Instituto Nacional de Matemática Pura e Aplicada

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

Carlos Eduardo Silva de Moura

Advisor: **Adrian Pizzinga**

Rio de Janeiro - January, 2014

To my family, girlfriend and friends, with love...

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Adrian Pizzinga and Jorge Zubelli for the continuous support of my master study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research and writing of this master thesis. I could not have imagined having a better advisor and mentor for my master study.

Besides my advisors, I would like to thank the rest of my teachers in IMPA: Prof. Alcides Neto, Prof. Paulo Cezar Carvalho, Prof. Luca Mertens, Prof. Rodrigo Novinski, Prof. Beatriz Mendes, Prof. Max de Souza, Prof. Claudia Sagastizábal, Prof. Sergei Vieira, Prof. Hugo de La Cruz and Prof. Gustavo Silva Araujo for their encouragement, insightful comments, and hard questions.

My sincere thanks also goes to Giuliano Lorenzoni, Guilherme Abry, Elsio Oliveira and all people that worked with me in this period.

I would like to thank all my friends for support and frendship.

I would like to thank my girlfriend Jessica for encouragement and support.

I would like to thank my siblings Marcos and Ana for friendship and happy moments.

Last but not the least, I would like to thank my family: my parents Margarida and George, for giving support to me at the first place and all the love during all my life.

Abstract

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

Key words: Kalman filter, mean-reverting conditional probabilities, pairs trading, spread, state space models, statistical arbitrage.

Resumo

Dentre as muitas estratégias no mercado financeiro, uma das mais populares em estudos acadêmicos é a estratégia denominada pairs trading. Ela consiste em uma estratégia de arbitragem estatística, que procura identificar e explorar ineficiências, de dois ativos financeiros relacionados no longo prazo. Quando um desvio deste equilíbrio entre os preços é significativo, um lucro pode ser obtido mediante aplicação de tais estratégias. Neste trabalho, é proposta uma estratégia de pairs trading inteiramente baseada em modelos de espaço de estados adequados para a série temporal do spread formado entre dois ativos. Uma vez estimado o modelo de espaço de estado adequado para o spread, são calculadas as probabilidades condicionais de que o spread retorne à sua média de longo prazo. A estratégia é executada quando são observados altos valores destas probabilidades: o spread é comprado ou vendido. Três aplicações com dados reais do mercado brasileiro e americano são oferecidas e indicam que uma carteira muito básica, que consiste em um único spread (par entre ativos), teve resultados melhores do que alguns dos principais benchmarks de mercado.

Palavras-chaves: arbitragem estatística, filtro de Kalman, modelo em espaço de estado, pairs trading, probabilidades condicionais de reversão à média, spread.

brazil

Contents

Introduction

Pairs trading is a type of statistical arbitrage strategy that has been firstly implemented in the mid 1980's by Nunzio Tartaglia and his group at Morgan Stanley (cf. [39]). Nowadays, pairs trading is widely used by investment banks and hedge funds. In general terms, a pairs trading aims at identifying and exploiting market inefficiencies observed with two long-term related assets, the two assets are said to form a *pair* –, mostly by using statistical methods. When a significant deviation of the prices between the two assets is detected, a trading position is carried out: the higher priced asset is sold (this is the so-called *short position* by market practitioners) and lower priced asset is bought (that is: a *short position* is taken), with the hope that mispricing will correct to the long term equilibrium value (cf. [11] and [39]).

In this work, we consider two linear state space models appropriate for modeling *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 models proposed by [11]. Such model, which has a Gaussian linear space state 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. [5], [6], [19] and [12]), whose particular specifications are also dealt with in this work under appropriate linear state forms. We shall prove that this second class of models, even though lacking finance theoretical support, encompasses the former proposal by [11] as a particular case. Moving on, we develop a methodology for calculating conditional probabilities (given past and actual spread data) that the spread will return to its long-term mean by k-steps ahead (the frequency can be daily or intra-daily), whenever it deviates somehow from the long-term mean at a given time instant. For such, we propose an alternative augmented state-space form for a given model, formerly selected and estimated with spread data, and with this enlarged state space form we apply the Kalman filter k-steps ahead prediction (see, for instance, [20] and [9]) to obtain conditional mean vectors and covariance matrices of the k future spreads. The latter is all that is needed for calculating the conditional probabilities previously mentioned. The quantitative strategy we shall pursue here is activated according to the rule: if the spread is found to be considerably below (above) its long-term mean and the conditional probability that the spread will increases above (decreases below) its long-term mean by k-steps ahead is reasonably large, buy (sell) the spread.

The dissertation is organized as follows. Chapter 1 briefly reviews the literature on pairs trading, without claiming exhaustiveness. Chapter 2 discusses pair trading from the statistical arbitrage standpoint, enumerating some of its main practical features. Chapter 3 presents the two aforementioned linear state space models models, discusses their mathematical properties and embeds each of them into the state space modeling/Kalman filter framework. Chapter 4 formally discusses how the conditional probabilities that the spread will mean-revert are calculated, the corresponding computational issues and describes step-by-step how the quantitative strategy shall be implemented. Chapter 5 offers three applications to real data from the US and Brazilian markets and compares the performances of the proposed strategy with the main benchmarks and with a former pairs trading strategy already tackled by market practitioners. Analysis regarding computational efforts for estimation and goodness-of-fit is included. Chapter 6 offers a discussion about the main results obtained in the former chapter and makes some comments regarding the use of the methodology in real scenarios. The appendices review the main Kalman filter techniques used in the work and provide the proofs of the technical results.

Chapter 1

Pairs Trading: a glimpse at the literature

This chapter briefly discusses earlier works on pairs trading strategies, focusing mainly on spread modeling. A common feature to each of such models – which, to some extent, shall be also pursued by the models of this paper – consists of recognizing the spread, associated with a pair of stocks (cf. the naive definition of "pair" given in Chapter), as some kind of mean-reverting stochastic process, whose parameters are estimated with financial market data. The last paper reviewed in this chapter, by its turn, is an empirical investigation that uses a simple standard deviation strategy to show that pairs trading can be profitable after costs.

Elliot et al. [11] developed a Gaussian linear state space models for the mean reversion behavior of the spread between paired stocks in 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 was based on a version of the expectation-maximization (EM) algorithm previously developed in [10]. The pairs trading strategy proposed is this: 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 so the trader could take a *short/long* position in the spread portfolio. Therefore, a profit is expected whenever a price correction occurs.

Triantafyllopoulos & Montana [37] have extended the modeling framework proposed by Elliot et al. in several ways. First, they introduced time-varying autoregressive (or mean-reverting) parameters, something that potentially allows the model to adapt itself to sudden changes in the data. Second, they developed and implemented a Bayesian approach for estimating the parameters, providing an on-line estimation scheme. Lastly, they advocated a procedure known as flexible least squares (FLS) to estimate the coefficient of co-integration coefficient recursively, unveiling possible time-varying co-integration relationship between the two asset prices.

Vidyamurthy [39] exploited the pairs trading universe in his book. He gives a good background about the theme and discusses several techniques to choose pairs trading, focusing on co-integration tests. Moreover, the author explains how pairs trading works and surveys some methods for dealing with the problem in real settings – for instance: common trends/cointegration models, arbitrage pricing theory (APT), distance measure and state space models/Kalman filter.

It is also worth citing the paper by Avellaneda & Lee [4]. These authors employed principal components analysis and sectors Exchange Trade Funds (ETF) for extracting risk factors. For each method, they modeled the corresponding residuals as mean-reverting processes.

Finally, Gatev et al. [16] studied pairs trading strategy in the U.S. equity market with daily data over the period from 1962 through 2002. In their study, stocks from companies that had at least one day out of business have been discarded. A pair formation for each stock was found by minimizing the squared deviations between the two normalized daily price series, where dividends were reinvested. The basic strategy consisted of opening a position in a pair when prices diverge by more than two historical standard deviations and unwind the position whenever the prices cross each other – and, should prices do not cross after the end of trading interval, gains and losses are calculated at the end of the last trading day. Latter, the performance of this strategy by [16] for the Brazilian stock market case was addressed by [29]. 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 contained between 1.5 and 2 standard deviations.

Chapter 2

Statistical Arbitrage Strategies

Quoting [22], *"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 statistical *mispricing* of one or more assets, based on their regular behavior. Despite the use of the term *"arbitrage"*, such class is not riskless. One of the simplest albeit very popular strategy that fits in with the definition of statistical arbitrage is pairs trading (cf. [11]). Other types of statistical arbitrage are discussed in [39] and [30].

Following [39], 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 1980. Pairs-trading is based on the arbitrage pricing theory (APT) (cf. [32]). Informally speaking, if two stocks have similar characteristics, then the prices of both assets must be more or less the same; that is, they maintain some degree of equilibrium. When 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 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 about 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. [11]). In order for the trade to be potentially profitable, and therefore be executable, 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. [28] and [25]). The main risks involved in a pairs trading are: (1) the *divergence risk*: the long-term equilibrium relation between the assets may change or even vanish; and (2) the *horizon risk*: the spread does not converge in a given horizon of time, hence forcing the traders to close your position before the convergence, due to worsened mispricing or margin call (cf. [13]). Additional details about pairs trading can be found in [30] and [39].

Chapter 3

Proposed Models

3.1 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*. Rigorously, two time series $Y_t \sim \mathrm{I}(1)$ and $X_t \sim \mathrm{I}(1)$ are said to be *cointegrated* iff $aY_t + bX_t \sim \mathrm{I}(0)$ for some $a \neq 0$ and $b \neq 0$ – the notation $\mathrm{I}(d)$ means "integrated of order d". This definition shall be enough for the aims of this work. For richer expositions on the theme and more general definitions, see [21], [19] and [12].

Consider now

$$S_t = \log(P_{t1}) - [\alpha + \beta \log(P_{t2})] , \tag{3.1.1}$$

where P_{t1} and P_{t2} 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 $\log(P_{t1})$ and $\log(P_{t2})$ are cointegrated, the *spread* S_t is stationary – that is: $S_t \sim \mathrm{I}(0)$. In such case, α is the mean of 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 the main ingredient for achieving success in a pairs trading. Such value is zero, in view of α in Eq.(3.1.1). Any expressive deviations from this value can be traded against.

3.2 Unobserved component models: the stochastic spread approach

Following [11], in this section we assume that the 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:

$$\begin{aligned} S_t &= x_t + D\epsilon_t \\ x_t - x_{t-1} &= a - bx_{t-1} + C\eta_t \end{aligned} \tag{3.2.1}$$

where $a \in \Re$, $0 < b < 2$, $C > 0$, $(\epsilon_t, \eta_t)' \sim \text{NID}\,(0, I_2)$. Adapting Eqs.(A.1.1) of Appendix A.1 in order to obtain an appropriate state space representation for the model in Eqs.(3.2.1), just define $Z_t = 1$, $d_t = 0$, $H_t = D^2$, $T_t = B \equiv 1 - b$, $c_t = A$, $R_t = 1$ and $Q_t = C^2$. Then the Kalman filter formulae in Eqs.(A.1.2) of Appendix A.1 turn to

$$\begin{aligned} &\upsilon_t = S_t - a_{t|t-1}, & &F_t = P_{t|t-1} + D^2, & & \\ &K_t = BP_{t|t-1}F_t^{-1}, & &L_t = B - K_t, & &t = 1, \ldots, n. \\ &a_{t+1|t} = A + Ba_{t|t-1} + K_t\upsilon_t, & &P_{t+1|t} = BP_{t|t-1}L_t' + C^2 & & \end{aligned} \tag{3.2.2}$$

Eqs.(3.2.2) can be started under 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.

