

Introduction to Algorithmic Trading Strategies Lecture 7

Small Mean Reverting Portfolio

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

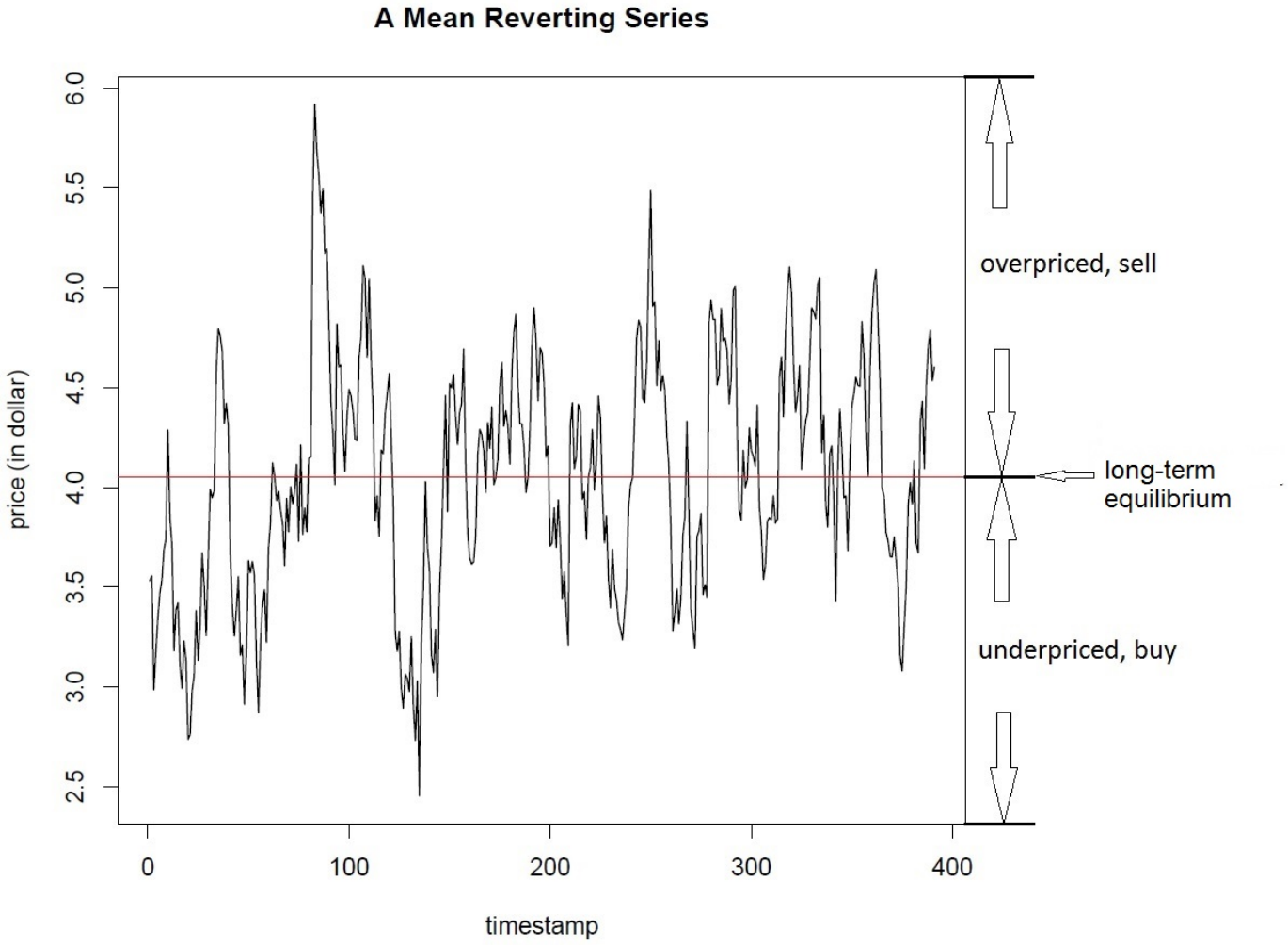
References

- A d'Aspremont. Identifying Small Mean Reverting Portfolios. Stanford University. 2008.
- Box, G. E. & Tiao, G. C. "A canonical analysis of multiple time series", Biometrika 64(2), 355. 1977.
- Dattorro, J. "Example 4.6.0.0.12, Semidefinite programming," Convex Optimization & Euclidean Distance Geometry. 2010.
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. Least Angle Regression. The annuals of Statistics, Vol. 32, No. 2, 407-499. 2004.

Outline

- Why small portfolios?

