# 3

# Mean-Reverting Portfolios

## *Tradeoffs between Sparsity and Volatility*

Marco Cuturi[1] and Alexandre d'Aspremont[2]
[1] *Kyoto University, Japan*
[2] *CNRS - Ecole Normale supérieure, France*

Mean-reverting assets are one of the holy grails of financial markets: if such assets existed, they would provide trivially profitable investment strategies for any investor able to trade them, thanks to the knowledge that such assets oscillate predictably around their long-term mean. The modus operandi of cointegration-based trading strategies (Tsay, 2005, §8) is to create first a portfolio of assets whose aggregate value mean-reverts, and then to exploit that knowledge by selling short or buying that portfolio when its value deviates from its long-term mean. Such portfolios are typically selected using tools from cointegration theory (Engle and Granger, 1987; Johansen, 1991), whose aim is to detect combinations of assets that are stationary and therefore mean-reverting. We argue in this chapter that focusing on stationarity only may not suffice to ensure profitability of cointegration-based strategies. While it might be possible to create synthetically, using a large array of financial assets, a portfolio whose aggregate value is stationary and therefore mean-reverting, trading such a large portfolio incurs in practice important trade or borrow costs. Looking for stationary portfolios formed by many assets may also result in portfolios that have a very small volatility and that require significant leverage to be profitable. We study in this chapter algorithmic approaches that can mitigate these effects by searching for maximally mean-reverting portfolios that are sufficiently sparse and/or volatile.

## 3.1 Introduction

Mean-reverting assets, namely assets whose price oscillates predictably around a long-term mean, provide investors with an ideal investment opportunity. Because of their tendency

*Financial Signal Processing and Machine Learning*, First Edition.
Edited by Ali N. Akansu, Sanjeev R. Kulkarni and Dmitry Malioutov.

to pull back to a given price level, a naive contrarian strategy of buying the asset when its price lies below that mean, or selling short the asset when it lies above that mean, can be profitable. Unsurprisingly, assets that exhibit significant mean reversion are very hard to find in efficient markets. Whenever mean reversion is observed in a single asset, it is almost always impossible to profit from it: the asset may typically have very low volatility, be illiquid, or be hard to short-sell, or its mean reversion may occur at a time scale (months, years) for which the borrow cost of holding or shorting the asset may well exceed any profit expected from such a contrarian strategy.

### *3.1.1 Synthetic Mean-Reverting Baskets*

Since mean-reverting assets rarely appear in liquid markets, investors have focused instead on creating synthetic assets that can mimic the properties of a single mean-reverting asset, and trading such synthetic assets as if they were a single asset. Such a synthetic asset is typically designed by combining long and short positions in various liquid assets to form a *mean-reverting portfolio*, whose aggregate value exhibits significant mean reversion.

Constructing such synthetic portfolios is, however, challenging. Whereas simple descriptive statistics and unit-root test procedures can be used to test whether a single asset is mean-reverting, building mean-reverting portfolios requires finding a proper vector of algebraic weights (long and short positions) that describes a portfolio that has a mean-reverting aggregate value. In that sense, mean-reverting portfolios are made by the investor and cannot be simply chosen among tradable assets. A mean-reverting portfolio is characterized both by the pool of assets the investor has selected (starting with the dimension of the vector) and by the fixed nominal quantities (or weights) of each of these assets in the portfolio, which the investor also needs to set. When only two assets are considered, such baskets are usually known as long-short trading pairs. We consider in this paper baskets that are constituted by more than two assets.

### *3.1.2 Mean-Reverting Baskets with Sufficient Volatility and Sparsity*

A mean-reverting portfolio must exhibit sufficient mean reversion to ensure that a contrarian strategy is profitable. To meet this requirement, investors have relied on cointegration theory (Engle and Granger, 1987; Johansen, 2005; Maddala and Kim, 1998) to estimate linear combinations of assets that exhibit stationarity (and therefore mean reversion) using historical data. We argue in this chapter, as we did in earlier publications (Cuturi and d'Aspremont, 2013; d'Aspremont, 2011), that mean-reverting strategies cannot, however, only rely on this approach to be profitable. Arbitrage opportunities can only exist if they are large enough to be traded without using too much leverage or incurring too many transaction costs. For mean-reverting baskets, this condition translates naturally into a first requirement that the gap between the basket valuation and its long-term mean is large enough on average, namely that the basket price has sufficient variance or volatility. A second desirable property is that mean-reverting portfolios require trading as few assets as possible to minimize costs, namely that the weights vector of that portfolio is sparse. We propose in this work methods that maximize a proxy for mean reversion, and that can take into account at the same time constraints on variance and sparsity.

We propose first in Section 3.2 three proxies for mean reversion. Section 3.3 defines the basket optimization problems corresponding to these quantities. We show in Section 3.4 that each of these problems translate naturally into semidefinite relaxations that produce either exact or approximate solutions using sparse principal component analysis (PCA) techniques. Finally, we present numerical evidence in Section 3.5 that taking into account sparsity and volatility can significantly boost the performance of mean-reverting trading strategies in trading environments where trading costs are not negligible.

## 3.2 Proxies for Mean Reversion

Isolating stable linear combinations of variables of multivariate time series is a fundamental problem in econometrics. A classical formulation of the problem reads as follows: given a vector valued process $x = (x_t)_t$ taking values in $\mathbb{R}^n$ and indexed by time $t \in \mathbb{N}$, and making no assumptions on the stationarity of each individual component of $x$, can we estimate one or many directions $y \in \mathbb{R}^n$ such that the univariate process $(y^T x_t)$ is stationary? When such a vector $y$ exists, the process $x$ is said to be cointegrated. The goal of cointegration techniques is to detect and estimate such directions $y$. Taking for granted that such techniques can efficiently isolate sparse mean-reverting baskets, their financial application can be either straightforward using simple event triggers to buy, sell, or simply hold the basket (Tsay, 2005, §8.6), or more elaborate optimal trading strategies if one assumes that the mean-reverting basket value is a Ohrstein–Ullenbeck process, as discussed in Elie and Espinosa, (2011), Jurek and Yang (2007), and Liu and Timmermann (2010).

