PAIRS TRADING

by

Robert J. Elliott
Haskayne School of Business
University of Calgary
Calgary, Alberta, Canada T2N 1N4
relliott@ucalgary.ca

John van der Hoek
Department of Applied Mathematics
The University of Adelaide
Adelaide, South Australia 5005
jvanderh@maths.adelaide.edu.au

and

William P. Malcolm
National ICT Australia
The Australian National University
Canberra, ACT. Australia 0200
paul.malcolm@nicta.com.au

October 25, 2004

Abstract

"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.

1. Introduction

Pairs Trading is a trading or investment strategy used to exploit financial markets that are out of equilibrium. Litterman [7] 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 thresh-hold 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 [6] and in two recent books by Vidyamurthy [12] and Whistler [13]. Reverre [9] 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 [8]).

2. The Spread Model

2.1 The State Process

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

$$x_{k+1} - x_k = (a - bx_k)\,\tau + \sigma\sqrt{\tau}\,\varepsilon_{k+1} \tag{2.1}$$

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 ε_{k+1} is independent of $x_0, x_1, ..., x_k$. The process mean reverts to $\mu = a/b$ with "strength" b. clearly

$$x_k \sim \mathcal{N}(\mu_k, \sigma_k^2) \tag{2.2}$$

where

$$\mu_k = \frac{a}{b} + \left[\mu_0 - \frac{a}{b}\right](1 - b\tau)^k \tag{2.3}$$

and

$$\sigma_k^2 = \frac{\sigma^2\tau}{1 - (1 - b\tau)^2}\left[1 - (1 - b\tau)^{2k}\right] + \sigma_0^2(1 - b\tau)^{2k} \tag{2.4}$$

It is easy to show that

$$\mu_k \to \frac{a}{b} \quad \text{as} \quad k \to \infty \tag{2.5}$$

and

$$\sigma_k^2 \to \frac{\sigma^2 \tau}{1-(1-b\tau)^2} \quad \text{as} \quad k \to \infty \tag{2.6}$$

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

We can also write (2.1) as

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1} \tag{2.7}$$

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 $\{X(t) \mid t \geq 0\}$ satisfies the stochastic differential equation:

$$dX(t) = (a - b\ X(t))\ dt + \sigma\ dW(t) \tag{2.8}$$

where $\{W(t) \mid t \geq 0\}$ 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:

$$y_k = x_k + D\ \omega_k \tag{2.9}$$

where $\{\omega_k\}$ are iid Gaussian $\mathcal{N}(0,1)$ and independent of the $\{\varepsilon_k\}$ in (2.1) and $D > 0$. We may assume that $0 \leq C < D$, which should be the case for small values of τ.

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

$$\widehat{x}_k = \mathbf{E}[x_k |\ \mathcal{Y}_k] \tag{2.10}$$

which are "best" estimates of the hidden state process through the observed process. In order to make the estimate (2.10), 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 $y_k > \widehat{x}_{k|k-1} = \mathbf{E}[x_k|\mathcal{Y}_{k-1}]$ 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 $\widehat{x}_{k|k-1}$ by some threshold value. A corresponding short trade could be entered when $y_k < \widehat{x}_{k|k-1}$.

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 $\widehat{x}_{k|k-1}$ may or may not lead to a trade if thresholds must be met. How are thresholds set? See Vidyamurthy [12] for some possibilities. When is the pairs trade unwound? There are various possibilities: the next trading time (see Reverre [9] example) or when the spread corrects sufficiently. The price application and data bases used are often proprietary in industry applications. The machinery we present provides then some useful tools appropriate for pairs trading.

Another explicit strategy could make use of First-Passage Times results (see [5] and references cited therein) for the (standardized) Ornstein-Uhlenbeck process

$$dZ(t) = -Z(t)\ dt + \sqrt{2}\ dW(t). \tag{2.11}$$

Let

$$T_{0,c} = \inf\{t \geq 0, \quad Z(t) = 0 \mid Z(0) = c\}. \tag{2.12}$$

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

$$f_{0,c}(t) = \sqrt{\frac{2}{\pi}} \frac{|c|\ e^{-t}}{(1 - e^{-2t})^{3/2}} \exp\left(-\frac{c^2\ e^{-2t}}{2(1 - e^{-2t})}\right) \tag{2.13}$$

