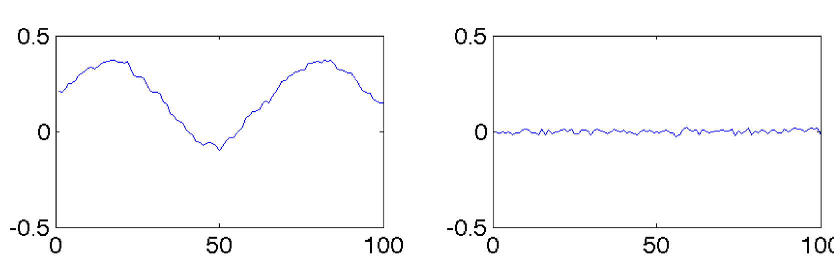
Mean Reversion with a Variance Threshold

Marco Cuturi[1], Alexandre d'Aspremont[2]

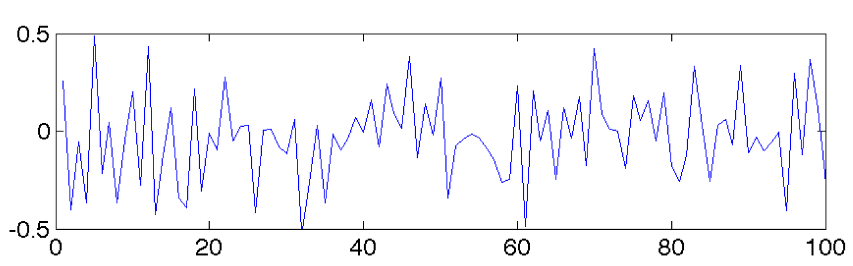
[1] Kyoto University (Japan), [2] CNRS - Ecole Polytechnique (France)

This Poster in 1 Minute

Classic Cointegration Problem: Given a multivariate time series $\mathbf{x}_t \in \mathbb{R}^n$, find α such that $y^T x_t$ is **stationary**.



Our formulation: Find cointegrated relationship such that $y^T x_t$ ***also*** has **fast mean-reversion** and **sufficient variance**.



Motivation: Makes a lot of sense in **financial applications**. We expect it can also be applied to other fields, such as anomaly detection.

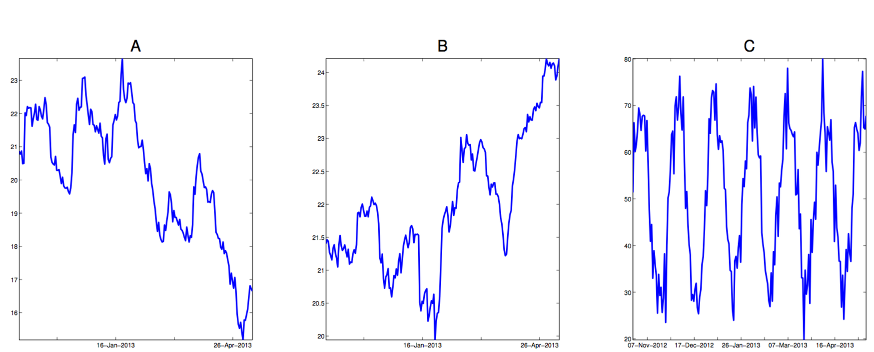
Approach: Formulate natural criteria that take into account **both** mean-reversion and variance.

Optimization: These criteria are **not convex**. We approximate them (and solve them exactly in some cases) using semidefinite programming and the $\mathcal{S}$-lemma.

