Identifying Small Mean Reverting Portfolios

Alexandre d'Aspremont
ORFE, Princeton University

Support from NSF, DHS and Google.

Introduction

Mean reversion:

- Classic case of statistical arbitrage.
- Highlights long-term structural relationships in the data.
- We could replace mean-reversion by momentum throughout the talk.

Sparse portfolios:

- Better interpretability.
- Less transaction costs.

Mean reversion

- Let S_{ti} be the value at time t of an asset S_i for $i = 1, \ldots, n$ and $t = 1, \ldots, m$.
- We form portfolios P_t of these assets with coeffiicients x_i, modeled by an Ornstein-Uhlenbeck process:

$$dP_t = \lambda(\bar{P} - P_t)dt + \sigma dZ_t \quad \text{with } P_t = \sum_{i=1}^{n} x_i S_{ti}$$

where Z_t is a standard Brownian motion.

- **Objective:** maximize the mean reversion coefficient λ of P_t by adjusting the coefficients x, while imposing $\|x\| = 1$ and $\mathbf{Card}(x) \leq k$.

Outline

- **Canonical decomposition**
- Sparse generalized eigenvalue problems
- Estimation and trading
- Numerical results

Canonical decomposition

- In a discrete setting, we assume that the asset prices follow a (stationary) autoregressive process with:

$$S_t = AS_{t-1} + Z_t \tag{1}$$

  where S_{t-1} is the lagged portfolio process, $A \in \mathbf{R}^{n\times n}$ and Z_t is a vector of i.i.d. Gaussian noise with zero mean and covariance $\Sigma \in \mathbf{S}^n$, independent of S_{t-1}.

- Take $n = 1$ in equation (1):

$$\mathbf{E}[S_t^2] = \mathbf{E}[(AS_{t-1})^2] + \mathbf{E}\left[Z_t^2\right]$$

  which can be rewritten as $\sigma_t^2 = \sigma_{t-1}^2 + \Sigma$.

- Box & Tiao (1977) then measure the **predictability** of stationary series by:

$$\lambda = \frac{\sigma_{t-1}^2}{\sigma_t^2}. \tag{2}$$

Canonical decomposition

- Consider a portfolio $P_t = x^T S_t$ with $x \in \mathbf{R}^n$, using (1) we know that

$$x^T S_t = x^T A S_{t-1} + x^T Z_t,$$

so its predicability can be measured as:

$$\lambda_x = \frac{x^T A \Gamma A^T x}{x^T \Gamma x}$$

where $\Gamma = \mathbf{E}[SS^T]$.

- The portfolio with maximum (respectively minimum) predictability will be the eigenvector corresponding to the largest (respectively smallest) eigenvalue of the matrix:

$$\Gamma^{-1} A \Gamma A^T. \tag{3}$$

- We then only need to estimate A...

Canonical decompositions

- The **Box-Tiao procedure** finds linear combinations of the assets ranked in order of predictability by computing the eigenvectors of the matrix:

$$\left(S^T S\right)^{-1}\left(\hat{S}_t^T \hat{S}_t\right) \tag{4}$$

where is $\hat{S}_t$ is the least squares estimate computed above.

- The **Johansen procedure:** following Bewley, Orden, Yang & Fisher (1994), we rewrite equation (1) as:

