Mean Reversion with a Variance Threshold

Marco Cuturi MCUTURI@I.KYOTO-U.AC.JP
Graduate School of Informatics, Kyoto University

Alexandre d'Aspremont ALEXANDRE.DASPREMONT@M4X.ORG
CNRS - Ecole Polytechnique

Abstract

Starting from a sample path of a multivariate stochastic process, we study several techniques to isolate linear combinations of the variables with a maximal amount of mean reversion, while constraining the variance of the combination to be larger than a given threshold. We show that many of the optimization problems arising in this context can be solved exactly using semidefinite programming and a variant of the $\mathcal{S}$-lemma. In finance, these methods can be used to isolate statistical arbitrage opportunities, *i.e.* mean reverting baskets with enough variance to overcome market friction. In a more general setting, mean reversion and its generalizations can also be used as a proxy for stationarity, while variance simply measures signal strength.

1. Introduction

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.

Here, we use financial applications as the main testing ground for these techniques. Stationary processes typically exhibit significant mean-reversion, *i.e.* a tendency to pull back to their mean, and mean reversion creates statistical arbitrage opportunities: a simple strategy of buying the asset below the mean and selling it short above will produce positive returns on average. Of course, stationary assets are not common in financial markets, assets that exhibit fast mean reversion even less so, hence arbitrageurs often resort to creating synthetic assets that have this property. Such synthetic assets are usually long-short *baskets* that are built by combining positive (long) and negative (short) positions in different liquid assets. The size of each of these positions (vector y in the previous paragraph) is usually computed using cointegration methods.

Since the original work of Engle & Granger (1987), several techniques have been proposed to form cointegrated baskets under various modeling assumptions and we refer the reader to (Maddala & Kim, 1998) and (Johansen, 2005) for a more complete survey. Financial applications (Tsay, 2005, §8.6) and in particular optimal trading strategies for mean reverting baskets were discussed in (Jurek & Yang, 2007; Liu & Timmermann, 2010; Elie & Espinosa, 2011), while the problem of isolating sparse mean reverting baskets was discussed in (d'Aspremont) using a criterion derived in (Box & Tiao, 1977).

Mean-reverting strategies cannot, however, only rely on mean-reversion to be profitable. Arbitrage opportunities can only exist if they are large enough to be traded without using too much leverage or incurring too much transaction costs. For mean-reverting baskets, this condition translates naturally into the 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. Here, we argue that classical cointegration techniques are ill-suited for identifying statistical arbitrage opportunities because they focus exclusively on mean-reversion without considering variance. In contrast, the methods we develop here maximize a proxy

Appearing in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013.

for mean reversion, while constraining variance to be higher than a certain threshold.

We use three different criteria as proxies for mean reversion: *predictability*; the *portmanteau* statistic; the *crossing* statistic. To our knowledge, the latter two criteria were never considered before as criteria to estimate cointegrated relationships, and for a good reason: the problem of optimizing these criteria over normalized basket weights is nonconvex. We show however that these problems can be efficiently approximated by semidefinite programs. These relaxations are exact in some of the settings detailed below, and explicit tightness or uniform approximation results related to the $\mathcal{S}$-lemma (Ben-Tal et al., 2009) control the quality of the solutions.

The paper is organized as follows. We focus first in Section 2 on various measures and proxies for mean reversion. Section 3 defines the basket optimization problems corresponding these quantities. We show in Section 4 that each of these problems translate naturally into semidefinite relaxations which produce either exact or approximate solutions. Section 5 briefly summarizes the complexity of solving the resulting semidefinite programs. Finally, we present numerical evidence in Section 6.

2. Criteria and estimators

Throughout this paper, 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 $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ and each $\mathbf{x}_t \in \mathbb{R}^n$, we write A_k for the *empirical* counterpart of $\mathcal{A}_k$ computed from $\mathbf{x}$,

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

Given $y \in \mathbb{R}^n$, we now define three measures which can all be interpreted as proxies for the mean reversion of $y^T x_t$. **Predictability** – defined for stationary processes by Box & Tiao (1977) and generalized for non-stationary processes by Bewley et al. (1994) – measures how close to noise the series is. The **portmanteau** statistic (Ljung & 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.

