b \neq 0$. Here, the notation $I(d)$ means ‘integrated of order d ’. This definition is sufficient for the scope of this paper. For richer expositions on the theme and more general definitions, please refer to Harvey (1993), Hamilton (1994) and Enders (2004).

Consider

$$S_t = \log(P_{t,1}) - [\alpha + \beta \log(P_{t,2})], \quad (1)$$

where $P_{t,1}$ and $P_{t,2}$ are the prices of assets \mathcal{A}_1 and \mathcal{A}_2 in time t , respectively. The time frequency can be daily or some kind of intraday frequency (second, minute, hour, etc.). If $\log(P_{t,1})$ and $\log(P_{t,2})$ are cointegrated, the *spread* S_t is stationary—that is, $S_t \sim I(0)$. In this case, α is the mean of the cointegration relationship, β is the cointegration coefficient, and \mathcal{A}_1 and \mathcal{A}_2 form a *pair*.

Cointegration, once verified, suggests that S_t would wander around an equilibrium value. This is actually almost indispensable for pairs trading. A suitable choice of α in equation (1) renders a value of zero. Any expressive deviations from this value can be traded against.

4.2. Unobserved component models: the stochastic spread approach

Following Elliott *et al.* (2005), in this subsection, we assume that the *observed* spread S_t , associated with a given pair of assets \mathcal{A}_1 and \mathcal{A}_2 , is a noisy realization of the *unobserved* or *actual* mean-reverting spread x_t :

$$\begin{aligned} S_t &= x_t + D\epsilon_t \\ x_t - x_{t-1} &= a - bx_{t-1} + C\eta_t, \end{aligned} \quad (2)$$

where $a \in \mathbb{R}$, $0 < b < 2$, $D > 0$, $C > 0$ and $(\epsilon_t, \eta_t)' \sim \text{NID}(0, I_2)$. In order to attain an appropriate state space representation for the model in equation (2), the second equation in the latter must be rephrased as $x_{t+1} = a + (1-b)x_t + \eta_t^*$, where $\eta_t^* = C\eta_{t+1}$. To adapt equation (A1) of appendix 1 to accommodate equation (2)—with the second equation replaced by its equivalent form—define $Z_t = 1$, $d_t = 0$, $H_t = D^2$, $T_t = B \equiv 1 - b$, $c_t = a$, $R_t = 1$ and $Q_t = C^2$. The Kalman filter formulae in equation (A2) of appendix 1 become

$$\begin{aligned} v_t &= S_t - a_{t|t-1}, & F_t &= P_{t|t-1} + D^2, \\ K_t &= B P_{t|t-1} F_t^{-1}, & L_t &= B - K_t, \quad t = 1, \dots, n. \\ a_{t+1|t} &= a + B a_{t|t-1} + K_t v_t, & P_{t+1|t} &= B P_{t|t-1} L_t' + C^2 \end{aligned} \quad (3)$$

Equation (3) can have the initial conditions $a_{1|0} = a/(1-B)$ and $P_{1|0} = C^2/(1-B^2)$. Notice that the latter are precisely unconditional first- and second-order moments of the stationary process x_t .

The model proposed by Elliott *et al.* (2005) has three interesting features. The first is that it has support in finance theory, as it can be viewed as a discrete-time version of the

Ohrstein-Uhlenbeck continuous time stochastic process (see Rampertshammer 2007). The second is that it recognizes a mean-reverting behaviour for the spread. The last good property is a consequence of the next result, the proof of which is provided in appendix 2:

PROPOSITION 1 *If S_t follows the unobserved component model by Elliott et al. given in equation (2), then $S_t \sim \text{ARMA}(1,1)$.*

This last proposition, besides encapsulating this proposal by Elliott *et al.* (2005) in a more general class of mean-reverting statistical models (next subsection), suggests a procedure for selecting/discarding equation (2) as a probabilistic description of some spread time series: if one obtains evidence from the spread data that the latter will not be adequately fitted by any ARMA(1,1) model, then the proposal by Elliott *et al.* is necessarily misspecified for being considered in a pairs trading scheme.

4.3. ARMA models: generalizing the stochastic spread approach

Because of their mean-reverting behaviour, stationary ARMA dynamics can be always considered as valid approaches for modelling the spread S_t . For instance, we could assume that $S_t \sim \text{ARMA}(2,2)$; that is,

$$S_t = \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}, \quad (4)$$

where $\epsilon_t \sim \text{NID}(0, \sigma^2)$ and $(\phi_1, \phi_2)'$ are such that the polynomial $p(z) = 1 - \phi_1 z - \phi_2 z^2$, $\forall z \in \mathbb{C}$, has its two roots outside the unit circle. The latter assumption on the coefficients ϕ_1 and ϕ_2 is a sufficient condition for S_t to be a stationary process (see Brockwell and Davis 1991, 2003, Hamilton 1994). The same restrictions could be imposed on the moving average coefficients θ_1 and θ_2 in order to guarantee that S_t is *invertible*—that is, ϵ_t can be written as a function of Y_t, Y_{t-1}, \dots , by means of an AR(∞) representation for S_t (again, see Brockwell and Davis 1991, 2003, Hamilton 1994). Fortunately, such questions regarding invertibility are immaterial under the state space modelling/Kalman filter framework, as in the latter, both likelihood function evaluation and forecasting attainable tasks are independent of the invertibility question, as cleverly discussed by Hamilton (1994) in chapters 4, 5 and 13.

We can use equation (A1) of appendix 1 to accommodate the model in equation (4) and any other stationary ARMA(p, q) model under a state space representation. Although there is no unique way of performing such a conversion and the literature has been frequently proposing several state space forms for ARIMA models (to cite a few books: Harvey 1989, Brockwell and Davis 1991, 2003, Hamilton 1994, Durbin and Koopman 2001), in this paper, the following alternative for the ARMA(2,2) model given in equation (4) will be used in the sequel:

$$\begin{aligned} Z_t &= [1 \ 0 \ 0 \ 0 \ 0], \quad d_t = 0, \quad H_t = 0, \\ T_t &= \begin{bmatrix} \phi_1 & \phi_2 & 1 & \theta_1 & \theta_2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad c_t = \begin{bmatrix} \phi_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \end{aligned}$$

$$R_t = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, Q_t = \sigma^2.$$

The Kalman filter formulae in equation (A1) (cf. appendix 1) are applied with the matrices above and can be initialized under the initial conditions $a_1 = \left(\frac{\phi_0}{1 - \phi_1 - \phi_2}, \frac{\phi_0}{1 - \phi_1 - \phi_2}, 0, 0, 0 \right)'$ and $\text{vec}(P_1) = (I - T \otimes T)^{-1} \text{vec}(RQR')$.

5. A new pairs trading strategy

In this section, we discuss the main elements of a quantitative pairs trading strategy based entirely on the estimation of the state space models proposed in section 4. First, in section 5.1, we provide theoretical details on how the conditional probabilities that the spread will return to its long-term mean, k -steps ahead from a given time instant t , are defined. In section 5.2, we explore the practical issues for effectively calculating the aforementioned probabilities in an online fashion: once an appropriate state space model is estimated using maximum likelihood (see appendix 1), the implementation of the usual Kalman filter prediction equations given in equation (A2) to an augmented version of the model is appropriate. Finally, in section 5.3, the quantitative strategy is described step by step, and the content derived in sections 5.1 and 5.2 is merged with the trading rule that involves buying or selling the spread accordingly.

5.1. Mean-reverting conditional probabilities p_{up} and p_{down} : theory

The main ingredient for success: to achieve, from a statistical/probabilistic standpoint, the minimum confidence that a future observed value of the spread will not take very long to cross back to some long-term value (for instance, its unconditional mean) once the spread observed on some time t is somewhat distant from that same long-term value. If such a task can be accomplished, one might buy (or sell) the spread on that time t whenever chances are that he or she can make a profit.