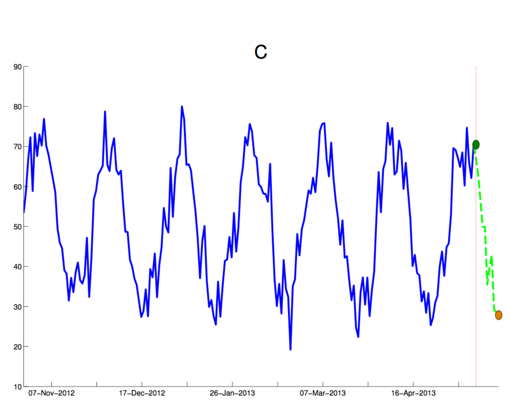
Experiments: We illustrate that, on stock volatility data, **Mean-reversion** $\rightarrow$ statistical arbitrage opportunities. **Sufficient variance** $\rightarrow$ lower transaction costs.

1. Mean-Reversion & Cointegration

Loose Definition of Mean Reversion: **Tendency of a stochastic process to revert (pull back) to its mean.**



Mean-reversion = Statistical Arbitrage Opportunity

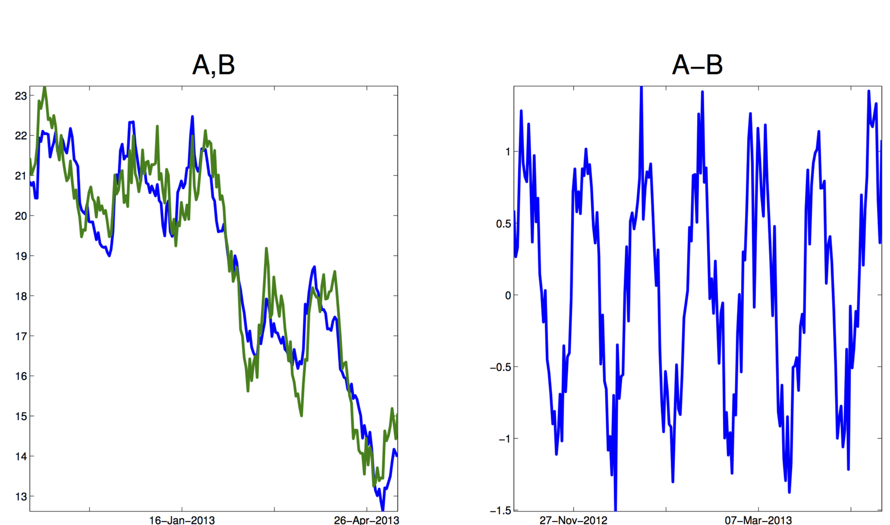


Which assets are mean-reverting?

- **Stationary** processes are mean-reverting,
- Arbitraging stationary assets is therefore desirable.

In practice, very few assets are stationary. Those who are tend to revert to their means very slowly.

However, **combining** assets can result in stationarity: *"pair-trades"* when $n = 2$, *"baskets"* when $n \geq 3$.



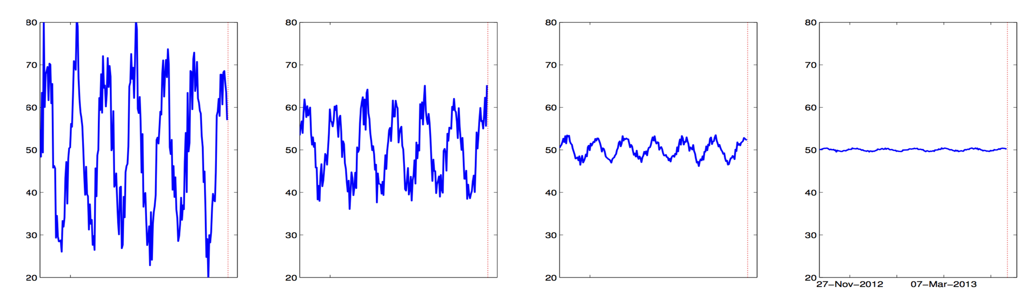
Finding weights y such that $y^T x_t$ is stationary
=
cointegration theory (econometrics, VAR modeling)

2. Stationarity is not enough

Problem 1: slow mean-reversion is bad.

Slow mean-reversion $\rightarrow$ smaller expected arbitrage.