2.1. Predictability

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

$$x_t = \hat{x}_{t-1} + \varepsilon_t, \quad (2)$$

where $\hat{x}_{t-1}$ is a predictor of x_t built upon past values of the process recorded up to $t-1$, and ε_t is a vector of 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 & Tiao, 1977) starts as follows.

Univariate case. Suppose $n = 1$ and thus $\Sigma \in \mathbb{R}_+$, Equation (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 σ^2 and $\hat{\sigma}^2$ of x_t and $\hat{x}_t$ respectively. Box & 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 almost pure noise, when it is large however, $\hat{x}_{t-1}$ dominates the noise and x_t is almost perfectly predictable.

Multivariate case. Suppose $n > 1$ and consider now the projected process $(y^T x_t)_t$ with weights $y \in \mathbb{R}^n$. Using (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

$$\lambda(y) \stackrel{\text{def}}{=} \frac{y^T \hat{\mathcal{A}}_0 y}{y^T \mathcal{A}_0 y}, \quad (3)$$

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 λ solving

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

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 $\mathcal{A}_0^{-1/2} \hat{\mathcal{A}}_0 \mathcal{A}_0^{-1/2}$.

Estimation of $\lambda(y)$. All of the quantities used to define λ above need to be estimated from sample paths. $\mathcal{A}_0$ can be estimated by A_0 following Equation (1). All other quantities depend on the predictor $\hat{x}_{t-1}$. Box & Tiao assume that x_t follows a vector autoregressive

model of order p – VAR(p) in short notation – 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 coefficients. Estimating $\mathcal{H}_k$ from the sample path $\mathbf{x}$, Box & Tiao solve for the optimal basket by inserting these estimates in the generalized eigenvalue problem above. 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

$$\min_y \hat{\lambda}(y),\ \hat{\lambda}(y) \stackrel{\text{def}}{=} \frac{y^T \left(H_1 A_0 H_1^T\right) y}{y^T A_0 y} = \frac{y^T \left(A_1 A_0^{-1} A_1^T\right) y}{y^T A_0 y},$$

which is equivalent to computing the smallest eigenvector of the matrix $A_0^{-1/2} A_1 A_0^{-1} A_1^T A_0^{-1/2}$.

The machinery of Box & 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 in the following two criteria that do without such modeling assumptions.

2.2. Portmanteau criterion

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

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

where $\mathbf{E}[x_t x_{t+i}]/\mathbf{E}[x_t^2]$ is the ith 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

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

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):

$$\hat{\phi}_p(y) = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{y^T A_i y}{y^T A_0 y} \right)^2 \tag{4}$$

2.3. Crossing statistics

Kedem & 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

$$\gamma(x) = \mathbf{E}\left[\frac{\sum_{t=2}^{T} \mathbb{1}_{\{x_t x_{t-1} \leq 0\}}}{T-1} \right], \tag{5}$$

A result known as the cosine formula states that if x_t is an autoregressive process of order one AR(1), namely if $|a| < 1$, ε_t is i.i.d. standard Gaussian noise and

$$x_t = a x_{t-1} + \varepsilon_t$$

then (Kedem & 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. Optimal baskets

Given a centered multivariate process $\mathbf{x}$, we form its covariance matrix A_0 and its p autocovariances $(A_1, \ldots, A_p)$. Because $y^T A y = y^T (A + A^T) y/2$, we can replace the matrices A_i by their symmetric part. We focus in this section on baskets that exhibit both mean reversion and sufficient volatility, that is have a variance that exceeds a given threshold $\nu > 0$. Note that for the variance of $(y^T x_t)$ to exceed a level ν, the largest eigenvalue of A_0 must necessarily be larger than ν, which we always assume in what follows. To highlight the central role of the covariance matrix A_0, we rename it to $B \stackrel{\text{def}}{=} A_0$ in the rest of the paper.