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

$$\widehat{t} = \frac{1}{2} \ln\left[1 + \frac{1}{2}\left(\sqrt{(c^2 - 3)^2 + 4c^2} + c^2 - 3\right)\right] \tag{2.14}$$

We can also write (2.8) in the form

$$dX(t) = -\rho\ (X(t) - \mu)\ dt + \sigma\ dW(t) \tag{2.15}$$

where $\rho = b$ and $\mu = \frac{a}{b}$. When

$$X(0) = \mu + c\ \frac{\sigma}{\sqrt{2\rho}} \tag{2.16}$$

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

$$T = \frac{1}{\rho}\ \widehat{t} \tag{2.17}$$

where $\widehat{t}$ is given by (2.14). Use the Ornstein-Uhlenbeck process as an approximation to (2.7) with $a = \frac{A}{\tau}$, $b = \frac{1-B}{\tau}$ and $\sigma = \frac{C}{\sqrt{\tau}}$ with the calibrated values A, B, C. Choose a value of $c > 0$. Enter a pair trade when $y_k \geq \mu + c\ \frac{\sigma}{\sqrt{2\rho}}$ and unwind the trade at time T later. A corresponding pair trade would be performed when $y_k \leq \mu - c\ \frac{\sigma}{\sqrt{2\rho}}$ 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**:

$$x_{k+1} = A + Bx_k + C\varepsilon_{k+1} \tag{3.1}$$

and the **observation equation**:

$$y_k = x_k + D\ \omega_k \tag{3.2}$$

for $k = 0, 1, 2, ...$

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

$$\mu_k = \widehat{x}_k \equiv \widehat{x}_{k|k} = \mathbf{E}[x_k|\mathcal{Y}_k] \tag{3.3}$$

using the Kalman Filter (see [1], for a reference probability style proof). Let

$$R_k = \Sigma_{k|k} \equiv \mathbf{E}[(x_k - \widehat{x}_k)^2|\mathcal{Y}_k] \tag{3.4}$$

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

$$\widehat{x}_{k+1|k} = A + B\mu_k = A + B\widehat{x}_{k|k} \tag{3.5}$$

$$\Sigma_{k+1|k} = B^2\ \Sigma_{k|k} + C^2 \tag{3.6}$$

$$\mathcal{K}_{k+1} = \Sigma_{k+1|k}/(\Sigma_{k+1|k} + D^2) \tag{3.7}$$

$$\widehat{x}_{k+1} = \widehat{x}_{k+1|k+1} = \widehat{x}_{k+1|k} + \mathcal{K}_{k+1}\left[y_{k+1} - \widehat{x}_{k+1|k}\right] \tag{3.8}$$

$$R_{k+1} = \Sigma_{k+1|k+1} = D^2\mathcal{K}_{k+1} = \Sigma_{k+1|k} - \mathcal{K}_{k+1}\Sigma_{k+1|k} \tag{3.9}$$

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

Remark

As $k \to \infty$, R_k converges (monotonically) to the positive root R of $B^2R^2 + (C^2 + D^2 - B^2D^2)R - C^2D^2 = 0$ provided $B^2 \neq 0$, $C^2D^2 \neq 0$. We cannot say very much about limiting values of $\widehat{x}_k$ except it is exponentially forgetting of $\widehat{x}_0$. However these comments are not very important as we will only assume the model (3.1) 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 $y_0, y_1, ..., y_N$. 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 [1])

$$\mathcal{L}_N(\vartheta) = \mathbf{E}_0\left[\frac{dP_\vartheta}{dP_0} \;\Big|\; \mathcal{Y}_N\right] \tag{3.10}$$

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

$$\hat{\vartheta} = \arg\max_{\vartheta\in\Theta} \mathcal{L}_N(\vartheta) \tag{3.11}$$

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

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

Step 1 [The E-step]

Compute (with $\tilde{\vartheta} = \hat{\vartheta}_j$)

$$Q(\vartheta, \tilde{\vartheta}) = \mathbf{E}_{\tilde{\vartheta}}\left[\log\ \frac{dP_\vartheta}{dP_{\tilde{\vartheta}}} \;\Big|\; \mathcal{Y}_N\right] \tag{3.12}$$

Step 2 [The M-step]

Find

$$\vartheta_{j+1} \in \arg\max_{\vartheta\in\Theta} Q(\vartheta, \hat{\vartheta}_j) \tag{3.13}$$

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 in Shumway and Stoffer [10] and [11] and is an off-line calculation and makes use of smoother estimators for the Kalman Filter.

