



## Quantitative Finance

Publication details, including instructions for authors and subscription information:  
<http://www.tandfonline.com/loi/rquf20>

### Pairs trading

Robert J. Elliott<sup>a</sup>, John Van Der Hoek<sup>b</sup> & William P. Malcolm<sup>c</sup>

<sup>a</sup> Haskayne School of Business, University of Calgary, Calgary, Alberta, Canada T2N 1N4

<sup>b</sup> Department of Applied Mathematics, The University of Adelaide, Adelaide, South Australia 5005, Australia

<sup>c</sup> National ICT Australia, The Australian National University, Canberra, ACT, Australia 0200

<sup>d</sup> Department of Applied Mathematics, The University of Adelaide, Adelaide, South Australia 5005, Australia E-mail:

Published online: 18 Feb 2007.

To cite this article: Robert J. Elliott, John Van Der Hoek & William P. Malcolm (2005) Pairs trading, Quantitative Finance, 5:3, 271-276, DOI: [10.1080/14697680500149370](https://doi.org/10.1080/14697680500149370)

To link to this article: <http://dx.doi.org/10.1080/14697680500149370>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Pairs trading

ROBERT J. ELLIOTT<sup>†</sup>, JOHN VAN DER HOEK<sup>\*‡</sup>  
 and WILLIAM P. MALCOLM<sup>§</sup>

<sup>†</sup>Haskayne School of Business, University of Calgary, Calgary, Alberta, Canada T2N 1N4

<sup>‡</sup>Department of Applied Mathematics, The University of Adelaide, Adelaide, South Australia 5005, Australia

<sup>§</sup>National ICT Australia, The Australian National University, Canberra, ACT, Australia 0200

(Received 27 December 2004; in final form 11 April 2005)

‘Pairs Trading’ is an investment strategy used by many Hedge Funds. Consider two similar stocks which trade at some spread. If the spread widens short the high stock and buy the low stock. As the spread narrows again to some equilibrium value, a profit results. This paper provides an analytical framework for such an investment strategy. We propose a mean-reverting Gaussian Markov chain model for the spread which is observed in Gaussian noise. Predictions from the calibrated model are then compared with subsequent observations of the spread to determine appropriate investment decisions. The methodology has potential applications to generating wealth from any quantities in financial markets which are observed to be out of equilibrium.

**Keywords:** Pairs trading; Hedge funds; Spreads

## 1. Introduction

Pairs Trading is a trading or investment strategy used to exploit financial markets that are out of equilibrium. Litterman (2003) explains the philosophy of Goldman Sachs Asset Management as one of assuming that while markets may not be in equilibrium, over time they move to a rational equilibrium, and the trader has an interest to take maximum advantage from deviations from equilibrium. Pairs Trading is a trading strategy consisting of a long position in one security and a short position in another security in a predetermined ratio. If the two securities are stocks from the same financial sector (like two mining stocks), one may take this ratio to be unity. This ratio may be selected in such a way that the resulting portfolio is market neutral, a portfolio with zero beta to the market portfolio. This portfolio is often called a spread. We shall model this spread (or the return process for this spread) as a mean-reverting process which we calibrate from market observations. This model will allow us to make predictions for this spread. If observations are larger (smaller) than the predicted value (by some threshold value) we take a long (short) position in the portfolio and we unwind the position and make a profit when the spread reverts. A brief history and

discussion of pairs trading can be found in Gatev *et al.* (1999) and in two recent books by Vidyamurthy (2004) and Whistler (2004). Reverre (2001) discusses a classical study of pairs trading involving Royal Dutch and Shell stocks. Pairs trading is also regarded as a special form of Statistical Arbitrage and is sometimes discussed under this topic. The idea of pairs trading can be applied to any equilibrium relationship in financial markets, or to (market neutral) portfolios of securities some held short and others held long (see Nicholas (2000)).