### *3.2.1 Related Work and Problem Setting*

Engle and Granger (1987) provided in their seminal work a first approach to compare two nonstationary univariate time series $(x_t, y_t)$, and test for the existence of a term $\alpha$ such that $y_t - \alpha x_t$ becomes stationary. Following this seminal work, several techniques have been proposed to generalize that idea to multivariate time series. As detailed in the survey by Maddala and Kim (1998, §5), cointegration techniques differ in the modeling assumptions they require on the time series themselves. Some are designed to identify only one cointegrated relationship, whereas others are designed to detect many or all of them. Among these references, Johansen (1991) proposed a popular approach that builds upon a vector autoregression (VAR) model, as surveyed in Johansen (2004, 2005). These approaches all discuss issues that are relevant to econometrics, such as detrending and seasonal adjustments. Some of them focus more specifically on testing procedures designed to check whether such cointegrated relationships exist or not, rather than on the robustness of the estimation of that relationship itself. We follow in this work a simpler approach proposed by d'Aspremont (2011), which is to trade off interpretability, testing, and modeling assumptions for a simpler optimization framework that can be tailored to include other aspects than only stationarity. d'Aspremont (2011) did so by adding regularizers to the predictability criterion proposed by Box and Tiao (1977). We follow in this chapter the approach we proposed in Cuturi and d'Aspremont (2013) to design mean reversion proxies that do not rely on any modeling assumption.

Throughout this chapter, we write $\mathbf{S}_n$ for the $n \times n$ cone of positive definite matrices. We consider in the following a multivariate stochastic process $x = (x_t)_{t \in \mathbb{N}}$ taking values in $\mathbb{R}^n$.

We write $\mathcal{A}_k = \mathbf{E}[x_t x_{t+k}^T], k \geq 0$ for the lag-$k$ autocovariance matrix of $x_t$ if it is finite. Using a sample path $\mathbf{x}$ of $(x_t)$, where $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_T)$ and each $\mathbf{x}_t \in \mathbb{R}^n$, we write $A_k$ for the *empirical* counterpart of $\mathcal{A}_k$ computed from $\mathbf{x}$,

$$A_k \stackrel{\text{def}}{=} \frac{1}{T-k-1} \sum_{t=1}^{T-k} \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_{t+k}^T, \; \tilde{\mathbf{x}}_t \stackrel{\text{def}}{=} \mathbf{x}_t - \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t. \tag{3.1}$$

Given $y \in \mathbb{R}^n$, we now define three measures that can all be interpreted as proxies for the mean reversion of $y^T x_t$. *Predictability* –defined for stationary processes by Box and Tiao (1977) and generalized for nonstationary processes by Bewley *et al*. (1994) – measures how close to noise the series is. The *portmanteau* statistic of Ljung and Box (1978) is used to test whether a process is white noise. Finally, the *crossing statistic* (Ylvisaker, 1965) measures the probability that a process crosses its mean per unit of time. In all three cases, low values for these criteria imply a fast mean reversion.

## *3.2.2 Predictability*

We briefly recall the canonical decomposition derived in Box and Tiao (1977). Suppose that $x_t$ follows the recursion:

$$x_t = \hat{x}_{t-1} + \varepsilon_t, \tag{3.2}$$

where $\hat{x}_{t-1}$ is a predictor of $x_t$ built upon past values of the process recorded up to $t-1$, and $\varepsilon_t$ is a vector of independent and identically distributed (i.i.d.) Gaussian noise with zero mean and covariance $\Sigma \in \mathbf{S}_n$ independent of all variables $(x_r)_{r<t}$. The canonical analysis in Box and Tiao (1977) starts as follows.

### 3.2.2.1 Univariate case

Suppose $n = 1$ and thus $\Sigma \in \mathbb{R}_+$; Equation (3.2 ) leads thus to

$$\mathbf{E}[x_t^2] = \mathbf{E}[\hat{x}_{t-1}^2] + \mathbf{E}[\varepsilon_t^2], \text{ thus } 1 = \frac{\hat{\sigma}^2}{\sigma^2} + \frac{\Sigma}{\sigma^2},$$

by introducing the variances $\sigma^2$ and $\hat{\sigma}^2$ of $x_t$ and $\hat{x}_t$, respectively. Box and Tiao measure the *predictability* of $x_t$ by the ratio

$$\lambda \stackrel{\text{def}}{=} \frac{\hat{\sigma}^2}{\sigma^2}.$$

The intuition behind this variance ratio is simple: when it is small, the variance of the noise dominates that of $\hat{x}_{t-1}$ and $x_t$ is dominated by the noise term; when it is large, $\hat{x}_{t-1}$ dominates the noise and $x_t$ can be accurately predicted on average.

### 3.2.2.2 Multivariate case

Suppose $n > 1$, and consider now the univariate process $(y^T x_t)_t$ with weights $y \in \mathbb{R}^n$. Using (3.2 ), we know that $y^T x_t = y^T \hat{x}_{t-1} + y^T \varepsilon_t$, and we can measure its predicability as

$$\lambda(y) \stackrel{\text{def}}{=} \frac{y^T \hat{\mathcal{A}}_0 y}{y^T \mathcal{A}_0 y}, \tag{3.3}$$

where $\hat{\mathcal{A}}_0$ and $\mathcal{A}_0$ are the covariance matrices of $x_t$ and $\hat{x}_{t-1}$, respectively. Minimizing predictability $\lambda(y)$ is then equivalent to finding the minimum generalized eigenvalue $\lambda$ solving

$$\det(\lambda\mathcal{A}_0 - \hat{\mathcal{A}}_0) = 0. \tag{3.4}$$

Assuming that $\mathcal{A}_0$ is positive definite, the basket with minimum predictability will be given by $y = \mathcal{A}_0^{-1/2}y_0$, where $y_0$ is the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathcal{A}_0^{-1/2}\hat{\mathcal{A}}_0\mathcal{A}_0^{-1/2}$.