Mean Reversion is Good



Construction Methods

- Distance method
- Cointegration

Shortcomings

- Dense
  - Hard to execute; need to trade all assets all at the same time.
  - Big transaction costs.
  - Hard to interpret the significance of the relationship.
- Cointegration
  - It gives an 'Yes'/'No' answer. Not sure what to do if the cointegration relationship is not stable, especially after entering a position.

As a Maximization Problem

- Find the portfolio weights $\{x_i\}$ such that the portfolio $P_t = \sum_{i=1}^{t} x_i S_{ti}$ is most mean reverting meaning that if $dP_t = \lambda\left(\bar{P} - P_t\right)dt + \sigma dZ_t$ then λ is maximized.
  - Sparsity: no more than k non-zeros in x, trading at most k assets.
- Note that unlike cointegration, this approach always give you a mean reverting portfolio and tells how mean reverting it is.

VAR(1)

- Assume that the prices follow a stationary VAR(1) process.
  - $S_t = S_{t-1}F + Z_t$
  - Z_t is $N(0, \Sigma)$
- Canonical Analysis
  - $n = 1$
- $\mathrm{E}(S_t) = \mathrm{E}(S_{t-1}F) + \mathrm{E}(Z_t)$
  - $\sigma_t^2 = \sigma_{t-1}^2 F^2 + \Sigma$

Estimation of F

- Regress S_t on S_{t-1}.
  - $\hat{F} = \left(S_{t-1}{}' S_{t-1}\right)^{-1} S_{t-1}{}' S_t$
  - But this $\hat{F}$ is dense. It does not highlight the dependence relationship between S_t and S_{t-1}.
- Sparse $\hat{F}$.
- Penalized Least Squares estimation, column by column.
  - $f_i = \underset{x \in R^n}{\operatorname{argmin}} \left\{ \|S_{it} - S_{t-1} x\|^2 + \gamma \|x\|_1 \right\}$
  - γ control the sparsity of x and hence f_i and hence $\hat{F}$.

Least Absolute Shrinkage and Selection Operator (LASSO)

- It is an automatic model building problem, a factor/model selection problem.
- $\min \sum_{i=1}^{n} \left(Y_i - \sum_{i=1}^{p} X_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$
  - Minimizing the L1 norm of β drives some β_j to zero.
- Forward Stagewise Linear Regression
  - Pick the covariate x_{j1} with the biggest correlation with y.
  - Compute the residual $r_1 = y - \beta_1 x_{j1}$.
  - Pick the covariate x_{j2} with the biggest correlation with r_1. Compute the residual. Repeat this step until k covariates are found.
  - Build a k-parameter linear model in the usual way, e.g., ordinary least squares.
  - This is a greedy search that may eliminate useful covariate in the 2nd step.

Least Angle Regression (LARS) algorithm

- Problem
  - $\min \sum_{i=1}^{n} \left(Y_i - \sum_{i=1}^{p} X_{ij} \beta_j \right)^2$ s.t., $\sum_{j=1}^{p} |\beta_j| \leq t$

Covariance Selection

- Covariance matrix represents relation between all variables while inverse covariance shows the relation of element with their neighbors.
- Zeros in X corresponds to conditionally independent variables, discovering structure in the model.
  - We want zeros in X to highlight the conditional independence of the assets.
- $\max_X \log\det X - \mathrm{Tr}(\Sigma X) - \rho\,\mathrm{Card}(X)$
- X: the inverse covariance matrix
- Solution: Graphical LASSO or LASSO.

Covariance Selection

- Robust and sparse estimation of the covariance matrix developed by Dempster (1972)
- Maximum-likelihood penalized by the cardinality of X

$$\max_x \log\det X - \mathrm{Tr}(\sum X) - \rho\,\mathrm{Card}(X)$$

  - in the variable $X \in S_n$, where $\boldsymbol{\Sigma}$ is the sample covariance matrix, *Card(X)* is the number of nonzero coefficients in X and $\boldsymbol{\rho}$>0 controls the trade-off between log-likelihood and sparsity
- Hard to solve numerically. D'Aspremont, Banerjee & El Ghaoui (2006) replaces the penalty by the ℓ_1 norm

$$\max_x \log\det X - \mathrm{Tr}(\sum X) - \rho \sum_{i,j=1}^{n} |X_{ij}|$$

Covariance Selection – Dual Problem Block-coordinate descent gradient method

- The dual problem is given by

$$\min -\log\det(\Sigma + U) - n$$

$$\text{subject to} \; |U_{ij}| \leq \rho, \quad i, j = 1, \ldots, n$$

  - in the variable $U \in S_n$

- We can decompose the matrices in blocks as follows

$$U = \begin{pmatrix} V & u \\ u^T & w \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} A & b \\ b^T & c \end{pmatrix}$$

  - where V is fixed, $A \in S_{(n-1)}$, $u,b \in R^{(n-1)}$ and $w,c \in R$
  - u represents the variables, i.e. the rows and columns we are updating at each iteration.