## 2. The spread model

### 2.1. The state process

Consider a *state process*  $\{x_k | k = 0, 1, 2, \dots\}$  where  $x_k$  denotes the value of some (real) variable at time  $t_k = k\tau$  for  $k = 0, 1, 2, \dots$ . We assume that  $\{x_k\}$  is mean reverting:

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\varepsilon_{k+1}, \quad (1)$$

where  $\sigma \geq 0$ ,  $b > 0$ ,  $a \in \mathcal{R}$  (which we may assume is non-negative without any loss of generality), and  $\{\varepsilon_k\}$  is iid Gaussian  $\mathcal{N}(0, 1)$ . Clearly, we assume that  $\varepsilon_{k+1}$  is independent of  $x_0, x_1, \dots, x_k$ . The process mean reverts to  $\mu = a/b$  with ‘strength’  $b$ . Clearly,

$$x_k \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad (2)$$

\*Corresponding author. Email: Jvanderh@maths.adelaide.edu.au

where

$$\mu_k = \frac{a}{b} + \left[ \mu_0 - \frac{a}{b} \right] (1 - b\tau)^k, \quad (3)$$

and

$$\sigma_k^2 = \frac{\sigma^2 \tau}{1 - (1 - b\tau)^2} \left[ 1 - (1 - b\tau)^{2k} \right] + \sigma_0^2 (1 - b\tau)^{2k}. \quad (4)$$

It is easy to show that

$$\mu_k \rightarrow \frac{a}{b} \quad \text{as } k \rightarrow \infty, \quad (5)$$

and

$$\sigma_k^2 \rightarrow \frac{\sigma^2 \tau}{1 - (1 - b\tau)^2} \quad \text{as } k \rightarrow \infty, \quad (6)$$

provided we have chosen  $\tau > 0$  and small so that  $|1 - b\tau| < 1$ .

We can also write (1) as

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1}, \quad (7)$$

with  $A = a\tau \geq 0$ ,  $0 < B = 1 - b\tau < 1$  and  $C = \sigma\sqrt{\tau}$ . We could also regard  $x_k \cong X(k\tau)$  where  $\{X(t) | t \geq 0\}$  satisfies the stochastic differential equation

$$dX(t) = (a - bX(t))dt + \sigma dW(t), \quad (8)$$

where  $\{W(t) | t \geq 0\}$  is a standard Brownian motion (on some probability space).

## 2.2. The observation process

We assume that we have an *observation process*  $\{y_k\}$  of  $\{x_k\}$  in Gaussian noise:

$$y_k = x_k + D\omega_k, \quad (9)$$

where  $\{\omega_k\}$  are iid Gaussian  $\mathcal{N}(0, 1)$  and independent of the  $\{\varepsilon_k\}$  in (1) and  $D > 0$ . We may assume that  $0 \leq C < D$ , which should be the case for small values of  $\tau$ .

We set  $\mathcal{Y}_k = \sigma\{y_0, y_1, \dots, y_k\}$  which represents the information from observing  $y_0, y_1, \dots, y_k$ . We will wish to compute the conditional expectation (filter):

$$\hat{x}_k = \mathbf{E}[x_k | \mathcal{Y}_k], \quad (10)$$

which are ‘best’ estimates of the hidden state process through the observed process. In order to make the estimate (10), we will need to estimate  $(A, B, C, D)$  or rather  $(A, B, C^2, D^2)$  from the observed data. We shall present various results for this below.

## 2.3. The application

We shall regard  $\{y_k\}$  as a model for the observed spread of two securities at time  $t_k$ . We assume the observed spread is a noisy observation of some mean-reverting state process  $\{x_k\}$ . The  $\{y_k\}$  could also model the returns of the spread portfolio as is often done in practice.

If  $y_k > \hat{x}_{k|k-1} = \mathbf{E}[x_k | \mathcal{Y}_{k-1}]$  the spread is regarded as too large, and so the trader could take a long position in the spread portfolio and profit when a correction

occurs. An alternative would be to initiate a long trade only when  $y_k$  exceeds  $\hat{x}_{k|k-1}$  by some threshold value. A corresponding short trade could be entered when  $y_k < \hat{x}_{k|k-1}$ .

Various decisions have to be made by the trader. What is a suitable pair of securities for pair trading? If our estimates for  $B$  reveal  $0 < B < 1$ , then this is consistent with the mean-reverting model we have described. Comparing  $y_k$  and  $\hat{x}_{k|k-1}$  may or may not lead to a trade if thresholds must be met. How are thresholds set? See Vidyamurthy (2004) for some possibilities. When is the pairs trade unwound? There are various possibilities: the next trading time (see Reverre (2001) example) or when the spread corrects sufficiently. The price application and data bases used are often proprietary in industry applications. The machinery we present then provides some useful tools appropriate for pairs trading.