Formally, the strategy that we build is strongly based upon the ability of calculating the conditional probability that the spread will revert to its long-term mean—or any other convenient value c to be chosen—by k steps ahead, given the past and actual spread data; that is:

$$\begin{aligned} p_{up}(t, k, c) &= P[(S_{t+1} > c) \cup (S_{t+2} > c) \cup \dots \cup (S_{t+k} > c) | \mathcal{F}_t] \\ &= 1 - P[(S_{t+1} \leq c, S_{t+2} \leq c, \dots, S_{t+k} \leq c) | \mathcal{F}_t] \\ &= 1 - F_{S_{t+1}, S_{t+2}, \dots, S_{t+k} | \mathcal{F}_t}(c, c, \dots, c), \\ p_{down}(t, k, c) &= P[(S_{t+1} < c) \cup (S_{t+2} < c) \cup \dots \cup (S_{t+k} < c) | \mathcal{F}_t] \\ &= P[(-S_{t+1} > -c) \cup (-S_{t+2} > -c) \cup \dots \cup (-S_{t+k} > -c) | \mathcal{F}_t] \\ &= 1 - P[-S_{t+1} \leq -c, -S_{t+2} \leq -c, \dots, -S_{t+k} \leq -c | \mathcal{F}_t] \\ &= 1 - F_{-S_{t+1}, -S_{t+2}, \dots, -S_{t+k} | \mathcal{F}_t}(-c, -c, \dots, -c), \end{aligned} \quad (5)$$

where \mathcal{F}_t is the σ -field generated by past and actual spread data; that is, $\mathcal{F}_t \equiv \sigma(S_1, \dots, S_{t-1}, S_t)$. If the assumption that a specific Gaussian linear state space model is appropriate for the spread (something that is to be checked for in practical implementations), the conditional distribution functions described in equation (5) correspond to

$$S_{t,k} \equiv \begin{bmatrix} S_{t+1} \\ S_{t+2} \\ \vdots \\ S_{t+k} \end{bmatrix} | \mathcal{F}_t \sim N(\mu_{t,k}, \Sigma_{t,k}), \quad (6)$$

where $\mu_{t,k} \equiv E(S_{t,k} | \mathcal{F}_t)$ and $\Sigma_{t,k} \equiv \text{Var}(S_{t,k} | \mathcal{F}_t)$. Using the notation established in appendix 1 for the key quantities related to the Kalman filter and also defining $P_{t+i,t+j|t} \equiv \text{Cov}(\alpha_{t+i}, \alpha_{t+j} | \mathcal{F}_t)$, for $i, j = 1, 2, \dots, k$ and $i < j$ (recall that $P_{t+i,t+j|t} = P'_{t+j,t+i|t}$), it follows that each entry of $\mu_{t,k}$ is given by

$$\begin{aligned} E(S_{t+i} | \mathcal{F}_t) &= E(Z_{t+i}\alpha_{t+i} + d_{t+i} + \epsilon_{t+i} | \mathcal{F}_t) \\ &= Z_{t+i}E(\alpha_{t+i} | \mathcal{F}_t) + d_{t+i} + E(\epsilon_{t+i} | \mathcal{F}_t) \\ &= Z_{t+i}E(\alpha_{t+i} | \mathcal{F}_t) + d_{t+i} + E(\epsilon_{t+i}) \\ &= Z_{t+i}a_{t+i|t} + d_{t+i}. \end{aligned} \quad (7)$$

Regarding $\Sigma_{t,k}$, its diagonal and off-diagonal blocks are, respectively, given by

$$\begin{aligned} \text{Var}(S_{t+i} | \mathcal{F}_t) &= Z_{t+i}\text{Var}(\alpha_{t+i} | \mathcal{F}_t)Z'_{t+i} + \text{Var}(\epsilon_{t+i} | \mathcal{F}_t) \\ &\quad + Z_{t+i}\text{Cov}(\alpha_{t+i}, \epsilon_{t+i} | \mathcal{F}_t) + \text{Cov}(\epsilon_{t+i}, \alpha_{t+i} | \mathcal{F}_t)Z'_{t+i} \\ &= Z_{t+i}P_{t+i|t}Z'_{t+i} + H_{t+i}, \end{aligned} \quad (8)$$

$$\begin{aligned} \text{Cov}(S_{t+i}, S_{t+j} | \mathcal{F}_t) &= \text{Cov}(Z_{t+i}\alpha_{t+i} + d_{t+i} + \epsilon_{t+i}, Z_{t+j}\alpha_{t+j} + d_{t+j} + \epsilon_{t+j} | \mathcal{F}_t) \\ &= Z_{t+i}\text{Cov}(\alpha_{t+i}, \alpha_{t+j} | \mathcal{F}_t)Z'_{t+j} + Z_{t+i}\text{Cov}(\epsilon_{t+i}, \epsilon_{t+j} | \mathcal{F}_t) \\ &\quad + \text{Cov}(\epsilon_{t+i}, \alpha_{t+j} | \mathcal{F}_t)Z'_{t+j} + \text{Cov}(\epsilon_{t+i}, \epsilon_{t+j} | \mathcal{F}_t) \\ &= Z_{t+i}P_{t+i,t+j|t}Z'_{t+j}. \end{aligned} \quad (9)$$

5.2. Mean-reverting conditional probabilities p_{up} and p_{down} : practical evaluation

For each t , the first- and second-order conditional moments displayed in equations (7) and (8) are obtained from the Kalman filter in equation (A2) applied with the data subset $\{S_1, S_2, \dots, S_t\}$ enlarged with k missing values after the last spread S_t : $\{S_1, S_2, \dots, S_t, \text{NaN}, \text{NaN}, \dots, \text{NaN}\}$, where the acronym 'NaN' stands for 'Not available Number'. Following Durbin and Koopman (2001, section 4.9), this indicates that, under the state space modelling approach, forecasting is a particular case of missing value estimation. On the other hand, equation (9) depends on an additional implementation of Kalman recursions other than those revisited in appendix 1—specifically, the ones derived in Durbin and Koopman (2001, section 4.5), with appropriate adaptations for the case of missing values. To avoid this computational effort, which is not always available as a ready-to-use option offered by commercial softwares and is not considered in the usual Kalman filter codes suggested in textbooks, we propose an alternative in this paper. Our proposal

makes use of already-implemented formulae known to time series analysts.

The building block for routinely evaluating equations (7), (8) and (9) for each time t involves using an augmented state space form equivalent to a given time series model formerly selected and estimated using the spread data. In this paper, the models considered are those previously discussed in sections 4.2 and 4.3. This task consists of adding k new blocks to the state vector in equation (A1) of appendix 1, and each one has the same dimensions as those of the original state vector. Formally:

$$Y_t = [Z_t \ 0 \ \dots \ 0] \begin{bmatrix} \alpha_t \\ \alpha_{t-1} \\ \vdots \\ \alpha_{t-k} \end{bmatrix} + d_t + \varepsilon_t,$$

$$\begin{bmatrix} \alpha_{t+1} \\ \alpha_t \\ \vdots \\ \alpha_{t-(k-1)} \end{bmatrix} = \begin{bmatrix} T_k & 0 & \dots & 0 \\ I & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \alpha_{t-1} \\ \vdots \\ \alpha_{t-k} \end{bmatrix} + \begin{bmatrix} c_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} R_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_t, \quad (10)$$

where Z_t , d_t , T_t , R_t and c_t are the system matrices of the original model. With this enlarged state space form, we apply the Kalman filter k -steps-ahead prediction for a given time t to obtain first- and second-order conditional moments of $(\alpha_{t+1}, \dots, \alpha_{t+k})'$; with these quantities, the calculation of the first- and second-order moments displayed in equations (7), (8) and (9) becomes straightforward.

Denote the vectors of the unknown parameters associated with equations (10) and (A1) by ψ^{\ddagger} and ψ^{\dagger} , respectively, and the corresponding likelihood functions by \mathcal{L}^{\ddagger} and \mathcal{L}^{\dagger} . Because the augmented model does not include any new parameters, it trivially follows that $\psi^{\ddagger} = \psi^{\dagger}$. Even though it is not that easy to claim the same for the maximum likelihood estimators obtained under \mathcal{L}^{\ddagger} and \mathcal{L}^{\dagger} , the next proposition, the proof of which is in appendix 3, asserts that it is indeed the case:

PROPOSITION 2 $\hat{\psi}^{\ddagger} \equiv \arg \max \mathcal{L}(\psi^{\ddagger}) = \arg \max \mathcal{L}^{\ddagger}(\psi^{\ddagger}) \equiv \hat{\psi}^{\dagger}$.

This result and its proof are admittedly inspired by theorem 2 of Atherino *et al.* (2010), and we decided to include them here in detail, with proper adaptations of the former proof, so that this paper is more self-contained.

The interpretation of proposition 2 is that there are no changes in the maximum likelihood estimation when considering the augmented model in equation (10); hence, one does not need to use the latter to estimate the parameters, which would result in additional and unnecessary computational endeavour. Instead, the estimation of unknown parameters can be accomplished using the original model in equation (A1) of appendix 1, and the estimates obtained are used with the augmented model. From a practical standpoint, this result is important in the applications of section 6 for speeding up the calculation of the probabilities in equation (5).

Finally, once $\mu_{t,k}$ and $\Sigma_{t,k}$ in equation (6) are calculated, the conditional probabilities in equation (5) are evaluated using standard numerical multiple integration algorithms, which have been adapted for multivariate normal distributions

framework—see for instance Drezner and Wesolowsky (1989), Genz (1992, 2004) and Drezner (1994).

5.3. The strategy

Assuming that a particular state space model has been already estimated with available time series data from the spread process S_t —the latter is associated with a pair of assets \mathcal{A}_1 and \mathcal{A}_2 —, that the numerical devices discussed in section 5.2 have been implemented, and that the capital is invested at some low-risk fixed income market, we now propose our trading rule. It can be split into two mutually exclusive situations:

- If the observed value of S_t is found to be minimally lower than (let us say for δ units) a long-term value c , which is the same as that used in equation (5) and previously fixed to a particular value (for instance: $c = 0$, should one choose the spread mean), and p_{up} in equation (5) is found to be greater than some ‘large’ value p_{up}^* , use the capital to buy the spread.
- If the observed value of S_t is found to be minimally larger than c (without loss of generality, consider the same amount δ) and p_{down} in equation (5) is found to be greater than some ‘large’ value p_{down}^* , use the capital to sell the spread.