This model proposed by Elliot et al. has three interesting features. The first is that it has at least some support from finance theory, since it can be viewed as a discrete time version of the Ohrstein-Uhlenbeck continuous time stochastic process – see [33]. The second is that it recognizes a mean reverting behavior for the spread. The last good property is a consequence of the next result, the proof of which is in Appendix A.2:

Proposition 1. *If S_t follows the unobserved component model by Elliot et al. given in Eqs.(3.2.1), then $S_t \sim \mathrm{ARMA}\,(1,1)$ with restrictions.*

This last proposition, besides encapsulating this proposal by Elliot et al. in a more general class of mean reverting statistical models (next section), suggests a procedure for selecting/discarding Eqs.(3.2.1) as a probabilistic description of some spread time series: if one obtains evidences from the data that the latter shall not be adequately fitted by any ARMA(1,1) model, then the proposal by Elliot et al. is necessarily misspecified for being considered in a pairs trading scheme.

3.3 ARMA models: generalizing the stochastic spread approach

Because of their mean reverting behavior, stationary autoregressive-moving average (ARMA) dynamics can be always considered as valid attempts for modeling the spread S_t. For instance, one could assume that $S_t \sim \mathrm{ARMA}\,(2,2)$, that is,

$$S_t = \phi_0 + \phi_1 S_{t-1} + \phi_2 S_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}, \tag{3.3.1}$$

where $\epsilon_t \sim \mathrm{NID}\,(0, \sigma^2)$ and $(\phi_1, \phi_2)'$ are such that the polynomial $p\,(z) = 1 - \phi_1 z - \phi_2 z^2$, $\forall z \in \Re$, has its two roots outside the unit circle. The latter assumption on the coefficients ϕ_1 and ϕ_2 is a sufficient condition to S_t be a stationary process – see [5], [6] and [19]. The same restrictions could be imposed to 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 $Y_t, Y_{t-1}, \ldots,$ by means of an AR (∞) representation for S_t – again, see [5], [6] and [19]. Fortunately, such question regarding invertibility is immaterial under the state space modeling/Kalman filter framework, since the latter always makes both likelihood function evaluation and forecasting attainable tasks independently of the invertibility question, as cleverly discussed by [19], Chaps. 4, 5 and 13.

One can use Eqs.(A.1.1) of Appendix A.1 to accommodate the model in Eq.(3.3.1), and any other stationary ARMA(p, q) model, under state space representations. Although there is no unique way of doing such conversion and the literature has been frequently offering and defending several state space forms for ARIMA models – to cite a few books: [20], [5], [6], [19] and [9] –, in this work the following alternative – for the ARMA$(2, 2)$ model given in Eq.(3.3.1) – shall be used in the sequel:

$$Z_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \; d_t = 0, \; H_t = 0,$$

$$T_t = \begin{bmatrix} \phi_1 & \phi_2 & 1 & \theta_1 & \theta_2 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \; c_t = \begin{bmatrix} \phi_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; R_t = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; Q_t = \sigma^2.$$

The Kalman filter formulae in Eqs.(A.1.1)(cf.Appendix A.1) is carried out with the matrices above and can be initialized under the initial conditions $a_1 = \left(\dfrac{\phi_0}{1 - \phi_1 - \phi_2}, \dfrac{\phi_0}{1 - \phi_1 - \phi_2}, 0, 0, 0 \right)'$ and $\text{vec}\,(P_1) = (I - T \otimes T)^{-1} \, \text{vec}\,(RQR)\,.$

Chapter 4

A new pairs trading strategy

In this chapter, we discuss the main elements of a quantitative pairs trading strategy entirely based on the estimation of state space models proposed in Chapter 3. Firstly, in Section 4.1, we give theoretical details on how conditional probabilities that the spread will return to its long-term mean, by k-steps ahead from a given time instant t, are defined. Moving on, in Section 4.2 we explore the practical matters for effectively calculating the aforementioned probabilities in a on-line fashion – as it will be shown, once an appropriate state space model is estimated by maximum likelihood (see Appendix A.1), the implementation of the usual Kalman filter prediction equations given in Eqs.(A.1.2) to an augmented version of the model shall be everything needed. Finally, in Section 4.3 the quantitative strategy is described step-by-step, where the content derived in Sections 4.1 and 4.2 are merged with the trading rule that involves buying or selling the spread accordingly.

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

The main target for success: to achieve, from a statistical/probabilistic standpoint, a minimum confidence that a future observed value of the spread will not take much long to cross back 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 task is accomplishable, one might buy (or sell) the spread on that time t, whenever chances are that he or she will be able to make a profit soon.

Formally, the strategy that we now begin to build is strongly based upon the ability of calculating conditional probabilities 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:

$$\begin{aligned}
p_{up}(t,k,c) &= P[(S_{t+1} > c) \cup (S_{t+2} > c) \cup \cdots \cup (S_{t+k} > c) | \mathcal{F}_t] \\
&= 1 - P[(S_{t+1} \leq c, S_{t+2} \leq c, \ldots, S_{t+k} \leq c | \mathcal{F}_t] \\
&= 1 - F_{S_{t+1}, S_{t+2}, \ldots, S_{t+k} | \mathcal{F}_t}(c, c, \ldots, c), \\
p_{down}(t,k,c) &= P[(S_{t+1} < c) \cup (S_{t+2} < c) \cup \cdots \cup (S_{t+k} < c) | \mathcal{F}_t] \\
&= P[(-S_{t+1} > -c) \cup (-S_{t+2} > -c) \cup \cdots \cup (-S_{t+k} > -c) | \mathcal{F}_t] \\
&= 1 - P[-S_{t+1} \leq -c, -S_{t+2} \leq -c, \cdots, -S_{t+k} \leq -c | \mathcal{F}_t] \\
&= 1 - F_{-S_{t+1}, -S_{t+2}, \cdots, -S_{t+k} | \mathcal{F}_t}(-c, -c, \cdots, -c),
\end{aligned} \tag{4.1.1}$$

where $\mathcal{F}_t$ is the σ-field generated by past and actual spread data; that is $\mathcal{F}_t \equiv \sigma(S_1, \ldots, S_{t-1}, S_t)$. If the assumption that a specific Gaussian linear state space model is appropriate for the spread (something that needs checking in practical implementations), the conditional distribution functions described in Eqs.(4.1.1) correspond to

$$S_{t,k} \equiv \begin{bmatrix} S_{t+1} \\ S_{t+2} \\ \vdots \\ S_{t+k} \end{bmatrix} |\mathcal{F}_t \sim \mathrm{N}\left(\mu_{t,k}, \Sigma_{t,k}\right), \tag{4.1.2}$$

where $\mu_{t,k} \equiv \mathrm{E}\left(S_{t,k}|\mathcal{F}_t\right)$ and $\Sigma_{t,k} \equiv \mathrm{Var}\left(S_{t,k}|\mathcal{F}_t\right)$. Sticking to the notation established in Appendix A.1 for key quantities related to the Kalman filter and also defining $P_{t+i,t+j|t} \equiv \mathrm{Cov}(\alpha_{t+i}, \alpha_{t+j}|\mathcal{F}_t)$, for $i, j = 1, 2, \ldots, k$ 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

$$\begin{aligned} \mathrm{E}(S_{t+i}|\mathcal{F}_t) &= \mathrm{E}(Z_{t+i}\alpha_{t+i} + d_{t+i} + \epsilon_{t+i}|\mathcal{F}_t) \\ &= Z_{t+i}\mathrm{E}(\alpha_{t+i}|\mathcal{F}_t) + d_{t+i} + \mathrm{E}(\epsilon_{t+i}|\mathcal{F}_t) \\ &= Z_{t+i}\mathrm{E}(\alpha_{t+i}|\mathcal{F}_t) + d_{t+i} + \mathrm{E}(\epsilon_{t+i}) \\ &= Z_{t+i}a_{t+i|t} + d_{t+i}. \end{aligned} \tag{4.1.3}$$

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

$$\begin{aligned} \mathrm{Var}(S_{t+i}|\mathcal{F}_t) &= Z_{t+i}\mathrm{Var}(\alpha_{t+i}|\mathcal{F}_t)Z'_{t+i} + \mathrm{Var}(\epsilon_{t+i}|\mathcal{F}_t) \\ &\quad + Z_{t+i}\mathrm{Cov}(\alpha_{t+i}, \epsilon_{t+i}|\mathcal{F}_t) + \mathrm{Cov}(\epsilon_{t+i}, \alpha_{t+i}|\mathcal{F}_t)Z'_{t+i} \\ &= Z_{t+i}P_{t+i|t}Z'_{t+i} + H_{t+i}, \end{aligned} \tag{4.1.4}$$

$$\tag{4.1.5}$$

$$
\begin{aligned}
\mathrm{Cov}(S_{t+i}, S_{t+j}|\mathcal{F}_t) &= \mathrm{Cov}(Z_{t+i}\alpha_{t+i} + d_{t+i} + \epsilon_{t+i}, Z_{t+j}\alpha_{t+j} + d_{t+j} + \epsilon_{t+j}|\mathcal{F}_t) \\
&= Z_{t+i}\mathrm{Cov}(\alpha_{t+i}, \alpha_{t+j}|\mathcal{F}_t)Z'_{t+j} + Z_{t+i}\mathrm{Cov}(\epsilon_{t+i}, \epsilon_{t+j}|\mathcal{F}_t) \qquad (4.1.6) \\
&+ \mathrm{Cov}(\epsilon_{t+i}, \alpha_{t+j}|\mathcal{F}_t)Z'_{t+i} + \mathrm{Cov}(\epsilon_{t+i}, \epsilon_{t+j}|\mathcal{F}_t) \\
&= Z_{t+i}P_{t+i,t+j|t}Z'_{t+j}.
\end{aligned}
$$

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

For each t, the first- and second-order conditional moments displayed in Eqs.(4.1.3) and (4.1.4) are trivially obtained from the Kalman filter in Eqs.(A.1.2) applied with the data subset $\{S_1, S_2, \ldots, S_t\}$ enlarged with k missing values after the last spread S_t; that is, one has to consider $\{S_1, S_2, \ldots, S_t, .\mathrm{NaN}, .\mathrm{NaN}, \ldots, .\mathrm{NaN}\}$, where the acronym ".NaN" means "Not Available Number" and appears k times exactly right after S_t. Following [9], Sec.4.9, this is the recognition that, under the state space modeling approach, forecasting is a particular case of missing values estimation. On the other hand, to be calculated, Eqs.(4.1.6) shall depend on additional implementation of Kalman recursions other than those revisited in Appendix A.1 – specifically, those derived in [9], Sec.4.5, with appropriate adaptations for the case of missing values. In order to avoid this computational effort, which is not always available as a ready-to-use option offered by commercial softwares and neither is considered in usual Kalman filter codes suggested in textbooks, in this work we propose an alternative. Our proposal will make use of already implemented formulae known to time series analysts.