#### 3.2.2.3 Estimation of $\lambda(y)$

All of the quantities used to define $\lambda$ above need to be estimated from sample paths. $\mathcal{A}_0$ can be estimated by $A_0$ following Equation (3.1). All other quantities depend on the predictor $\hat{x}_{t-1}$. Box and Tiao assume that $x_t$ follows a vector autoregressive model of order $p$ –VAR(p), in short –and therefore $\hat{x}_{t-1}$ takes the form,

$$\hat{x}_{t-1} = \sum_{k=1}^{p} \mathcal{H}_k x_{t-k},$$

where the $p$ matrices $(\mathcal{H}_k)$ contain each $n \times n$ autoregressive coefficient. Estimating $\mathcal{H}_k$ from the sample path $\mathbf{x}$, Box and Tiao solve for the optimal basket by inserting these estimates in the generalized eigenvalue problem displayed in Equation (3.4). If one assumes that $p = 1$ (the case $p > 1$ can be trivially reformulated as a VAR(1) model with adequate reparameterization), then

$$\hat{\mathcal{A}}_0 = \mathcal{H}_1\mathcal{A}_0\mathcal{H}_1^T \text{ and } \mathcal{A}_1 = \mathcal{A}_0\mathcal{H}_1,$$

and thus the Yule–Walker estimator (Lütkepohl, 2005, §3.3) of $\mathcal{H}_1$ would be $H_1 = A_0^{-1}A_1$. Minimizing predictability boils down to solving in that case

$$\min_y \hat{\lambda}(y), \hat{\lambda}(y) \overset{\text{def}}{=} \frac{y^T(H_1A_0H_1^T)y}{y^TA_0y} = \frac{y^T(A_1A_0^{-1}A_1^T)y}{y^TA_0y},$$

which is equivalent to computing the smallest eigenvector of the matrix $A_0^{-1/2}A_1A_0^{-1}A_1^TA_0^{-1/2}$ if the covariance matrix $A_0$ is invertible.

The machinery of Box and Tiao to quantify mean reversion requires defining a model to form $\hat{x}_{t-1}$, the conditional expectation of $x_t$ given previous observations. We consider now two criteria that do without such modeling assumptions.

### *3.2.3 Portmanteau Criterion*

Recall that the *portmanteau* statistic of order $p$ (Ljung and Box 1978) of a centered univariate stationary process $x$ (with $n = 1$) is given by

$$\mathrm{por}_p(x) = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\mathbf{E}[x_tx_{t+i}]}{\mathbf{E}[x_t^2]}\right)^2$$

where $\mathbf{E}[x_t x_{t+i}]/\mathbf{E}[x_t^2]$ is the $i$th-order autocorrelation of $x_t$. The portmanteau statistic of a white noise process is by definition 0 for any $p$. Given a multivariate ($n > 1$) process $x$, we write

$$\phi_p(y) = \mathrm{por}_p(y^T x) = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{y^T \mathcal{A}_i y}{y^T \mathcal{A}_0 y}\right)^2,$$

for a coefficient vector $y \in \mathbb{R}^n$. By construction, $\phi_p(y) = \phi_p(ty)$ for any $t \neq 0$, and in what follows, we will impose $\|y\|_2 = 1$. The quantities $\phi_p(y)$ are computed using the following estimates (Hamilton, 1994, p. 110):

$$\hat{\phi}_p(y) = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{y^T A_i y}{y^T A_0 y}\right)^2. \tag{3.5}$$

### *3.2.4 Crossing Statistics*

Kedem and Yakowitz (1994, §4.1) define the *zero crossing rate* of a univariate ($n = 1$) process $x$ (its expected number of crosses around 0 per unit of time) as

$$\gamma(x) = \mathbf{E}\left[\frac{\sum_{t=2}^{T} \mathbf{1}_{\{x_t x_{t-1} \leq 0\}}}{T-1}\right]. \tag{3.6}$$

A result known as the cosine formula states that if $x_t$ is an autoregressive process of order 1 (AR(1)), namely if $|a| < 1$, $\varepsilon_t$ is i.i.d. standard Gaussian noise and $x_t = a x_{t-1} + \varepsilon_t$, then (Kedem and Yakowitz, 1994, §4.2.2):

$$\gamma(x) = \frac{\arccos(a)}{\pi}.$$

Hence, for AR(1) processes, minimizing the first-order autocorrelation $a$ also directly maximizes the crossing rate of the process $x$. For $n > 1$, since the first-order autocorrelation of $y^T x_t$ is equal to $y^T \mathcal{A}_1 y$, we propose to minimize $y^T \mathcal{A}_1 y$ and ensure that all other absolute autocorrelations $|y^T \mathcal{A}_k y|$, $k > 1$ are small.

## 3.3 Optimal Baskets

Given a centered multivariate process $\mathbf{x}$, we form its covariance matrix $A_0$ and its $p$ autocovariances $(A_1, \cdots, A_p)$. Because $y^T A y = y^T (A + A^T) y/2$, we symmetrize all autocovariance matrices $A_i$. We investigate in this section the problem of estimating baskets that have maximal mean reversion (as measured by the proxies proposed in Section 3.2), while being at the same time sufficiently volatile and supported by as few assets as possible. The latter will be achieved by selecting portfolios $y$ that have a small “0-norm,” namely that the number of nonzero components in $y$,

$$\|y\|_0 \stackrel{\text{def}}{=} \#\{1 \leq i \leq d | y_i \neq 0\},$$

is small. The former will be achieved by selecting portfolios whose aggregated value exhibits a variance over time that exceeds a given threshold $\nu > 0$. Note that for the variance of $(y^T x_t)$ to exceed a level $\nu$, the largest eigenvalue of $A_0$ must necessarily be larger than $\nu$, which we always assume in what follows. Combining these two constraints, we propose three different mathematical programs that reflect these tradeoffs.

### 3.3.1 Minimizing Predictability

Minimizing Box–Tiao's predictability $\hat{\lambda}$ defined in Section 3.2.2, while ensuring that both the variance of the resulting process exceeds $\nu$ and the vector of loadings is sparse with a 0-norm equal to $k$, means solving the following program:

$$\begin{array}{ll} \text{minimize} & y^T M y \\ \text{subject to} & y^T A_0 y \geq \nu, \\ & \|y\|_2 = 1, \\ & \|y\|_0 = k, \end{array} \tag{P1}$$

in the variable $y \in \mathbb{R}^n$ with $M \stackrel{\text{def}}{=} A_1 A_0^{-1} A_1^T$, where $M, A_0 \in \mathbf{S}_n$. Without the normalization constraint $\|y\|_2 = 1$ and the sparsity constraint $\|y\|_0 = k$, problem (P1) is equivalent to a generalized eigenvalue problem in the pair $(M, A_0)$. That problem quickly becomes unstable when $A_0$ is ill-conditioned or $M$ is singular. Adding the normalization constraint $\|y\|_2 = 1$ solves these numerical problems.