The items above deserve some qualification and complementation. First, the meaning of the expression ‘buy the spread’ is that the lower priced asset (in this case, \mathcal{A}_1 —see equation (1)) is bought and the other asset is sold. The expression ‘sell the spread’ can be analogously explained. Second, notice that either the first situation (long position on the spread) or the second (short position on the spread) occurs when the spread deviates from the long-term value for more than or less than δ , the latter being a threshold that guarantees a minimum profitable trade after costs. Third, because their values are fixed, p_{up}^* and p_{down}^* necessarily reflect risk aversion, and one assuredly has the option of choosing different values for each one. Fourth, we can disable the position (either long or short) when either the spread hits the long-term value c or does not hit c by k time instants ahead—recall equation (5). When the position is disabled, the capital is immediately shifted back to the previous fixed income market. Finally, even though the two situations in which the spread is supposed to be bought or sold are mutually exclusive, these are certainly not exhaustive: indeed, if none of the conditions required for each of them are met, the capital remains invested at the very fixed income market until one of the two ‘triggers’ is activated.

The latter trading rule is repeated each time the observed value of S_t and the appropriate mean-reverting conditional probability (p_{up} or p_{down}) concurrently meet the required conditions. The choices for the parameters δ , p_{up}^* , p_{down}^* , c and k considered in this paper will be given in the applications of section 6. When the strategy described is used, the main risk involved is related to specific fundamental changes: the prices of \mathcal{A}_1 and \mathcal{A}_2 may diverge, which means that the spread, which is not stationary anymore, does not reach its former long-term value c . The parameter k has the function of mitigating such a divergence risk. Another aspect is that the target return must always be higher than the one corresponding to the fixed income

market because it is the opportunity cost inherent to this strategy. This issue is allegedly addressed using the parameter δ .

6. Applications

This section presents the results of applying models from section 4 and the pairs trading strategy derived in section 5 with real data from the US and Brazilian markets. In section 6.1, we describe the data used in the estimations and justify our choice of the stocks as candidates to form pairs. For each case, an effort is made to examine the expected equilibrium between the pair of stock prices in light of the existing economic relation between both firms. In section 6.2, we present the results on cointegration tests (which statistically confirm the economic insights), model estimation and goodness-of-fit, and the strategy performances.

6.1. The data & economic justifications

All the financial time series used in the implementations were obtained from Bloomberg Professional service. Two of them, considered in one of the two exercises offered here, consist of daily stock prices of two securities, Exxon Mobil Corporation (traded in the NYSE with the symbol XOM) and Southwest Airlines Co (traded in the NYSE with the symbol LUV). ExxonMobil Corporation is the world's largest traded international oil and gas company and has its headquarters located in Texas in the US. Southwest Airlines Co operates passenger airlines that provide scheduled air transportation services in the United States. For these two stocks, the period considered ranges from 22 September 2011 to 26 March 2013, which has been split into two parts, 22 September 2011 to 20 September 2012 (in-sample) and 21 September 2012 to 26 March 2013 (out-of-sample). Two other series, corresponding to the second exercise, are daily stock prices of Vale (traded in the stock exchange BMF&BOVESPA in São Paulo with the symbol VALE5) and Bradespar (traded in the stock exchange BMF&BOVESPA in São Paulo with the symbol BRAP4). Vale is the second largest mining company in the world and the largest private company in Brazil. It is the largest producer of iron ore in the world and the second largest producer of nickel. Bradespar is an investment company that seeks to create value for its shareholders through relevant interests in companies that are leaders in their operational areas. Currently, Bradespar holds a stake in Vale, acting directly through senior management, with members on the Board of Directors and Advisory Committees. We have used available data for these two stocks from 29 August 2011 to 3 April 2013. As performed previously with the two stocks from the US market, this whole period was divided into two parts, one ranging from 29 August 2011 to 20 September 2012 and the other containing the remainder data. In view of the definition of the pair given and discussed in section 4, the stocks described above have been chosen mainly because, in view of the details given above, XOM and LUV—similar to VALE5 and BRAP4—are supposedly long-term related.

Table 1. Engle–Granger cointegration tests with the pairs (in-sample analysis).

Pairs	Dicker–Fuller test*
XOM-LUV	−3.006**
VALE5-BRAP4	−4.059**

*Critical values considered have been taken from MacKinnon (2010).

**Pair was considered stationary at a 5% level.

Additionally, the following asset class indexes have been used in the evaluation of strategy results:

- **Libor—1 year:** This indicator stands for London Inter-bank Offered Rate. It is the rate that banks use to borrow from and lend to one another in the wholesale money markets in London.
- **Standard and Poor's 500 Index (S&P):** This is a capitalization-weighted index of 500 stocks representing all major industries and is designed to measure the performance of the broad domestic economy through changes in the aggregate market.
- **Inter-bank deposit certificate (CDI):** This indicator is the overnight rate in Brazil. As such, it plays the same role as Libor. Despite being a market rate, the CDI is closely tied to the interest rate, which is fixed by the Brazilian Central Bank for monetary policy decisions.
- **Bovespa Index (Ibovespa):** This is the main indicator of the Brazilian stock market's average performance. The relevance of this index is due to several reasons: one is the integrity of its historical series, which has been regularly calculated without any methodological change since its inception in 1968.

6.2. Results

We begin by checking whether XOM-LUV and VALE5-BRAP4 show degrees of mutual equilibrium in the periods considered. This is assessed by testing cointegration hypotheses (see section 4.1 of section 4). We used the two-step Engle Granger cointegration test, which is essentially an augmented Dickey–Fuller unit root test performed with the ordinary least squares (OLS) residuals (this is the second step), obtained after regressing one time series on the other (this is the first step); the critical values for the unit root test must be conveniently modified—*cf.* Engle and Granger (1987), Enders (2004, chapter 6), and MacKinnon (2010). Once the cointegration hypothesis is not rejected, the spread to be considered in upcoming analyses is simply the OLS residuals—recall equation (1) in section 4.1 of section 4. The Engle–Granger tests were implemented in EViews 4.0 with the in-sample parts of the data (see previous section 6.1 for details). From table 1, we see that the data provide enough evidence in favour of cointegration for both XOM-LUV and VALE5-BRAP4, supporting the previous fundamental/economic conjectures of section 6.1.

We now examine the information presented in table 2, which still considers the in-sample data-sets formally defined for both pairs in section 6.1. This contains information on the goodness of fit for three parsimonious ARMA(p, q) models and the model proposed by Elliott *et al.*, along with some diagnostics

performed using the standardized residuals $v_t^S = \frac{v_t}{\sqrt{F_t}}$, where v_t and $\sqrt{F_t}$ are obtained from equation (A2). MATLAB 7.6.0 was used for the implementations. The unknown parameters were estimated using maximum likelihood, and we adopted the exact log-likelihood function displayed in equation (A3) (see appendix 1). First, we see that, for each of the models estimated using spreads from both the US and Brazilian markets, the data are reproduced by each of the four models almost under similar capabilities according to Pseudo R^2 and MSE measures. However, AIC and BIC criteria reveal that the AR(1) model, which is the simplest option, shows a slightly better complexity/fit relation. Before addressing the diagnostics, it is worth noting that if a given linear Gaussian state space model is adequate for the data at hand, the standardized residuals must behave like the observed values of i.i.d. standard normal random variables. Regarding serial dependence, Ljung–Box tests for both level and squared standardized residuals showed good results for all the models and spreads from both markets. Regarding the normality assumption, the Jarque–Bera normality test and the coverage Kupiec tests agreed on revealing adequacy for the pair XOM-LUV. On the other hand, even though the Kupiec tests suggested that the standardized residuals from all the four models estimated using the VALES-BRAP4 spread seem to come from a probability distribution similar to the standard normal distribution in terms of the tails, the Jarque–Bera test unveiled discrepancies. Therefore, some care must be exercised in interpreting and even using the conditional probabilities p_{up} and p_{down} in equation (5) for trading decisions: p_{up} and p_{down} might not be ‘tail’ probabilities.

We now discuss our pairs trading strategy performances. This time, as opposed to previous tasks (cointegration testing, parameter estimations and goodness-of-fit analysis), we also consider the out-of-sample parts of the data-sets for both pairs. Therefore, additionally to address performances during the period spanning the first year, we investigate the ability of our strategy to make profits as compared with other investment alternatives during a period spanning about six months without re-estimating any parameter. This should be viewed as an assessment of how robust our proposed methodology as a whole might be in real scenarios when it may perhaps take some time to update/calibrate the statistical models for the spread time series.

The parameter c is set to zero, which is the long-term mean of the spreads, as these are precisely the OLS residual time series from the cointegration regressions. The parameter δ is set to 0.5% to overcome operating costs, due to slippage (this is the difference between the trade expected price and the trade actual price) and transaction. In view of these two choices for c and δ , a position to buy (sell) spread is open if and only if the spread is less (greater) than $-\delta$ ($+\delta$). Finally, for the conditional probabilities p_{up} and p_{down} , their threshold values p_{up}^* and p_{down}^* are both set to 80% and the parameter k is fixed at 25, meaning that the strategy will be closed if, once the spread is bought or sold, the pair does not return to its long-term mean in 25 days at the current market prices, with the latter being an event with a conditional probability of 20% at the most.