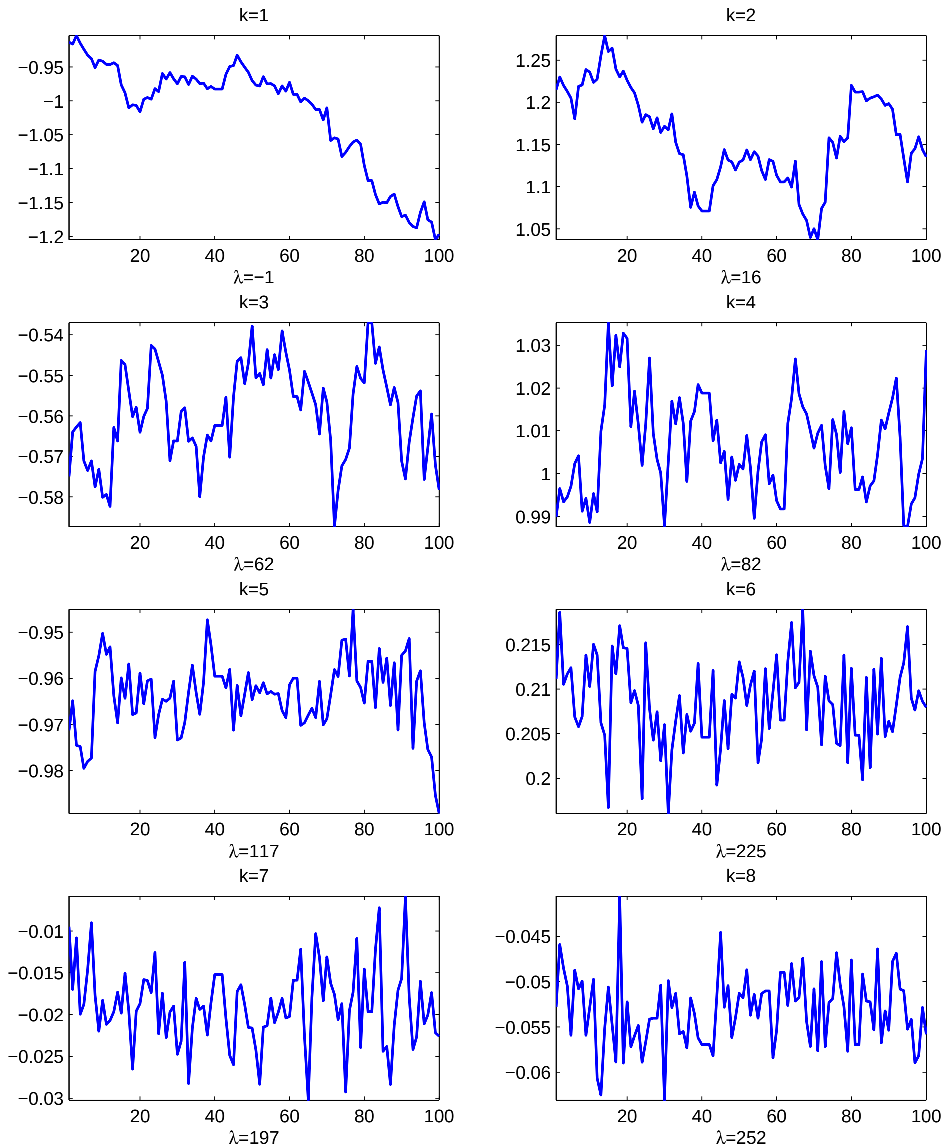
$$\Delta S_t = Q S_{t-1} + Z_t$$

where $Q = A - \mathbf{I}$. The basis of cointegrating portfolios is then found by solving the following generalized eigenvalue problem:

$$\lambda S_{t-1}^T S_{t-1} - S_{t-1}^T \Delta S_t (\Delta S_t^T \Delta S_t)^{-1} \Delta S_t^T S_{t-1} \tag{5}$$

in the variable $\lambda \in \mathbf{R}$.

Mean-reversion: canonical decompositions



Box & Tiao (1977) canonical decomposition on 100 days of U.S. swap rate data (in percent), ranked in decreasing order of predictability. The mean reversion coefficient λ is listed below each plot.

Mean-reversion: related works

- Fama & French (1988), Poterba & Summers (1988) model and test for market predictability in excess returns.
- Cointegration techniques, (see Engle & Granger (1987), and Alexander (1999) for a survey of applications in finance) are usually used to extract mean reverting portfolios.
- Several authors focused on the optimal investment problem when expected returns are mean reverting, with Kim & Omberg (1996) and Campbell & Viceira (1999) or Wachter (2002) among others, obtaining closed-form solutions in some particular cases.
- Liu & Longstaff (2004) study the optimal investment problem in the presence of a "textbook" finite horizon arbitrage opportunity, modeled as a Brownian bridge. Jurek & Yang (2006) study this same problem when the arbitrage horizon is indeterminate. Gatev, Goetzmann & Rouwenhorst (2006) also studied the performance of pairs trading, which are classic examples of structurally mean-reverting portfolios.
- The LTCM meltdown in 1998 focused a lot of attention on the impact of leverage limits and liquidity, see Grossman & Vila (1992) or Xiong (2001) for a discussion.

Sparse methods

- ℓ_1 regularized regression (LASSO): Tibshirani (1996).
- Feature selection: ℓ_1 penalized support vector machines.
- Compressed sensing: Candès & Tao (2005), Donoho & Tanner (2005).
- Basis pursuit: Chen, Donoho & Saunders (2001), . . .
- Sparse PCA and covariance selection: d'Aspremont, El Ghaoui, Jordan & Lanckriet (2007) and d'Aspremont, Banerjee & El Ghaoui (2006).