Another explicit strategy could make use of First-Passage Times results (see Finch (2004) and references cited therein) for the (standardized) Ornstein–Uhlenbeck process

$$dZ(t) = -Z(t)dt + \sqrt{2}dW(t). \quad (11)$$

Let

$$T_{0,c} = \inf\{t \geq 0, Z(t) = 0 | Z(0) = c\}, \quad (12)$$

which has a probability density function  $f_{0,c}$ . It is known explicitly that

$$f_{0,c}(t) = \sqrt{\frac{2}{\pi}} \frac{|c| e^{-t}}{(1 - e^{-2t})^{3/2}} \exp\left(-\frac{c^2 e^{-2t}}{2(1 - e^{-2t})}\right) \quad (13)$$

for  $t > 0$ . Now  $f_{0,c}$  has a maximum value at  $\hat{t}$  given by

$$\hat{t} = \frac{1}{2} \ln \left[ 1 + \frac{1}{2} \left( \sqrt{(c^2 - 3)^2 + 4c^2} + c^2 - 3 \right) \right]. \quad (14)$$

We can also write (8) in the form

$$dX(t) = -\rho(X(t) - \mu)dt + \sigma dW(t), \quad (15)$$

where  $\rho = b$  and  $\mu = a/b$ . When

$$X(0) = \mu + c \frac{\sigma}{\sqrt{2\rho}}, \quad (16)$$

the most likely time  $T$  at which  $X(T) = \mu$  is given by

$$T = \frac{1}{\rho} \hat{t}, \quad (17)$$

where  $\hat{t}$  is given by (14). Use the Ornstein–Uhlenbeck process as an approximation to (7) with  $a = A/\tau$ ,  $b = (1 - B)/\tau$  and  $\sigma = C/\sqrt{\tau}$  with the calibrated values  $A, B, C$ . Choose a value of  $c > 0$ . Enter a pair trade when  $y_k \geq \mu + c(\sigma/\sqrt{2\rho})$  and unwind the trade at time  $T$  later. A corresponding pair trade would be performed when  $y_k \leq \mu - c(\sigma/\sqrt{2\rho})$  and unwound at time  $T$  later.

Other methods based on up- and down-crossing results for AR(1) processes could also be considered. Corresponding results like those for the Ornstein–Uhlenbeck processes are not known.

### 3. Filtering and estimation results

We assume some underlying probability space  $(\Omega, \mathcal{F}, P)$  whose details need not concern the trader, except that  $P$  represents the real world probability.

#### 3.1. Kalman filtering

We have a *state equation*

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1}, \quad (18)$$

and the *observation equation*

$$y_k = x_k + D\omega_k, \quad (19)$$

for  $k = 0, 1, 2, \dots$

Given  $(A, B, C, D)$  we can compute

$$\mu_k = \hat{x}_k \equiv \hat{x}_{k|k} = \mathbf{E}[x_k | \mathcal{Y}_k] \quad (20)$$

using the Kalman Filter (see Elliott *et al.* (1995) for a reference probability style proof). Let

$$R_k = \Sigma_{k|k} \equiv \mathbf{E}[(x_k - \hat{x}_k)^2 | \mathcal{Y}_k]. \quad (21)$$

Then  $(\hat{x}_k, R_k)$  are determined recursively as follows:

$$\hat{x}_{k+1|k} = A + B\mu_k = A + B\hat{x}_{k|k}, \quad (22)$$

$$\Sigma_{k+1|k} = B^2 \Sigma_{k|k} + C^2, \quad (23)$$

$$\mathcal{K}_{k+1} = \Sigma_{k+1|k} / (\Sigma_{k+1|k} + D^2), \quad (24)$$

$$\hat{x}_{k+1} = \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + \mathcal{K}_{k+1}[y_{k+1} - \hat{x}_{k+1|k}], \quad (25)$$

$$R_{k+1} = \Sigma_{k+1|k+1} = D^2 \mathcal{K}_{k+1} = \Sigma_{k+1|k} - \mathcal{K}_{k+1} \Sigma_{k+1|k}. \quad (26)$$

For initialization we could take  $\hat{x}_0 = y_0$  and  $R_0 = D^2$ .

**Remark:** As  $k \rightarrow \infty$ ,  $R_k$  converges (monotonically) to the positive root  $R$  of  $B^2 R^2 + (C^2 + D^2 - B^2 D^2)R - C^2 D^2 = 0$  provided  $B^2 \neq 0$ ,  $C^2 D^2 \neq 0$ . We cannot say very much about limiting values of  $\hat{x}_k$  except it is exponentially forgetting of  $\hat{x}_0$ . However, these comments are not very important as we will only assume the model (18) holds over a short time horizon for a given set of values on  $(A, B, C, D)$ .