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

$$\widehat{x}_{k|N} = \mathbf{E}[x_k|\mathcal{Y}_N] \tag{3.14}$$

$$\Sigma_{k|N} = \mathbf{E}[(x_k - \widehat{x}_{k|N})^2|\mathcal{Y}_N] = \mathbf{E}[(x_k - \widehat{x}_{k|N})^2] \tag{3.15}$$

$$\Sigma_{k-1,k|N} = \mathbf{E}[(x_k - \widehat{x}_{k|N})(x_{k-1} - \widehat{x}_{k-1|N})] \tag{3.16}$$

These smoothers can be computed by:

$$\mathcal{J}_k = \frac{B\Sigma_{k|k}}{\Sigma_{k+1|k}} \tag{3.17}$$

$$\widehat{x}_{k|N} = \widehat{x}_{k|k} + \mathcal{J}_k\left[\widehat{x}_{k+1|N} - (A + B\widehat{x}_{k|k})\right] \tag{3.18}$$

$$\Sigma_{k|N} = \Sigma_{k|k} + \mathcal{J}_k^2\left[\Sigma_{k+1|N} - \Sigma_{k+1|k}\right] \tag{3.19}$$

$$\Sigma_{k-1,k|N} = \mathcal{J}_{k-1}\Sigma_{k|k} + \mathcal{J}_k\mathcal{J}_{k-1}\left[\Sigma_{k,k+1|N} - B\Sigma_{k|k}\right] \tag{3.20}$$

$$\Sigma_{N-1,N|N} = B(1 - \mathcal{K}_N)\Sigma_{N-1|N-1} \tag{3.21}$$

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

$$\widehat{A} = \frac{\alpha\gamma - \delta\beta}{N\alpha - \delta^2} \tag{3.22}$$

$$\widehat{B} = \frac{N\beta - \gamma\delta}{N\alpha - \delta^2} \tag{3.23}$$

$$\widehat{C^2} = \frac{1}{N}\sum_{k=1}^{N}\left[(x_k - \widehat{A} - \widehat{B}x_{k-1})^2 | \mathcal{Y}_N\right] \tag{3.24}$$

$$\widehat{D^2} = \frac{1}{N+1}\sum_{k=0}^{N}\left[(y_k - x_k)^2 | \mathcal{Y}_N\right] \tag{3.25}$$

where

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

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

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

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

and the right hand sides of (3.24) and (3.25) are readily computed in terms of smoothers:

$$\begin{aligned}\widehat{C^2} = \frac{1}{N}\sum_{k=1}^{N}[&\Sigma_{k|N} + \widehat{x}_{k|N}^2 + \widehat{A}^2 + \widehat{B}^2\Sigma_{k-1|N} + \widehat{B}^2(\widehat{x}_{k-1|N})^2 - 2\widehat{A}\widehat{x}_{k|N} \\ &+ 2\widehat{A}\widehat{B}\widehat{x}_{k-1|N} - 2\widehat{B}\Sigma_{k-1,k|N} - 2\widehat{B}\widehat{x}_{k|N}\widehat{x}_{k-1|N}]\end{aligned}$$

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

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 $y_0, y_1, ..., y_N$, then with y_{N+1} we simply provide the code with input $y_1, y_2, ..., y_{N+1}$. The Shumway and Stoffer Algorithm has been widely tested.

3.2.2 Elliott and Krishnamurthy (1999) filter approach

This approach at 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 the equations (3.1) and (3.2). The important advantages of this filter-based EM-Algorithm compared with the (standard) smoother based EM-Algorithm include (i) substantially reduced memory requirements (ii) ease of parallel implementation on a multiprocessor system (see Elliott and Krishnamurthy [3]). The details of this approach are discussed in Elliott et al [4] where computational issues and convergence are discussed.

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 $\hat{\vartheta}_{j+1} = (\widehat{A}, \widehat{B}, \widehat{C^2}, \widehat{D^2})$.