Outline

- Canonical decomposition
- **Sparse generalized eigenvalue problems**
- Estimation and trading
- Numerical results

Sparse generalized eigenvalue problems

Both canonical decompositions involve solving a **generalized eigenvalue problem** of the form:

$$\det(\lambda B - A) = 0 \tag{6}$$

in the variable $\lambda \in \mathbf{R}$, where $A, B \in \mathbf{S}^n$. This is usually solved using a QZ decomposition. The largest solution of this problem can be written in variational form as:

$$\lambda^{\max} = \max_{x \in \mathbf{R}^n} \frac{x^T A x}{x^T B x}.$$

Here however, we seek to maximize (or minimize) that ratio while constraining the cardinality of the (portfolio) coefficient vector x and solve instead:

$$\begin{array}{ll} \text{maximize} & x^T A x / x^T B x \\ \text{subject to} & \mathbf{Card}(x) \leq k \\ & \|x\| = 1, \end{array} \tag{7}$$

where $k > 0$ is a given constant and $\mathbf{Card}(x)$ is the number of nonzero coefficients in x.

Sparse generalized eigenvalue problems

- Solving generalized eigenvalue problems is **easy**: takes $O(n^3)$ operations.
- Solving sparse generalized eigenvalue problems is **hard**: equivalent to subset selection which is NP-Hard.

Here, we seek good approximate solutions to:

$$\begin{array}{ll} \text{maximize} & x^T A x / x^T B x \\ \text{subject to} & \mathbf{Card}(x) \leq k \\ & \|x\| = 1, \end{array}$$

using two algorithms:

- **Greedy search**: Incrementally scan all variables.
- **Semidefinite relaxation**: form a tractable convex relaxation.

Greedy Search

- Define:

$$I_k = \{i \in [1,n] : \; x_i \neq 0\},$$

- We build approximate solutions **recursively** in k. When $k = 1$, we can simply find I_1 as:

$$I_1 = \mathop{\rm argmax}_{i \in [1,n]} A_{ii}/B_{ii}.$$

- Given I_k, we add one variable with index i_{k+1} to produce the largest increase in predictability:

$$\max_{\{x \in \mathbf{R}^n : \; \mathrm{supp}(x) = I_k \cup \{i\}\}} \frac{x^T A x}{x^T B x}.$$

- The complexity of computing solutions for all k is in $O(n^4)$.

Semidefinite relaxation

Start from our original problem:

$$\begin{array}{ll} \text{maximize} & x^T A x / x^T B x \\ \text{subject to} & \mathbf{Card}(x) \leq k \\ & \|x\| = 1, \end{array}$$

with variable $x \in \mathbf{R}^n$, and rewrite it in terms of $X = xx^T \in \mathbf{S}_n$:

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(AX)/\mathbf{Tr}(BX) \\ \text{subject to} & \mathbf{Card}(X) \leq k^2 \\ & \mathbf{Tr}(X) = 1 \\ & X \succeq 0, \;\; \mathbf{Rank}(X) = 1, \end{array}$$

in the variable $X \in \mathbf{S}_n$. This program is **equivalent** to the first one.

Semidefinite relaxation

$$\begin{array}{ll}\text{maximize} & \mathbf{Tr}(AX)/\mathbf{Tr}(BX)\\ \text{subject to} & \mathbf{Card}(X) \leq k^2\\ & \mathbf{Tr}(X) = 1\\ & X \succeq 0, \;\; \mathbf{Rank}(X) = 1,\end{array}$$

- Since $\mathbf{Card}(u) = q$ implies $\|u\|_1 \leq \sqrt{q}\|u\|_2$, we can replace the nonconvex constraint $\mathbf{Card}(X) \leq k^2$, by a weaker but **convex** constraint: $\mathbf{1}^T|X|\mathbf{1} \leq k$.
- We drop the rank constraint to get the following quasi-convex program:

$$\begin{array}{ll}\text{maximize} & \mathbf{Tr}(AX)/\mathbf{Tr}(BX)\\ \text{subject to} & \mathbf{1}^T|X|\mathbf{1} \leq k\\ & \mathbf{Tr}(X) = 1\\ & X \succeq 0,\end{array}$$

in the variable $X \in \mathbf{S}_n$.