3.1. Minimizing predictability

Minimizing Box-Tiao's predictability $\hat{\lambda}$ defined in §2.1 while ensuring that the variance of the resulting process exceeds ν, means solving the following QCQP:

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

in the variable $y \in \mathbb{R}^n$ with $M \stackrel{\text{def}}{=} A_1 B^{-1} A_1^T$ and $B = A_0$, where $M, B \in \mathbf{S}_n$. Without the normalization constraint $\|y\|_2 = 1$, problem (P1) is equivalent to a generalized eigenvalue problem in the pair (M, B).

That problem quickly becomes unstable when B is ill-conditioned or M is singular. Adding the normalization constraint $\|y\|_2 = 1$ resolves those problems yet does not affect the relaxation results that follow in §4.

3.2. Minimizing the portmanteau statistic

Using a similar formulation, we can also minimize the order p portmanteau statistic defined in §2.2 while ensuring a minimal variance level ν by solving:

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

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 and impose a variance larger than ν. We will see in what follows that (P2) can be solved exactly using a semidefinite relaxation when $p = 1$. Also, while solving (P2) exactly is hard when $p > 1$, semidefinite relaxations produce tractable solutions with uniform approximation bounds.

3.3. Minimizing the crossing statistic

Following the results in §2.3, maximizing the crossing rate while keeping the rest of the autocorrelogram low,

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

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

4. Semidefinite relaxations

In this section, we detail convex relaxations to the problems detailed above in Section §3.

4.1. Exact solutions for predictability

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

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

using the lifting argument of (Lovász & Schrijver, 1991), i.e. writing $Y = yy^T$, so (P1) becomes

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

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

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

which is a semidefinite program in $Y \in \mathbf{S}_n$. We call $Y^\star$ the optimum solution to this problem. By construction, the optimal value of (SDP1) is an upper bound on that of (P1). Here however, the two problems are in fact equivalent. 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

$$\begin{gathered} \left\{(y^T A y, y^T B y) : y \in \mathbb{R}^n,\ \|y\|_2 = 1\right\} = \\ \{(\mathbf{Tr}(AY), \mathbf{Tr}(BY)) : Y \in \mathbf{S}_n,\ \mathbf{Tr}\, Y = 1,\ Y \succeq 0\} \end{gathered}$$

We refer the reader to (Barvinok, 2002, §II.13) for a more complete discussion of this result. This means that for any solution $Y^\star$ of the relaxation (SDP1) there exists a vector $y^\star$ which satisfies $\|y\|_2^2 = \mathbf{Tr}(Y^\star) = 1$, $y^{\star T} B y^\star = \mathbf{Tr}(BY^\star)$ 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 & Vandenberghe (2004, App. B) show how to explicitly extract such a solution $y^\star$ from a matrix $Y^\star$ solving (SDP1). We detail this in §5.

4.2. Portmanteau: exact solution when $p = 1$

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

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

a semidefinite program in $Y \in \mathbf{S}_n$. When $p = 1$ the objective becomes $|\,\mathbf{Tr}(A_1 Y)|$ and this program can be further simplified to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \mathbf{Tr}(A_1 Y) \leq t \\ & \mathbf{Tr}(A_1 Y) \geq -t \\ & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}\, Y = 1, Y \in \mathbf{S}_n \end{array} \tag{SDP2bis}$$

which is a semidefinite program in the variables $Y \in \mathbf{S}_n$ and $t \in \mathbb{R}$. Using Brickman's theorem as above, we can recover a vector $y^\star$ which solves (P2) from an optimal solution $Y^\star$ to problem (SDP2bis). When $p > 1$ this tightness result does not hold and we detail uniform approximation results in what follows.

4.3. Portmanteau: approximations for $p > 1$

Program (P2) is a nonconvex quadratically constrained quadratic program which is hard to solve exactly. However, randomization arguments (Nemirovski et al., 1999; Nemirovski, 2007; So, 2009; Ben-Tal et al., 2009) show that if we call OPT the solution of the original problem in (P2) and SDP the solution to its relaxation in (SDP2), we have uniform approximation bounds with

$$\text{SDP} \leq \text{OPT} \leq \text{SDP}\ c \log p$$

where $c > 0$ is an absolute constant. These results are constructive and good approximate solutions to (P2) can be constructed from optimal solutions to (SDP2) using randomization.