The building block for routinely evaluating Eqs.(4.1.3), (4.1.4) and (4.1.6) for each time t is to use an augmented state space form equivalent to a given time series model formerly selected and estimated with the spread data. In this paper, the models considered are those previosly discussed in Sections 3.2 and 3.3. This task consists of figuring up k new blocks to the state vector in Eqs.(A.1.1), each one having the same dimension of the original state vector. Formally:

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

$$\begin{bmatrix} \alpha_{t+1} \\ \alpha_t \\ \vdots \\ \alpha_{t-(k-1)} \end{bmatrix} = \begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \alpha_t \\ \alpha_{t-1} \\ \vdots \\ \alpha_{t-k} \end{bmatrix} + \begin{bmatrix} c_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} R_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_t, \tag{4.2.1}$$

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 in 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 Eqs.(4.1.3), (4.1.4) and (4.1.6) becomes straightforward.

Denote the vectors of unknown parameters associated with Eqs.(4.2.1) and (A.1.1) by $\psi^{\tilde{\dagger}}$ and $\psi^{\dagger}$ respectively, and the corresponding likelihood functions by $\mathcal{L}^{\tilde{\dagger}}$ and $\mathcal{L}^{\dagger}$. Since the augmented model does not carry off any new parameters, it trivially follows that $\psi^{\tilde{\dagger}}=\psi^{\dagger}$. Even though it is not that easy to claim the same for the maximum likelihood estimators obtained under $\mathcal{L}^{\tilde{\dagger}}$ and $\mathcal{L}^{\dagger}$, the next proposition, whose proof is in Appendix A.3, asserts that it is indeed the case:

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

This result and its proof are admittedly inspired on Theorem 2 of [3], but we decided to include them here in detail, with proper adaptations from the former proof, in order to make this paper more self-contained.

The interpretation of Proposition 2 is that there are no changes in maximum likelihood estimation when considering the augmented model in Eqs.(4.2.1); hence, does not need to use the latter to estimate the parameters, something that would create additional and unnecessary computational endeavor. Instead, the estimation of unknown parameters can be accomplished using the original model in Eqs.(A.1.1) and the estimates obtained shall be used with the augmented model. From a practical standpoint, this result shall prove to be a key one in the applications of Chapter 5 for speeding up the calculation of the probabilities in Eqs.(4.1.1).

Finally, once $\mu_{t,k}$ and $\Sigma_{t,k}$ in Eq.(4.1.2) are calculated, the conditional probabilities in Eqs.(4.1.1) shall be evaluated through standard numerical multiple integration algorithms, which have been adapted for multivariate normal distributions framework – see for instance [8], [17], [18] and [7].

4.3 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 $\mathcal{A}_2$ – and that the numerical devices discussed in Section 4.2 have been implemented, we are now able to propose our trading rule. Summarily, this can be split in two mutually exclusive situations:

- If the observed value of S_t is found minimally below (let us say: for δ units) a long term value c, the very same used along Eqs.(4.1.1) and previously fixed in a particular value (for instance: $c = 0$, should one choose the spread mean), *and* p_{up} in Eq.(4.1.1) is found above some "large" value p_{up}^*, buy the spread.

- If the observed value of S_t is found minimally above c (without loosing generality, consider the same amount δ) *and* p_{down} in Eq.(4.1.1) is found above some "large" value p_{down}^*, sell the spread.

The items above deserve some qualification. Firstly, the meaning of the expression "buy the spread" is: the lower priced asset (in this case, $\mathcal{A}_1$ – see Eq.(3.1.1) is bought and the other asset is sold. The expression "sell the spread" could be analogously explained. Secondly, it is worth noticing that either the first situation (long position) or the second (short position) shall occur when the spread deviates more than or less than δ, the latter being a threshold that guarantees a minimum profitable trade after costs. Thirdly, since their values are priory set, p_{up}^* and p_{down}^* necessarily reflect risk aversion and one does have the option of choosing different values for each one. Fourthly, the position (either long or short) shall be disabled whenever the spread hits the long-term value c, or when it does not return in k-steps ahead – recall Eqs.(4.1.1). Finally, even though the two situations are mutually exclusive, these are certainly not exhaustive: indeed, if the conditions required for each of them are not met, the capital remains invested in some fixed income market until one of the "triggers" is activated. The choices for the parameters δ, p_{up}^*, p_{down}^*, c and k considered in this paper will be given in the applications of Chapter 5.

When the strategy just described is adopted, the main risk one might be exposed to is that related to specific fundamental changes: the prices of $\mathcal{A}_1$ and $\mathcal{A}_2$ may diverge, which means that the spread, not stationary anymore, is not hitting its former long term value c. Actually, the parameter k has precisely the function of mitigating such divergence risk. Another aspect is that the target return must always be higher than the return earned in the fixed income market because it is the opportunity cost inherent to this strategy. The parameter δ is present here to try to controls such point.

Chapter 5

Applications

This section presents the results of applying models from Chapter 3 and the pairs trading strategy derived in Chapter 4 with real data from the US and Brazilian markets. In Section 5.1 we describe the data used in the estimations and justify our choice of the stocks as candidates to form pairs. In Section 5.2 we present the results regarding co-integration tests, model estimation and goodness-of-fit, and the strategy performances.

5.1 The data and some computational details

All the financial time series used in the implementations have been obtained from Bloomberg Professional service. Four of them, considered in two of the three 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. The other exercise in the US equity market consist of modeling daily stock prices of two ETFs:

Market Vectors Gold Miners ETF (traded in the NYSE Arca with the symbol GDX) and SPDR Gold Shares (traded in the NYSE Arca with the symbol GLD). Market Vectors Gold Miners seeks to replicate as closely as possible, before fees and expenses, the price and yield performance of the NYSE Arca Gold Miners Index. SPDR Gold Shares seeks to replicate the performance of the price of gold bullion. ETF (Exchange Traded Fund) is a security that tracks an index, a commodity or a basket of assets like an index fund, but is traded as a stock on an exchange. For these four stocks, the period considered goes from September 22nd, 2011 till September 20th, 2012. Two other series, corresponding to the third exercise, are daily stock prices of Vale (traded in the stock exchange BMF&BOVESPA in Sao Paulo with the symbol VALE5) and Bradespar (traded in the stock exchange BMF&BOVESPA in Sao 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 of nickel. Bradespar is an investment company seeeking 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 in senior management, with members on the Board of Directors and Advisory Committees. We have used available data for these two stocks from August 29th, 2011 till September 20th, 2012. In view of the definition of pair given and discussed in Chapter 3, the stocks described above have been chosen mainly because, in view of their details given above, XOM and LUV, GDX and GLD like VALE5 and BRAP4 are supposedly long-term related.

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

- Libor - 1 year: This indicator stands for London Interbank 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 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, this play the very same role as Libor does. Despite of being a market rate, the CDI is closely tied to the interest rate, which is fixed by the Brazilian Central Bank on the course of monetary policy decisions.

- Bovespa Index (Ibovespa): This is the main indicator of the Brazilian stock market's average performance. The relevance of this index comes from several reasons; one of them is the integrity of its historical series, which have been regularly calculated without any methodological change since its inception in 1968.

5.2 Results

We begin by checking whether XOM-LUV, GDX-GLD and VALE5-BRAP4 show degrees of mutual equilibrium in the periods considered. This is assessed by testing co-integration hypotheses. See Section 3.1. For such, we used the two-step Engle Granger co-integration 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. [14], [12], Chap.6, and [27]. Once the co-integration hypothesis is not rejected, the spread to be considered in upcoming analyses shall be simply the OLS residuals – recall Eq.(3.1.1) in Section 3.1. The Engle-Granger test results were obtained from EViews 4.0. Looking at Table 5.1, we see that the data give enough evidence in favor of co-integration for both XOM-LUV and VALE5-BRA4. The pair GDX-GLD can also be considered co-integrated at a 10% level.

Table 5.1: Engle-Granger cointegration tests with the pairs.

Pairs	Dicker-Fuller Test*
XOM-LUV	-3.006**
VALE5-BRAP4	-4.059**
GDX-GLD	-1.952 ***

* Critical values considered have been taken from [27].

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

*** Pair was considered stationary at a 10% level.

Moving on, we now examine the information depicted in Table 5.2, 5.3, 5.4. This contains information concerning goodness-of-fit for three parsimonious ARMA (p, q) models and the model proposed by Elliot et al., along with some diagnostics performed with the standardized residuals $\upsilon_t^S = \frac{\upsilon_t}{\sqrt{F_t}}$, where υ_t and $\sqrt{F_t}$ are obtained from Eqs.(A.1.2). The implementations have been carried out using MATLAB 7.6.0. The unknown parameters were estimated by maximum likelihood, in which we adopted the exact log-likelihood function displayed in Eq.(A.1.3). See Appendix A.1. First, we see that, for each of models estimated with spreads from both US and Brazilian markets, the data are reproduced by each of the models almost under similar capabilities according to Pseudo R^2 and MSE measures. However, AIC and BIC criteria do reveal that the AR (1) model, which is the simplest option, shows the best complexity/fit relation.

Before addressing the diagnostics, it is worth bearing in mind that, if a given linear Gaussian state space model is adequate for the data at hand, the standardized residuals must behave like 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 models and spreads from both markets. 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 models estimated with VALE5-BRAP4 and GDX-GLD spread seem to come from a probability distribution similar to the standard normal distribution as regards 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 Eqs.(4.1.1) in trading decisions – indeed: p_{up} and p_{down} are not “tail” probabilities.

Table 5.2: Results from the estimations with the pair XOM-LUV (p-values in parentheses).

	XOM-LUV			
Attribute	AR(1)	AR(2)	ARMA(1,1)	Elliot
Log-likelihood	-989.044	-989.044	-989.044	-989.081
Pseudo R^2	0.896	0.896	0.896	0.902
MSE x 10^{-4}	2.299	2.298	2.299	2.161
AIC	7.865	7.873	7.873	7.882
BIC	7.893	7.915	7.915	7.938
LR Kupiec test (superior)*	0.077	0.077	0.077	0.434
	(0.781)	(0.781)	(0.781)	(0.510)
LR Kupiec test (inferior)*	0.434	0.434	0.434	0.434
	(0.510)	(0.510)	(0.510)	(0.510)
Ljung-Box test 1(20 lags) **	13.718	13.727	13.726	13.684
	(0.845)	(0.844)	(0.844)	(0.846)
Ljung-Box test 2 (20 lags) ***	29.557	26.679	29.669	29.582
	(0.148)	(0.145)	(0.145)	(0.152)
Jarque-Bera test	0.709	0.706	0.706	0.703
	(0.685)	(0.687)	(0.689)	(0.688)
Mean****	0.069	0.068	0.068	0.085
Variance****	0.999	0.999	0.999	0.997

* These are likelihood ratio unconditional coverage tests proposed by [24]. The first and second tests check standard residual violations of 95% and 5% standard normal distribution quantiles (that is: 1,65 and -1,65) respectively.