Semidefinite relaxation

Starting from the quasi-convex program:

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(AX)/\mathbf{Tr}(BX) \\ \text{subject to} & \mathbf{1}^T|X|\mathbf{1} \leq k \\ & \mathbf{Tr}(X) = 1 \\ & X \succeq 0, \end{array}$$

we change variables:

$$Y = \frac{X}{\mathbf{Tr}(BX)}, \quad z = \frac{1}{\mathbf{Tr}(BX)}$$

and solve:

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(AY) \\ \text{subject to} & \mathbf{1}^T|Y|\mathbf{1} - kz \leq 0 \\ & \mathbf{Tr}(Y) - z = 0 \\ & \mathbf{Tr}(BY) = 1 \\ & Y \succeq 0, \end{array} \tag{8}$$

which is a semidefinite program in the variables $Y \in \mathbf{S}_n$ and $z \in \mathbf{R}_+$ and can be solved using standard SDP solvers such as SDPT3 by Toh, Todd & Tutuncu (1999).

Performance

Greedy algorithm:

- The optimal solutions of problem (7) might not have an increasing support set sequence $I_k \subset I_{k+1}$.
- However, the cost of this method is relatively low: with each iteration costing $O(k^2(n-k))$, the complexity of computing solutions for all k is in $O(n^4)$.
- This recursive procedure can also be repeated both forward and backward to improve the quality of the solution.
- Stability issues.

Semidefinite relaxation:

- Higher complexity.
- ℓ_1 penalization makes it potentially more stable.

Outline

- Canonical decomposition
- Sparse generalized eigenvalue problems
- **Estimation and trading**
- Numerical results

Estimation and trading

By integrating P_t over a time increment Δt we get:

$$P_t = \bar{P} + e^{-\lambda \Delta t}(P_{t-\Delta t} - \bar{P}) + \sigma \int_{t-\Delta t}^{t} e^{\lambda(s-t)} dZ_s,$$

so we can **estimate** λ and σ by simply regressing P_t on $P_{t-\Delta t}$ and a constant. We have the following estimators for the parameters of P_t:

$$\begin{aligned}
\hat{\mu} &= \frac{1}{N}\sum_{i=0}^{N} P_{t_i} \\
\hat{\lambda} &= -\frac{1}{\Delta t}\log\left(\frac{\sum_{i=1}^{N}(P_{t_i}-\hat{\mu})(P_{t_{i-1}}-\hat{\mu})}{\sum_{i=1}^{N}(P_{t_i}-\hat{\mu})(P_{t_i}-\hat{\mu})}\right) \\
\hat{\sigma}^2 &= \frac{2\lambda}{(1-e^{-2\lambda\Delta t})(N-2)}\sum_{i=1}^{N}\left((P_{t_i}-\hat{\mu}) - e^{-\lambda\Delta t}(P_{t_{i-1}}-\hat{\mu})\right)^2
\end{aligned}$$

Estimation and trading

Trading O.U. processes: two classic strategies.

- **Threshold:** Invest when the spread $|\bar{P} - P_t|$ crosses a certain threshold, cf. Gatev et al. (2006).
- **Linear:** Under log-utility, the optimum strategy is **linear**:

$$N = \frac{\lambda(\bar{P} - P_t) - rP_t}{\sigma^2} W_t$$

where N is the number of units of portfolio the agent holds and W_t the investor's wealth at time t. See Jurek & Yang (2006).

A few remarks:

- None of these results account for transaction costs.
- Jurek & Yang (2006) also find the optimal strategy for CRRA utility defined over wealth at a finite horizon and Epstein-Zin utility defined over intermediate cash flows.
- Similar results hold with proportional fund flows, cf. Jurek & Yang (2006).