4.4. Approximate solutions for crossing stats

As above, we can write a semidefinite relaxation for problem (P3), and solve the semidefinite program

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

The same randomization arguments show that this relaxation produces solutions which are suboptimal by a factor at most $c \log p$ (except when $p = 1$ where it is exact, since the right-hand side term weighted by μ disappears).

5. Algorithms

In this section, we detail how to efficiently solve the semidefinite relaxations (SDP1) and (SDP2).

5.1. Predictability

We can form the dual of problem (SDP1) by writing the Lagrangian

$$\begin{aligned} L(Y, w) &= \mathbf{Tr}(MY) + w\,\mathbf{Tr}(BY) - w\nu \\ &= \mathbf{Tr}(Y(M + wB)) - w\nu \end{aligned}$$

in the variables $Y \in \mathbf{S}_n$ and $w \in \mathbb{R}$. Minimizing this Lagrangian over the set $\{Y : \mathbf{Tr}\, Y = 1, Y \succeq 0\}$ and using

$$\lambda_{\min}(M + wB) = \min_{\mathbf{Tr}\, Y = 1,\, Y \succeq 0} \mathbf{Tr}(Y(M + wB))$$

means the dual of problem (SDP1) is written

$$\text{maximize } \lambda_{\min}(V + wC) - w\nu \quad (6)$$

in the variable $w \in \mathbb{R}$. The Karush-Kuhn-Tucker (KKT) optimality conditions (Boyd & Vandenberghe, 2004, §5.9.2) for this pair of problems are then given by

$$\left\{ \begin{array}{l} (M + wB)Y = \lambda_{\min}(M + wB)Y \\ \mathbf{Tr}(BY) \geq \nu,\ \mathbf{Tr}\, Y = 1,\ Y \succeq 0, \end{array} \right. \quad (7)$$

Two scenarios arise depending on the multiplicity of $\lambda_{\min}(M + wB)$:

- *Nondegenerate case.* Suppose that $\lambda_{\min}(M+wB)$ is simple at the optimal w, the KKT conditions mean that Y must be rank one, hence can be written $Y = yy^T$ for some $y \in \mathbb{R}$. As above, this y is then an optimal solution to problem (P1).
- *Degenerate case.* Let $Y \in \mathbf{S}_n$ and $w \in \mathbb{R}$ be optimal solutions to (SDP1) and (6) respectively. If the eigenspace associated with $\lambda_{\min}(M+wM)$ has dimension $k > 1$, we let $U \in \mathbb{R}^{n \times k}$ be an orthonormal basis of that subspace. Writing $Y = UWU^T$ for some $W \in \mathbf{S}_k$ with $W \succeq 0$ yields

  $$\mathbf{Tr}(U^T AUW) \geq \nu \quad \text{and} \quad \mathbf{Tr}(W) = 1. \quad (8)$$

  The procedure in (Boyd & Vandenberghe, 2004, App.B) shows how to construct a vector $y \in \mathbb{R}^k$ such that

  $$y^T U^T AUy \geq \nu \quad \text{and} \quad \|y\|_2 = 1,$$

  in at most k explicit steps. The vector $Uy \in \mathbb{R}^n$ then solves (P1) since $\|Uy\|_2 = 1$, $y^T U^T AUy \geq \nu$ and $y^T U^T (M + wB)Uy = \lambda_{\min}(M + wB)$ by construction.

The result above shows that in the nondegenerate case (typical here), it suffices to solve the minimum eigenvalue maximization problem in (6) to get a solution to (P1). We'll see below that the complexity of solving that program is in fact very low. In the degenerate case, if k is the multiplicity of the maximum eigenvalue $\lambda_{\min}(M + wB)$ at the optimal w, we need to compute the matrix U defined above at a cost of $O(kn^2)$ and solve the semidefinite feasibility problem in (8) to find Y (Ben-Tal & Nemirovski, 2001, §6.6.3).