### 3.3.2 Minimizing the Portmanteau Statistic

Using a similar formulation, we can also minimize the order $p$ portmanteau statistic defined in Section 3.2.3 while ensuring a minimal variance level $\nu$ by solving:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} (y^T A_i y)^2 \\ \text{subject to} & y^T A_0 y \geq \nu, \\ & \|y\|_2 = 1, \\ & \|y\|_0 = k, \end{array} \tag{P2}$$

in the variable $y \in \mathbb{R}^n$, for some parameter $\nu > 0$. Problem (P2) has a natural interpretation: the objective function directly minimizes the portmanteau statistic, while the constraints normalize the norm of the basket weights to one, impose a variance larger than $\nu$, and impose a sparsity constraint on $y$.

### 3.3.3 Minimizing the Crossing Statistic

Following the results in Section 3.2.4, maximizing the crossing rate while keeping the rest of the autocorrelogram low,

$$\begin{array}{ll} \text{minimize} & y^T A_1 y + \mu \sum_{k=2}^{p} (y^T A_k y)^2 \\ \text{subject to} & y^T A_0 y \geq \nu, \\ & \|y\|_2 = 1, \\ & \|y\|_0 = k, \end{array} \tag{P3}$$

in the variable $y \in \mathbb{R}^n$, for some parameters $\mu, \nu > 0$. This will produce processes that are close to being AR(1) while having a high crossing rate.

## 3.4 Semidefinite Relaxations and Sparse Components

Problems (P1), (P2), and (P3) are not convex and can be in practice extremely difficult to solve, since they involve a sparse selection of variables. We detail in this section convex relaxations to these problems that can be used to derive relevant suboptimal solutions.

### *3.4.1 A Semidefinite Programming Approach to Basket Estimation*

We propose to relax problems (P1), (P2), and (P3) into semidefinite programs (SDPs) (Vandenberghe and Boyd, 1996). We show that these SDPs can handle naturally sparsity and volatility constraints while still aiming at mean reversion. In some restricted cases, one can show that these relaxations are tight, in the sense that they solve exactly the programs described above. In such cases, the true solution $y^\star$ of some of the programs above can be recovered using their corresponding SDP solution $Y^\star$.

However, in most of the cases we will be interested in, such a correspondence is not guaranteed, and these SDP relaxations can only serve as a guide to propose solutions to these hard nonconvex problems when considered with respect to vector $y$. To do so, the optimal solution $Y^\star$ needs to be *deflated* from a large rank $d \times d$ matrix to a rank one matrix $yy^T$, where $y$ can be considered a good candidate for basket weights. A typical approach to deflate a positive definite matrix into a vector is to consider its eigenvector with the leading eigenvalue. Having sparsity constraints in mind, we propose to apply a heuristic grounded on sparse PCA (d'Aspremont *et al.*, 2007; Zou *et al.*, 2006). Instead of considering the lead eigenvector, we recover the leading *sparse* eigenvector of $Y^\star$ (with a 0-norm constrained to be equal to $k$). Several efficient algorithmic approaches have been proposed to solve approximately that problem; we use the SpaSM (sparse statistiscal modeling) toolbox (Sjöstrand *et al.*, 2012) in our experiments.

### *3.4.2 Predictability*

We can form a convex relaxation of the predictability optimization problem (P1) over the variable $y \in \mathbb{R}^n$,

$$\begin{array}{ll} \text{minimize} & y^T M y \\ \text{subject to} & y^T A_0 y \geq \nu \\ & \|y\|_2 = 1, \\ & \|y\|_0 = k, \end{array}$$

by using the lifting argument of Lovász and Schrijver (1991), (i.e., writing $Y = yy^T$) to solve now the problem using a semidefinite variable $Y$, and by introducing a sparsity-inducing regularizer on $Y$ that considers the $L_1$ norm of $Y$,

$$\|Y\|_1 \overset{\text{def}}{=} \sum_{ij} |Y_{ij}|,$$

so that Problem (P1) becomes (here $\rho > 0$),

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(MY) + \rho\|Y\|_1 \\ \text{subject to} & \mathbf{Tr}(A_0 Y) \geq \nu \\ & \mathbf{Tr}(Y) = 1, \mathbf{Rank}(Y) = 1, Y \succeq 0. \end{array}$$

We relax this last problem further by dropping the rank constraint, to get

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(MY) + \rho\|Y\|_1 \\ \text{subject to} & \mathbf{Tr}(A_0 Y) \geq \nu \\ & \mathbf{Tr}(Y) = 1, Y \succeq 0 \end{array} \tag{SDP1}$$

which is a convex semidefinite program in $Y \in \mathbf{S}_n$.

### 3.4.3 Portmanteau

Using the same lifting argument and writing $Y = yy^T$, we can relax problem (P2) by solving

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} \mathbf{Tr}(A_i Y)^2 + \rho\|Y\|_1 \\ \text{subject to} & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}(Y) = 1, Y \succeq 0, \end{array} \tag{SDP2}$$

a semidefinite program in $Y \in \mathbf{S}_n$.

### 3.4.4 Crossing Stats

As above, we can write a semidefinite relaxation for problem (P3):

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(A_1 Y) + \mu \sum_{i=2}^{p} \mathbf{Tr}(A_i Y)^2 + \rho\|Y\|_1 \\ \text{subject to} & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}(Y) = 1, Y \succeq 0. \end{array} \tag{SDP3}$$

#### 3.4.4.1 Tightness of the SDP Relaxation in the Absence of Sparsity Constraints

Note that for the crossing stats criterion (with $p = 1$ and no quadratic term in $Y$), the original problem P3 and its relaxation SDP3 are equivalent, taking for granted that no sparsity constraint is considered in the original problems and $\mu$ is set to 0 in the relaxations. These relaxations boil down to an SDP that only has a linear objective, a linear constraint, and a constraint on the trace of $Y$. In that case, Brickman (1961) showed that the range of two quadratic forms over the unit sphere is a convex set when the ambient dimension $n \geq 3$, which means in particular that for any two square matrices $A, B$ of dimension $n$,

$$\begin{aligned} &\{(y^T Ay, y^T By) : y \in \mathbb{R}^n, \|y\|_2 = 1\} \\ &\quad = \{(\mathbf{Tr}(AY), \mathbf{Tr}(BY)) : Y \in \mathbf{S}_n, \mathbf{Tr} Y = 1, Y \succeq 0\}. \end{aligned}$$