We introduce various quantities:

$$H_k^0 = \sum_{l=0}^{k} x_l^2, \qquad \widehat{H_k^0} = \mathbf{E}[H_k^0 | \mathcal{Y}_k]$$

$$H_k^1 = \sum_{l=1}^{k} x_l x_{l-1}, \qquad \widehat{H_k^1} = \mathbf{E}[H_k^1 | \mathcal{Y}_k]$$

$$H_k^2 = \sum_{l=0}^{k} x_{l-1}^2, \qquad \widehat{H_k^2} = \mathbf{E}[H_k^2 | \mathcal{Y}_k]$$

$$J_k = \sum_{l=0}^{k} x_l y_l, \qquad \widehat{J_k} = \mathbf{E}[J_k | \mathcal{Y}_k]$$

$$I_k^0 = \sum_{l=0}^{k} x_l, \qquad \widehat{I_k^0} = \mathbf{E}[I_k^0 | \mathcal{Y}_k]$$

$$I_k^1 = \sum_{l=0}^{k} x_{l-1}, \qquad \widehat{I_k^1} = \mathbf{E}[I_k^1 | \mathcal{Y}_k]$$

$$Y_k = \sum_{l=0}^{k} y_l^2$$

If $\mathbf{E} = \mathbf{E}_{\hat{\vartheta}_j}$ which means using $\hat{\vartheta}_j = (A, B, C^2, D^2)$ in the dynamics (3.1), (3.2), then $\hat{\vartheta}_{j+1} = (\widehat{A}, \widehat{B}, \widehat{C^2}, \widehat{D^2})$ is given through:

$$\widehat{A} = \left[1 - \frac{(\widehat{I_N^1})^2}{\widehat{H_N^2}} \right]^{-1} \left[\widehat{I_N^0} - \frac{\widehat{H_N^1}\widehat{I_N^1}}{\widehat{H_N^2}} \right] \tag{3.26}$$

$$\widehat{B} = \frac{1}{\widehat{H_N^2}} \left[\widehat{H_N^1} - \widehat{A}\widehat{I_N^1} \right] \tag{3.27}$$

$$\widehat{C^2} = \frac{1}{T}\left[\widehat{H_N^0} + T\widehat{A}^2 + \widehat{H_N^2}\widehat{B}^2 - 2\widehat{A}\widehat{I_N^0} + 2\widehat{A}\widehat{B}\widehat{I_N^1} - 2\widehat{B}\widehat{H_N^1}\right] \tag{3.28}$$

$$\widehat{D^2} = \frac{1}{T+1}\left[Y_N - 2\widehat{J_N} + \widehat{H_N^0}\right] \tag{3.29}$$

We now provide recurrences for computing the quantities in (3.26)-(3.29). Given ϑ_j we use the Kalman-Filter calculations (3.5) - (3.9) to determine the values of μ_k and R_k from which we have ($M = 0, 1, 2$)

$$\widehat{H_k^M} = a_k^M + b_k^M \mu_k + d_k^M [R_k + \mu_k^2] \tag{3.30}$$

$$\widehat{J_k} = \overline{a}_k + \overline{b}_k \mu_k \tag{3.31}$$

$$\widehat{I_k^0} = s_k^0 + t_k^0 \mu_k \tag{3.32}$$

$$\widehat{I_k^1} = s_k^1 + t_k^1 \mu_k \tag{3.33}$$

$$Y_k = Y_{k-1} + y_k^2 \tag{3.34}$$

where the various coefficients are determined as follows:

Set

$$\xi_k = \frac{1}{R_k} + \frac{B^2}{C^2} \tag{3.35}$$

$$\Sigma_k = \frac{1}{\xi_k}\frac{B^2}{C^2} \tag{3.36}$$

$$S_k = \frac{1}{\xi_k}\left[\frac{\mu_k}{R_k} - \frac{AB}{C^2}\right] \tag{3.37}$$

then

$$\begin{aligned} a_0^0 &= 0, \quad b_0^0 = 0, \quad d_0^0 = 1 \\ a_{k+1}^0 &= a_k^0 + b_k^0 S_k + d_k^0\left[S_k^2 + \xi_k^{-1}\right] \\ b_{k+1}^0 &= b_k^0 \Sigma_k + S_k + 2d_k^0 \Sigma_k S_k \\ d_{k+1}^0 &= 1 + d_k^0 \Sigma_k^2 \end{aligned} \tag{3.38}$$