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

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

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

Table 5.3: Results from the estimations with the pair GDX-GLD (p-values in parentheses).

	GDX-GLD		
Attribute	AR(1)	ARMA(1,1)	Elliot
Log-likelihood	-885.036	-885.124	-885.126
Pseudo R^2	0.935	0.935	0.935
MSE x 10^{-4}	3.325	3.323	3.337
AIC	7.040	7.048	7.056
BIC	7.068	7.091	7.112
LR Kupiec test (superior)*	0.434	0.434	0.434
	(0.510)	(0.510)	(0.510)
LR Kupiec test (inferior)*	0.296	0.987	0.296
	(0.587)	(0.320)	(0.587)
Ljung-Box test 1(20 lags) **	28.829	29.411	28.853
	(0.091)	(0.080)	(0.091)
Ljung-Box test 2 (20 lags) ***	12.664	12.606	12.707
	(0.891)	(0.894)	(0.890)
Jarque-Bera test	299.843	309.519	299.517
	(0.000)	(0.000)	(0.000)
Mean****	-0.023	-0.024	-0.019
Variance****	1.002	1.002	1.002

* These are likelihood ratio unconditional coverage tests proposed by [24]. The first and second tests check standard residual violations of 95% and 5% standard normal distribution quantiles (that is: 1,65 and -1,65) respectively.

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

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

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

Table 5.4: Results from the estimations with the pair VALE5-BRAP4 (p-values in parentheses).

	VALE5-BRAP4			
Attribute	AR(1)	AR(2)	ARMA(1,1)	Elliot
Log-likelihood	-1053.502	-1060.960	-1068.300	-1068.310
Pseudo R^2	0.767	0.780	0.788	0.789
MSE x 10^{-4}	0.905	0.857	0.8130	0.8110
AIC	8.377	8.444	8.502	8.510
BIC	8.405	8.486	8.544	8.566
LR Kupiec test (superior)*	0.987	0.987	0.015	0.015
	(0.320)	(0.320)	(0.903)	(0.903)
LR Kupiec test (inferior)*	2.952	0.434	0.434	0.434
	(0.086)	(0.510)	(0.510)	(0.510)
Ljung-Box test 1(20 lags) **	29.706	23.628	18.040	18.035
	(0.075)	(0.259)	(0.585)	(0.585)
Ljung-Box test 2 (20 lags) ***	13.891	13.694	16.381	16.473
	(0.836)	(0.832)	(0.693)	(0.687)
Jarque-Bera test	22.602	24.308	24.880	24.914
	(0.002)	(0.001)	(0.001)	(0.001)
Mean****	-0.019	-0.029	-0.045	-0.049
Variance****	1.004	1.003	1.002	1.002

* These are likelihood ratio unconditional coverage tests proposed by [24]. The first and second tests check standard residual violations of 95% and 5% standard normal distribution quantiles (that is: 1,65 and -1,65) respectively.

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

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

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

Now, we start the discussion about the pairs trading strategy performances. The parameter c is set to zero, which is the long-term mean of the spreads, since these are precisely the OLS residual time series from co-integration regressions. Regarding the parameter value δ: since operating costs, due to slippage (this is the difference between the trade expected price and the trade actual price) and transaction, are being ignored here, δ is set to 0.5% to overcome this flaw. 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 -δ (+δ). Finally, regarding 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 was defined arbitrarily as 25 days, 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 after 25 days at current market prices – the latter being an event with conditional probability of 20% at the most (whenever model assumptions are satisfied).

Table 5.5 and Figures 5.1, 5.2, 5.3 and 5.4 display the results corresponding to the pair XOM-LUV for the four linear state space models already under investigation. These also show results from traditional benchmarks of the USA financial market already detailed in Section 5.1, and the performance of something 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 [16], consists of 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 â in case of prices do not converge after the end of the trading interval, gains and losses are calculated at the end of the last trading day. Analyzing Table 5.5, Sharpe ratios, calculated here for being used as the main criterion for choosing the best strategy (since these measure return performances adjusted to market risk, cf. [34] and [35]), indicate that the best trading options in the period considered have been our pairs trading strategy implemented with AR (2) and ARMA (1, 1) models, both presenting the same cumulative return and historical volatility. Cumulative and average returns corresponding to these two models are larger than the others, except for the plain strategy. However, due to its quite larger volatility, the latter has a worse Sharpe ratio. Additionally, low correlations with the stock index(S&P) are observed – the latter have been previously expected, since this type of strategy is supposedly market neutral. Figure 5.1, 5.2, 5.3 and 5.4 depict cumulative returns for the four state space model, together with cumulative returns from the market indices and the plain strategy, corroborating and illustrating the findings from Table 5.5.

Table 5.5: USA market data: performance measures from four different models for the spread and from three benchmarks

	XOM-LUV				Benchmarks		
Measures	AR(1)	AR(2)	ARMA(1,1)	ELLIOT	LIBOR	S&P	Plain strategy
Average return	0.066%	0.077%	0.077%	0.066%	0.0004%	0.077%	0.080%
Volatility*	0.645%	0.595%	0.595%	0.645%	0.00004%	1.034%	0.751%
Cumulative return	15.648%	18.606%	18.606%	15.648%	0.095%	17.573%	19.086%
Sharpe ratio	1.601	2.066	2.066	1.602	-	1.122	1.678
Correlation**	0.160	0.150	0.150	0.160	0.021	1.000	0.006

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

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

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

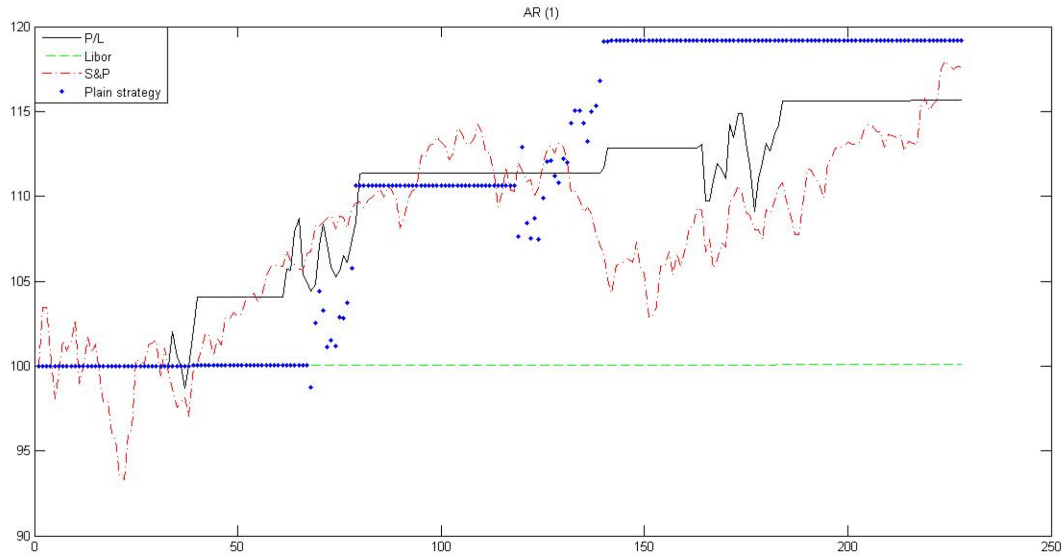

Figure 5.1: AR(1) Model. Comparison among cumulative returns: strategy P/L with the pair XOM-LUV, Libor, S&P and plain strategy.

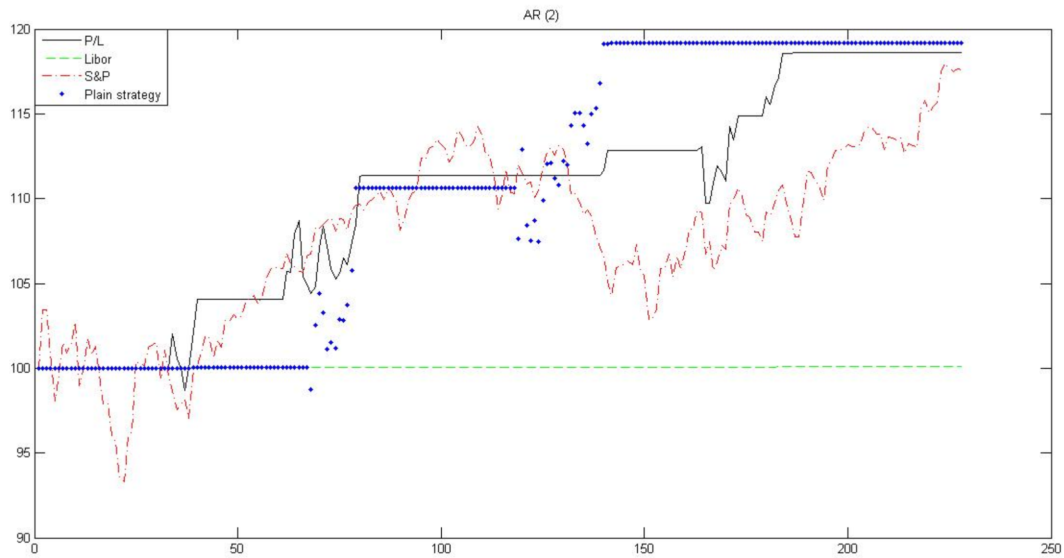

Figure 5.2: AR(2) Model. Comparison among cumulative returns: strategy P/L with the pair XOM-LUV, Libor, S&P and plain strategy.

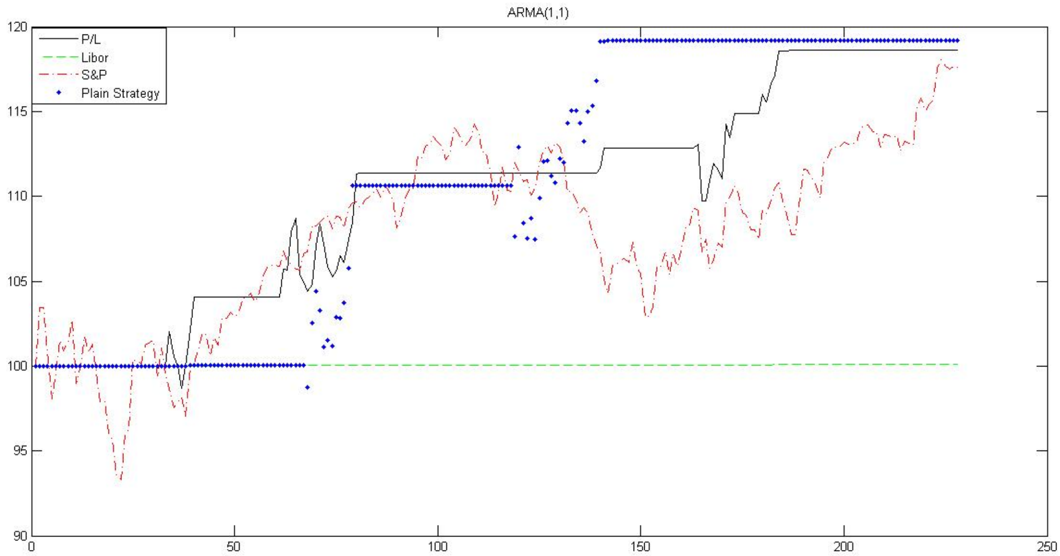

Figure 5.3: ARMA(1,1) Model. Comparison among cumulative returns: strategy P/L with the pair XOM-LUV, Libor, S&P and plain strategy.