Covariance Selection – Dual Problem Block-coordinate descent gradient method

- The dual problem in blocks becomes

$$\min -\log\det(A+V) - \log[(w+c) - (b+u)^T (A+V)^{-1}(b+u)] - n$$
$$\text{subject to} \; |w| \le \rho, \; |u_{ij}| \le \rho, \quad i, j = 1, \ldots, n$$

- At each iteration, the main step is then a box constrained quadratic program of the form

$$\min (b+u)^T (A+V)^{-1}(b+u)$$
$$\text{subject to} \; |u_{ij}| \le \rho, \quad i, j = 1, \ldots, n$$

  - can be solved using SeDuMi by Sturm (1999)

Block-coordinate descent gradient method Algorithm

- Pick the row and colum to update;
- Compute $(A + V)^{-1;}$
- Update the row and column previously picked with the solution of the box constrained QP
- After each step, check the folowing convergence condition

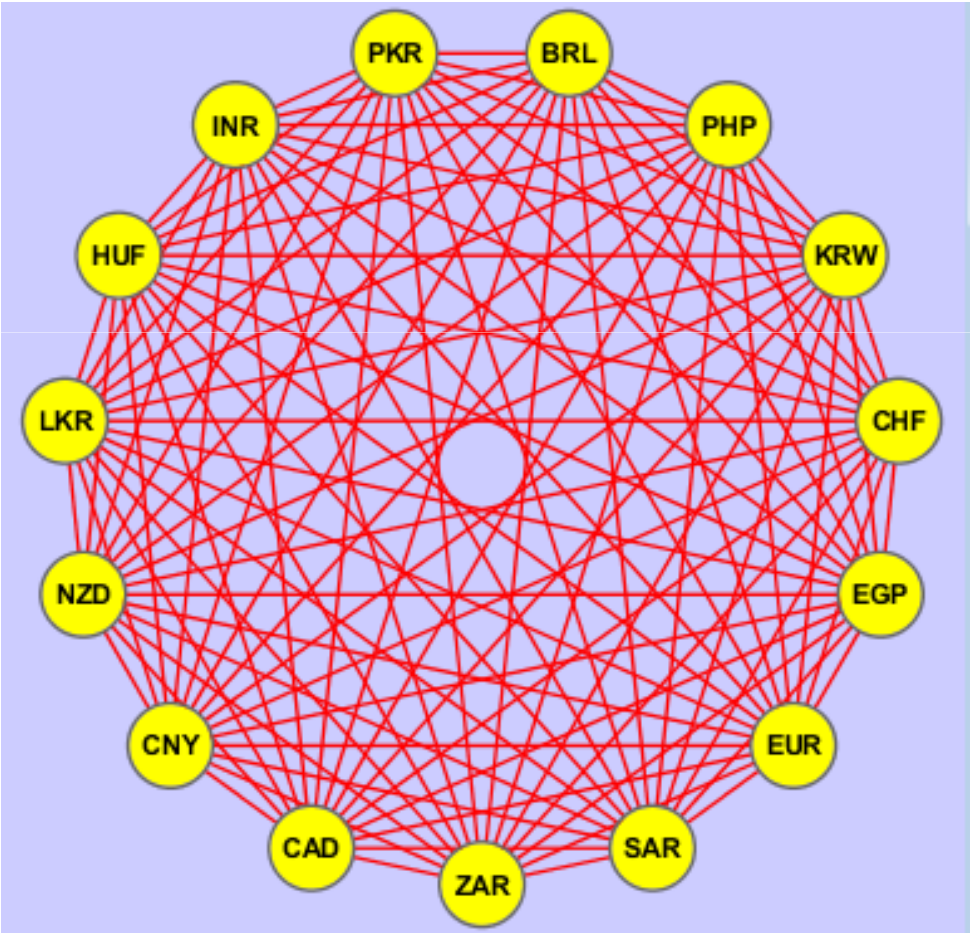
$$\mathrm{Tr}(\sum X) - n + \rho \sum_{i,j=1}^{n} \left| X_{ij} \right| \leq \varepsilon$$