We refer the reader to Barvinok 2002 (§II.13) for a more complete discussion of this result. As remarked in Cuturi and d'Aspremont (2013), the same equivalence holds for P1 and SDP1. This means that, in the case where $\rho, \mu = 0$ and the 0-norm of $y$ is *not* constrained, for any solution $Y^\star$ of the relaxation (SDP1) there exists a vector $y^\star$ that satisfies $\|y\|_2^2 = \mathbf{Tr}(Y^\star) = 1$, $y^{\star T} A_0 y^\star = \mathbf{Tr}(BY^\star)$, and $y^{\star T} M y^\star = \mathbf{Tr}(MY^\star)$, which means that $y^\star$ is an optimal solution of the original problem (P1). Boyd and Vandenberghe (2004, App. B) show how to explicitly

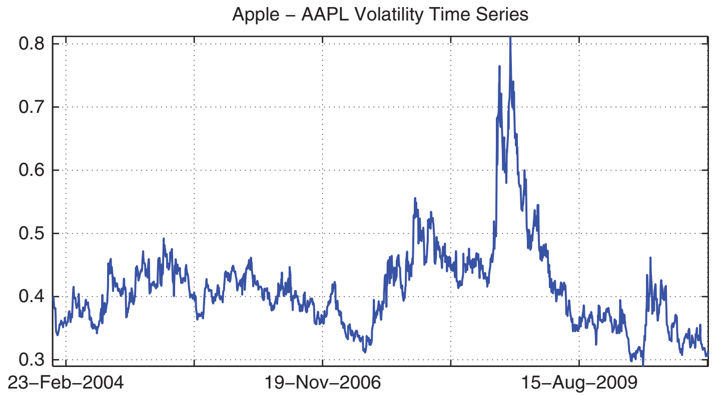


**Figure 3.1** Option implied volatility for Apple between January 4, 2004, and December 30, 2010.

extract such a solution $y^\star$ from a matrix $Y^\star$ solving (SDP1). This result is, however, mostly anecdotical in the context of this chapter, in which we look for sparse and volatile baskets: using these two regularizers breaks the tightness result between the original problems in $\mathbb{R}^d$ and their SDP counterparts.

## 3.5 Numerical Experiments

In this section, we evaluate the ability of our techniques to extract mean-reverting baskets with sufficient variance and small 0-norm from a universe of tradable assets. We measure performance by applying to these baskets a trading strategy designed specifically for mean-reverting processes. We show that, under realistic trading costs assumptions, selecting sparse and volatile mean-reverting baskets translates into lower incurred costs and thus improves the performance of trading strategies.

### *3.5.1 Historical Data*

We consider daily time series of option implied volatilities for 210 stocks from January 4, 2004, to December 30, 2010. A key advantage of using option implied volatility data is that these numbers vary in a somewhat limited range. Volatility also tends to exhibit regime switching, and hence can be considered piecewise stationary, which helps in extracting structural relationships. We plot a sample time series from this dataset in Figure 3.1 that corresponds to the implicit volatility of Apple's stock. In what follows, we mean by *asset* the implied volatility of any of these stocks whose value can be efficiently replicated using option portfolios.

### 3.5.2 Mean-reverting Basket Estimators

We compare the three basket selection techniques detailed here, *predictability, portmanteau*, and *crossing statistic*, implemented with varying targets for both sparsity and volatility, with two cointegration estimators that build upon PCA (Maddala and Kim, 1998, §5.5.4).. By the label PCA, we mean in what follows the eigenvector with the smallest eigenvalue of the covariance matrix $A_0$ of the process (Stock and Watson, 1988). By sPCA, we mean the sparse eigenvector of $A_0$ with 0-norm $k$ that has the smallest eigenvalue, which can be simply estimated by computing the leading sparse eigenvector of $\lambda I - A_0$ where $\lambda$ is bigger than the leading eigenvalue of $A_0$. This sparse principal component of the covariance matrix $A_0$ should not be confused with our utilization of sparse PCA in Section 3.4.1 as a way to recover a vector solution from the solution of a positive semidefinite problem. Note also that techniques based on principal components do not take explicitly variance levels into account when estimating the weights of a co-integrated relationship.

### 3.5.3 Jurek and Yang (2007) Trading Strategy

While option implied volatility is not directly tradable, it can be synthesized using baskets of call options, and we assimilate it to a tradable asset with (significant) transaction costs in what follows. For baskets of volatilities isolated by the techniques listed above, we apply the Jurek and Yang (2007) strategy for log utilities to the basket process recording out of sample performance. Jurek and Yang proposed to trade a stationary autoregressive process $(x_t)_t$ of order 1 and mean $\mu$ governed by the equation $x_{t+1} = \rho x_t + \sigma \varepsilon_t$, where $|\rho| < 1$, by taking a position $N_t$ in the asset $x_t$, which is proportional to

$$N_t = \frac{\rho(\mu - x_t)}{\sigma^2} W_t \tag{3.7}$$

In effect, the strategy advocates taking a long (resp. short) position in the asset whenever it is below (resp. above) its long-term mean, and adjust the position size to account for the volatility of $x_t$ and its mean reversion speed $\rho$. Given basket weights $y$, we apply standard AR estimation procedures on the in-sample portion of $y^T \mathbf{x}$ to recover estimates for $\hat{\rho}$ and $\hat{\sigma}$ and plug them directly in Equation (3.7). This approach is illustrated for two baskets in Figure 3.2.

### 3.5.4 Transaction Costs

We assume that fixed transaction costs are negligible, but that transaction costs per contract unit are incurred at each trading date. We vary the size of these costs across experiments to show the robustness of the approaches tested here to trading costs fluctuations. We let the transaction cost per contract unit vary between 0.03 and 0.17 cents by increments of 0.02 cents. Since the average value of a contract over our dataset is about 40 cents, this is akin to considering trading costs ranging from about 7 to about 40 base points (BPs), that is, 0.07 to 0.4%.