#### 3.2. Estimation of model

We now provide estimates for  $\vartheta \equiv (A, B, C^2, D^2)$  based on observations  $y_0, y_1, \dots, y_N$ . We use the EM-Algorithm to find  $\hat{\vartheta}$  by an iteration that provides a stationary value of the likelihood function based on the observations. In fact, let (see Elliott *et al.* (1995))

$$\mathcal{L}_N(\vartheta) = \mathbf{E}_0 \left[ \frac{dP_{\vartheta}}{dP_0} \middle| \mathcal{Y}_N \right] \quad (27)$$

be the likelihood function for  $\vartheta \in \Theta$ . The maximum likelihood estimate solves

$$\hat{\vartheta} = \arg \max_{\vartheta \in \Theta} \mathcal{L}_N(\vartheta). \quad (28)$$

The EM-Algorithm is an iterative method to compute  $\hat{\vartheta}$ .

If  $\hat{\vartheta}_0$  is an initial estimate, the EM-Algorithm provides  $\hat{\vartheta}_j, j = 1, 2, \dots$ , as a sequence of estimates.

**Step 1** (the E-step): Compute (with  $\tilde{\vartheta} = \hat{\vartheta}_j$ )

$$Q(\vartheta, \tilde{\vartheta}) = \mathbf{E}_{\tilde{\vartheta}} \left[ \log \frac{dP_{\vartheta}}{dP_{\tilde{\vartheta}}} \middle| \mathcal{Y}_N \right]. \quad (29)$$

**Step 2** (the M-step): Find

$$\vartheta_{j+1} \in \arg \max_{\vartheta \in \Theta} Q(\vartheta, \hat{\vartheta}_j). \quad (30)$$

In the literature there are basically two procedures to implement the EM-Algorithm.

**3.2.1. Shumway and Stoffer (1982) smoother approach.** This method is described by Shumway and Stoffer (1982, 2000) and is an off-line calculation and makes use of smoother estimators for the Kalman Filter.

We define smoothers (for  $k \leq N$ ):

$$\hat{x}_{k|N} = \mathbf{E}[x_k | \mathcal{Y}_N], \quad (31)$$

$$\Sigma_{k|N} = \mathbf{E}[(x_k - \hat{x}_{k|N})^2 | \mathcal{Y}_N] = \mathbf{E}[(x_k - \hat{x}_{k|N})^2], \quad (32)$$

$$\Sigma_{k-1, k|N} = \mathbf{E}[(x_k - \hat{x}_{k|N})(x_{k-1} - \hat{x}_{k-1|N})]. \quad (33)$$

These smoothers can be computed by

$$\mathcal{J}_k = \frac{B \Sigma_{k|k}}{\Sigma_{k+1|k}}, \quad (34)$$

$$\hat{x}_{k|N} = \hat{x}_{k|k} + \mathcal{J}_k [\hat{x}_{k+1|N} - (A + B\hat{x}_{k|k})], \quad (35)$$

$$\Sigma_{k|N} = \Sigma_{k|k} + \mathcal{J}_k^2 [\Sigma_{k+1|N} - \Sigma_{k+1|k}], \quad (36)$$

$$\Sigma_{k-1, k|N} = \mathcal{J}_{k-1} \Sigma_{k|k} + \mathcal{J}_k \mathcal{J}_{k-1} [\Sigma_{k, k+1|N} - B \Sigma_{k|k}], \quad (37)$$

$$\Sigma_{N-1, N|N} = B(1 - \mathcal{K}_N) \Sigma_{N-1|N-1}, \quad (38)$$

where initial values for this backward recursion  $\hat{x}_{N|N}$  and  $\Sigma_{N|N}$  are obtained from the Kalman Filter along with other estimates. Given  $\vartheta_j = (A, B, C^2, D^2)$  and initial values for the Kalman Filter  $\hat{x}_0 = \hat{x}_{0|N}$  and  $\Sigma_{0|0} = \Sigma_{0|N}$  which are the smoothers from the previous step ( $j-1$ ). The updates  $\vartheta_{j+1} = (\hat{A}, \hat{B}, \hat{C}^2, \hat{D}^2)$  are computed as follows:

$$\hat{A} = \frac{\alpha\gamma - \delta\beta}{N\alpha - \delta^2}, \quad (39)$$

$$\hat{B} = \frac{N\beta - \gamma\delta}{N\alpha - \delta^2}, \quad (40)$$

$$\hat{C}^2 = \frac{1}{N} \sum_{k=1}^N \left[ (x_k - \hat{A} - \hat{B}x_{k-1})^2 \middle| \mathcal{Y}_N \right], \quad (41)$$

$$\hat{D}^2 = \frac{1}{N+1} \sum_{k=0}^N [(y_k - x_k)^2 | \mathcal{Y}_N], \quad (42)$$

where

$$\alpha = \sum_{k=1}^N \mathbf{E}[x_{k-1}^2 | \mathcal{Y}_N] = \sum_{k=1}^N [\Sigma_{k-1|N} + \hat{x}_{k-1|N}^2],$$

$$\beta = \sum_{k=1}^N \mathbf{E}[x_{k-1}x_k | \mathcal{Y}_N] = \sum_{k=1}^N [\Sigma_{k-1, k|N} + \hat{x}_{k-1|N}\hat{x}_{k|N}],$$

$$\gamma = \sum_{k=1}^N \hat{x}_{k|N},$$

$$\delta = \sum_{k=1}^N \hat{x}_{k-1|N} = \gamma - \hat{x}_{N|N} + \hat{x}_{0|N},$$

and the right-hand sides of (41) and (42) are readily computed in terms of smoothers:

$$\hat{C}^2 = \frac{1}{N} \sum_{k=1}^N \left[ \Sigma_{k|N} + \hat{x}_{k|N}^2 + \hat{A}^2 + \hat{B}^2 \Sigma_{k-1|N} + \hat{B}^2 (\hat{x}_{k-1|N})^2 - 2\hat{A}\hat{x}_{k|N} + 2\hat{A}\hat{B}\hat{x}_{k-1|N} - 2\hat{B}\Sigma_{k-1,k|N} - 2\hat{B}\hat{x}_{k|N}\hat{x}_{k-1|N} \right],$$

$$\hat{D}^2 = \frac{1}{N+1} \sum_{k=0}^N \left[ y_k^2 - 2y_k\hat{x}_{k|N} + \Sigma_{k|N} + \hat{x}_{k|N}^2 \right].$$

The disadvantage of this algorithm is that, as new values of observations are given, the whole algorithm must be repeated off-line. However, if we have written a code for this estimation based on  $N+1$  observations  $y_0, y_1, \dots, y_N$ , then with  $y_{N+1}$  we simply provide the code with input  $y_1, y_2, \dots, y_{N+1}$ . The Shumway and Stoffer algorithm has been widely tested.

**3.2.2. Elliott and Krishnamurthy (1999) filter approach.** This approach to the implementation of the EM-Algorithm uses filtered quantities and can be performed on-line. This was based on a new class of finite-dimensional recursive filters for linear dynamic systems, which can be adapted to equations (18) and (19). The important advantages of this filter-based EM-Algorithm compared with the (standard) smoother based EM-Algorithm include (i) substantially reduced memory requirements, and (ii) ease of parallel implementation on a multiprocessor system (see Elliott and Krishnamurthy (1997, 1999)). The details of this approach are discussed in Elliott *et al.* (in press), where computational issues and convergence are reported.

As in section 3.2.1, we start with  $\hat{\vartheta}_j = (A, B, C^2, D^2)$  and initial values for the Kalman Filter and the next estimate  $\hat{\vartheta}_{j+1} = (\hat{A}, \hat{B}, \hat{C}^2, \hat{D}^2)$ .

We introduce various quantities:

$$\begin{aligned} H_k^0 &= \sum_{l=0}^k x_l^2, & \widehat{H}_k^0 &= \mathbf{E}[H_k^0 | \mathcal{Y}_k], \\ H_k^1 &= \sum_{l=1}^k x_l x_{l-1}, & \widehat{H}_k^1 &= \mathbf{E}[H_k^1 | \mathcal{Y}_k], \\ H_k^2 &= \sum_{l=0}^k x_l^2 x_{l-1}^2, & \widehat{H}_k^2 &= \mathbf{E}[H_k^2 | \mathcal{Y}_k], \\ J_k &= \sum_{l=0}^k x_l y_l, & \widehat{J}_k &= \mathbf{E}[J_k | \mathcal{Y}_k], \\ I_k^0 &= \sum_{l=0}^k x_l, & \widehat{I}_k^0 &= \mathbf{E}[I_k^0 | \mathcal{Y}_k], \\ I_k^1 &= \sum_{l=0}^k x_l x_{l-1}, & \widehat{I}_k^1 &= \mathbf{E}[I_k^1 | \mathcal{Y}_k], \\ Y_k &= \sum_{l=0}^k y_l^2. \end{aligned}$$