  - Matlab code available on d'Aspremont's website

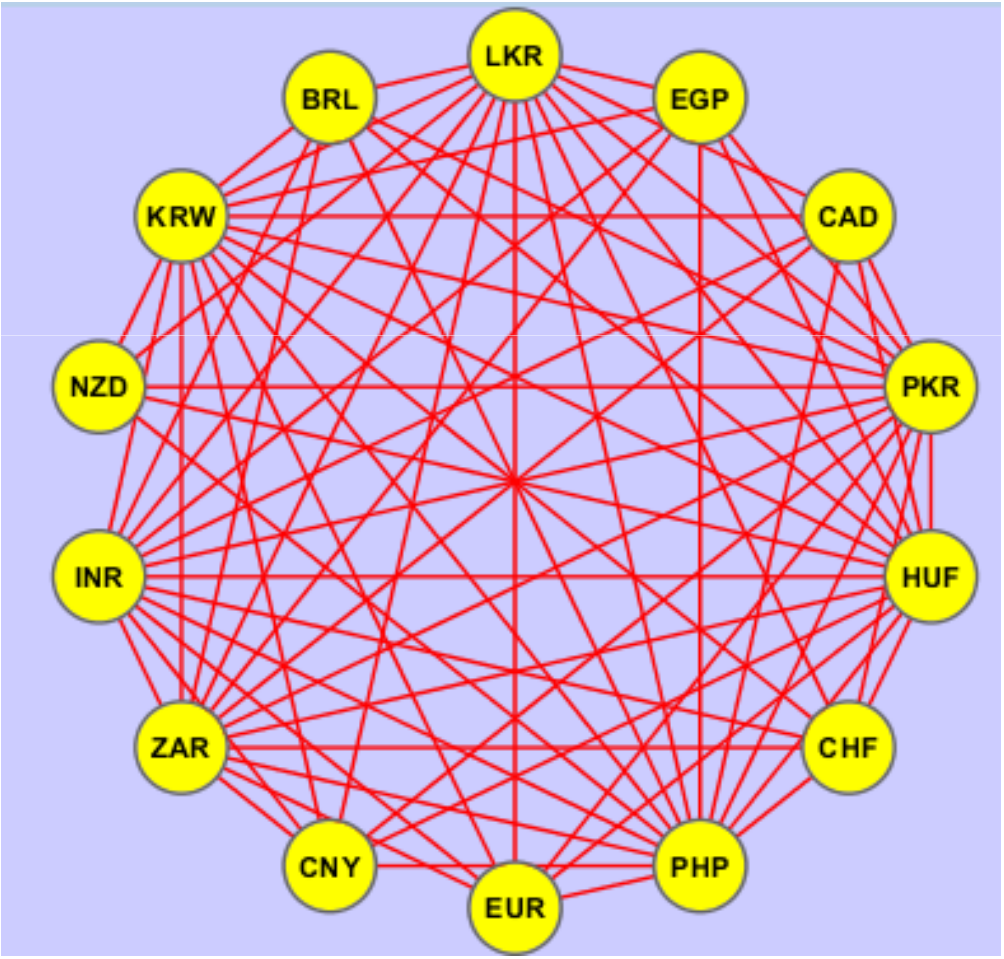
Covariance Selection Example

- Each component equal to 0 in the inverse covariance matrix represents two variables that are conditionally independent
  - Financial meaning of the inverse covariance matrix: idiosyncratic components of asset prices dynamics
- Example
  - 15 currencies vs. USD
  - daily data from Jan 2008 to Dec 2009
  - graphs done using Cytoscape

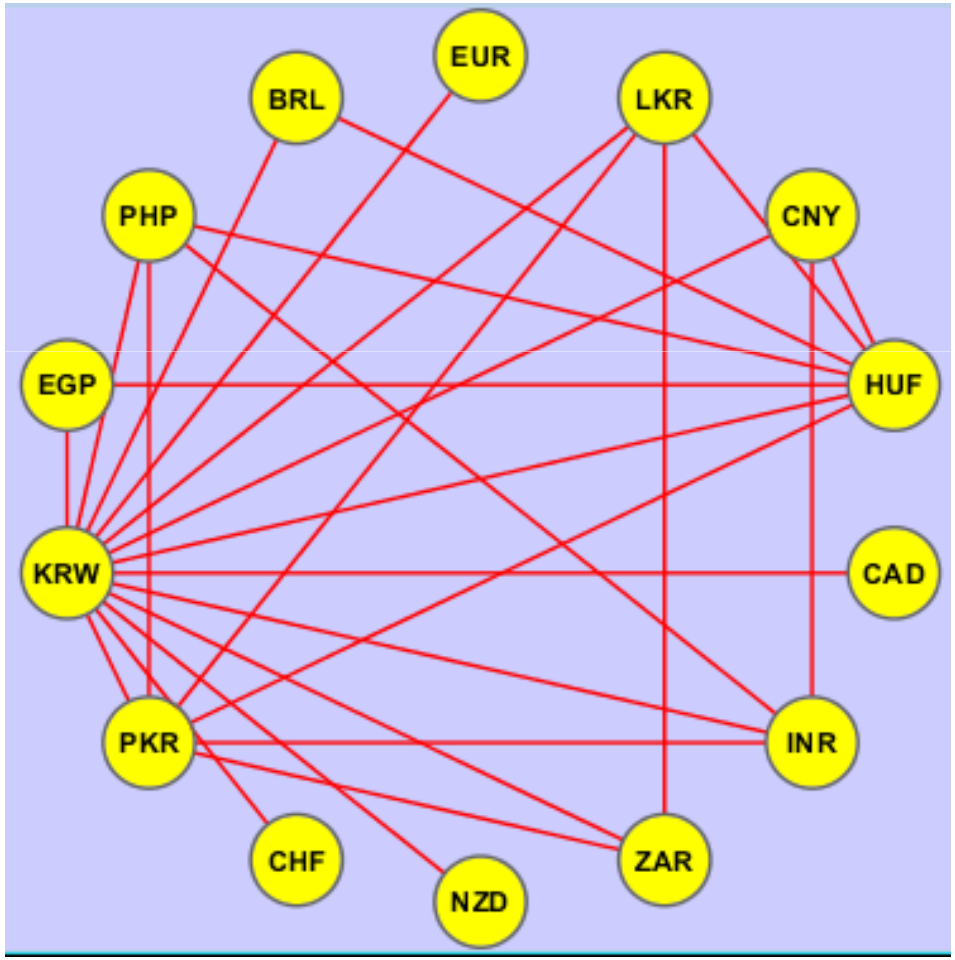
Covariance Selection
Initial inverse covariance matrix