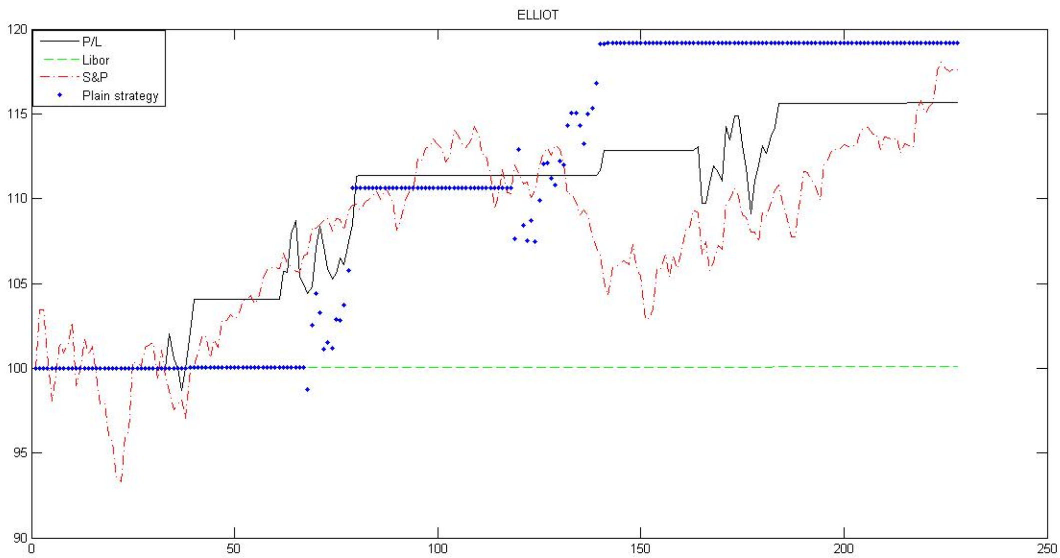

Figure 5.4: ELLIOT Model. Comparison among cumulative returns: strategy P/L with the pair XOM-LUV, Libor, S&P and plain strategy.

Now, analysing the results from the pair GDX-GLD in Table 5.6 and Figures 5.5, 5.6, 5.7, we see that the best performances, relying once again on Sharpe ratio comparisons, are those corresponding to model plain strategy. This has also shown low cumulative return and small volatility in the period. The low volatility is due to the small number of negotiations (only one) in the interval trading. Notice also that, in terms of cumulative returns, the S&P and the Elliot et al. model had the best performance and the S&P is showing a higher risk-return ratio in the period considered.

Table 5.6: USA market data: performance measures from four different models for the spread and from three benchmarks

	GDX-GLD			Benchmarks		
Measures	AR(1)	ARMA(1,1)	ELLIOT	LIBOR	S&P	Plain strategy
Average return	0.047%	0.047%	0.065%	0.0004%	0.077%	0.019%
Volatility*	1.388%	1.388%	1.364%	0.00004%	1.034%	0.242%
Cumulative return	8.772%	8.772%	13.485%	0.095%	17.573%	4.121%
Sharpe ratio	0.415	0.415	0.651	-	1.122	1.105
Correlation**	0.068	0.068	0.062	0.021	1.000	0.181

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

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

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

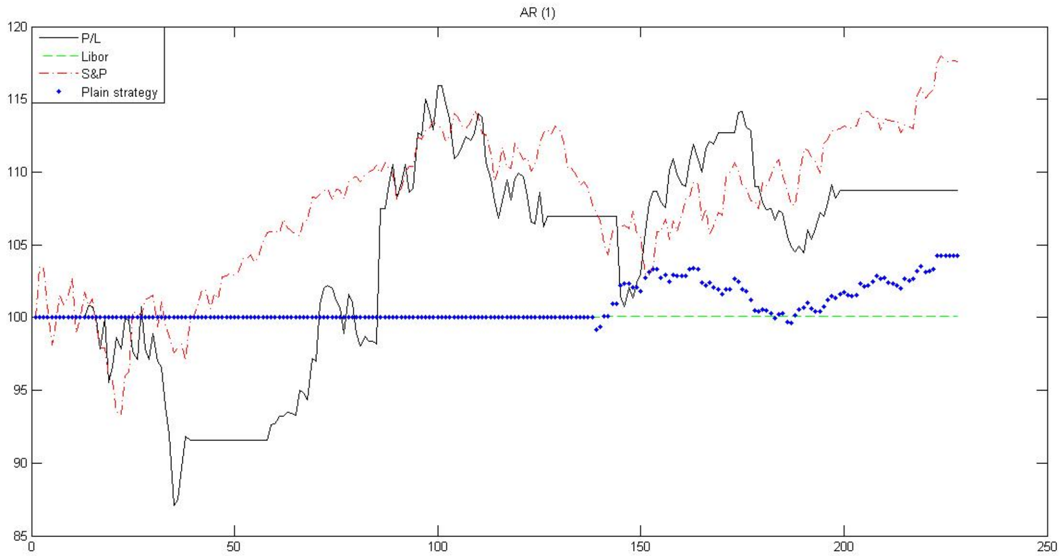

Figure 5.5: AR(1) Model. Comparison among cumulative returns: strategy P/L with the pair GDX-GLD, Libor, S&P and plain strategy.

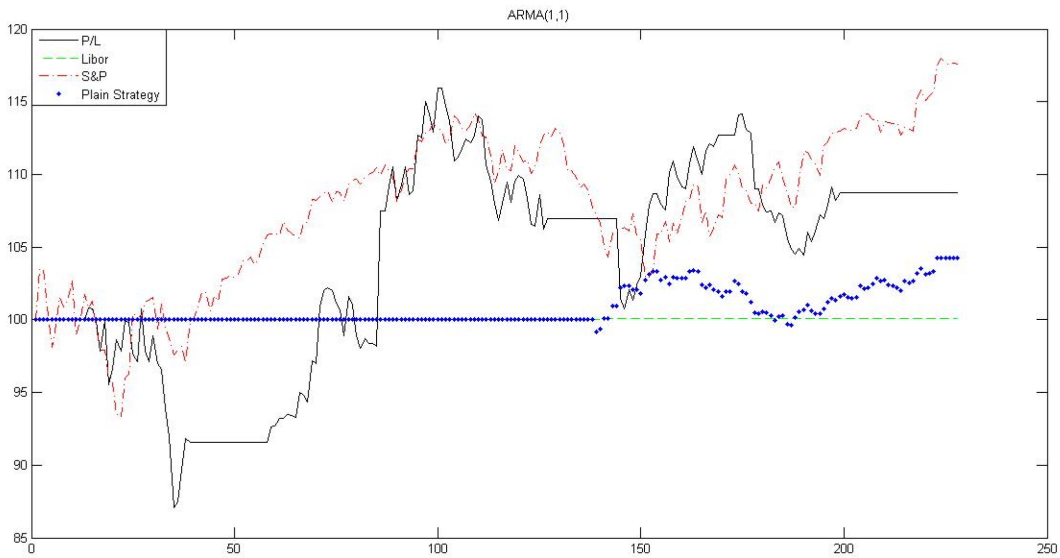

Figure 5.6: ARMA(1,1) Model. Comparison among cumulative returns: strategy P/L with the pair GDX-GLD, Libor, S&P and plain strategy.

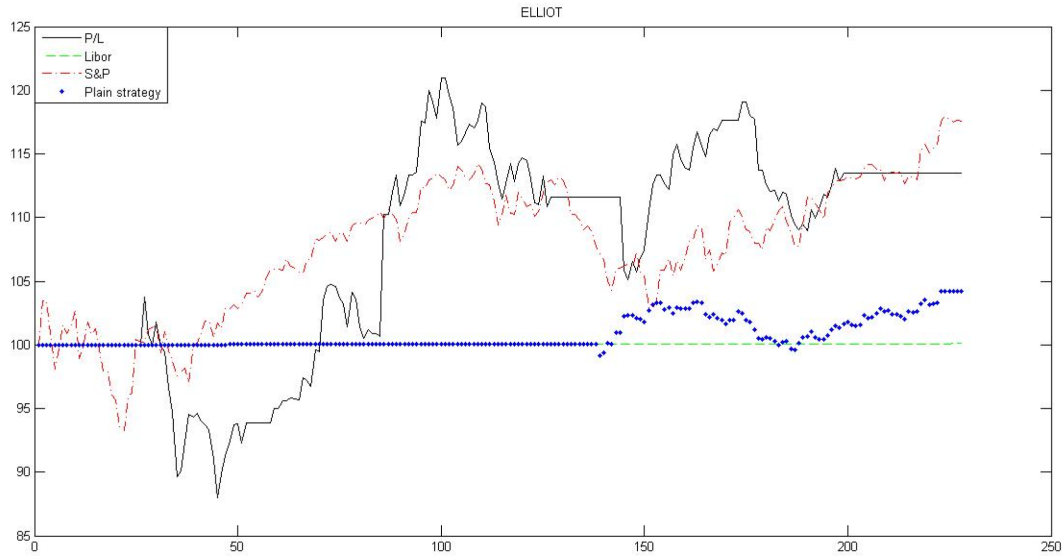

Figure 5.7: ELLIOT Model. Comparison among cumulative returns: strategy P/L with the pair GDX-GLD, Libor, S&P and plain strategy.

Likewise, both Table 5.7 and Figures 5.8, 5.9, 5.10 and 5.11 show the results for the pair VALE5-BRAP4. The best performances, relying once again on Sharpe ratio comparisons, are those corresponding to models AR (1) and AR (2) – these have also shown low correlations with the Ibovespa domestic stock index. In Figures 5.8, 5.9, 5.10 and 5.11, it is suggested that cumulative returns coming from our pairs trading strategy, implemented with the best two models, maintained an upward trend with relatively low volatility, probably corroborating the best Sharpe ratios. On the other hand, even though did Ibovespa present larger final return in the period considered amongst all the investment alternatives, one should notice its huge risky behavior (compare volatilities in Table 5.4), which have certainly contributed for some temporary losses. This can be seen on the downward and persistent reverses for this index in Figures 5.8, 5.9, 5.10 and 5.11. Notice also that, in terms of cumulative returns, Ibovespa has been the worst investment option for several moments in the period considered.

Table 5.7: Brazilian market data: performance measures from four different models for the spread and from three benchmarks.

	VALE5-BRAP4				Benchmarks		
Measures	AR(1)	AR(2)	ARMA(1,1)	ELLIOT	CDI	IBOVESPA	Plain strategy
Average return	0.084%	0.083%	0.063%	0.063%	0.036%	0.095%	0.035%
Volatility*	0.9505%	0.933%	0.881%	0.881%	0.005%	1.494%	0.261%
Cumulative return	19.878%	19.592%	9.872%	9.872%	8.550%	21.064%	8.202%
Sharpe ratio	0.789	0.784	0.0994	0.0994	-	0.5547	-0.0884
Correlation **	0.036	0.050	0.033	0.033	0.038	1.000	0.009
%CDI ***	232.49%	229.15%	115.46%	115.46%	100.00%	246.36%	95.93%

* This the standard deviation calculated with the daily returns.