Proposition 1. *Let $\epsilon > 0$ be a target precision, the complexity of solving the dual problem in (6)*

$$\textit{maximize } \lambda_{\min}(M + wB) - w\nu$$

in the variable $w \in \mathbb{R}$, grows as $O(n^2 \log_2(1/\epsilon))$.

Proof. The function $\lambda_{\min}(M + wB)$ is convex in w, hence we can minimize it by bisection. At each iteration, forming a gradient amounts to computing a leading eigenvector using iterative algorithms such as the power or Lanczos methods (see Golub & Van Loan (1996, Chap. 8-9) for example), at a cost of $O(n^2)$. Reaching a target precision ϵ then requires $O(n^2 \log_2(1/\epsilon))$ flops. ■

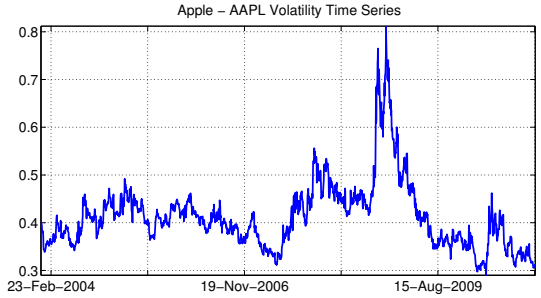

Figure 1. **Option implied volatility** for Apple between January 4 2004 and December 30 2010.

5.2. Portmanteau statistic

We first compute a dual of problem (SDP2) by rewriting the problem as

$$\max_{\substack{\|y\|_2\leq 1,\\ y\in\mathbb{R}^p, w\in\mathbb{R}}} \ \min_{\substack{\mathbf{Tr}\, Y=1,\\ Y\succeq 0}} \mathbf{Tr}\left(Y\left(\sum_{k=1}^{p} y_k A_k + wC\right)\right) - w\nu$$

because this is a convex saddle-point problem where one of the feasible sets is compact, we can get a dual by switching the min and the max, to get

$$\max_{\|y\|_2\leq 1} \ \lambda_{\min}\left(\sum_{k=1}^{p} y_k A_k + wC\right) - w\nu \qquad (9)$$

in the variables $y \in \mathbb{R}^p$ and $w \in \mathbb{R}$. Using a smoothing argument, Nesterov (2007) showed that given a bound μ on the Euclidean norm of the solution, the complexity of solving (9) using a first-order method grows as

$$\frac{\mu\|A\|n^3\sqrt{\log n}}{\epsilon}$$

where ϵ is the target precision and

$$\|A\| = \max_{\|(y,w)\|=1} \left\|\sum_{k=1}^{p} y_k A_k + wC\right\|_2$$

is computed from the autocovariance operators. A similar bound holds for (SDP3).

6. Numerical experiments

In this section, we evaluate the ability of our techniques to extract mean-reverting baskets with sufficient variance from tradeable 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 mean-reverting baskets with sufficient variance translates into lower incurred costs and thus improves the performance of trading strategies.

6.1. 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, hence can be considered piecewise stationary, which helps in extracting structural relationships. We illustrate a sample time series from this dataset in Figure 1 corresponding to Apple's stock.

6.2. Mean-reverting Basket Estimators

We compare the three basket selection techniques detailed here – **predictability**, **portmanteau** and **crossing statistic** ($p = 3$) – with three classical cointegration estimators: that which arises from the **Johansen** VEC model (Johansen, 1991), orthogonal least-squares (**OLS**) estimation – equivalent to selecting the eigenvector with the smallest eigenvalue of the variance matrix A_0 (Maddala & Kim, 1998, §6.7.1) – and the fully modified OLS (**FM-OLS**) procedure described by Phillips (1995). None of these classical techniques takes into account variance when estimating the weights of a co-integrated relationship.