Covariance Selection
rho = 0.05

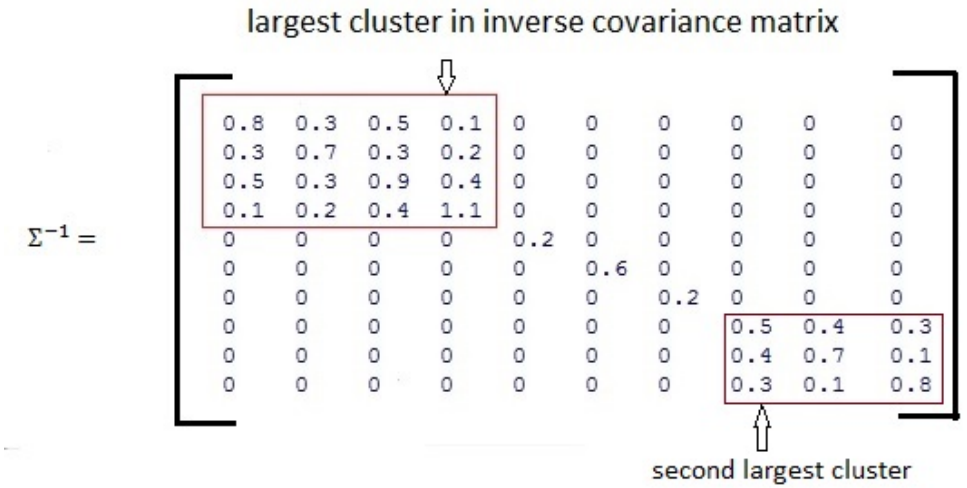


Covariance Selection
rho = 0.5



Clusters

- Select the right financial instruments
  - Based on conditional dependence (inverse covariance)
    - The chosen instruments should be conditional dependent on each other, but conditional independent to the rest.
  - Identify clusters in the inverse covariance matrix



Predictability

- $\nu = \frac{\sigma_{t-1}^2 F^2}{\sigma_t^2}$
- ν small → variance of S_{t-1} is shadowed by the contribution from Σ → S_t is mostly noise, not predictable
- ν big → variance of S_{t-1} dominates that of S_t and the contribution from Σ is small → S_t is more predictable
- Hence ν is a good measure for mean reversion (low predictability), and is a good proxy for λ.

Portfolio Predictability

- $P_t = xS_t = S_{t-1}Fx + Z_t x$
- $\nu = \frac{xF'\Gamma xF}{x'\Gamma x}$

Problem Formulation (Greedy Search)

- $\min_x \frac{x'F'\Gamma Fx}{x'\Gamma x}$
- s.t.,
  - $\|x\|_0 \leq k$
  - $\|x\| = 1$
- $A = F'\Gamma F$
- $B = \Gamma$

Greedy Search

- $I_k = \{i \in [1, n] \mid x_i \neq 0\}$, the set of non-zero indices.
- When $k = 1$, all but one index of x_i is empty, ${x_i}' A x_i = A_{ii}$.
  - $I_1 = \underset{i \in [1,n]}{\operatorname{argmin}} A_{ii}/B_{ii}$.
- Recursion. Given I_k, find I_{k+1}.
  - For each index i not in I_k, solve the minimization problem.
    - With a fixed set of indices,
      - Find z the vector corresponding to the smallest eigenvalue in $\Gamma^{-1/2} A^T \Gamma A \Gamma^{-1/2}$.
      - $x = \Gamma^{-1/2} z$.
  - Add the index i of the smallest objective value to I_k.
- Simple.
- Solution not optimal.