Table 3 and figure 1 display the results corresponding to the pair XOM-LUV for the four linear state space models already under investigation. They also show the results of

the traditional benchmarks of the US financial market already detailed in section 6.1, and the performance of what we term the ‘plain strategy’: the ‘spread’ for this strategy is defined as the ratio between the highest and lowest price assets, and the trading strategy, formerly addressed by Gatev *et al.* (2006), involves opening a position with two assets whenever their corresponding spread deviates more than two historical (sample) standard deviations and unwinding the position when it returns to the spread historical mean. In case the prices do not converge at the end of the trading interval, gains and losses are calculated at the end of the last trading day. All the returns observed from the spreads considered here in this paper, for both our strategy and the plain strategy, have been calculated according to the directions given in appendix 4. From table 3, Sharpe ratios, calculated here as the main criterion for choosing the best strategy (as these measure return performances adjusted to the market risk *cf.* Sharpe (1966, 1994), indicate that the best trading options for the whole period considered (including the out-of-sample part as well) are the ones related to our pairs trading strategy implemented using the AR(2) and ARMA(1,1) models, which both present the same cumulative return, historical volatility and maximum drawdown. The Sharpe ratio for the plain strategy has a negative value and is therefore not shown. The cumulative and average returns corresponding to the AR(2) and ARMA(1,1) models are larger than the other investment opportunities, except for the stock index (S&P), which showed a strong upward trend in the out-of-sample period, as illustrated in each panel of figure 1 by the corresponding return lines during the time instants after observation 250. Economic explanations for this excellent performance of the US stock market in the mentioned period would include the US economy expansion in the first quarter of 2013 and an agreement reached by the US federal government regarding the US debt ceiling. However, due to its quite larger volatility, the S&P had a worse Sharpe ratio and a larger maximum drawdown. Additionally, both our strategy and the plain strategy with the S&P displayed low correlations: the plain strategy exhibited a better performance, as the latter and the S&P were virtually uncorrelated. This evidence was previously expected, as the type of quantitative strategy considered is one that is supposedly market neutral. On the other hand, based on the ability to make profits when a trading position on the spread is opened, our strategy proved to be considerably superior to the plain strategy, as gains were achieved 90% of the times with the former (see the fifth performance measure in table 3). Figure 1 depicts cumulative returns for the four state space models, together with cumulative returns from the market indices and the plain strategy, corroborating and illustrating the findings presented in table 3.

Likewise, both table 4 and figure 2 present the results for the pair VALES-BRAP4. The best performance, relying once again on Sharpe ratio comparisons (which were negative for both the Ibovespa domestic stock index and the plain strategy and for two models considered with our strategy), is that corresponding to the AR(1) model. Additionally, like all the other models and the plain strategy, the AR(1) model has also shown almost no correlation at all with Ibovespa. In figure 2, it is suggested that cumulative returns of our pairs trading strategy, implemented with this best AR(1) model, maintained an upward trend with relatively low volatility, probably

Table 2. Results from in-sample estimations (p -values in parentheses).

	XOM-LUV				VALE5-BRAP4			
Attribute	AR(1)	AR(2)	ARMA(1,1)	Elliott	AR(1)	AR(2)	ARMA(1,1)	Elliott
Log-likelihood	−989.044	−989.044	−989.044	−989.081	−1053.502	−1060.960	−1068.300	−1068.310
Pseudo R^2	0.896	0.896	0.896	0.902	0.767	0.780	0.788	0.789
$MSE \times 10^{-4}$	2.299	2.298	2.299	2.161	0.905	0.857	0.8130	0.8110
AIC	7.865	7.873	7.873	7.882	8.377	8.444	8.502	8.510
BIC	7.893	7.915	7.915	7.938	8.405	8.486	8.544	8.566
LR Kupiec test (superior)*	0.077	0.077	0.077	0.434	0.987	0.987	0.015	0.015
	(0.781)	(0.781)	(0.781)	(0.510)	(0.320)	(0.320)	(0.903)	(0.903)
LR Kupiec test (inferior)*	0.434	0.434	0.434	0.434	2.952	0.434	0.434	0.434
	(0.510)	(0.510)	(0.510)	(0.510)	(0.086)	(0.510)	(0.510)	(0.510)
Ljung-Box test 1(20 lags)**	13.718	13.727	13.726	13.684	29.706	23.628	18.040	18.035
	(0.845)	(0.844)	(0.844)	(0.846)	(0.075)	(0.259)	(0.585)	(0.585)
Ljung-Box test 2 (20 lags)***	29.557	26.679	29.669	29.582	13.891	13.694	16.381	16.473
	(0.148)	(0.145)	(0.145)	(0.152)	(0.836)	(0.832)	(0.693)	(0.687)
Jarque–Bera test	0.709	0.706	0.706	0.703	22.602	24.308	24.880	24.914
	(0.685)	(0.687)	(0.689)	(0.688)	(0.002)	(0.001)	(0.001)	(0.001)
Mean****	0.069	0.068	0.068	0.085	−0.019	−0.029	−0.045	−0.049
Variance****	0.999	0.999	0.999	0.997	1.004	1.003	1.002	1.002

*These are likelihood ratio unconditional coverage tests proposed by Kupiec (1995). The first and second tests check the standard residual violations of 95 and 5% standard normal distribution quantiles (that is, 1.65 and -1.65), respectively.

**This test has been performed using the standardized residuals.

***This test has been performed using the squared standardized residuals.

****These sample statistics have been calculated using the standardized residuals.

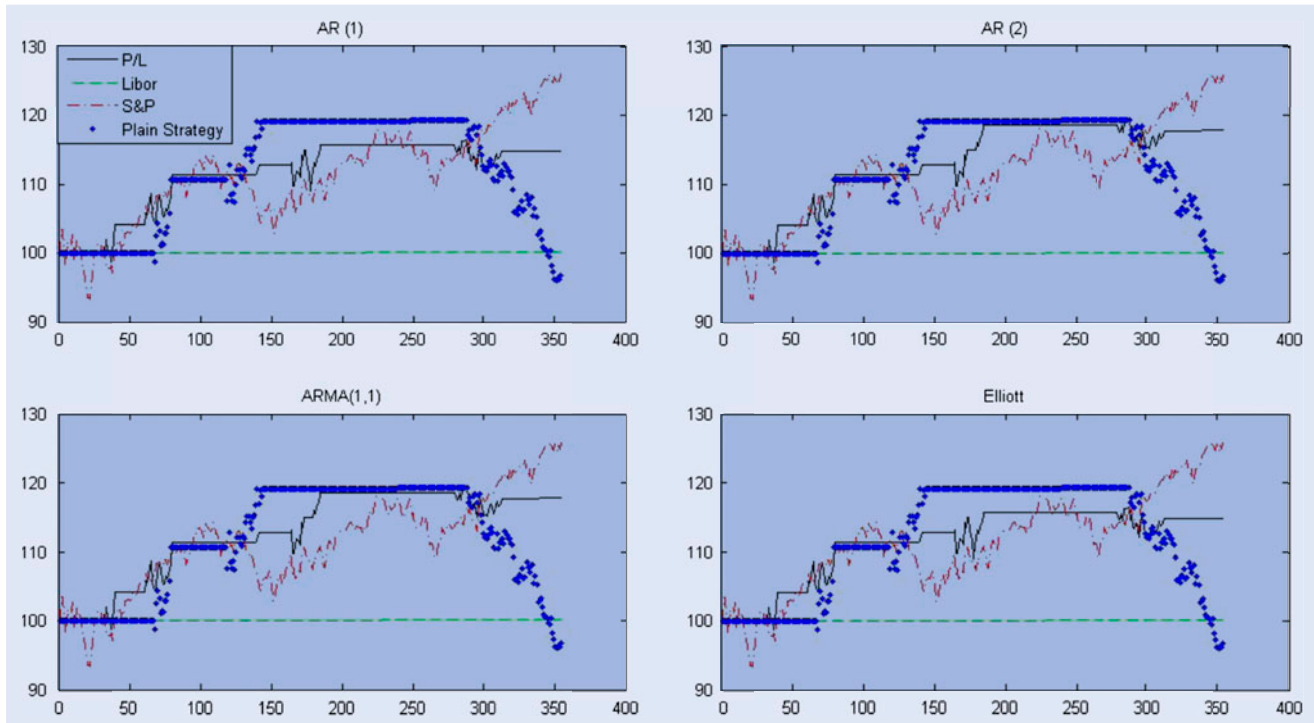


Figure 1. Comparison of the cumulative returns: strategy P/L with the pair XOM-LUV, Libor, S&P and plain strategy (whole period analysis).

corroborating the best Sharpe ratio. In terms of Ibovespa, we observe that even though this benchmark did present at specific times the largest returns amongst all the investment alternatives in the period considered, its huge risky behaviour (comparing the volatilities and maximum drawdowns in table 4) is noteworthy and has certainly contributed to some temporary losses and a worse cumulative return at the very end of the out-of-sample period. This can also be seen from the downward

and persistent reversals of this index in figure 2. Finally, in terms of the efficiency indicator given in table 4, our strategy has clearly outperformed the plain strategy: similar to the first exercise with the US market, the percentages of success in trading positions were in tune with the nominal threshold value of 80% for the conditional probabilities p_{up} and p_{down} .