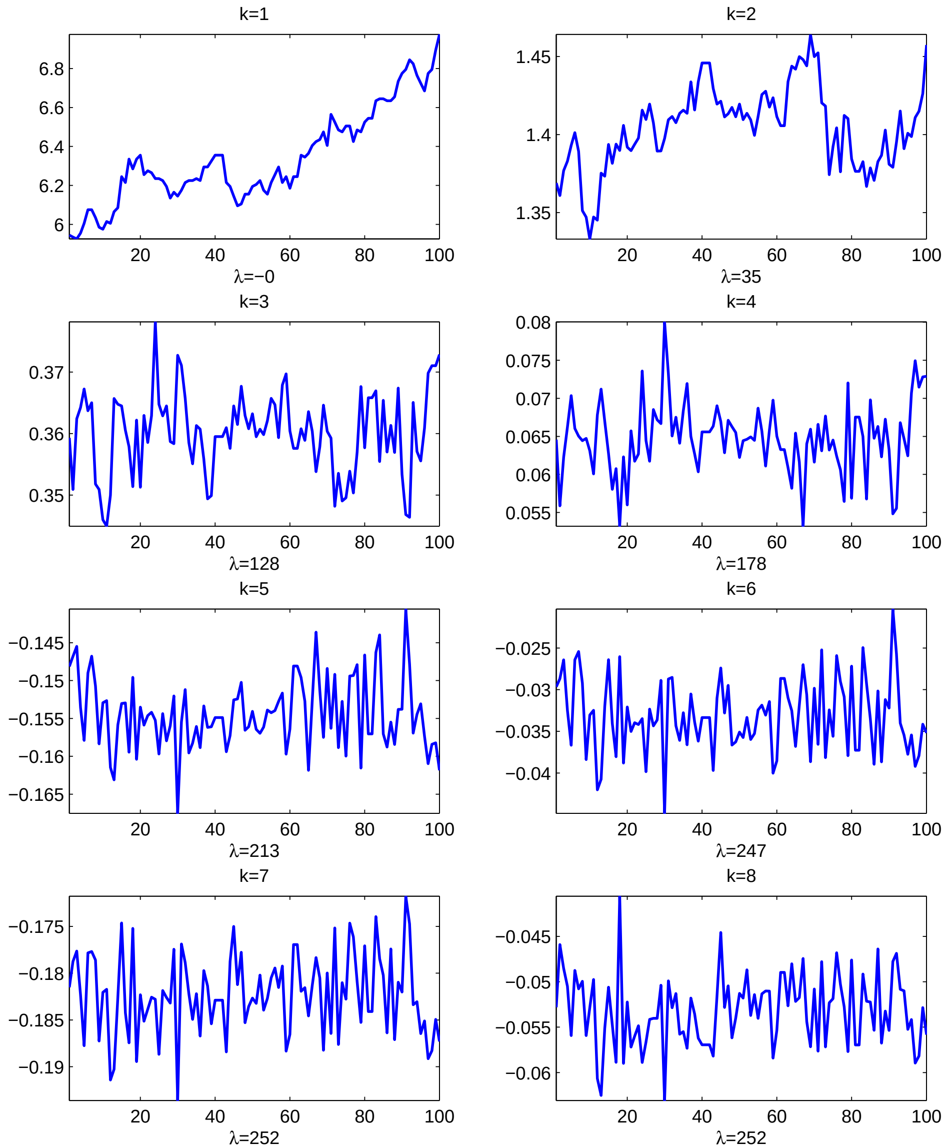
Outline

- Canonical decomposition
- Sparse generalized eigenvalue problems
- Estimation and trading
- **Numerical results**