Primal SDP Problem

- $\min_X \mathrm{Tr}(CX)$
- s.t.,
  - $\mathrm{Tr}(A_i X) = b_i$
  - $X \succcurlyeq 0$
- C, X, A_i are all symmetric matrices of n-by-n.
- Convex set: pick any two points and draw a line. The line lies entirely inside the set. No dents in the perimeter.
- Convex cone K: $x \in K$ implies that $\alpha x \in K$ for any scalar $\alpha \geq 0$.

Linear Programming

- Linear Programming is a special case.
- C, A_i are diagonal matrices.
- $\min_x c'x$
- s.t.,
  - $Ax = b$
  - $x \geq 0$
- Note: Quadratic Programming is another special case of SDP.

Problem Formulation (SDP)

- $\min_x \frac{x'Ax}{x'Bx}$
- s.t.,
  - $\|x\|_0 \leq k$
  - $\|x\| = 1$
- $A = F'\Gamma F$
- $B = \Gamma$

Equivalent Problem

- $X = xx'$
- $\min_{X \in S^n} \frac{\mathrm{Tr}(AX)}{\mathrm{Tr}(BX)}$
- s.t.,
  - $\|X\|_0 \leq k^2$
  - $\mathrm{Tr}(X) = 1$
  - $X \succcurlyeq 0$
  - Rank of X =1

Convex Relaxation

- $\min_{X \in S^n} \frac{\mathrm{Tr}(AX)}{\mathrm{Tr}(BX)}$
- s.t.,
  - $\mathbf{1}'|\boldsymbol{X}|\mathbf{1} \leq \boldsymbol{k}$
  - $\mathrm{Tr}(X) = 1$
  - $X \succcurlyeq 0$

Relaxed SDP Problem Problems

- Need to transform $\mathbf{1}'|\boldsymbol{X}|\mathbf{1} \leq \boldsymbol{k}$ into the standard form.
- The transformation will introduce $N(N+1)$ constraints.
- The relaxation may return an X that has a bigger rank and/or cardinality, violating the constraints.

Equivalent SDP Problem

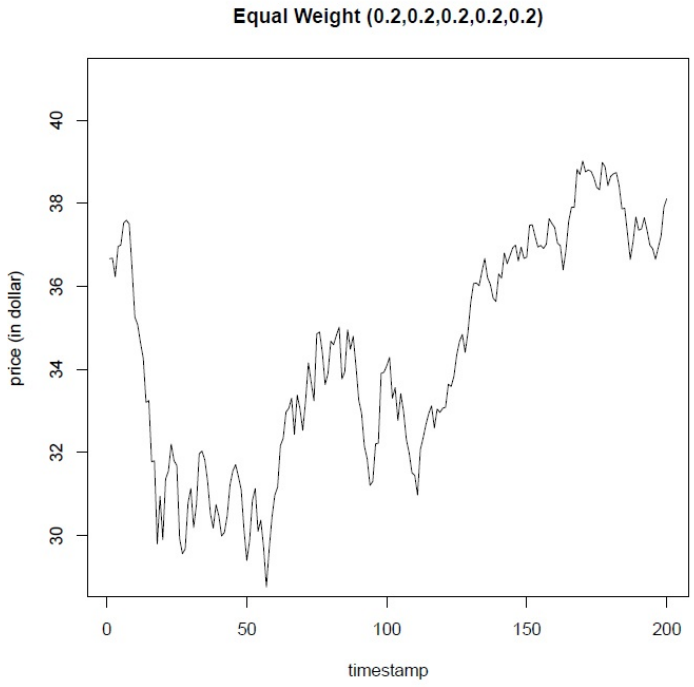
- $\min_{Y \in S^n} \mathrm{Tr}(AY)$
- s.t.,
  - $\|Y\|_0 \leq k$
  - $\mathrm{Tr}(BY) = 1$
  - $Y \succcurlyeq 0$
  - Rank of $Y = 1$
- Change of variable
  - $Y = \frac{X}{\mathrm{Tr}(BX)}$

Handling Rank and Cardinality

- Iteratively solve this system of three equations, starting with
  - $W = D = 0$
- $\min_{Y \in S^n} Tr(AY) - w_1 Tr(WY) - w_2 Tr(DY)$
  - $\mathrm{Tr}(BY) = 1$
  - $Y \succcurlyeq 0$
- $\min_{W \in S^n} Tr(WY^*)$
  - $0 \leq W \leq I$
  - $tr(W) = n - 1$
  - Can be solved analytically.
- $\min_{D \in Diagonal} Tr(DY^*)$
  - $0 \leq D \leq I$
  - $Tr(D) = n - k$
  - Can be solved analytically.