Problem 2: small variance is bad.



Large variance $\rightarrow$ larger arbitrage expected per trade.

Both problems lead to more leverage = higher risk.

Both issues are **not** addressed by classic **cointegration methods**, which focus exclusively on **stationarity**.

When $n \gg 1$, when estimating y from finite samples, small variance can also mean overfitting.

3. Criteria

- $\mathcal{A}_k = \mathbf{E}[x_t x_{t+k}^T], k \geq 0$ when finite.
- When $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ and each $\mathbf{x}_t \in \mathbb{R}^n$,

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

To **quantify mean-reversion**, three different proxies:

1. Portmanteau (Ljung and Box, 1978)

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

(norm of autocorrelogram)

2. Crossing Stats (Kedem and Yakowitz, 1994)
Number of times $y^T x_t$ crosses its mean is a decreasing function of $y^T \mathcal{A}_1 y$ assuming $y^T \mathcal{A}_k y \approx 0, k > 1$.

3. Predictability (Box and Tiao, 1977) Suppose

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

where $\hat{x}_{t-1}$ is a predictor of x_t; ε_t i.i.d. Gaussian $(0, \Sigma)$.

n=1: $\mathbf{E}[x_t^2] = \mathbf{E}[\hat{x}_{t-1}^2] + \mathbf{E}[\varepsilon_t^2]$, thus $1 = \frac{\hat{\sigma}^2}{\sigma^2} + \frac{\Sigma}{\sigma^2}$,
Box and Tiao measure the *predictability* of x_t by the ratio

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

n>1: Consider the process $(y^T x_t)_t$ with $y \in \mathbb{R}^n$. We can measure the predicability of $y^T x_t$ as

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

where $\hat{\mathcal{A}}_0$ and $\mathcal{A}_0$ are covariance matrices of x_t and $\hat{x}_{t-1}$.

To **Quantify variance**

$$\mathbf{var}(y^T x_t) = y^T \mathcal{A}_0 y > \nu.$$

Mean Reversion with Variance Threshold

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} \left(y^T A_i y\right)^2 \\ \text{subject to} & y^T A_0 y \geq \nu \\ & \|y\|_2 = 1, \end{array} \qquad \text{(P1)}$$

$$\begin{array}{ll} \text{minimize} & y^T A_1 y + \mu \sum_{k=2}^{p} \left(y^T A_k y\right)^2 \\ \text{subject to} & y^T A_0 y \geq \nu \\ & \|y\|_2 = 1, \end{array} \qquad \text{(P2)}$$

$$\begin{array}{ll} \text{minimize} & y^T M y \\ \text{subject to} & y^T A_0 y \geq \nu \\ & \|y\|_2 = 1, \end{array} \qquad \text{(P3)}$$

4. SDP Relaxations

SDP Formulation Writing $Y = yy^T$,

Brickman (1961): when $n \geq 3$,

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

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{p} \mathbf{Tr}(A_i Y)^2 \\ \text{subject to} & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}(Y) = 1,\ Y \succeq 0, \end{array} \qquad \text{(SDP1)}$$