If  $\mathbf{E} = \mathbf{E}_{\hat{\vartheta}_j}$ , which means using  $\hat{\vartheta}_j = (A, B, C^2, D^2)$  in the dynamics (18), (19), then  $\hat{\vartheta}_{j+1} = (\hat{A}, \hat{B}, \hat{C}^2, \hat{D}^2)$  is given through

$$\hat{A} = \left[ 1 - \frac{(\widehat{I}_N^1)^2}{\widehat{H}_N^2} \right]^{-1} \left[ \widehat{I}_N^0 - \frac{\widehat{H}_N^1 \widehat{I}_N^1}{\widehat{H}_N^2} \right], \quad (43)$$

$$\hat{B} = \frac{1}{\widehat{H}_N^2} \left[ \widehat{H}_N^1 - \hat{A} \widehat{I}_N^1 \right], \quad (44)$$

$$\hat{C}^2 = \frac{1}{T} \left[ \widehat{H}_N^0 + T \hat{A}^2 + \widehat{H}_N^2 \hat{B}^2 - 2\hat{A} \widehat{I}_N^0 + 2\hat{A} \hat{B} \widehat{I}_N^1 - 2\hat{B} \widehat{H}_N^1 \right], \quad (45)$$

$$\hat{D}^2 = \frac{1}{T+1} \left[ Y_N - 2\widehat{J}_N + \widehat{H}_N^0 \right]. \quad (46)$$

We now provide recurrences for computing the quantities in (43)–(46). Given  $\vartheta_j$  we use the Kalman Filter calculations (22)–(26) to determine the values of  $\mu_k$  and  $R_k$ , from which we have ( $M = 0, 1, 2$ )

$$\widehat{H}_k^M = a_k^M + b_k^M \mu_k + d_k^M [R_k + \mu_k^2], \quad (47)$$

$$\widehat{J}_k = \bar{a}_k + \bar{b}_k \mu_k, \quad (48)$$

$$\widehat{I}_k^0 = s_k^0 + t_k^0 \mu_k, \quad (49)$$

$$\widehat{I}_k^1 = s_k^1 + t_k^1 \mu_k, \quad (50)$$

$$Y_k = Y_{k-1} + y_k^2, \quad (51)$$

where the various coefficients are determined as follows.

Set

$$\xi_k = \frac{1}{R_k} + \frac{B^2}{C^2}, \quad (52)$$

$$\Sigma_k = \frac{1}{\xi_k} \frac{B^2}{C^2}, \quad (53)$$

$$S_k = \frac{1}{\xi_k} \left[ \frac{\mu_k}{R_k} - \frac{AB}{C^2} \right], \quad (54)$$

then

$$\begin{aligned} a_0^0 &= 0, & b_0^0 &= 0, & d_0^0 &= 1, \\ a_{k+1}^0 &= a_k^0 + b_k^0 S_k + d_k^0 [S_k^2 + \xi_k^{-1}], \\ b_{k+1}^0 &= b_k^0 \Sigma_k + S_k + 2d_k^0 \Sigma_k S_k, \\ d_{k+1}^0 &= 1 + d_k^0 \Sigma_k^2, \end{aligned} \quad (55)$$

$$\begin{aligned} a_0^1 &= 0, & b_0^1 &= 0, & d_0^1 &= 0, \\ a_{k+1}^1 &= a_k^1 + b_k^1 S_k + d_k^1 [S_k^2 + \xi_k^{-1}], \\ b_{k+1}^1 &= b_k^1 \Sigma_k + S_k + 2d_k^1 \Sigma_k S_k, \\ d_{k+1}^1 &= \Sigma_k + d_k^1 \Sigma_k^2, \end{aligned} \quad (56)$$

$$\begin{aligned} a_0^2 &= 0, & b_0^2 &= 0, & d_0^2 &= 0, \\ a_{k+1}^2 &= a_k^2 + b_k^2 S_k + (d_k^2 + 1) [S_k^2 + \xi_k^{-1}], \\ b_{k+1}^2 &= b_k^2 \Sigma_k + 2(d_k^2 + 1) \Sigma_k S_k, \\ d_{k+1}^2 &= (1 + d_k^2) \Sigma_k^2, \end{aligned} \quad (57)$$