6.3. Jurek & 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 & Yang, 2007) strategy for log utilities to the basket process recording out of sample performance. Jurek & Yang propose to trade a stationary autoregressive process $(x_t)_t$ of order 1 and mean μ 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

$$N_t = \frac{\rho(\mu - x_t)}{\sigma^2} W_t \qquad (10)$$

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 ρ. 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 (10). This approach is illustrated for two baskets in Figure 2.

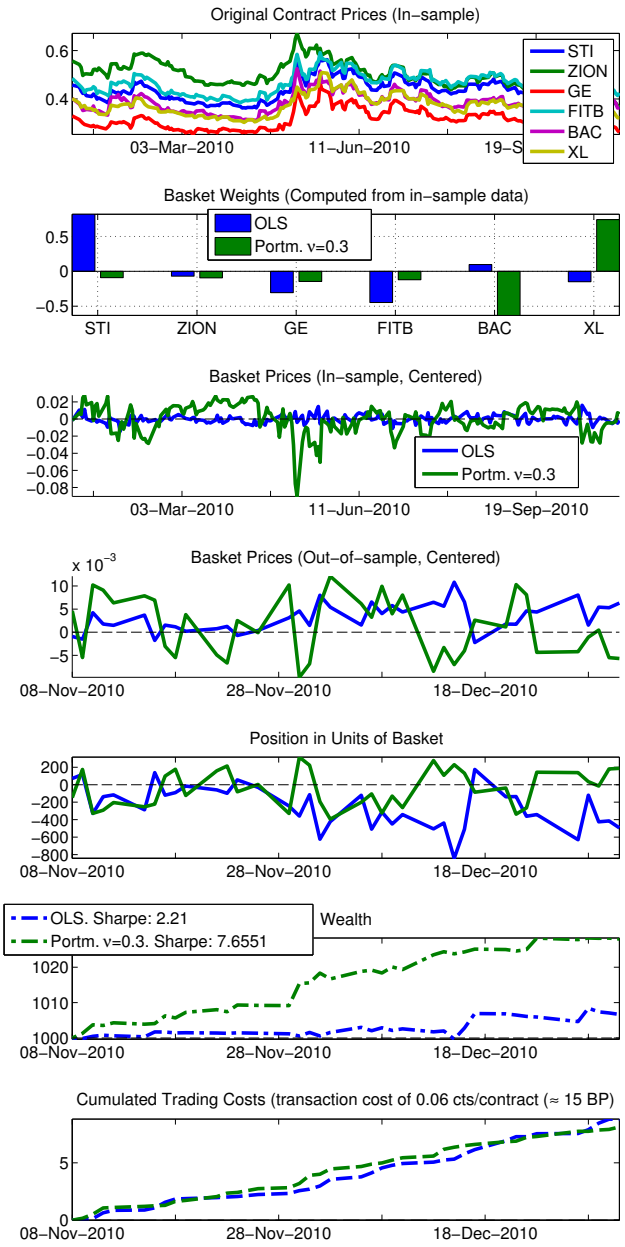

Figure 2. **Two sample trading experiments, using either the OLS or Portmanteau estimator** (From top to bottom) Pool of 6 consumer finance related volatility time-series; Basket weights estimated with in-sample data using either least-squares or the Portmanteau estimator with $\nu = 0.3$, *i.e.*a constraint on the basket's variance to be larger than $0.3\times$ the median variance of all 6 assets; basket prices in-sample; basket prices out-of-sample; trading position in units of baskets; cumulative wealth and Sharpe of the strategy; cumulative transaction costs. Note how the threshold ν forces the estimation of a co-integrated basket with higher variance than what would have been found otherwise by classical cointegration techniques. This situation translates into a more volatile basket that requires less leverage to achieve a comparable performance.

6.4. Transaction Costs

We assume that fixed transaction costs are negligible, but that transaction costs per contract unit are incurred at each trading date, varying the size of these costs across experiments. We let the transaction cost per contract unit vary between 0 and 0.14 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 0 to 35 Base Points (BP), that is 0 to 0.35%.