Finally, table 5 shows the computational gain, in terms of estimation time, due to proposition 2 of this paper. Even though

Table 3. USA market data: performance measures from four different models for the spread and three benchmarks (whole period analysis).

	XOM-LUV				Benchmarks		
Measures	AR(1)	AR(2)	ARMA(1,1)	ELLIOTT	LIBOR	S&P	Plain strategy
Average return	0.047%	0.054%	0.054%	0.047%	0.0004%	0.07%	−0.006%
Volatility*	0.590%	0.556%	0.556%	0.590%	0.00005%	0.936%	0.788%
Cumulative return	14.815%	17.753%	17.753%	14.815%	0.150%	25.907%	−3.323%
Maximum drawdown**	−5.036%	−3.953%	−3.953%	−5.036%	0.000%	−9.936%	−19.491%
Efficiency***	90%	90%	90%	90%	—	—	67%
Sharpe ratio	1.322	1.685	1.685	1.322	—	1.464	—
Correlation****	0.183	0.176	0.176	0.183	0.053	1.000	0.017

*This is the standard deviation calculated using the daily returns.

**The maximum drawdown is a market risk measure for a given portfolio. It is the difference, observed in the period being analysed, between the highest peak and the lowest bottom in the value of the portfolio (see Karatzas and Shreve (1997) and Magdon-Ismail *et al.* (2004)).

***This indicator is the percentage of the total number of times when the strategy activated has resulted in profits.

****Correlation between the daily returns from the strategy P/L and the equity market (S&P).

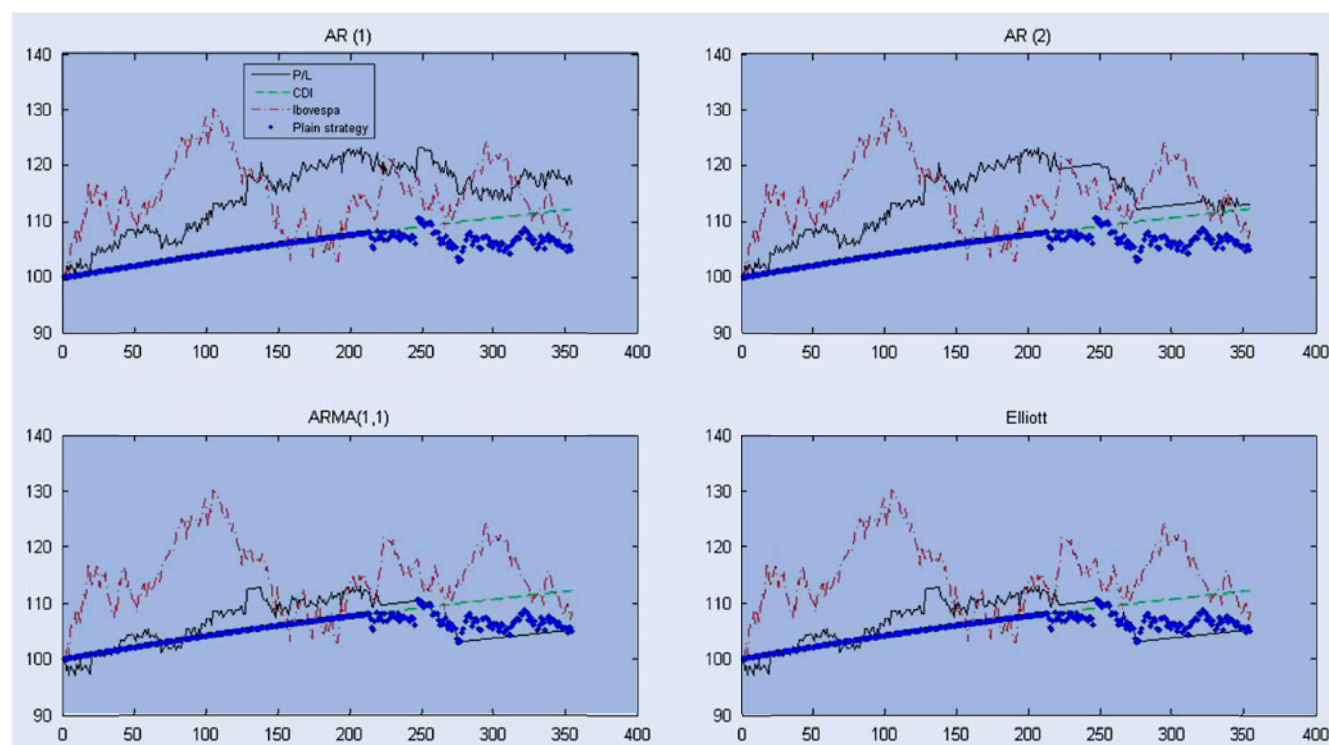


Figure 2. Comparison of cumulative returns: strategy P/L with the pair VALE5-BRAP4, CDI, IBOVESPA and plain strategy (whole period analysis).

Table 4. Brazilian market data: performance measures of four different models for the spread and three benchmarks (whole period analysis).

	VALE5-BRAP4				Benchmarks		
Measures	AR(1)	AR(2)	ARMA(1,1)	ELLIOTT	CDI	IBOVESPA	Plain strategy
Average return	0.048%	0.038%	0.029%	0.029%	0.033%	0.032%	0.015%
Volatility*	0.923%	0.815%	0.730%	0.730%	0.006%	1.362%	0.531%
Cumulative return	16.746%	13.155%	5.245%	5.245%	12.304%	8.428%	5.260%
Maximum drawdown**	−7.737%	−9.718%	−8.867%	−8.867%	−0.045%	−20.853%	−6.931%
Efficiency***	77%	82%	82%	82%	—	—	0%
Sharpe ratio	0.256	0.056	—	—	—	—	—
Correlation****	−0.027	−0.016	−0.010	−0.010	0.066	1.000	−0.012
%CDI*****	136.104%	106.918%	42.631%	42.631%	—	68.505%	40.822%

*This is the standard deviation calculated with the daily returns.

**The maximum drawdown is a market risk measure for a given portfolio. It is the difference, observed in the period being analysed, between the highest peak and the lowest bottom in the value of the portfolio—see Karatzas and Shreve (1997) and Magdon-Ismail *et al.* (2004).

***This indicator is the percentage of the total number of times when the strategy activated has resulted in profits.

****Correlation between the daily returns from the strategy P/L and the equity market (IBOVESPA).

*****Ratio between accumulated returns from the strategy P/L and the CDI in percentual terms.

Table 5. Computational times (seconds) for maximum likelihood estimation of the models with the pair VALE5-BRAP4 (in-sample analysis).

Models	Original model	Augmented model ($k = 10$)	Augmented model ($k = 15$)	Augmented model ($k = 20$)	Augmented model ($k = 25$)
ELLIOT	2.481	6.113	14.049	44.603	152.091
AR(1)	0.579	1.125	2.250	7.513	24.086
AR(2)	0.939	1.737	4.026	12.725	41.557
ARMA(1,1)	0.891	2.004	5.716	18.154	58.531

the information corresponds to model estimations with a portfolio that has only a pair of assets on a daily basis, it is plausible to assume that the augmented model would also be excessively time consuming. If we had adopted and implemented the modelling and pairs trading strategy proposed in this paper with intraday high-frequency data, the estimation times would have been increased in the case of a portfolio containing several pairs. For instance, the augmented model with $k = 25$ for Elliott's model required almost 3 min for estimation; the original model took less than three seconds.

7. Discussion

In this paper, we have developed a new pairs trading strategy based on linear state space models and the Kalman Filter. As opposed to other approaches found in the literature, neither point forecasts nor confidence bands constitute the basis for decisions on trading operations; instead, we examine the conditional probability that the value of the mispriced spread will mean-revert eventually to some pre-established horizon.