where programmers are warned to distinguish here between superscript 2 and squared terms

$$\begin{aligned}\bar{a}_0 &= 0, & \bar{b}_0 &= y_0, \\ \bar{a}_{k+1} &= \bar{a}_k + \bar{b}_k S_k, \\ \bar{b}_{k+1} &= y_{k+1} + \bar{b}_k \Sigma_k,\end{aligned}\quad (58)$$

$$\begin{aligned}s_0^0 &= 0, & t_0^0 &= 1, \\ s_{k+1}^0 &= s_k^0 + t_k^0 S_k, \\ t_{k+1}^0 &= 1 + t_k^0 \Sigma_k,\end{aligned}\quad (59)$$

$$\begin{aligned}s_0^1 &= 0, & t_0^1 &= 0, \\ s_{k+1}^1 &= s_k^1 + t_k^1 S_k, \\ t_{k+1}^1 &= 1 + t_k^1 \Sigma_k.\end{aligned}\quad (60)$$

**Remarks:** (a) Given  $\hat{\vartheta}_j$  the calculation of  $\hat{\vartheta}_{j+1}$  is computed by the steps: initialize the Kalman Filter with  $\mu_0 = y_0$  and  $R_0 = D^2$ . If the  $\mu_k$  and  $R_k$  have been calculated, the various coefficients may now be calculated using (52)–(54) and then (55)–(60). Find  $\mu_{k+1}$  and  $R_{k+1}$  from the Kalman Filter equations. Continue until  $k = N$ . Then compute the quantities in (47)–(51) (at  $k = N$ ) and then  $\hat{\vartheta}_{j+1}$  from (43)–(46). Some initial guess for  $\hat{\vartheta}_0$  must be made, and then iterations are concluded when the values for  $\hat{\vartheta}_j$  have converged sufficiently. Call this  $\hat{\vartheta}(N)$ . (b) If  $\hat{\vartheta}(N) = (\hat{A}, \hat{B}, \hat{C}^2, \hat{D}^2)$ , we should check that  $\hat{A} > 0$  and  $0 < \hat{B} < 1$ , else the pairs trading algorithm should not be used with this data. (c) The procedure described in (a) could be regarded as an initialization, and need not be repeated in subsequent steps, where only one iteration should suffice to update the coefficients in the model.

### 3.3. Implementation of the EM-Algorithm

We will assume that model (18), (19) holds over  $N$  periods. The values of  $\hat{\vartheta}(N)$  and  $\mu_N$  are computed based on the observations  $y_0, y_1, \dots, y_N$  and a trade may be initiated as described in section 2, and possibly unwound at  $t = N + 1$  (or according to some other criterion). Based on  $\hat{\vartheta}(N)$ ,  $\mu_{N+1}$  is computed based on the data  $y_1, y_2, \dots, y_{N+1}$  (the most recent  $N + 1$  values with the Kalman Filter initialized at  $\mu_1 = y_1$  and  $R_1 = \hat{D}^2(N)$ ) and a trade initiated.  $\hat{\vartheta}(N + 1)$  is calculated with one iteration using section 3.2.1 or 3.2.2 and using the Kalman Filter based on data  $y_1, y_2, \dots, y_{N+1}$ . The procedure is then repeated.

## 4. Numerical examples

Here we will provide some simulation and calibration results which demonstrate that the Shumway and Stoffer algorithm provides a consistent and robust estimating algorithm for the model. Studies based on Elliott and Krishnamurthy are given by Elliott *et al.* (in press). Some initial experiments have also been performed with real data with a hedge fund.

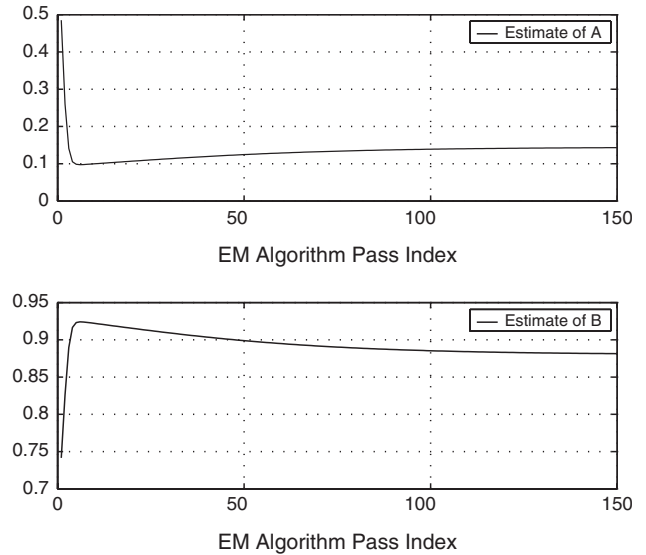


Figure 1. Convergence of the maximum likelihood estimates  $A$  and  $B$ .