Numerical Results

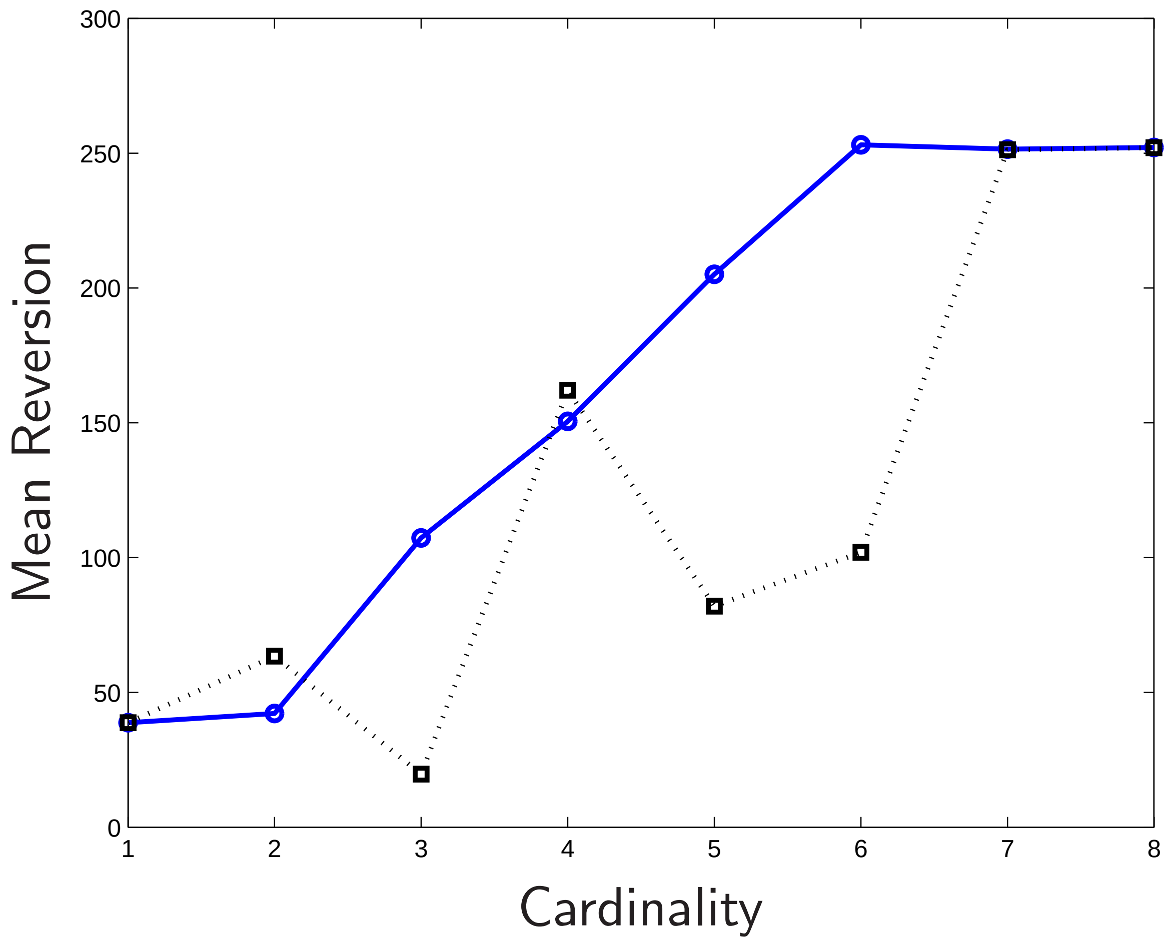
- U.S. swap rate data for maturities 1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y and 30Y from 1998 until 2005.
- Use greedy algorithm to compute optimally mean reverting portfolios of increasing cardinality for time windows of 200 days and repeat the procedure every 50 days.
- Update portfolios daily using linear rule.

Numerical Results



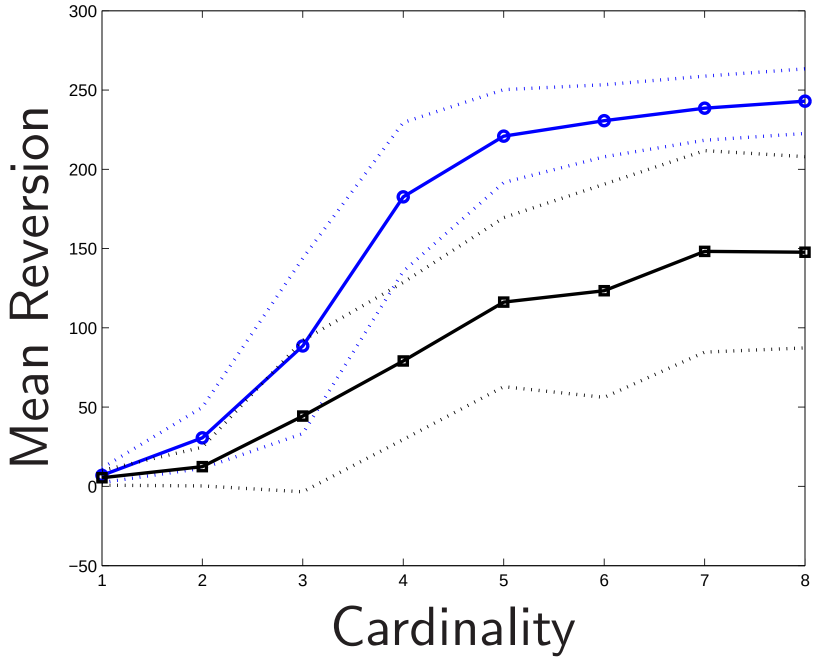
Sparse canonical decomposition on 100 days of U.S. swap rate data (in percent). The number of nonzero coefficients in each portfolio vector is listed as k on top of each subplot, the mean reversion coefficient λ is listed below each one.