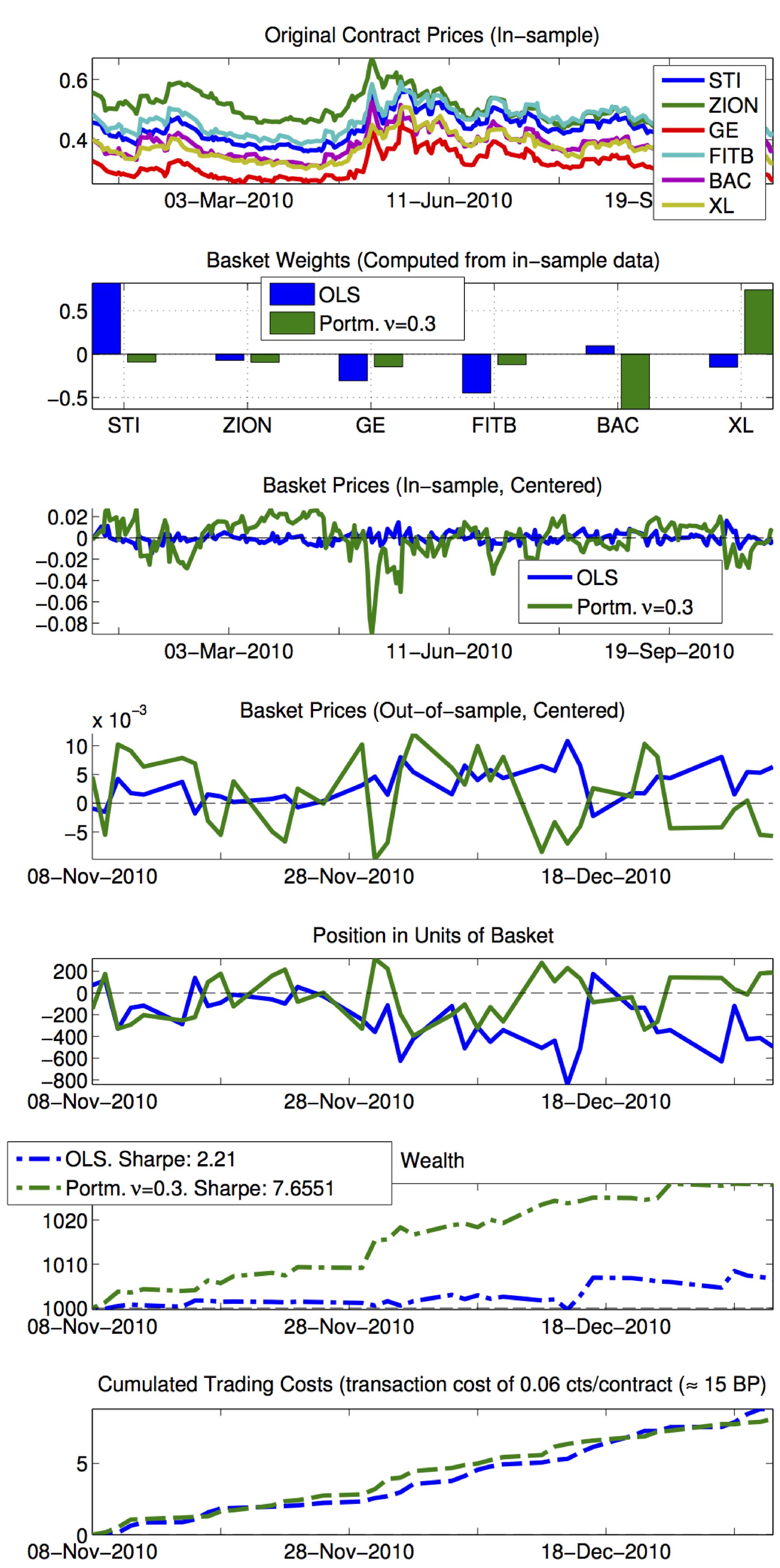
$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(A_1 Y) + \mu \sum_{i=2}^{p} \mathbf{Tr}(A_i Y)^2 \\ \text{subject to} & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}(Y) = 1,\ Y \succeq 0 \end{array} \qquad \text{(SDP2)}$$

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}(MY) \\ \text{subject to} & \mathbf{Tr}(BY) \geq \nu \\ & \mathbf{Tr}(Y) = 1,\ Y \succeq 0, \end{array} \qquad \text{(SDP3)}$$

Exact solutions when $p = 1$, **approximation** (randomization, leading eigenvector) when $p > 1$.

5. Experiments

Data: implied volatility data for 217 stocks.
Sample Trade Episode: using our approach and OLS



Results: 20 time windows, results on most