$$\begin{aligned} a_0^1 &= 0, \quad b_0^1 = 0, \quad d_0^1 = 0 \\ a_{k+1}^1 &= a_k^1 + b_k^1 S_k + d_k^1\left[S_k^2 + \xi_k^{-1}\right] \\ b_{k+1}^1 &= b_k^1 \Sigma_k + S_k + 2d_k^1 \Sigma_k S_k \\ d_{k+1}^1 &= \Sigma_k + d_k^1 \Sigma_k^2 \end{aligned} \tag{3.39}$$

$$\begin{aligned} a_0^2 &= 0, \quad b_0^2 = 0, \quad d_0^2 = 0 \\ a_{k+1}^2 &= a_k^2 + b_k^2 S_k + (d_k^2 + 1)\left[S_k^2 + \xi_k^{-1}\right] \\ b_{k+1}^2 &= b_k^2 \Sigma_k + 2(d_k^2 + 1)\Sigma_k S_k \\ d_{k+1}^2 &= (1 + d_k^2)\Sigma_k^2 \end{aligned} \tag{3.40}$$

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

$$\begin{aligned} \overline{a}_0 &= 0, \quad \overline{b}_0 = y_0 \\ \overline{a}_{k+1} &= \overline{a}_k + \overline{b}_k S_k \\ \overline{b}_{k+1} &= y_{k+1} + \overline{b}_k \Sigma_k \end{aligned} \tag{3.41}$$

$$\begin{aligned} s_0^0 &= 0, \quad t_0^0 = 1 \\ s_{k+1}^0 &= s_k^0 + t_k^0 S_k \\ t_{k+1}^0 &= 1 + t_k^0 \Sigma_k \end{aligned} \tag{3.42}$$

$$\begin{aligned} s_0^1 &= 0, \quad t_0^1 = 0 \\ s_{k+1}^1 &= s_k^1 + t_k^1 S_k \\ t_{k+1}^1 &= 1 + t_k^1 \Sigma_k \end{aligned} \tag{3.43}$$

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 μ_k and R_k have been calculated, the various coefficients may now be calculated using (3.35)-(3.37) and then (3.38)-(3.43). Find μ_{k+1} and R_{k+1} from the Kalman-Filter equations. Continue till $k = N$. Then compute the quantities in (3.30) - (3.34) (at $k = N$) and then $\hat{\vartheta}_{j+1}$ from (3.26) - (3.29). 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 $\hat{\vartheta}(N) = (\widehat{A}, \widehat{B}, \widehat{C^2}, \widehat{D^2})$, we should check that $\widehat{A} > 0$ and $0 < \widehat{B} < 1$, else the pairs trading algorithm should not be used with this data.

(c) The procedure described in (a) could be regarded as an initialization, it need not be repeated in subsequent steps, when only one iteration should suffice to update the coefficients in the model.