6.5. 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. All 210 stocks are divided into 13 different groups depending on their economic sector, resulting in 13 asset pools whose size varies between 3 assets and 43 assets. Because all combinations of stocks are not necessarily mean-reverting, we select smaller candidate pools of n assets through a greedy backward-forward selection scheme, where $2 \leq n \leq 8$. We score each pool by the mean-reversion speed obtained, the faster the better. We use the OLS estimator for this computational intensive phase because it is the fastest to compute. We keep the best 50 asset pools of each window, and use each technique separately on these candidates. One such pool was, for instance, composed of the stocks {`STI,ZION,GE,FITB,BAC,XL`} history in 2010. Figure 2 examines this pool in detail, and shows the results of two trading experiments.

6.6. Results

In Figures 3 and 5, we plot the average of the Sharpe ratio and the total return (computed during the 38 days trading period) over the $20 \times 50 = 1,000$ baskets estimated in our experimental set versus transaction costs. In all cases, we have set the variance bound ν to be 0.3 times the median of all variances of assets available in a given asset pool. In both figures, we observe that returns and Sharpe ratio decrease faster for the three classical cointegration methods than for the three techniques detailed here. These empirical observations agree with the intuition of this paper: cointegration techniques can produce synthetic baskets with high mean-reversion but low variance. Trading an asset with low variance translates in practice into high trading costs and thus badly performing trading strategies. The three techniques detailed in this paper

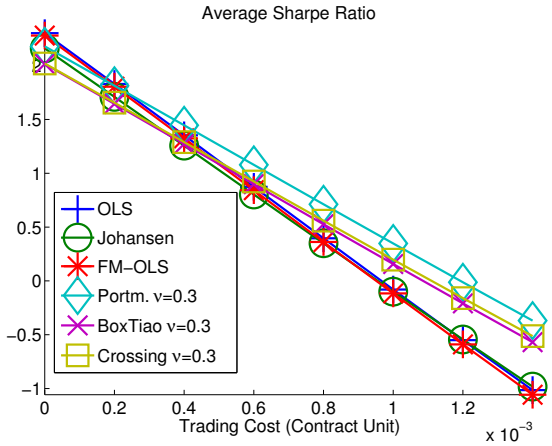

Figure 3. **Average Sharpe ratio vs. Transaction Costs.** The Sharpe ratios are computed during the out-of-sample trade experiment, using the Jurek & Yang (2007) trading strategy using 50 pre-selected pools of assets during each of the 20 trading periods. Transaction costs range from 0 cents/contract to 0.14 cents per contract, *i.e.* approximately 0.35% or 35 basis points.

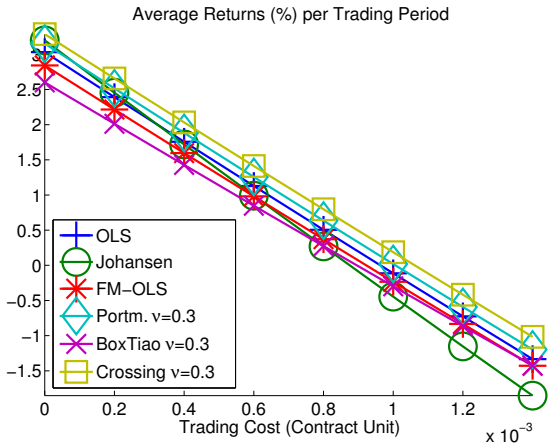

Figure 4. **Average returns vs. Transaction Costs** using the same setup considered in Figure 3

manage instead to achieve a trade-off between desirable mean-reversion properties with sufficient variance to allow for lower overall transaction costs. Finally, the bell-shaped curves of Figure 5 show the importance of setting a variance threshold ν within a reasonable range as trading costs increase. Indeed, in a typical trading environment (where costs are between 10 or 20 BP), Figure 5 shows that trading off some mean-reversion to gain variance instead is needed to remain profitable.