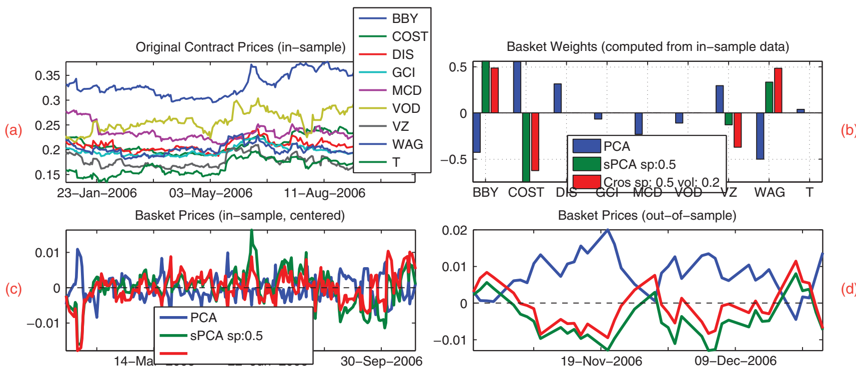


**Figure 3.2** Three sample trading experiments, using the PCA, sparse PCA, and crossing statistics estimators. (a) Pool of 9 volatility time series selected using our fast PCA selection procedure. (b) Basket weights estimated with in-sample data using the eigenvector of the covariance matrix with the smallest eigenvalue, the smallest eigenvector with a sparsity constraint of $k = \lfloor 0.5 \times 9 \rfloor = 4$, and the crossing statistics estimator with a volatility threshold of $\nu = 0.2$, (i.e., a constraint on the basket's variance to be larger than $0.2\times$ the median variance of all 8 assets). (c) Using these 3 procedures, the time series of the resulting basket price in the in-sample part (c) and out-of-sample parts (d) are displayed. (e) Using the Jurek and Yang (2007) trading strategy results in varying positions (expressed as units of baskets) during the out-sample testing phase. (f) Transaction costs that result from trading the assets to achieve such positions accumulate over time. (g) Taking both trading gains and transaction costs into account, the net wealth of the investor for each strategy can be computed (the Sharpe ratio over the test period is displayed in the legend). Note how both sparsity and volatility constraints translate into portfolios composed of fewer assets, but with a higher variance.

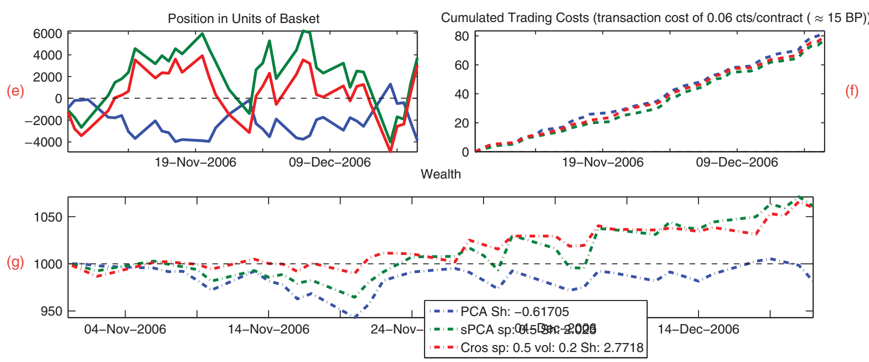


**Figure 3.2** (*continued*)

### 3.5.5 Experimental Setup

We consider 20 sliding windows of one year (255 trading days) taken in the history, and consider each of these windows independently. Each window is split between 85% of days to estimate and 15% of days to test-trade our models, resulting in 38 test-trading days. We do not recompute the weights of the baskets during the test phase. The 210 stock volatilities (assets) we consider are grouped into 13 subgroups, depending on the economic sector of their stock. This results in 13 sector pools whose size varies between 3 assets and 43 assets. We look for mean-reverting baskets in each of these 13 sector pools.

Because all combinations of stocks in each of the 13 sector pools may not necessarily be mean-reverting, we select smaller candidate pools of $n$ assets through a greedy backward-forward minimization scheme, where $8 \leq n \leq 12$. To do so, we start with an exhaustive search of all pools of size 3 within the sector pool, and proceed by adding or removing an asset using the PCA estimator (the smallest eigenvalue of the covariance matrix of a set of assets). We use the PCA estimator in that backward-forward search because it is the fastest to compute. We score each pool using that PCA statistic, the smaller meaning the better. We generate up to 200 candidate pools per each of the 13 sector pools. Out of all these candidate pools, we keep the best 50 in each window, and then use our cointegration estimation approaches separately on these candidates. One such pool was, for instance, composed of the stocks $\{\texttt{BBY}, \texttt{COST}, \texttt{DIS}, \texttt{GCI}, \texttt{MCD}, \texttt{VOD}, \texttt{VZ}, \texttt{WAG}, \texttt{T}\}$ observed during the year 2006. Figure 3.2 provides a closeup on that universe of stocks, and shows the results of three trading experiments using PCA, sparse PCA, or the Crossing Stats estimator to build trading strategies.

### 3.5.6 Results

#### 3.5.6.1 Robustness of Sharpe Ratios to Costs

In Figure 3.3, we plot the average of the Sharpe ratio over the 922 baskets estimated in our experimental set versus transaction costs. We consider different PCA settings as well as our three estimators using, in all three cases, the variance bound $\nu$ to be 0.3 times the median of all variances of assets available in a given asset pool, and the 0-norm to be equal to 0.3 times the size of the universe (itself between 8 and 12). We observe that Sharpe ratios decrease the fastest for the naive PCA-based method, this decrease being somewhat mitigated when adding a constraint on the 0-norm of the basket weights obtained with sparse PCA. Our methods require, in addition to sparsity, enough volatily to secure sufficient gains. These empirical observations agree with the intuition of this chapter: simple cointegration techniques can produce synthetic baskets with high mean reversion, large support, and low variance. Trading a portfolio with low variance that is supported by multiple assets translates in practice into high trading costs, which can damage the overall performance of the strategy. Both sparse PCA and our techniques manage instead to achieve a tradeoff between desirable mean reversion properties and, at the same time, control for sufficient variance and small basket size to allow for lower overall transaction costs.