** 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.

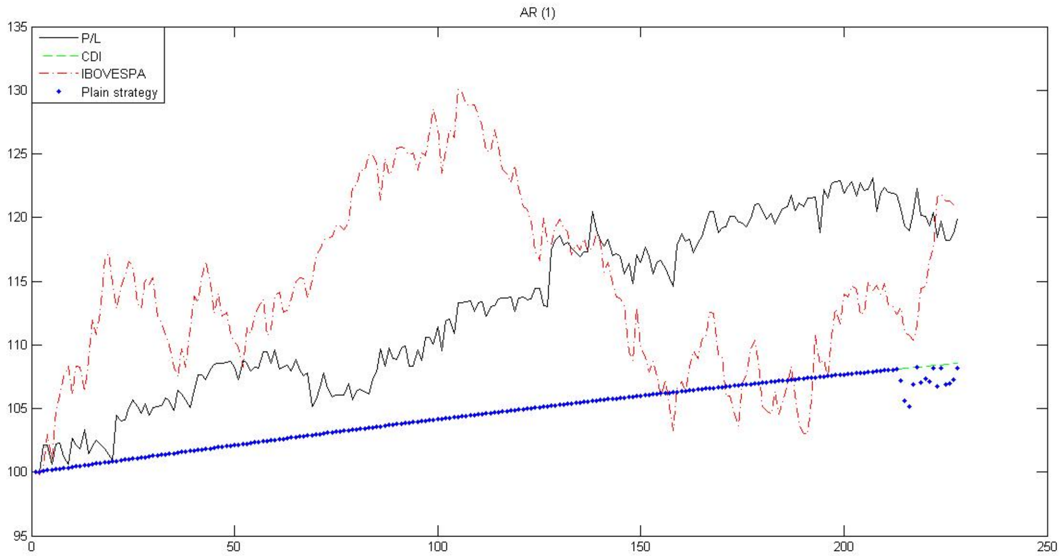

Figure 5.8: AR(1) Model. Comparison among cumulative returns: strategy P/L with the pair VALE5-BRAP4, CDI, IBOVESPA and plain strategy.

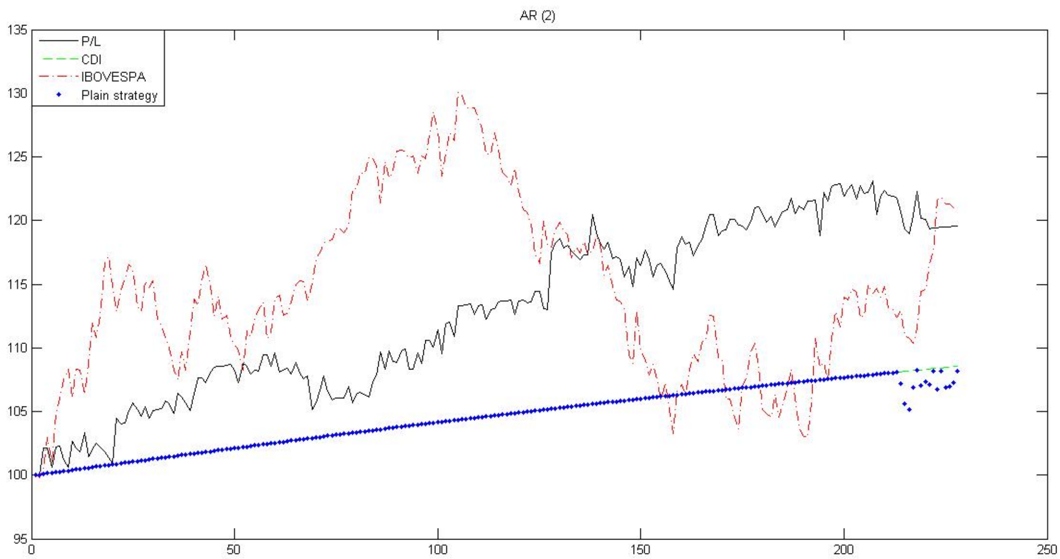

Figure 5.9: AR(2) Model. Comparison among cumulative returns: strategy P/L with the pair VALE5-BRAP4, CDI, IBOVESPA and plain strategy.

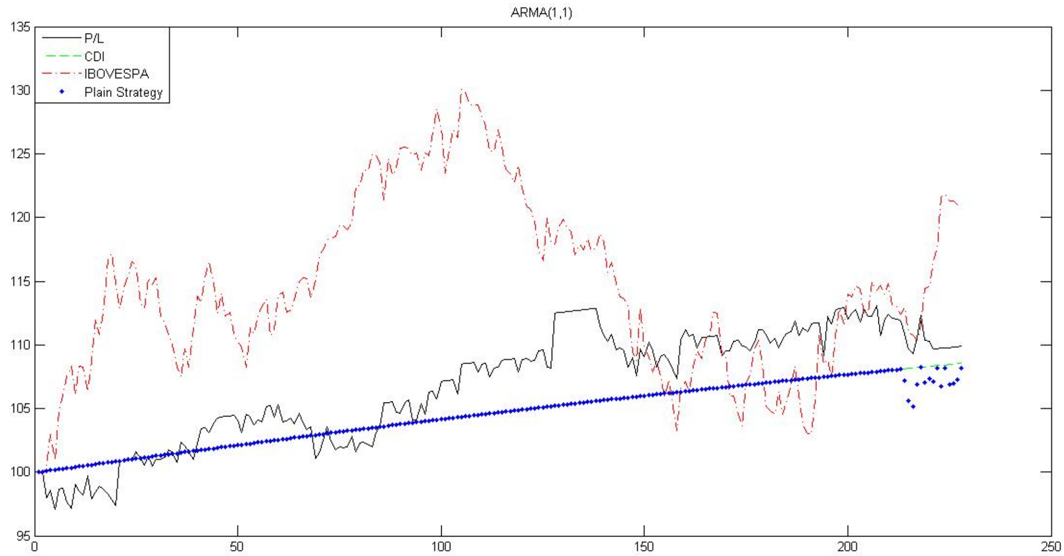

Figure 5.10: ARMA(1,1) Model. Comparison among cumulative returns: strategy P/L with the pair VALE5-BRAP4, CDI, IBOVESPA and plain strategy.

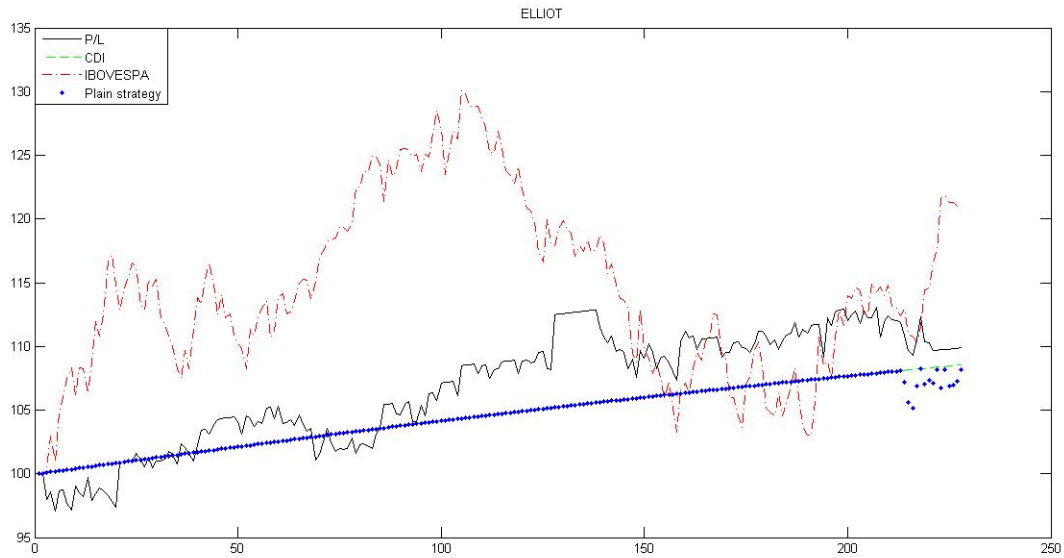

Figure 5.11: ELLIOT Model. Comparison among cumulative returns: strategy P/L with the pair VALE5-BRAP4, CDI, IBOVESPA and plain strategy.

Finally, Table 5.8 shows the computational gain, in terms of estimation time, due to the Proposition 2 of this work. Even though the information corresponding to model estimations with a portfolio with just a pair of assets in a daily basis, it is plausible to assume that the augmented model would also be excessively time consuming, should we have adopted and implemented the modeling and pairs trading strategy proposed in this work with intraday high frequency data â estimation times would have been even increased in case of a portfolio containing several pairs. For instance, the augmented model with $k = 25$ for the Elliot's model required almost three minutes to be estimated; the original model took less than three seconds.

Table 5.8: Computational times (seconds) for maximum likelihood estimation of the models with the pair VALE5-BRAP4.

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

Chapter 6

Conclusion

In this dissertation we 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 any basis for decisions regarding trading operations; instead, we look at the conditional probability that the value of the mispriced spread will mean-revert eventually by some pre-established horizon. The evidences gathered from the three applications detailed along Chapter 5, even though limited (and therefore far from being conclusive), suggest that this change of direction in usual pairs trading paradigms might work well in practice. At the end of this work, 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 that further investigations about how the parameters c, k and δ, which have been held constant in the examples of this work, must be set (notice that nothing prevents them from being estimated â or, should one prefer, optimized under usual back-testing schemes). One might 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. Since transactions fees vary according to the type of investment, analysis on these latter would help to identify how well our strategy would be suitable. More details on these costs in the Brazilian market can be found www.bmfbovespa.com.br, and for the New York Stock Exchange, one is referred to nyse.nyx.com.

Moving on, we now take a closer look at the question of distributional assumptions, as strong violations of normality can make the quantities p_{up} and p_{down} pretty unreliable as proxies for the true conditional probability of mean-reverting. An alternative for dealing with such inconvenient situation is to rely on Monte Carlo simulation of future trajectories of the spread S_t k steps ahead. For the case of the ARMA models, this would require modeling the error term with the aid of standardized residuals. A second alternative, which releases one from choosing/modeling error distributions (but is quite more demanding in computational terms), is to adopt some bootstrap procedure to estimate the mean-reverting conditional probabilities. Wall & Stoffer [40] and Rodriguez & Ruiz [31] are two papers of a certainly large list of references on bootstrapping state space models â these two papers seem to have methodologies that shall address the aims being discussed here.

Lastly, we discuss the use of our strategy in high-frequency data. The analysis of these data are complicated by irregular temporal spacing, daily patterns and price discreteness (cf. [1], Ch.7). Another major characteristic of high frequency data is the strong intra-day seasonal behavior of the volatility, as pointed by [15], Ch.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 work.