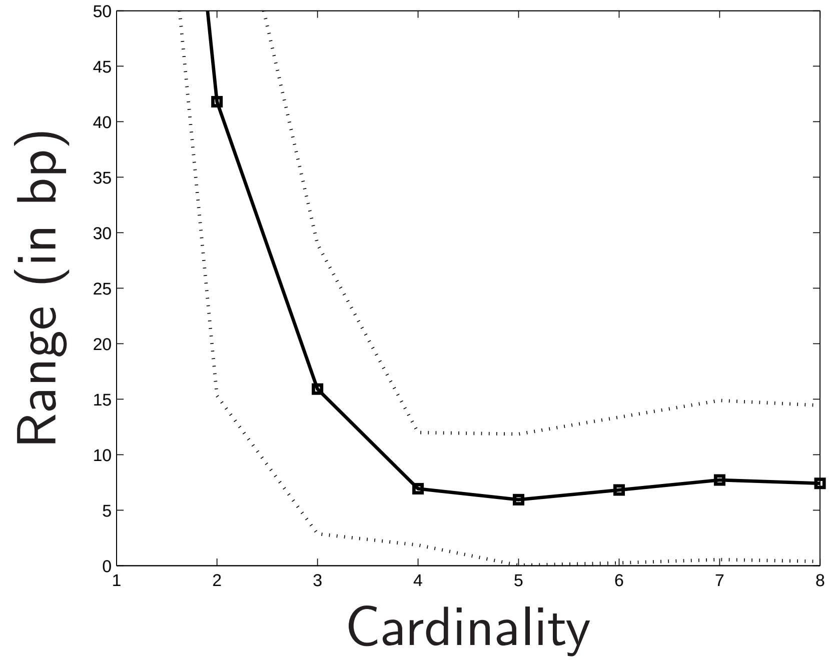
Numerical Results



Mean reversion coefficient λ versus portfolio cardinality (number of nonzero coefficients) using the greedy search (solid line) and the semidefinite relaxation (dashed line) on U.S. swap rate data.