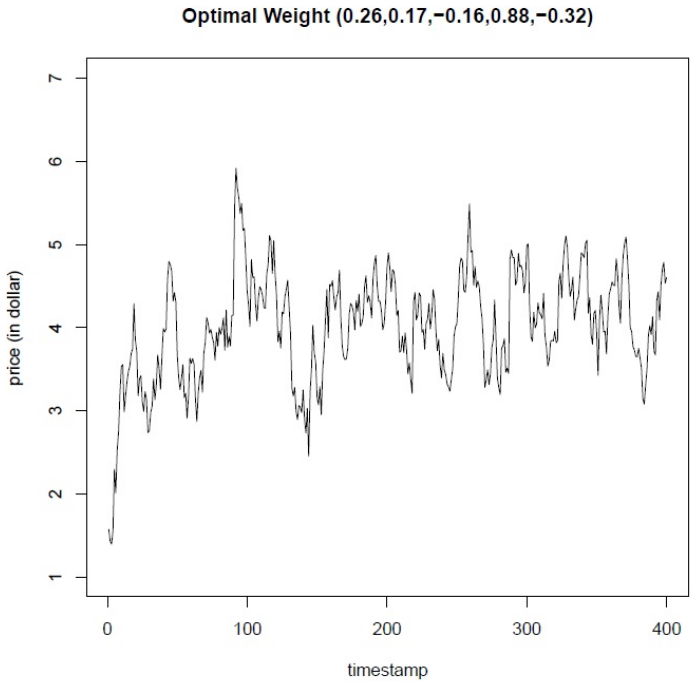
Optimal Weights

- Even if the assets are correctly chosen, without the correct weights, the portfolio is still not mean reverting.



Determine optimal weights

Semidefinite programming



Utilize Mean Reverting Portfolio in Trading

- Given the portfolio can be modelled as a mean reverting OU process, dynamic spread trading is a stochastic optimal control problem.

- Objectives:
  - Given a fixed amount of capital, dynamically allocate capital over a risky mean reverting portfolio and a risk-free asset over a finite time horizon to maximize a general constant relative risk aversion (CRRA) utility function of the terminal wealth
  - Allocate capital amongst several mean reverting portfolios

Utilize Mean Reverting Portfolio in Trading

- Given the portfolio can be modelled as a mean reverting OU process, dynamic spread trading is a stochastic optimal control problem.

- Mathematics involved:
  - Ornstein-Uhlenbeck (OU) estimation
  - Hamilton-Jacobi-Bellman (HJB) equation
  - ODE (Riccatti differential equations)
  - Portfolio allocation (mean variance analysis)

Proposed Procedure

- Step 1: Split a large pool of risky assets into sufficiently small clusters;
- Step 2: Search each of the clusters for optimal *small* mean reverting portfolios;
- Step 3: For each portfolio identified in Step 2, determine the optimal pairs trading strategy;
- Step 4: Dynamically allocate money among mean reverting portfolios.

Case Study – S&P 500 Stocks

- Trading periods
  - From 2010-Jan-01 to 2014-Dec-31, 10 sub-trading period, length of every sub-period is 6 month;
  - Re-select the portfolio before each sub-trading period begins;
  - For every sub-trading period, 2 years stock data are used to select the portfolio.
  - Once portfolio is chosen, the weights of stocks are unchanged within the 6 month period
  - Rebalance when the difference between current portfolio price and last trade price is larger than a certain threshold.



Case Study – S&P 500 Stocks

- Assumptions
  - Margin requirement
    - For short selling, 50% of the short selling value needs to be deposited into the margin account.
    - 50% of values of stocks in long position can be used as collateral for short margin account.
  - Cost
    - Transaction cost: 0.2% per transaction
    - Borrowing cost: 8% p.a. (short selling)
    - Risk-free rate: 0.55% p.a. (1-year LIBOR)



Case Study – S&P 500 Stocks

- An example, the last sub-trading period, 2014/07/02 - 2014/12/31:
  - In-sample fitting
    - 2004/07/01-2014/07/01 Identify the largest cluster: 69 stocks are selected from S&P 500 stocks (*covariance selection*);
    - 2012/07/01-2014/07/01 Calculate the optimal combination of stocks: 5 stocks are selected from the largest cluster and their weights are optimized (*LASSO, semidefinite programming*) ;
    - Fit an OU model to the portfolio constructed (*OU estimation*);
  - Out-of-sample trading (2014/07/02-2014/12/31)
    - Calibration frequency: rebalance when state changes more than 1 unit, state is (current price - mean)/s.d.;
    - Determine the optimal capital to be allocated to the portfolio and the risk-free asset (*ODE solving, portfolio optimization*);