The economic motivation behind this strategy is akin to the intuitive argument used in commodities or in interest rate theory: if the price of one asset would outperform the other with very high probability over long periods of time, then going long on the former and short on the latter would generate an imbalance seen by the market - *cf.* Bessembinder *et al.* (1995), Pindyck (2001) and Nielsen and Schwartz (2004). To dwell on this financial matter and be more specific in terms of the economic intuition of our model, we must recall that the spread process S_t fluctuates around a zero equilibrium that is expected to be reached within a certain precision radius δ . At the root of such a claim is the appropriate testing to which a trader should submit a 'potential pair' of assets to confirm co-integration (see Fasn (2013a, 2013b)). Therefore, *ex ante* we have a well-established return to a close-to-zero historical value. Associated with that long-term equilibrium assumption, and given that the spread is below (above) the threshold, we have a certain probability p_{up} (respec. p_{down}) that it will mean-revert. The use of a mean-reverting model, like those considered in this paper, with a fixed divergence risk parameter k corresponds to an *interim* observation period for which the Kalman filter allows us to accumulate information. As is well-known, Kalman filters have extremely desirable properties, as they provide conditional expectations for Gaussian linear state space models (*cf.* Harvey 1989, Durbin and Koopman 2001). Therefore, for the models considered in this paper, the Kalman filter is the best estimator to adapt to the information flow. Bringing together these points, we can conclude that the use of the augmented model given in equation (10) offers a clear indication of whether we should go short or long on the spread position or do nothing but remain on the low-risk fixed income

market. In other words, the model in equation (10), which is the building block of the quantitative strategy proposed in this paper, serves ultimately as a way of incorporating the extra information provided by the data as well as of informed traders (*cf.* Baruci 2003, chapter 7, and references therein) in an algorithmic and consistent decision mechanism. As more practical exercises are still to be made, we do recognize that the empirical evidence shared in this paper is limited in order to confirm these economic perspectives. However, the two applications detailed in section 6 already prove that our strategy can be efficiently implemented and suggest that this change of direction in the usual pairs trading paradigms might work well.

At the end of this paper, we address some points potentially relevant and in tune with the financial market reality for the case of implementing the strategy under real scenarios. We start by suggesting further investigation on the parameters c , k and δ , which have been held constant in the examples in this paper (notice that nothing prevents them from being estimated or, should one prefer, optimized under the usual back-testing schemes). We may also enhance the use of such parameters. For instance, the parameter δ , although designed here to simultaneously take into account the transaction costs from both long and short positions, might be doubled: a δ_1 for one type of position and a δ_2 for the other. In the case of short positions, a very important cost, which anyone willing to adopt any pairs trading strategy (including ours) must pay attention to, is the rented asset cost. Because transactions fees vary according to the type of investment, analysis of the latter helps to identify how suitable our strategy is. More details on rented asset costs can be found in www.bmfbovespa.com.br for the Brazilian market and in nyse.nyx.com for the New York Stock Exchange.

Like the latter, practically every cost related to our strategy—such as commissions, the bid/ask spread, market impact and opportunity cost—should be classified as a *turnover cost*. Following Grinold and Kahn (1999, chapter 16), a turnover occurs every time one constructs or changes/rebalances a portfolio; possible motivations might include new information, risk control and, in our case, the perception (large p_{up} and p_{down} probabilities) that a specific change in the portfolio at the present time will lead to a profit in the very near future. Turnover costs are difficult to measure, increase with trade size and require quick execution and are therefore difficult to incorporate in the calculation of cumulative returns and other portfolio performance measures. Something that aggravates these problems, as is particularly relevant for this paper, is the fact that the two examples we consider are culled from two entirely different markets with different characteristics and regulations. Because turnovers by their very definition regularly occur with our strategy, we are certainly aware that accurate estimates of such costs would affect the realized

portfolio value. However, sticking to Grinold and Kahn's point of view, 'trading is itself a portfolio optimization problem, distinct from the portfolio construction problem'; therefore, 'optimal trading can lower transactions costs, though at the expense of additional short-term risk'. Bearing these last quotations in mind, we understand that to effectively combine such trading schemes for reducing (that is, optimizing) costs and a pairs trading strategy such as the one proposed in our paper deserves much more time and space. We leave this as a possible theme for upcoming papers on pairs trading.

We now dedicate some effort towards discussing the question of market neutrality. The starting point is to recall that virtually any portfolio return variation can possibly be explained by some market factors. Following the standard finance literature, the natural way of addressing this is to consider some type of factor model for the portfolio, amongst which we recall the CAPM & APT models (cf. Elton *et al.* 2014), the model by Fama and French (cf. Fama and French 1996), and the asset class factor model (cf. Sharpe 1992, de Roon *et al.* 2004). We understand that such modelling should consider the quite plausible assumption of time-varying coefficients, which would allow us to precisely see the real exposures on different financial risks and, thus, be more solid about unveiling the markets that a portfolio defined by our strategy would be neutral to. This time-varying coefficient assumption is justified here because our strategy involves time-varying trading positions on three different assets: the two assets forming the pair and a risk-free asset. Such extensions of the original factor models previously cited already exist and view the coefficients (the factor exposures) as latent stochastic processes: for instance, the reader is referred to the dynamic asset class factor models proposed in Swinkels and Van Der Sluis (2006), Pizzinga *et al.* (2011) and Marques *et al.* (2012). Statistically speaking, we are interested in the selection and estimation of stochastic coefficient regression models. Coincidentally, the method for implementing such tasks is proper linear state space modelling with the use of the Kalman filter. However, although it would require the same methodological framework of the statistical analysis already found in our paper, we would require other implementations (the stochastic coefficient regression models are quite different from, for example, the ARMA models considered in our paper). Due to space limitations and the huge relevance of market neutrality as a mainstream subject within the finance literature, we leave this for future research.

We now take a closer look at a more statistically oriented question: that of distributional assumptions. As strong violations of normality can make the quantities p_{up} and p_{down} quite unreliable as proxies for the true conditional probability of mean-reverting, an alternative for dealing with such inconvenient situations is to rely on Monte Carlo simulations of future trajectories of the spread S_t k steps ahead. For the ARMA models, this would require modelling the error term with the aid of standardized residuals. A second alternative, which releases one from choosing/modelling error distributions (but is much more demanding in computational terms), is to adopt some bootstrap procedure to estimate the mean-reverting conditional probabilities. Wall and Stoffer (2002) and Rodriguez and Ruiz (2009) are two papers from among a large list of references on bootstrapping state space models, and these two papers have methodologies that address the aims being discussed here.

Finally, we discuss the use of our strategy in high-frequency data. The analyses of these data are complicated due to irregular temporal spacing, intra-daily patterns and price discreteness (cf. Ait-Sahalia and Hansen 2010, chapter 7). Another major characteristic of high-frequency data is the strong intra-day seasonal behaviour of the volatility, as pointed out by Fouque *et al.* (2000, chapter 4). A data-generating process with strong seasonal patterns cannot be stationary. Therefore, controlling these periodical movements before fitting any time series model to the data should be a mandatory initial step. In light of these issues typically related to high-frequency situations, other state space models shall be combined with the pairs trading strategy proposed in this paper.

Acknowledgements

Our sincere thanks go to the referees, whose comments, requirements and suggestions were invaluable for improving this paper. We are also truly grateful to Cristiano Fernandes, Adrien Nguyen Huu and Paulo Cezar Carvalho for their very constructive comments. All remaining errors are ours.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Ait-Sahalia, Y. and Hansen, L., *Handbook of Financial Econometrics*, 2nd ed., 2010 (Springer: New York).
- Atherino, R., Pizzinga, A. and Fernandes, C., A row-wise stacking of the runoff triangle: State space alternatives for IBNR reserve prediction. *Astin Bull.*, 2010, **40**(2), 917–946.
- Avellaneda, M. and Lee, J.H., Statistical arbitrage in the US equities market. *Quant. Finance*, 2010, **10**(7), 761–782.
- Baronyan, S., Boduroglu, I. and Sener, E., Investigation of stochastic pairs trading strategies under different volatility regimes. *The Manchester School*, 2010, **2010** (supplement), 114–134.
- Baruci, E., *Financial Markets Theory*, 2003 (Springer: New York).
- Bertram, W.K., Optimal trading strategies for Itô diffusion processes. *Physica A*, 2009, **388**, 2865–2873.
- Bertram, W.K., Analytic solutions for optimal statistical arbitrage trading. *Physica A*, 2010, **389**, 2234–2243.
- Bessembinder, H., Coughenour, J., Seguin, P. and Smoller, M., Mean reversion in equilibrium asset prices: Evidence from the futures term structure. *J. Finance*, 1995, **50**, 361–375.
- Brockwell, P.J. and Davis, R.A., *Time Series: Theory and Methods*, 2nd ed., 1991 (Springer: New York).
- Brockwell, P.J. and Davis, R.A., *Introduction to Time Series and Forecasting*, 2nd ed., 2003 (Springer: New York).
- de Roon, F.A., Nijman, T.E. and Ter Horst, J.R., Evaluating style analysis. *J. Empirical. Finance*, 2004, **11**(1), 29–53.
- Drezner, Z., Computation of the trivariate normal integral. *Math. Comput.*, 1994, **63**, 289–294.
- Drezner, Z. and Wesolowsky, G.O., On the computation of the bivariate normal integral. *J. Stat. Comput. Simul.*, 1989, **35**, 101–107.
- Durbin, J. and Koopman, S.J., *Time Series Analysis by State Space Methods*, 2001 (Oxford Statistical Science Series: Oxford).
- Elliott, R.J. and Krishnamurthy, V., New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models. *IEEE Trans. Autom. Control*, 1999, **44**, 938–951.
- Elliott, R.J., van der Hoek, J. and Malcolm, W.P., Pairs trading. *Quant. Finance*, 2005, **5**(3), 271–276.