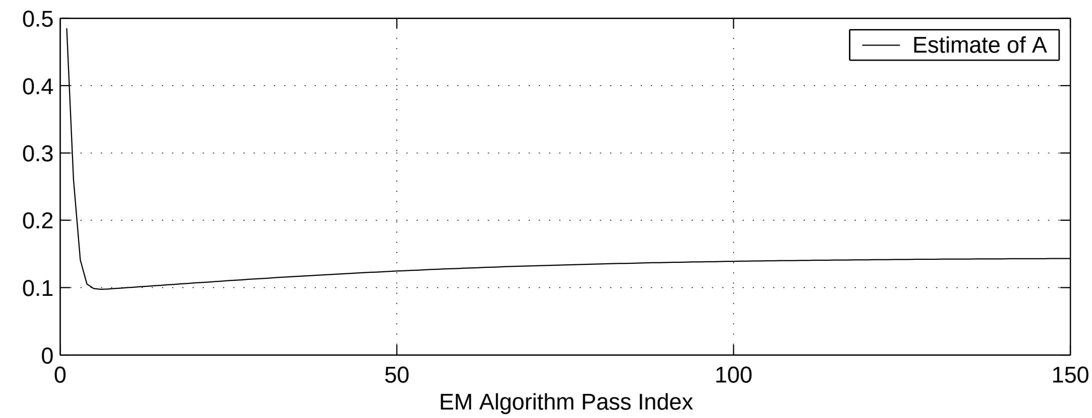
3.3 Implementation of the EM-Algorithm

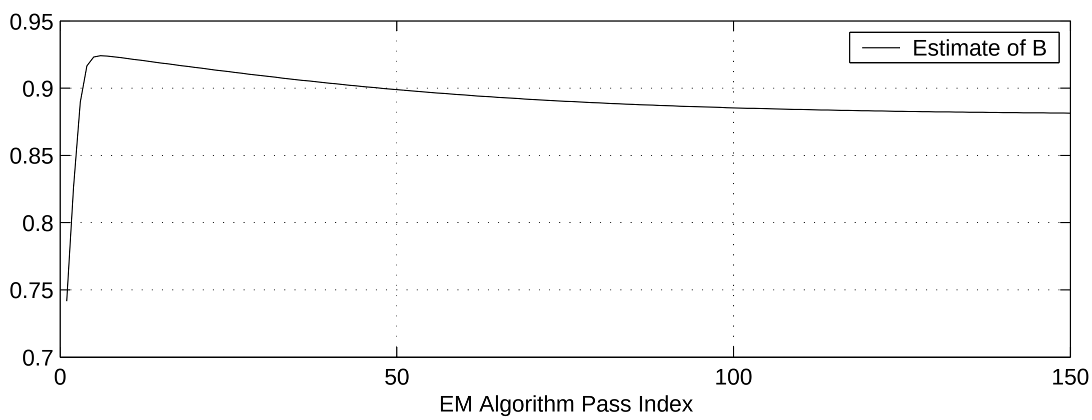
We will assume that model (3.1) (3.2) holds over N periods. The values of $\hat{\vartheta}(N)$ and μ_N are computed based on the observations $y_0, y_1, ..., y_N$ and a trade may be initiated as described in section 2, and possibly unwound at $t = N + 1$ (or according to some other criterion). Based on $\hat{\vartheta}(N)$, μ_{N+1} is computed based on the data $y_1, y_2, ..., y_{N+1}$ (the most recent $N + 1$ values with Kalman-Filter initialized at $\mu_1 = y_1$ and $R_1 = \widehat{D^2}(N)$) and a trade initiated. $\hat{\vartheta}(N + 1)$ is calculated with one iteration using sections 3.2.1 or 3.2.2 and using the Kalman-Filter based on data $y_1, y_2, ..., y_{N+1}$. The procedure is then repeated.