Case Study – S&P 500 Stocks

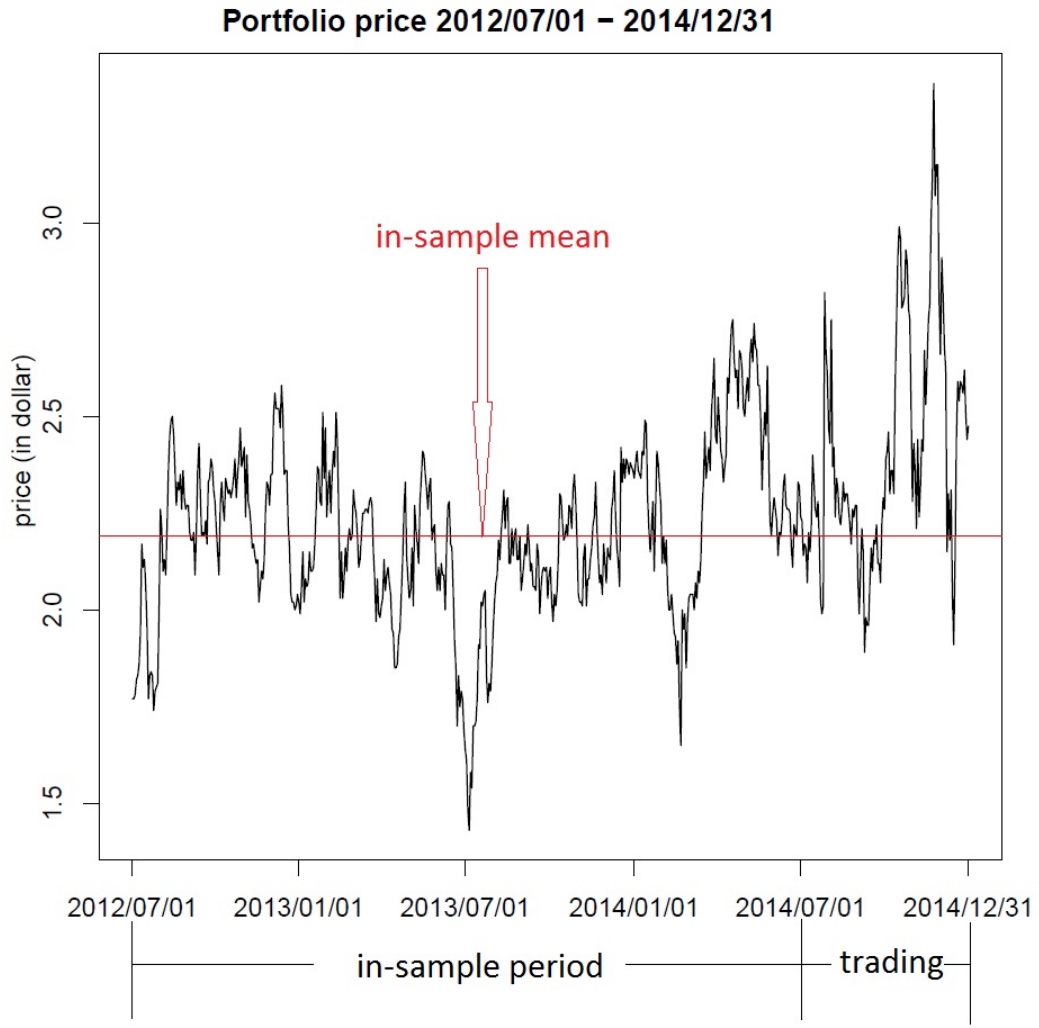
- Optimal portfolio

Stock	Sector	Weight	Avg. price
CSX Corp.	Industrials	0.13	24.1
D. R. Horton	Consumer Discretionary	0.07	21.1
Frontier Communications	Telecommunications Services	0.98	4.1
Gap (The)	Consumer Discretionary	-0.04	36.7
Marriott Int'l.	Consumer Discretionary	-0.11	43.5

- The portfolio price is volume weighted average price
  - If we want to buy N shares of the portfolio, we need to buy 0.13N shares of CSX, buy 0.07N shares of DHI, buy 0.98N shares of FTR, sell 0.04N shares of GPS and sell 0.11 shares of MAR.
  - Although the weight of FTR seems to be large, it is only because the price of FTR is significantly lower than other stocks. Total value of FTR does not dominate values of other stocks.



Case Study – S&P 500 Stocks



Case Study – S&P 500 Stocks 2010/01/01-2014/12/31