- Elton, E.J., Gruber, M.J., Brown, S.J. and Goetzmann, W.N., *iModern Portfolio Theory and Investment Analysis*, 9th ed., 2014 (John Wiley & Sons: Hoboken, NJ).
- Enders, W., *Applied Econometric Time Series*, 2nd ed., 2004 (John Wiley & Sons: Hoboken, NJ).
- Engelberg, J., Gao P. and Jagannathan R., An anatomy of Pairs trading: The role of idiosyncratic news, common information and liquidity, *Third Singapore International Conference on Finance*, 2009.
- Engle, R. and Granger, C., Co-Integration and error correction: Representation, estimation, and testing. *Econometrica* 1987, **55** (n°2), 251–276.
- Fama, E.F. and French, K.R., Multifactor explanations of asset pricing anomalies. *J. Finance*, 1996, **51**(1), 55–84.
- Fasen, V., Statistical estimation of multivariate Ornstein-Uhlenbeck processes and applications to co-integration. *J. Econometrics*, 2013a, **172**, 325–337.
- Fasen, V., Time series regression on integrated continuous-time processes with heavy and light tails. *Econometric Theory*, 2013b, **29**, 28–67.
- Fouque, P.J., Papanicolaou, G., Sircar, R. and K., *Derivatives in Financial Markets with Stochastic Volatility*, 2000 (Cambridge University Press: Cambridge).
- Gatev, E., Goetzmann, W. and Rouwenhorst, K., Pairs trading: Performance of a relative value arbitrage rule. *Rev. Financial Stud.*, 2006, **19**, 797–827.
- Genz, A., Numerical computational of multivariate normal probabilities. *J. Comput. Graphical Stat.*, 1992, **1**, 141–149.
- Genz, A., Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Stat. Comput.*, 2004, **14**(3), 251–260.
- Grinold, R.C. and Kahn, R.N., *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, 2nd ed., 1999 (McGraw-Hill Education: New York).
- Hamilton, J.D., *Time Series Analysis*, 1994 (Princeton University Press: Princeton, NJ).
- Harvey, A.C., *Forecasting, Structural Time Series Models and the Kalman Filter*, 1989 (Cambridge University Press: Cambridge).
- Harvey, A.C., *Time Series Models*, 2nd ed., 1993 (Harvester Wheatsheaf: Hemel Hempstead).
- Huck, N., Pairs selection and outranking: An application to the S&P 100 index. *Eur. J. Operational Res.*, 2009, **196**, 819–825.
- Huck, N., Pairs trading and outranking: The multi-step-ahead forecasting case. *Eur. J. Operational Res.*, 2010, **207**, 1702–1716.
- Jacobs, B. and Levy, K., *Market Neutral Strategies*, 2005 (John Wiley & Sons: Hoboken, NJ).
- Karatzas, I. and Shreve, S.E., *Brownian Motion and Stochastic Calculus*, 1997 (Springer: New York).
- Kaufman, P.J. *New Trading Systems and Methods*, 4th ed., 2005 (John Wiley & Sons: Hoboken, NJ).
- Kupiec, P., Techniques for verifying the accuracy of risk management models. *J. Derivatives*, 1995, **3**, 73–84.
- Mackinnon J.G., Critical Values for Cointegration Tests, Queen's Economics Department Working Paper No. 1227, Queen's University, 2010.
- Magdon-Ismael, M., Atiya, A.F., Pratap, A. and Abu-Mostafa, Y.S., On the maximum drawdown of a Brownian motion. *J. Appl. Probab.*, 2004, **41**, 147–161.
- Marques, R., Pizzinga, A. and Vereda, L., Restricted Kalman filter applied to dynamic style analysis of actuarial funds. *Appl. Stochastic Models Bus. Ind.*, 2012, **28**, 558–570.
- Mori, M. and Ziobrowski, A., Performance of Pairs trading strategy in the U.S. REIT market. *Real Estate Economics*, 2011, **39**(3), 409–428.
- Nielsen, M. and Schwartz, E., Theory of Storage and the Pricing of Commodity Claims. *Rev. Derivatives Res.*, 2004, **7**, 5–24.
- Nicholas, J.G., *Market-Neutral Investing: Long/Short Hedge Fund Strategies*, 2000 (Bloomberg Press: Princeton, NJ).
- Perlin, M.S., Evaluation of Pairs trading strategy at the Brazilian financial market. *J. Derivatives Hedge Funds*, 2007, **15**, 122–136.
- Pindyck, R., The dynamics of commodity spot and futures markets: A primer. *Energy J.*, 2001, **22**(3), 1–29.
- Pizzinga, A., Vereda, L. and Fernandes, C., A dynamic style analysis of exchange rate funds: The case of Brazil at the 2002 election. *Adv. Appl. Stat. Sci.*, 2011, **6**, 111–135.
- Pole, A., *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*, 2007 (John Wiley & Sons: Hoboken, NJ).
- Rampertshammer, S., *An Ornstein-Uhlenbeck Framework for Pairs Trading*, 2007 (Department of Mathematics and Statistics of the University of Melbourne). Unpublished Note.
- Rodriguez, A. and Ruiz, E., Bootstrap prediction intervals in state space models. *J. Time Ser. Anal.*, 2009, **30**(2).
- Ross, S., The arbitrage theory of capital asset pricing. *J. Economic Theory*, 1976, **13**, 341–360.
- Sharpe, F.W., Mutual fund performance. *J. Bus.*, 1966, **39**, 119–138.
- Sharpe, W.F., Asset allocation: Management style and performance measurement. *J. Portfolio Manage.* 1992, (winter), 7–19.
- Sharpe, F.W., The sharpe ratio. *J. Portfolio. Manage.*, 1994, **21**, 49–58.
- Shumway, R.H. and Stoffer, D.S., *Time Series Analysis and its Applications (With R Examples)*, 2nd ed., 2006 (Springer: New York).
- Swinkels, L. and van der Sluis, P.J., Return-based style analysis with time-varying exposures. *Eur. J. Finance*, 2006, **12**, 529–552.
- Tourin, A. and Yan, R., Dynamic pairs trading using the stochastic control approach. *J. Economic Dyn. Control*, 2013, **37**, 1972–1981.
- Triantafyllopoulos, K. and Montana, G., Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Comput. Manage. Sci.*, 2009, **8**, 23–49.
- Vidymurthy, G., *Pairs Trading, Quantitative Methods and Analysis*, 2004 (John Wiley & Sons: Hoboken, NJ).
- Wall, K. and Stoffer, S., A State space approach to bootstrapping conditional forecasts in ARMA models. *J. Time Ser. Anal.*, 2002, **23**(6).
- Wissner-Gross, A.D. and Freer, C.E., Relativistic statistical arbitrage. *Phys. Rev. E*, 2010, **82**, 056104.

Appendix 1. Linear state space models & the Kalman filter

By a *Gaussian linear state space model*, we mean the following measurement equation, state equation and initial state vector:

$$\begin{aligned} Y_t &= Z_t \alpha_t + d_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + c_t + R_t \eta_t, \quad \eta_t \sim \text{NID}(0, Q_t) \\ \alpha_1 &\sim N(a_1, P_1). \end{aligned} \quad (\text{A1})$$

The former equation is an affine function relating the observed p -variate time series Y_t to the generally unobserved m -variate state vector α_t , and the latter equation stipulates the state evolution through a Markovian structure. The random errors ϵ_t and η_t are independent (in time, between each other and of α_1). The system matrices Z_t , d_t , H_t , T_t , c_t , R_t and Q_t are deterministic or, at most, depend on the past value of Y_t . In the latter case, Harvey (1989, section 3.7), refers to equation (A1) as a *conditionally Gaussian* state space model.

For a given time series of size n and any time instants $t, j \in \{1, 2, \dots, n\}$, define $\mathcal{F}_j \equiv \sigma(Y_1, \dots, Y_j)$, $a_{t|j} \equiv E(\alpha_t | \mathcal{F}_j)$ and $P_{t|j} \equiv \text{Var}(\alpha_t | \mathcal{F}_j)$. *Kalman filtering* consists of recursive equations for these first- and second-order conditional moments, corresponding to one-step-ahead prediction ($j = t - 1$) and smoothing ($j = n$). The formulae corresponding to the predictions are given below:

$$\begin{aligned} v_t &= Y_t - Z_t a_{t|t-1} - d_t, \quad F_t = Z_t P_{t|t-1} Z_t' + H_t, \\ K_t &= T_t P_{t|t-1} Z_t' F_t^{-1}, \quad L_t = T_t - K_t Z_t, \quad t = 1, \dots, n, \\ a_{t+1|t} &= T_t a_{t|t-1} + c_t + K_t v_t, \quad P_{t+1|t} = T_t P_{t|t-1} L_t' + R_t Q_t R_t', \end{aligned} \quad (\text{A2})$$

The derivations of equation (A2) are found in Durbin and Koopman (2001). There are several other references on this subject that deserve mention, such as the books by Harvey (1989, 1993), Brockwell and Davis (1991, 2003), Hamilton (1994), and Shumway and Stoffer (2006).