Appendix A

Appendix

A.1 Linear state space models and the Kalman filter

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

$$\begin{aligned} Y_t &= Z_t\alpha_t + d_t + \epsilon_t \quad , \quad \epsilon_t \sim NID(0, H_t) \\ \alpha_{t+1} &= T_t\alpha_t + c_t + R_t\eta_t \quad , \quad \eta_t \sim NID(0, Q_t) \\ \alpha_1 &\sim N(a_1, P_1). \end{aligned} \tag{A.1.1}$$

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 gives 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 past of Y_t. In the latter case, [20] , Sec. 3.7, refers to Eq. (A.1.1) as a *conditionally Gaussian* state space model.

For a given time series of size n and any time instants $t,j \in \{1, 2, \ldots n\}$, define $\mathcal{F}_j \equiv \sigma(Y_1, \ldots, Y_j)$, $a_{t|j} \equiv \mathrm{E}(\alpha_t|\mathcal{F}_j)$ and $P_{t|j} \equiv \mathrm{Var}(\alpha_t|\mathcal{F}_j)$. The *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 prediction are given below:

$$\begin{aligned} v_t &= Y_t - Z_t a_{t|t-1} - d_t, & F_t &= Z_t P_{t|t-1} Z_t' + H_t, & & \\ K_t &= T_t P_{t|t-1} Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t, & t &= 1, \ldots, n, \\ a_{t+1|t} &= T_t a_{t|t-1} + c_t + K_t v_t, & P_{t+1|t} &= T_t P_{t|t-1} L_t' + R_t Q_t R_t', & & \end{aligned} \tag{A.1.2}$$

The derivations of Eqs.(A.1.2) are found in [9]. There are several other references on this subject that deserve mentioning, such as [20], [21], [5], [6], [19] and[36].

In practice, system matrices include unknown parameters that must be estimated. By grouping all unknown parameters of the model described in (A.1.1) in a vector ψ, and denoting its corresponding parametric space by Θ, one can obtain an *exact* log-likelihood function, using some outputs from Eqs.(A.1.2):

$$\log L(\psi) = -\frac{np}{2}\log - \frac{1}{2}\sum_{t=1}^{n}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \quad , \quad \forall \psi \in \Theta. \tag{A.1.3}$$

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, Eq.(A.1.3) should be viewed as a *quasi* log-likelihood function and $\hat{\psi}$, in its turn, as a *quasi* maximum likelihood estimator.

A.2 Proof of Proposition 1

From the second equation of Eqs.(3.2.1), it follows that

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

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

$$x_t = \frac{1}{(1-BL)}a + \frac{1}{(1-BL)}\eta_t^* = \frac{a}{(1-B)} + \frac{a}{(1-BL)}\eta_t^*, \tag{A.2.1}$$

where L is the usual lag operator (recall: $0 < b < 2$). Now place Eq.(A.2.1) in the first equation of Eqs.(A.2.1) to get

$$S_t = \frac{a}{(1-BL)} + \frac{1}{(1-BL)}\eta_t^* + D\epsilon_t = \frac{a}{(1-BL)} + \frac{1}{(1-BL)}\eta_t^* + D\epsilon_t^* \tag{A.2.2}$$

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

Applying the operator $(1-BL)$ on both sides of Eq.(A.2.2),

$$S_t^* \equiv (1-BL)S_t = a + \eta_t^* + \omega_t^* - B\omega_{t-1}^* \tag{A.2.3}$$

From Eq.(A.2.3), it is straighforward to see that

$$\gamma(0) = C^2 + (1+B^2)D^2,\ \gamma(1) = C^2 + (1-B)D^2,\ \gamma(k) = 0, k \geq 2, \tag{A.2.4}$$

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

From Eqs.(A.2.4) and [5], p.89, Prop.3.2.1, it follows that $S_t^* \sim \mathrm{MA}(1)$. □

A.3 Proof of Proposition 2

We have to prove that the likelihood function from models $\dagger$ (original) and $\tilde{\dagger}$ (augmented) are equal; in other words, $\mathcal{L}^{\dagger} = \mathcal{L}^{\tilde{\dagger}}$ over all the parametric space. For such, it is sufficient to show that $v_t^{\dagger} = v_t^{\tilde{\dagger}}$ for each $t = 1, \ldots, n$. Notice that

$$v_t^{\dagger} = y_t^{\dagger} - Z_t a_{t|t-1}^{\dagger} - d_t^{\dagger} \quad \text{and} \quad v_t^{\tilde{\dagger}} = y_t^{\tilde{\dagger}} - Z_t a_{t|t-1}^{\tilde{\dagger}} - d_t^{\tilde{\dagger}}, \tag{A.3.1}$$

where $a_{t|t-1}^{\tilde{\dagger}} \equiv \mathrm{E}(\alpha_t|\mathcal{F}_{t-1}^{\tilde{\dagger}})$ and $d_t^{\dagger} = d_t^{\tilde{\dagger}}$. Besides, under the augmented model in Eqs.(4.2.1), the recursive solution for the measurement equation, for an arbitrary $s = 1, \ldots, t-1$, is

$$Y_s^{\tilde{\dagger}} = \begin{bmatrix} Z_s & 0 & \cdots & 0 \end{bmatrix} \left\{ \left[\prod_{j=1}^{s-1} \begin{bmatrix} T_j & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} \right] \begin{bmatrix} \tilde{\alpha_1} \end{bmatrix} + \right.$$

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

$$+ \begin{bmatrix} Z_s & 0 & \cdots & 0 \end{bmatrix} \left\{ \begin{bmatrix} c_{s-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} R_{s-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \eta_{s-1} \right\} + d_s + \varepsilon_s, \quad \text{(A.3.2)}$$

where $\tilde{\alpha_1}$ is an initial state vector with appropriate 1^{st} and 2^{nd} moments.

Now, observe that

$$\prod_{j=1}^{s-1} \begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix} = \begin{bmatrix} \prod_{j=1}^{s-1} T_j & 0 & 0 & \cdots & 0 \\ \prod_{j=1}^{s-1} T_{j+1} & 0 & 0 & \cdots & 0 \\ \prod_{j=1}^{s-1} T_{j+2} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{s-1} & 0 & 0 & \cdots & 0 \\ I & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \text{(A.3.3)}$$

$$\sum_{j=1}^{s-2}\left[\prod_{k=j+1}^{s-1}\begin{bmatrix} T_k & 0 & \cdots & 0 \\ I & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix}\right]\left(\begin{bmatrix} R_j \\ 0 \\ \vdots \\ 0 \end{bmatrix}\eta_j + \begin{bmatrix} c_j \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) = \sum_{j=1}^{s-2}\begin{bmatrix} \prod_{k=j+1}^{s-1} T_k\left(R_j\eta_j + c_j\right) \\ \prod_{k=j+1}^{s-1} T_{k+1}\left(R_j\eta_j + c_j\right) \\ \vdots \\ T_{s-1}\left(R_j\eta_j + c_j\right) \\ \left(R_j\eta_j + c_j\right) \\ 0 \end{bmatrix}. \tag{A.3.4}$$

Placing (A.3.3) and (A.3.4) properly in (A.3.2) implies

$$Y_s^{\tilde{\dagger}} = Z_s\left\{\left[\prod_{j=1}^{s-1} T_j\right]\alpha_1 + \sum_{j=1}^{s-2}\left[\prod_{k=j+1}^{s-1} T_k\right]\left(R_j\eta_j + c_j\right) + c_{s-1} + R_{s-1}\eta_{s-1}\right\} + d_t + \varepsilon_s\ , \tag{A.3.5}$$

which coincides with the recursive solution of the measurement equation from the original model (A.1.1). Finally, combine Eq. (A.3.5) and Eq. (A.3.1) to obtain the result. □