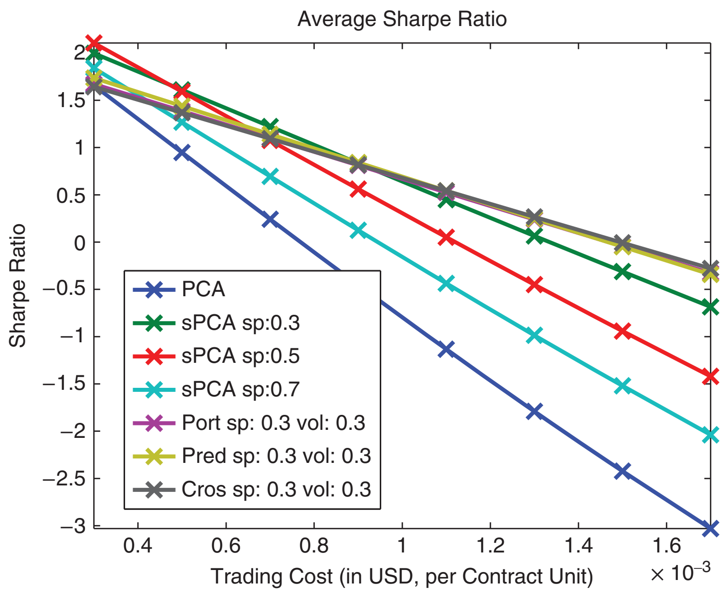


**Figure 3.3** Average Sharpe ratio for the Jurek and Yang (2007) trading strategy captured over about 922 trading episodes, using different basket estimation approaches. These 922 trading episodes were obtained by considering 7 disjoint time-windows in our market sample, each of a length of about one year. Each time-window was divided into 85% in-sample data to estimate baskets, and 15% outsample to test strategies. On each time-window, the set of 210 tradable assets during that period was clustered using sectorial information, and each cluster screened (in the in-sample part of the time-window) to look for the most promising baskets of size between 8 and 12 in terms of mean reversion, by choosing greedily subsets of stocks that exhibited the smallest minimal eigenvalues in their covariance matrices. For each trading episode, the same universe of stocks was fed to different mean-reversion algorithms. Because volatility time-series are bounded and quite stationary, we consider the PCA approach, which uses the eigenvector with the smallest eigenvalue of the covariance matrix of the time-series to define a cointegrated relationship. Besides standard PCA, we have also consider sparse PCA eigenvectors with minimal eigenvalue, with the size $k$ of the support of the eigenvector (the size of the resulting basket) constrained to be 30%, 50% or 70% of the total number of considered assets. We consider also the portmanteau, predictability and crossing stats estimation techniques with variance thresholds of $\nu = 0.2$ and a support whose size $k$ (the number of assets effectively traded) is targeted to be about 30% of the size of the considered universe (itself between 8 and 12). As can be seen in the figure, the sharpe ratios of all trading approaches decrease with an increase in transaction costs. One expects sparse baskets to perform better under the assumption that costs are high, and this is indeed observed here. Because the relationship between sharpe ratios and transaction costs can be efficiently summarized as being a linear one, we propose in the plots displayed in Figure 3.4 a way to summarize the lines above with two numbers each: their intercept (Sharpe level in the quasi-absence of costs) and slope (degradation of Sharpe as costs increase). This visualization is useful to observe how sparsity (basket size) and volatility thresholds influence the robustness to costs of the strategies we propose. This visualization allows us to observe how performance is influenced by these parameter settings.

#### 3.5.6.2 Tradeoffs between Mean Reversion, Sparsity, and Volatility

In the plots of Figure 3.4, this analysis is further detailed by considering various settings for $v$ (volatility threshold) and $k$. To improve the legibility of these results, we summarize, following the observation in Figure 3.3 that the relationship between Sharpes and transactions costs seems almost linear, each of these curves by two numbers: an intercept level (Sharpe ratio when costs are low) and a slope (degradation of Sharpe as costs increase). Using these two

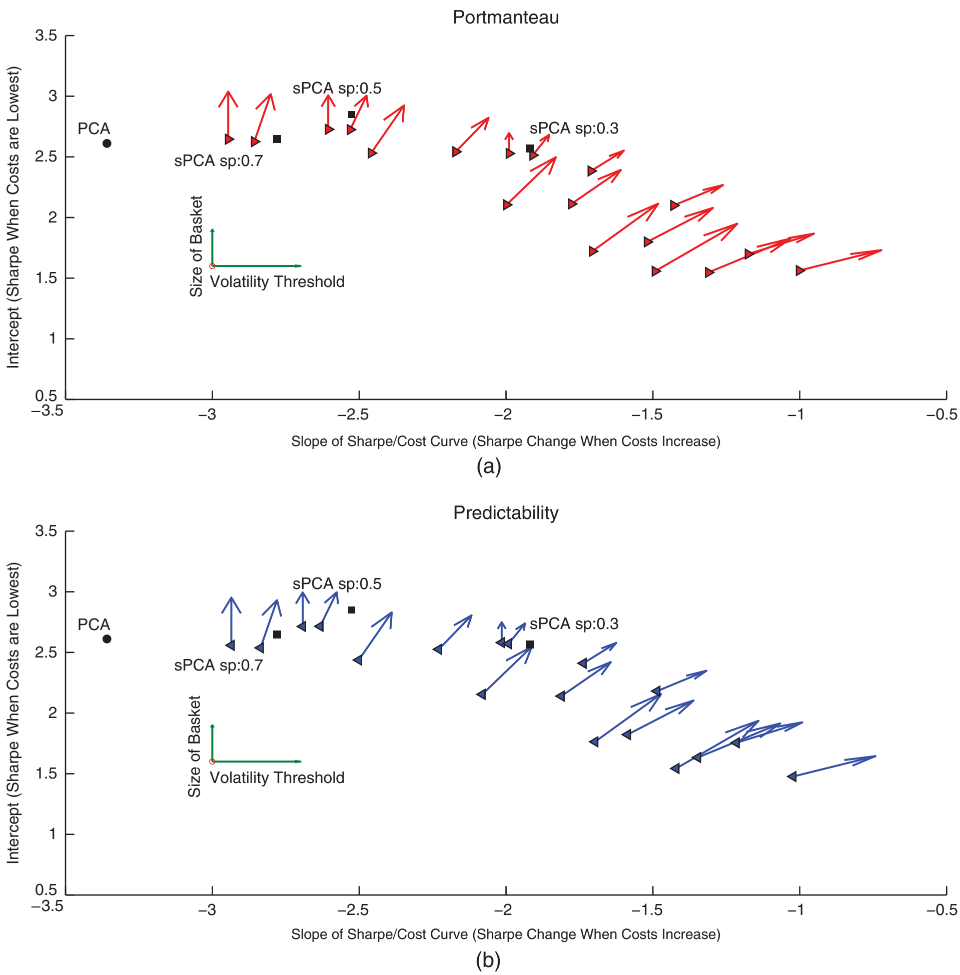


**Figure 3.4** Relationships between Sharpe in a low cost setting (intercept) in the $x$-axis and robustness of Sharpe to costs (slope of Sharpe/costs curve) of a different estimators implemented with varying volatility levels $v$ and sparsity levels $k$ parameterized as a multiple of the universe size. Each colored square in the figures above corresponds to the performance of a given estimator (Portmanteau in subfigure ($a$), Predictability in subfigure ($b$) and Crossing Statistics in subfigure ($c$)) using different parameters for $v \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and $u \in \{0.3, 0.5, 0.7\}$. The parameters used for each experiment are displayed using an arrow whose vertical length is proportional to $v$ and horizontal length is proportional to $u$.