In practice, system matrices include unknown parameters that must be estimated. By grouping all unknown parameters of the model

described in (A1) in a vector ψ , and denoting the corresponding parametric space by Θ , one can obtain an *exact* log-likelihood function using some outputs from equation (A2):

$$\log L(\psi) = -\frac{np}{2} \log -\frac{1}{2} \sum_{t=1}^n \left(\log |F_t| + v_t' F_t^{-1} v_t \right), \quad \forall \psi \in \Theta. \quad (\text{A3})$$

The maximum likelihood estimator of ψ is defined by $\hat{\psi} \equiv \arg \max_{\psi \in \Theta} \log L(\psi)$. When the normality assumption for $(\epsilon_t', \eta_t')'$ is violated, equation (A3) should be viewed as a *quasi* log-likelihood function and $\hat{\psi}$, in turn, as a *quasi* maximum likelihood estimator.

Appendix 2. Proof of proposition 1

From the second equation of equation (2), it follows that

$$x_t = a + (1-b)x_{t-1} + C\eta_t \equiv a + Bx_{t-1} + \eta_t^*,$$

where $\eta_t^* \sim N(0, C^2)$. Therefore, $(1-BL)x_t = x_t - Bx_{t-1} = a + \eta_t^*$, leading to

$$x_t = \frac{1}{(1-BL)}a + \frac{1}{(1-BL)}\eta_t^* = \frac{a}{(1-B)} + \frac{a}{(1-BL)}\eta_t^*, \quad (\text{B1})$$

where L is the usual lag operator (recall: $0 < b < 2$). Now, we substitute equation (B1) in the first equation of equation (2) to get

$$\begin{aligned} S_t &= \frac{a}{(1-BL)} + \frac{1}{(1-BL)}\eta_t^* + D\epsilon_t \\ &= \frac{a}{(1-BL)} + \frac{1}{(1-BL)}\eta_t^* + \epsilon_t^* \end{aligned} \quad (\text{B2})$$

where $\epsilon_t^* \sim N(0, D^2)$.

Applying the operator $(1-BL)$ on both sides of equation (B2),

$$S_t^* \equiv (1-BL)S_t = a + \eta_t^* + \epsilon_t^* - B\epsilon_{t-1}^* \quad (\text{B3})$$

From equation (B3), it is straightforward to see that

$$\gamma(0) = C^2 + (1+B^2)D^2, \quad \gamma(1) = -BD^2, \quad \gamma(k) = 0, \quad k \geq 2, \quad (\text{B4})$$

where $\gamma(k) = \text{Cov}(S_t^*, S_{t-k}^*)$, $k = 0, \pm 1, \pm 2, \dots$

From equation (B4) and Brockwell and Davis (1991, p. 89, proposition 3.2.1), it follows that $S_t^* \sim \text{MA}(1)$. \square

Appendix 3. Proof of proposition 2

We have to prove that the likelihood functions of models \dagger (original) and \ddagger (augmented) are equal; in other words, $\mathcal{L}^\dagger = \mathcal{L}^\ddagger$ over all the parametric space. It is sufficient to show that $v_t^\dagger = v_t^\ddagger$ for each $t = 1, \dots, n$. Notice that

$$v_t^\dagger = Y_t^\dagger - Z_t a_{t|t-1}^\dagger - d_t^\dagger \quad \text{and} \quad v_t^\ddagger = Y_t^\ddagger - Z_t a_{t|t-1}^\ddagger - d_t^\ddagger, \quad (\text{C1})$$

where $a_{t|t-1}^\ddagger \equiv E(\alpha_t | \mathcal{F}_{t-1}^\ddagger)$ and $d_t^\ddagger = d_t^\dagger$. Under the augmented model in equation (10) and using the convention that $\prod_{j=1}^k T_j = T_1 \cdots T_k$, i.e. the product is taken from left to right in increasing order, the recursive solution for the measurement equation, for an arbitrary

$s = 1, \dots, t-1$, is

$$\begin{aligned} Y_s^\ddagger &= [Z_s \ 0 \ \cdots \ 0] \left\{ \prod_{j=1}^{s-1} \begin{bmatrix} T_j & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} [\tilde{\alpha}_1] \right. \\ &\quad \left. + \sum_{j=1}^{s-2} \left[\prod_{k=j+1}^{s-1} \begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \left(\begin{bmatrix} R_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_j + \begin{bmatrix} c_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) \right] \right\} \\ &\quad + [Z_s \ 0 \ \cdots \ 0] \left\{ \begin{bmatrix} c_{s-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} R_{s-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_{s-1} \right\} + d_s + \varepsilon_s, \quad (\text{C2}) \end{aligned}$$

where $\tilde{\alpha}_1$ is an initial state vector with appropriate first and second moments.

Now, we observe that

$$\prod_{j=1}^{s-1} \begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} = \begin{bmatrix} \prod_{j=1}^{s-1} T_j & 0 & 0 & \cdots & 0 \\ \prod_{j=1}^{s-1} T_{j+1} & 0 & 0 & \cdots & 0 \\ \prod_{j=1}^{s-1} T_{j+2} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T_{s-1} & 0 & 0 & \cdots & 0 \\ I & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (\text{C3})$$

$$\begin{aligned} \sum_{j=1}^{s-2} \left[\prod_{k=j+1}^{s-1} \begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \left(\begin{bmatrix} R_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_j + \begin{bmatrix} c_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) \right] \\ = \sum_{j=1}^{s-2} \begin{bmatrix} \prod_{k=j+1}^{s-1} T_k (R_j \eta_j + c_j) \\ \prod_{k=j+1}^{s-1} T_{k+1} (R_j \eta_j + c_j) \\ \vdots \\ T_{s-1} (R_j \eta_j + c_j) \\ (R_j \eta_j + c_j) \\ 0 \end{bmatrix}. \quad (\text{C4}) \end{aligned}$$

Placing (C3) and (C4) properly in (C2) implies

$$\begin{aligned} Y_s^\ddagger &= Z_s \left\{ \left[\prod_{j=1}^{s-1} T_j \right] \alpha_1 + \sum_{j=1}^{s-2} \left[\prod_{k=j+1}^{s-1} T_k \right] (R_j \eta_j + c_j) \right. \\ &\quad \left. + c_{s-1} + R_{s-1} \eta_{s-1} \right\} + d_t + \varepsilon_s, \quad (\text{C5}) \end{aligned}$$

which coincides with the recursive solution of the measurement equation from the original model (A1). To conclude the proof, combine equations (C5) and (C1). \square

Appendix 4. Spread returns: plain strategy & our pairs trading strategy

The spread for the plain strategy, whenever it is activated, is defined as the ratio between the highest and lowest price assets of the pair. For instance, assume that in time t , the prices $P_{t,1}$ and $P_{t,2}$ of assets \mathcal{A}_1 and \mathcal{A}_2 are such that $P_{t,1} \geq P_{t,2}$. Then, the spread is $S_t = \frac{P_{t,1}}{P_{t,2}}$.

Using this definition, the daily return depends on how S_t deviates from its historical mean. If S_t is found to be higher than its historical mean by at least two historical standard deviations, the return is $\frac{P_{t+1,2}}{P_{t,2}} - \frac{P_{t+1,1}}{P_{t,1}}$; if the S_t takes on a value that is lower than its historical mean by at least two historical standard deviations, the return is, in turn, $\frac{P_{t+1,1}}{P_{t,1}} - \frac{P_{t+1,2}}{P_{t,2}}$.

On the other hand, Vidyamurthy (2004, pp. 80–82), derived a direct way of obtaining the return for the pairs trading strategy proposed in section 5 that is justified with some elements of the definition of a pair of assets (recall section 4.1). Assume that $\log(P_{t,1})$ and $\log(P_{t,2})$ are cointegrated—that is, \mathcal{A}_1 and \mathcal{A}_2 form a pair—with mean α and cointegration coefficient β . If the investor takes a long position on asset \mathcal{A}_1 and takes a short position on asset \mathcal{A}_2 (that is, the investor buys the spread) and if he or she maintains the position by at least $t + i$ ($i = 1, 2, \dots, k$, where k denotes divergence risk parameters that are previously set - cf. section 5.3), then the corresponding return from time t to $t + i$ is given by

$$\begin{aligned}
 & \log\left(\frac{P_{t+i,1}}{P_{t,1}}\right) - \beta \log\left(\frac{P_{t+i,2}}{P_{t,2}}\right) \\
 &= \log(P_{t+i,1}) - \log(P_{t,1}) - \beta(\log(P_{t+i,2}) - \log(P_{t,2})) \\
 &= \log(P_{t+i,1}) - \beta \log P_{t+i,2} - (\log(P_{t,1}) - \beta \log P_{t,2}) \quad (\text{D1}) \\
 &= \log(P_{t+i,1}) - \alpha - \beta \log P_{t+i,2} - (\log(P_{t,1}) - \alpha - \beta \log P_{t,2}) \\
 &= S_{t+i} - S_t;
 \end{aligned}$$

that is, once a long position is taken on the spread, the return for the latter is simply the difference between the spread value in times $t + i$ and t . If, in turn, the investor sells asset \mathcal{A}_2 and buys asset \mathcal{A}_1 (spread is being sold now), virtually the same derivation in equation (D1) would demonstrate that the spread return becomes $S_t - S_{t+i}$, which in turn corresponds to the negative of the return.

Copyright of Quantitative Finance is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.