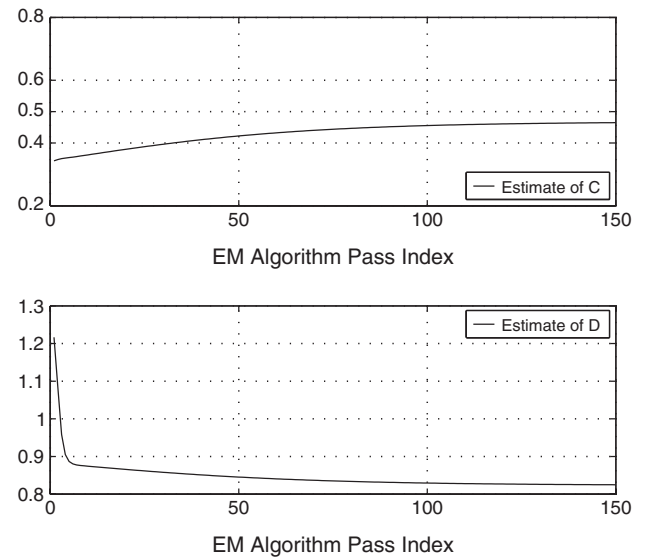


Figure 2. Convergence of the maximum likelihood estimates  $C$  and  $D$ .

To illustrate the typical performance of the Shumway and Stoffer EM algorithm, adapted to estimation of the set  $\{A, B, C, D\}$ , we consider a simulation with parameter values  $A = 0.20$ ,  $B = 0.85$ ,  $C = 0.60$  and  $D = 0.80$ . Our observation set contained 100 points. To initialize the EM algorithm, the following values were used:  $A = 1.20$ ,  $B = 0.50$ ,  $C = 0.30$  and  $D = 0.70$ , with  $\hat{x}_{0|0} = 0$  and  $\Sigma_{0|0} = 0.1$ . The EM algorithm was iterated 150 times. Figures 1 and 2 show convergence of the maximum likelihood estimates of all parameters.

## References

- Elliott, R.J., Aggoun, L. and Moore, J.B., *Hidden Markov Models*, 1995 (Springer: Berlin).
- Elliott, R.J. and Krishnamurthy, V., Exact finite-dimensional filters for maximum likelihood parameter estimation of

- continuous-time linear Gaussian systems. *SIAM Journal of Control and Optimization*, 1997, **35**, 1908–1923.
- Elliott, R.J. and Krishnamurthy, V., New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models. *IEEE Transactions of Automatic Control*, 1999, **44**, 938–951.
- Elliott, R.J., Malcolm, W.P. and van der Hoek, J., The numerical analysis of a filter based EM algorithm (in press).
- Finch, S., Ornstein–Uhlenbeck process. Unpublished Note. Available online at: <http://pauillac.inria.fr/algo/bsolve/constant/constant.html>.
- Gatev, E.G., Goetzmann, W.N. and Rouwenhorst, K.G., Pairs trading: performance of a relative average arbitrage rule. NBER Working Paper 7032, National Bureau of Economic Research Inc., 1999. Available online at: <http://www.nber.org/papers/w7032>.
- Litterman, B., *Modern Investment Management—An Equilibrium Approach*, chapter 1, 2003 (Wiley: New York).
- Nicholas, J.G., *Market Neutral Investing—Long/Short Hedge Fund Strategies*, 2000 (Bloomberg Professional Library, Bloomberg Press: Princeton, NJ, USA).
- Reverre, S., *The Complete Arbitrage Desk-book*, chapter 10, 2001 (McGraw Hill: New York).
- Shumway, R.H. and Stoffer, D.S., An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series*, 1982, **3**, 253–264.
- Shumway, R.H. and Stoffer, D.S., *Time Series Analysis and Its Applications*, 2000 (Springer: New York).
- Vidyamurthy, G., *Pairs Trading—Quantitative Methods and Analysis*, 2004 (Wiley: New York).
- Whistler, M., *Trading Pairs—Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies*, 2004 (Wiley: New York).