7. 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

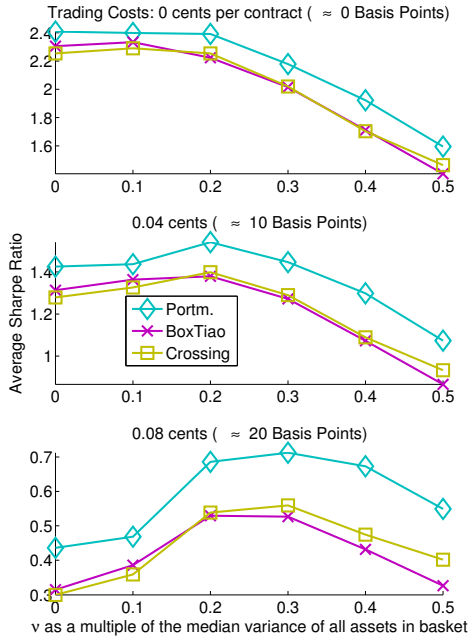

Figure 5. **Sharpe ratios vs. ν variance threshold for different trading costs scenarios.** In the absence of market frictions and free trades (above), choosing a high threshold is irrelevant. Whenever costs rise to reasonable levels, setting a variance threshold is critical for mean-reverting strategies to achieve satisfactory performance, as illustrated by the bell-shaped curves of the Sharpe ratio (middle, below)

to isolate a basket with optimal mean reversion from a multivariate sample path of an asset process, while constraining the variance (or signal strength) of the resulting basket to be above a certain level. We show that this bound on variance, together with our new criteria for mean reversion can significantly improve the performance of statistical arbitrage strategies.

Acknowledgements

We thank anonymous reviewers for their comments. MC acknowledges the support of the Japanese Society for the Promotion of Science grant 23700172 and AA acknowledges the support of the European Research Council (starting grant SIPA).

References

Barvinok, A. *A course in convexity.* American Mathematical Society, 2002.

Ben-Tal, A. and Nemirovski, A. *Lectures on modern convex optimization.* SIAM, 2001.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A.S. *Robust optimization.* Princeton University Press, 2009.

Bewley, R., Orden, D., Yang, M., and Fisher, L.A. Comparison of Box-Tiao and Johansen Canonical Estimators

of Cointegrating Vectors in VEC (1) Models. *Journal of Econometrics*, 64:3–27, 1994.

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

Boyd, Stephen and Vandenberghe, Lieven. *Convex Optimization*. Cambridge University Press, 2004.

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

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

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

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

Golub, G.H. and Van Loan, C.F. *Matrix computations*. Johns Hopkins Univ Pr, 1996.

Hamilton, J.D. *Time series analysis*, volume 2. Cambridge Univ Press, 1994.

Johansen, S. Cointegration: a survey. *Palgrave Handbook of Econometrics*, 1, 2005.

Johansen, Soren. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–80, November 1991.

Jurek, Jakub W. and Yang, Halla. Dynamic Portfolio Selection in Arbitrage. *SSRN eLibrary*, 2007. doi: 10.2139/ssrn.882536.

Kedem, B. and Yakowitz, S. *Time series analysis by higher order crossings*. IEEE press Piscataway, NJ, 1994.

Liu, J. and Timmermann, A. Optimal arbitrage strategies. Technical report, UC San Diego Working Paper, 2010.

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

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

Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

Maddala, GS and Kim, I.M. *Unit roots, cointegration, and structural change*. Cambridge Univ Pr, 1998.

Nemirovski, A. Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Mathematical programming*, 109(2):283–317, 2007.

Nemirovski, A., Roos, C., and Terlaky, T. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.

Nesterov, Y. Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming*, 110(2):245–259, 2007.

Phillips, P.C.B. Fully modified least squares and vector autoregression. *Econometrica: Journal of the Econometric Society*, pp. 1023–1078, 1995.

So, A.M.C. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical Programming*, pp. 2, 2009.

Tsay, Ruey S. *Analysis of financial time series*, volume 543. Wiley-Interscience, 2005.

Ylvisaker, N Donald. The expected number of zeros of a stationary gaussian process. *The Annals of Mathematical Statistics*, pp. 1043–1046, 1965.