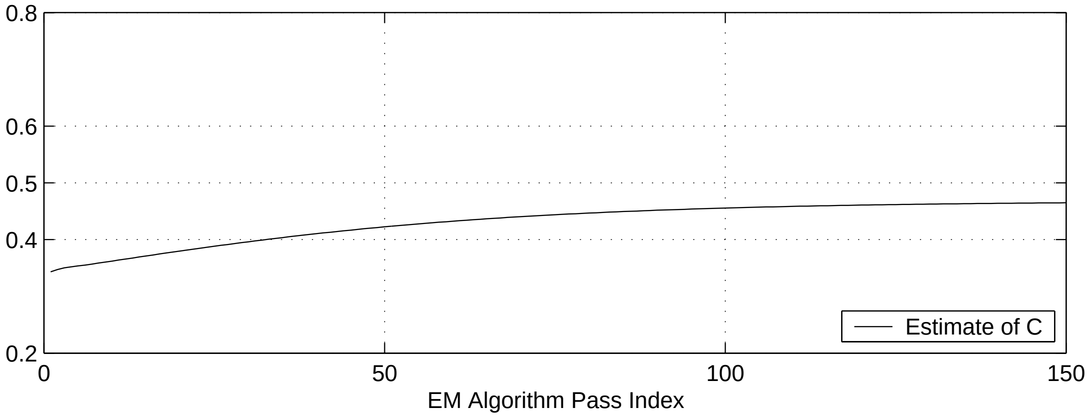
4. Numerical Examples

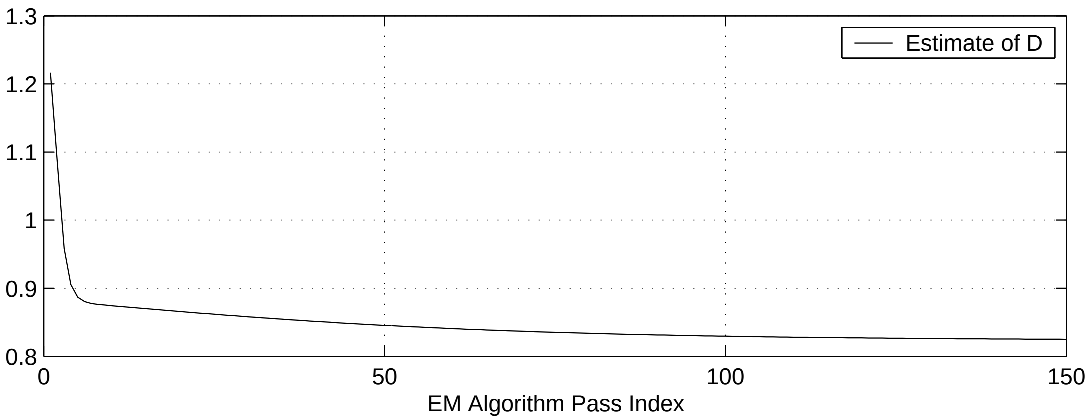
Here we will provide some simulation and calibration results which demonstrate that that the Shumway and Stoffer algorithm provides a consistent and robust estimating algorithm for the model. Studies based on Elliott and Krishnamurthy will be given in [4]. Some initial experiments have also been performed with real data with a hedge fund.

To illustrate 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 $\widehat{x}_{0|0} = 0$ and $\Sigma_{0|0} = 0.1$. The EM algorithm was iterated 150 times. The following figures show convergence of the maximum likelihood estimates of all parameters.









5. References

[1] Elliott, R.J., L. Aggoun and J.B. Moore. *Hidden Markov Models.* (Springer Verlag 1995)

[2] Elliott, R.J. and V. Krishnamurthy. *Exact Finite-dimensional Filters for Maximum Likelihood Parameter Estimation of Continuous-time Linear Gaussian Systems.* SIAM J. Control and Optimization 35 [6] (1997), 1908-1923.

[3] Elliott, R.J. and V. Krishnamurthy. *New Finite-dimensional Filters for Parameter Estimation of Discrete-time Linear Gaussian Models.* IEEE Transactions of Automatic Control 44 [5] (1999), 938-951.

[4] Elliott, R.J., W.P. Malcolm and J van der Hoek. *The numerical analysis of a filter based EM algorithm* (in progress 2004)

[5] Finch, Steven. *Ornstein-Uhlenbeck Process.* Unpublished Note. http://pauillac.inria.fr/algo/bsolve/constant/constant.html

[6] Gatev, Evan G., William N. Goetzmann and K. Geert Rouwenhorst. *Pairs Trading: Performance of a Relative Average Arbitrage Rule.* NBER Working Papers 7032, National Bureau of Economic Research Inc., 1999. (http://www.nber.org/papers/w7032)

[6] Litterman, Bob. *Modern Investment Management - An Equilibrium Approach.* (John Wiley & Sons, Inc. 2003) [Chapter 1]

[8] Nicholas, Joseph G. *Market Neutral Investing - Long/Short Hedge Fund Strategies.* (Bloomberg Professional Library, Bloomberg Press 2000)

[9] Reverre, Stephane. *The Complete Arbitrage Desk-book.* (McGraw Hill 2001) [Chapter 10]

[10] Shumway, R.H. and D.S. Stoffer. *An Approach to Time Series Smoothing and Forecasting using the EM Algorithm.* Journal of Time Series 3 [4] (1982), 253-264.

[11] Shumway, R.H. and D.S. Stoffer. *Time Series Analysis and Its Applications.* (Springer Verlag 2000)

[12] Vidyamurthy, Ganpathy. *Pairs Trading - Quantitative Methods and Analysis.* (John Wiley & Sons, Inc. 2004)

[13] Whistler, Mark. *Trading Pairs - Capturing Profits and Hedging Risk with Statistical Arbitrage Strategies.* (John Wiley & Sons, Inc. 2004)