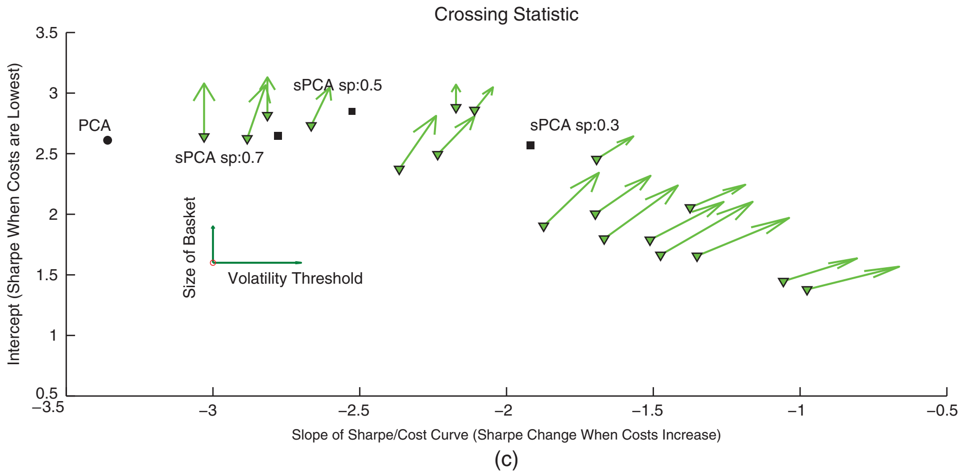


**Figure 3.4** (*continued*)

numbers, we locate all considered strategies in the intercept–slope plane. We first show the spectral techniques, PCA and sPCA, with different levels of sparsity, meaning that $k$ is set to $\lfloor u \times d \rfloor$, where $u \in \{0.3, 0.5, 0.7\}$ and $d$ is the size of the original basket. Each of the three estimators we propose is studied in a separate plot. For each, we present various results characterized by two numbers: a volatility threshold $v \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ and a sparsity level $u \in \{0.3, 0.5, 0.7\}$. To avoid cumbersome labels, we attach an arrow to each point: the arrow's length in the vertical direction is equal to $u$ and characterizes the size of the basket, and the horizontal length is equal to $v$ and characterizes the volatility level. As can be seen in these three plots, an interesting interplay between these two factors allows for a continuum of strategies that trade mean reversion (and thus Sharpe levels) for robustness to cost level.

## 3.6 Conclusion

We have described three different criteria to quantify the amount of mean reversion in a time series. For each of these criteria, we have detailed a tractable algorithm to isolate a vector of weights that has optimal mean reversion, while constraining both the variance (or signal strength) of the resulting univariate series to be above a certain level and its 0-norm to be at a certain level. We show that these bounds on variance and support size, together with our new criteria for mean reversion, can significantly improve the performance of mean reversion statistical arbitrage strategies and provide useful controls to adjust mean-reverting strategies to varying trading conditions, notably liquidity risk and cost environment.

## References

Barvinok, A. (2002). *A course in convexity*. Providence, RI: American Mathematical Society.

Bewley, R., Orden, D., Yang, M. and Fisher, L. (1994). Comparison of Box-Tiao and Johansen canonical estimators of cointegrating vectors in VEC (1). Models. *Journal of Econometrics*, 64, 3–27.

Box, G. and Tiao, G. (1977). A canonical analysis of multiple time series. *Biometrika*, 64(2), 355–365.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.

Brickman, L. (1961). On the field of values of a matrix. *Proceedings of the American Mathematical Society*, 12, 61–66.

Cuturi, M. and d'Aspremont, A. (2013). Mean reversion with a variance threshold. In *Proceedings of the International Conference in Machine Learning 2013*, June, Atlanta, GA.

d'Aspremont, A. (2011). Identifying small mean reverting portfolios. *Quantitative Finance*, 11(3), 351–364.

d'Aspremont, A., El Ghaoui, L., Jordan, M.I. and Lanckriet, G.R. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review*, 49(3), 434–448.

Elie, R., and Espinosa, G.E. (2011). Optimal stopping of a mean reverting diffusion: minimizing the relative distance to the maximum. hal-00573429.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251–276.

Hamilton, J. (1994). *Time series analysis*, Vol. 2. Cambridge: Cambridge University Press.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6), 1551–1580.

Johansen, S. (2004). Cointegration: overview and development. In *Handbook of financial time series* (ed. Andersen, T.G., Davis, R.A., Kreiß, J.P. and Mikosch, T.V.). Berlin: Springer.

Johansen, S. (2005). *Cointegration: a survey*. Palgrave Handbook of Econometrics. London: Palgrave.

Jurek, J.W. and Yang, H. (2007). Dynamic portfolio selection in arbitrage. EFA 2006 Meetings Paper. SSRN eLibrary, http://dx.doi.org/10.2139/ssrn.882536

Kedem, B. and Yakowitz, S. (1994). *Time series analysis by higher order crossings*. Piscataway, NJ: IEEE Press.

Liu, J. and Timmermann, A. (2010). *Optimal arbitrage strategies*. Technical report. San Diego: University of California, San Diego.

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303.

Lovász, L. and Schrijver, A. (1991). Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1(2), 166–190.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.

Maddala, G. and Kim, I. (1998). *Unit roots, cointegration, and structural change*. Cambridge: Cambridge University Press.

Sjöstrand, K., Clemmensen, L.H., Larsen, R. and Ersbøll, B. (2012). SpaSM: A MATLAB toolbox for sparse statistical modeling. *Journal of Statistical Software*, accepted.

Stock, J. and Watson, M. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83, 1097–1107.

Tsay, R.S. (2005). *Analysis of financial time series*. Weinheim: Wiley-Interscience.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM Review*, 38(1), 49–95.

Ylvisaker, N.D. (1965). The expected number of zeros of a stationary Gaussian process. *The Annals of Mathematical Statistics*, 36(3), 1043–1046.

Zou, H., Hastie, T. and Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2), 265–286.