A.4 Matlab Code

```
 function [para, sumll] = AJUSTEKFAR1(r,k,p,delta,prob)


dados = importdata('SPREAD_GG_1ANO.csv');

spread = dados.data;

tic


Y=spread;

n=length(Y);

pos = n-252+1;
```

```
Y = Y(pos:n,1:3);
%plot(Y);
%legend('Spread Observado');

para0 = [0,0];

[x,fval,exitflag,output] = fminunc(@loglikAR1,para0,optimset('Display','iter',
'MaxFunEvals',5000,'HessUpdate','bfgs'),Y,r);

para = x

sumll = fval

exitflag

output

    phi = 1/(1+exp(-para(1)));
    sigeta = exp(para(2));

%FKAR1(para,Y,nc,r);
FKAR1_Trading(para,Y,r,k,n,p,delta,prob);
toc

 end




function sumll = loglikAR1(para0,Y,r)
```

```
% Chute Inicial
n=length(Y);

phi = 1/(1+exp(-para0(1)));
sigeta = exp(para0(2));

Z = [1 zeros(1,r-1)];
T = [phi; eye(r-1) zeros(r-1,1)];
c = zeros(r,1);
R = [1; zeros(r-1,1)];
Q = sigeta;

at = pinv((eye(r)-T))*c;
Pt =reshape(pinv((eye((r)^2)-kron(T,T)))*vec2mat(R*Q*R',1),r,r);

ll = zeros(n,1) ;

for t=1:n

    Ft = Z*Pt*Z';
    Kt = T*Pt*Z'*inv(Ft);
    inov = Y(t,1) - Z*at;

    at = T*at + Kt*inov + c;

    Lt= T - Kt*Z;

    Pt = T*Pt*Lt'+ R*Q*R';
```

```
    ll(t) = - 0.5*(log(Ft)+ (inov^2)/(Ft));

end

      sumll = -sum(ll);

end



    function [at,Pt,Ft] = FKAR1_Trading(para,Y,r,k,n,p,delta,prob)

% p : posição do início do backtesting
%n : tamanho da amostra
%r : ordem do modelo AR
%K : número de passos

phi = 1/(1+exp(-para(1)));
sigeta = exp(para(2));

flag = 0; % Ativa o tipo de estratégia

% Matrizes do Sistema Modelo Aumentado

Z = [1 zeros(1,k-1)];
T = [phi zeros(1,k-1); eye(r+k-2) zeros(k-1,1)];
c = zeros(r+k-1,1);
R = [1; zeros(k-1,1)];
Q = sigeta;
```

```
a1 = pinv((eye(r+k-1)-T))*c;

P1 =reshape(pinv((eye((r+k-1)^2)-kron(T,T)))*vec2mat(R*Q*R',1),r+k-1,r+k-1);


% Dimensão dos vetores de saida


%RECURSOES DE KALMAN

%Condição Inicial


% Matrizes de Outputs

%yup - guarda as probabilidades de subir

%ydown - guarda as probabilidades de cair

%cota - guarda as cotas da estratégia e dos principais benchmarking


yup = zeros(n-p,4);

ydown = zeros(n-p,4);

cota = zeros(n-(p-1),3);


cota(1,:)=100;  % Cota inicial


for s=p:n


    at = zeros(r+k-1,1,s+k+1);

    Pt = zeros(r+k-1,r+k-1,s+k+1);

    at(:,:,1)= a1;

    Pt(:,:,1)= P1 ;


for t=1:s+k
```

```
    if (t<=s)

    Ft = Z*Pt(:,:,t)*Z';
    Kt = T*Pt(:,:,t)*Z'*inv(Ft);
    inov= Y(t,1) - Z*at(:,:,t);
    Lt= T - Kt*Z;
    at(:,:,t+1) = T*at(:,:,t) + Kt*inov +c;
    Pt(:,:,t+1) = T*Pt(:,:,t)*Lt'+ R*Q*R';

    else

    at(:,:,t+1) = T*at(:,:,t)+c;
    Pt(:,:,t+1) = T*Pt(:,:,t)*T'+ R*Q*R';

    end

end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Estrategia %%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% de %%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Trading %%%%%%%%%%%%%%%%%%%%%%%%
% c (ponto de ativação da estratégia)
% delta - desvio em relação ao equilibrio de longo prazo (c).
X=zeros(1,k);
c=0;
```

```
    if flag == 0

        contdown =0;
        contup=0;

        %Probabilidade de Cair
        ydown(t-p-(k-1),1) = Y(s,1);
        ydown(t-p-(k-1),2) = 1 - mvncdf(X,-at(:,:,s+k)',Pt(:,:,s+k));
        ydown(t-p-(k-1),3) = c + delta;

        yup(t-p-(k-1),1) = Y(s,1) ;
        yup(t-p-(k-1),2) = 1 - mvncdf(X,at(:,:,s+k)',Pt(:,:,s+k));
        yup(t-p-(k-1),3) = c + delta;

        cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1+Y(s,2));
        cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
        cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));

        if (ydown(t-p-(k-1),1) >= c + delta && ydown(t-p-(k-1),2)> prob)
            flag =1;
            contdown = contdown +1;
            ydown(t-p-(k-1),4)= contdown;

%              if contdown ~= 1
%
            cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1+(Y(s-1,1)- Y(s,1)));
            cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
            cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));
```

```
%
%               end

        elseif  (yup(t-p-(k-1),1) <= c - delta  && yup(t-p-(k-1),2)> prob )

           % Probabilidade de Subir
              flag=2;
              contup = contup + 1;
              yup(t-p-(k-1),4)= contup;

%               if contup ~= 1
%
              cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1 + (Y(s,1)- Y(s-1,1)));
              cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
              cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));
%
%               end
        end

     elseif (flag == 1)

               contdown  = contdown +1;
               %ydown(t-p-(k-1),2) = 1 - mvncdf(X,-at(:,:,s+k)',Pt(:,:,s+k));
               ydown(t-p-(k-1),1) = Y(s,1);
               ydown(t-p-(k-1),3) = c;
               ydown(t-p-(k-1),4)= contdown;

               yup(t-p-(k-1),1) = Y(s,1) ;
               %yup(t-p-(k-1),2) = 1 - mvncdf(X,at(:,:,s+k)',Pt(:,:,s+k));
```

```
   yup(t-p-(k-1),3)  = c;

   yup(t-p-(k-1),4)= contup;


if contdown <= k


    if ydown(t-p-(k-1),1)<= c


        %Desfazer a Estrategia

        cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1+(Y(s-1,1)- Y(s,1)));

        cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));

        cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));

        flag=0;


    else


        % Mantenho a estratégia porque o spread não voltou

        % e o número de dias é menor que k.


        cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1+(Y(s-1,1)- Y(s,1)));

        cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));

        cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));


    end


else


    flag =0;

    cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1+(Y(s-1,1)- Y(s,1)));

    cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
```

```
            cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));

        end

    else

            contup = contup + 1;
            yup(t-p-(k-1),1) = Y(s,1) ;
            %yup(t-p-(k-1),2) = 1 - mvncdf(X,at(:,:,s+k)',Pt(:,:,s+k));
            yup(t-p-(k-1),3) = c;
            yup(t-p-(k-1),4)= contup;

            %ydown(t-p-(k-1),2) = 1 - mvncdf(X,-at(:,:,s+k)',Pt(:,:,s+k));
            ydown(t-p-(k-1),1) = Y(s,1);
            ydown(t-p-(k-1),3) = c;
            ydown(t-p-(k-1),4)= contdown;

        if contup <= k

                if yup(t-p-(k-1),1)>= c

                    %Desfazer a Estrategia

                    cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1 + (Y(s,1)- Y(s-1,1)));
                    cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
                    cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));
                    flag=0;

                else
```

```
                        % Mantenho a estratégia porque o spread não voltou
                        % e o número de dias é menor que k.

                        cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1 + (Y(s,1)- Y(s-1,1)));
                        cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
                        cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));

                    end

            else

                flag=0;
                cota(t-p-(k-2),1)= cota(t-p-(k-1),1)*(1 + (Y(s,1)- Y(s-1,1)));
                cota(t-p-(k-2),2)= cota(t-p-(k-1),2)*(1+Y(s,2));
                cota(t-p-(k-2),3)= cota(t-p-(k-1),3)*(1+Y(s,3));

            end

        end

end

plot(cota(:,1),'blue');
hold on
plot(cota(:,2),'green');
hold on
plot(cota(:,3),'red');
legend('P/L','CDI','Ibovespa');
```

```
cota

yup

ydown


end
```

Bibliography

[1] AIT-SAHALIA, Y. and HANSEN, L. (2010). *Handbook of Financial Econometrics*. 2nd edition. Springer-Verlag.

[2] ANDERSON, B. D. O. and MOORE, J. B. (1979). *Optimal Filtering*. Prentice Hall.

[3] ATHERINO, R., PIZZINGA. A and FERNANDES; C. (2010)."A row-wise stacking of the runoff triangle: state space alternatives for IBNR reserve prediction". *Astin Bulletin*, vol.40(2), p.917-946.

[4] AVELLANEDA, M. and LEE, J. H. (2010). "Statistical arbitrage in the US equities market". *Quantitative Finance*, vol.10(7), p.761-782.

[5] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. 2nd edition. Springer-Verlag.

[6] BROCKWELL, P. J. and DAVIS, R. A. (2003). *Introduction to Time Series and Forecasting*. 2nd edition. Springer Texts in Statistics.

[7] DREZNER, Z. (1994). "Computation of the Trivariate Normal Integral". *Mathematics of Computation*, vol.63, p.289-294.

[8] DREZNER, Z. and WESOLOWSKY, G. O. (1989). "On the Computation of the Bivariate Normal Integral". *Journal of Statistical Computation and Simulation*, vol.35, p.101-107.

[9] DURBIN, J. and KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Methods.* Oxford Statistical Science Series.

[10] ELLIOT, R.J. and KRISHNAMURTHY, V. (1999). "New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models". *IEEE Transactions of Automatic Control*, vol.44, p.938-951.

[11] ELLIOT, R. J., VAN DER HOEK, J. and MALCOLM, W. P. (2005). "Pairs Trading". *Quantitative Finance*, vol.5(3), p.271-276.

[12] ENDERS, W. (2004). *Applied Econometric Time Series.* 2nd edition. John Wiley & Sons.

[13] ENGELBERG, J., GAO, P. and JAGANNATHAN, R. (2009). "An anatomy of Pairs trading: the role of idiosyncratic news, common information and liquidity". *Third Singapore International Conference on Finance.*

[14] ENGLE, R. and GRANGER, C. (1987) "Co-Integration and Error Correction: Representation, Estimation, and Testing". *Econometrica*, vol.55, n$^{\text{o}}$2, p.251-276.

[15] FOUQUE,P. J.,PAPANICOLAOU, G. and SIRCAR, R., K. (2000). *Derivatives in Financial Markets with Stochastic Volatility.* Cambridge University Press.

[16] GATEV, E., GOETZMANN, W. and ROUWENHORST, K. (2006). "Pairs Trading: Performance of a Relative Value Arbitrage Rule". *The Review of Financial Studies*, vol.19, p.797-827.

[17] GENZ, A. (1992) "Numerical Computational of Multivariate Normal Probabilities". *Journal of Computation and Graphical Statistics*, vol.1, p.141-149.

[18] GENZ, A. (2004) "Numerical Computation of Rectangular Bivariate and Trivariate Normal and t Probabilities". *Statistics and Computing*, vol.14, No. 3, p.251-260.

[19] HAMILTON, J. D. (1994). *Time Series Analysis.* Princeton University Press.

[20] HARVEY, A. C. (1989). *Forecasting, Structural Time Series Models and The Kalman Filter.* Cambridge University Press.

[21] HARVEY, A. C. (1993). *Time Series Models.* 2nd edition. Harvester Wheatsheaf.

[22] KAUFMAN, P. J. (2005). *New Trading Systems and Methods.* Fourth Edition. John Wiley & Sons, Inc.

[23] KORN, R. and KORN, E. (2001). *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics (Graduate Studies in Mathematics).* AMS.

[24] KUPIEC, P. (1995). "Techniques for Verifying the Accuracy of Risk Management Models". *Journal of Derivatives*, vol.3, p.73-84.

[25] JACOBS, B. and LEVY, K. (2005). *Market Neutral Strategies.* John Wiley & Sons.

[26] JOHNSON, B. (2010). *Algorithmic Trading & DMA An introduction to direcet acess trading strategies.* 4Myeloma Press.

[27] MACKINNON, J. G. (2010). "Critical Values for Cointegration Tests". *Queens Economics Department Working Paper No. 1227.* Queens University.

[28] NICHOLAS, J. G. (2000). *Market-Neutral Investing: Long/Short Hedge Fund Strategies.* Bloomberg Press.

[29] PERLIN, M. S. (2007). "Evaluation of Pairs Trading Strategy at the Brazilian Financial Market". *Journal of Derivatives & Hedge Funds*, vol.15, p.122-136.

[30] POLE, A. (2007). *Statistical Arbitrage: Algorithmic Trading Insights and Techniques.* John Wiley & Sons, Inc.

[31] RODRIGUEZ, A. and RUIZ, E. (2009). “Bootstrap Prediction Intervals in State Space Models ”. *The Journal of Time Series Analysis*, vol.30, No. 2.

[32] ROSS, S. (1976). “The Arbitrage Theory of Capital Asset Pricing”. *Journal of Economic Theory*, vol.13, p.341-360.

[33] RAMPERTSHAMMER, S. (2007). “An Ornstein-Uhlenbech Framework for Pairs Trading”. *Department of Mathematics and Statistics the University of Melbourne*, Unpublished Note.

[34] SHARPE, F. W. (1966). “Mutual Fund Performance”. *The Journal of Business*, vol.39, p.119-138.

[35] SHARPE, F. W. (1994). “The sharpe ratio”. *The Journal of Portfolio Management*, vol.21, p.49-58.

[36] SHUMWAY, R. H. and STOFFER, D. S. (2006). *Time Series Analysis and Its Applications (With R Examples).* 2nd edition. Springer.

[37] TRIANTAFYLLOPOULOS, K. and MONTANA, G. (2009). “Dynamic modeling of mean-reverting spreads for statistical arbitrage”. *Computational Management Science*, vol.8, p.23-49.

[38] TSAY, R. S. (2010). *Analysis of Financial Time Series.* 3rd edition, John Wiley & Sons, Inc.

[39] VIDYAMURTHY, G. (2004). *Pairs Trading, Quantitative Methods and Analysis.* John Wiley & Sons, Inc.

[40] WALL, K. and STOFFER, S. (2002). “A State Space Approach to Bootstrapping Conditional Forecasts in ARMA Models ”. *The Journal of Time Series Analysis*, vol.23, No. 6.