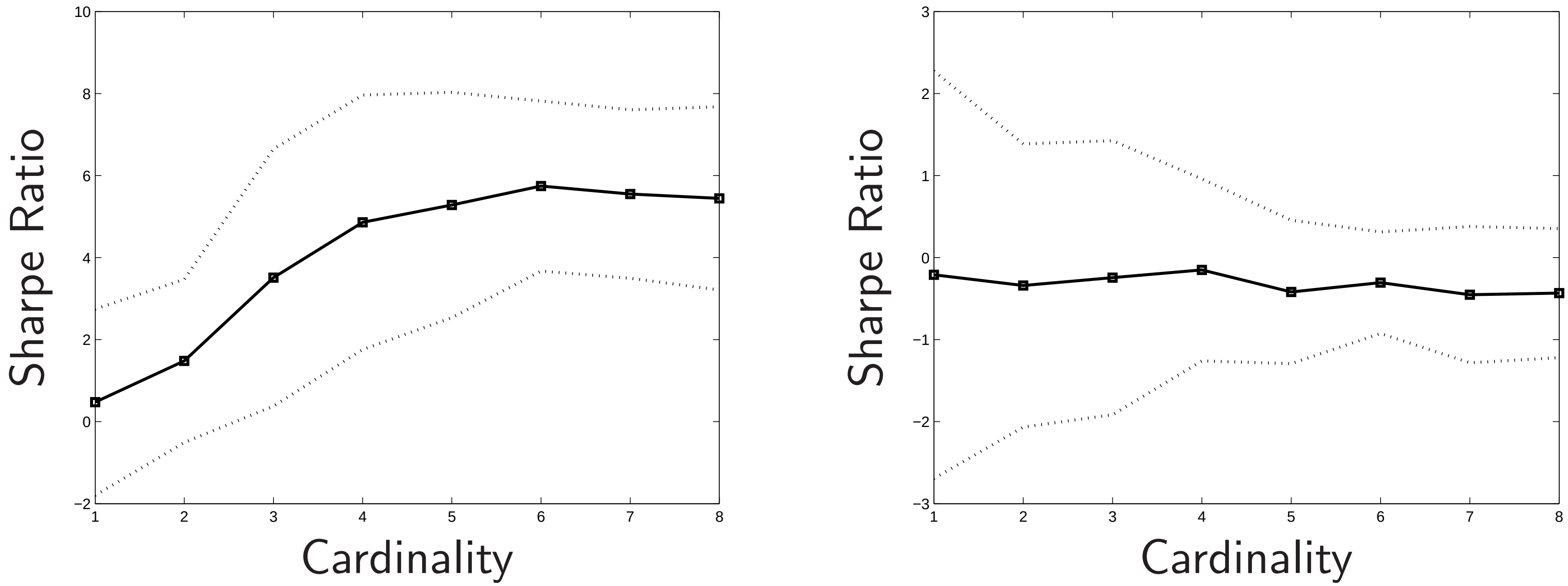
Numerical Results





Left: mean reversion coefficient λ versus portfolio cardinality (number of nonzero coefficients), in sample (blue circles) and out of sample (black squares) on U.S. swaps.
Right: out of sample portfolio **price range** (in basis points) versus cardinality (number of nonzero coefficients) on U.S. swap rate data. Dashed lines at plus and minus one standard deviation.

Numerical Results



Left: average out of sample **sharpe ratio** versus portfolio cardinality on U.S. swaps.
Right: idem, with **transaction costs** modeled as a Bid-Ask spread of 1bp. The dashed lines are at plus and minus one standard deviation.

References

Alexander, C. (1999), 'Optimal hedging using cointegration', *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **357**(1758), 2039–2058.

Bewley, R., Orden, D., Yang, M. & 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. E. & Tiao, G. C. (1977), 'A canonical analysis of multiple time series', *Biometrika* **64**(2), 355.

Campbell, J. & Viceira, L. (1999), 'Consumption and Portfolio Decisions When Expected Returns Are Time Varying', *The Quarterly Journal of Economics* **114**(2), 433–495.

Candès, E. J. & Tao, T. (2005), 'Decoding by linear programming', *Information Theory, IEEE Transactions on* **51**(12), 4203–4215.

Chen, S., Donoho, D. & Saunders, M. (2001), 'Atomic decomposition by basis pursuit.', *SIAM Review* **43**(1), 129–159.

d'Aspremont, A., Banerjee, O. & El Ghaoui, L. (2006), 'First-order methods for sparse covariance selection', *Arxiv math/0609812* .

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

Donoho, D. L. & Tanner, J. (2005), 'Sparse nonnegative solutions of underdetermined linear equations by linear programming', *Proceedings of the National Academy of Sciences* **102**(27), 9446–9451.

Engle, R. & Granger, C. (1987), 'Cointegration and error correction: representation, estimation and testing', *Econometrica* **55**(2), 251–276.

Fama, E. & French, K. (1988), 'Permanent and Temporary Components of Stock Prices', *The Journal of Political Economy* **96**(2), 246–273.

Gatev, E., Goetzmann, W. & Rouwenhorst, K. (2006), 'Pairs Trading: Performance of a Relative-Value Arbitrage Rule', *Review of Financial Studies* **19**(3), 797.

Grossman, S. & Vila, J. (1992), 'Optimal Dynamic Trading with Leverage Constraints', *The Journal of Financial and Quantitative Analysis* **27**(2), 151–168.

Jurek, J. & Yang, H. (2006), Dynamic portfolio selection in arbitrage, Technical report, Working Paper, Harvard Business School.

Kim, T. & Omberg, E. (1996), 'Dynamic Nonmyopic Portfolio Behavior', *The Review of Financial Studies* **9**(1), 141–161.

Liu, J. & Longstaff, F. (2004), 'Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities', *Review of Financial Studies* **17**(3).

Poterba, J. M. & Summers, L. H. (1988), 'Mean reversion in stock prices: Evidence and implications', *Journal of Financial Economics* **22**(1), 27–59.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the LASSO', *Journal of the Royal statistical society, series B* **58**(1), 267–288.

Toh, K. C., Todd, M. J. & Tutuncu, R. H. (1999), 'SDPT3 – a MATLAB software package for semidefinite programming', *Optimization Methods and Software* **11**, 545–581.

Wachter, J. (2002), 'Portfolio and Consumption Decisions under Mean-Reverting Returns: An Exact Solution for Complete Markets', *The Journal of Financial and Quantitative Analysis* **37**(1), 63–91.

Xiong, W. (2001), 'Convergence trading with wealth effects: an amplification mechanism in financial markets', *Journal of Financial Economics* **62**(2), 247–292.