Implementation of Statistical Arbitrage using PCA Factorization Approach

MTH9845 Algorithmic Trading

Yuxi (Asana) Liu

Shengquan Zhou

Dong Niu

CONTENTS

OVERVIEW OF STATISTICAL ARBITRAGE

Baruch College CUNY

PAIRS TRADING → **RISK FACTORS** → **OU RESIDUAL** → **SIGNAL GENERATION** → **BACK TESTING**

$$\frac{\mathrm{d}P_t}{P_t} = \alpha\,\mathrm{d}t + \sum_{j=1}^{n} \beta_j F_t^{(j)} + \mathrm{d}X_t$$

- Proper benchmarks as risk factors
- Market Neutrality
- **Drift is mean-reverting**
- α term is neglected

- Risk factor decomposition:

$$\frac{\mathrm{d}P_t}{P_t} = \alpha\,\mathrm{d}t + \boxed{\sum_{j=1}^{n} \beta_j F_t^{(j)}} + \mathrm{d}X_t$$

- β neutral:

$$\overline{\beta}_j = \sum_{i=1}^{N} \beta_{ij} Q_i = 0, \quad j = 1, 2, \ldots, m$$

- Can be done by PCA or ETF approach

- Each X_t is modeled as OU process (mean-reverting).
- The parameters are specific for each stock.
- Estimation is done by AR(1) model.

- *s-score* $s_i = \dfrac{X_i(t) - m_i}{\sigma_{eq,i}}$
- Trading Signals:

$$\begin{aligned}
\text{buy to open if } & s_i < -\bar{s}_{bo}\\
\text{sell to open if } & s_i > +\bar{s}_{so}\\
\text{close short position if } & s_i < +\bar{s}_{bc}\\
\text{close long position } & s_i > -\bar{s}_{sc}
\end{aligned}$$

PCA Factorization Approach

Correlation matrix computation

Stock returns: $R_{ik} = \dfrac{S_{i(t_0-(k-1)\Delta t)} - S_{i(t_0-k\Delta t)}}{S_{i(t_0-k\Delta t)}}$

Standardized returns: $Y_{ik} = \dfrac{R_{ik} - \overline{R}_i}{\overline{\sigma}_i}$

Empirical correlation: $\rho_{ij} = \dfrac{1}{M-1}\sum_{k=1}^{M} Y_{ik} Y_{jk}$

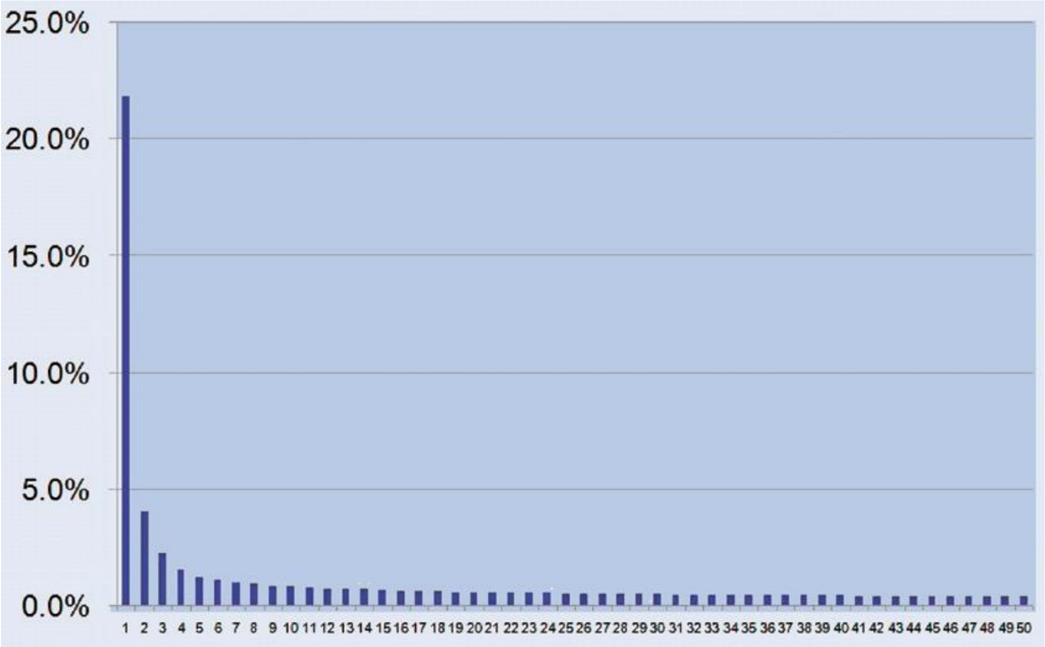
Eigenportfolio

Amount invested in each stocks: $Q_i^{(j)} = v_i^{(j)}/\overline{\sigma}_i$

Eigenportfolio returns: $F_{jk} = \sum_{i=1}^{N} \dfrac{v_i^{(j)}}{\overline{\sigma}_i} R_{ik}, \quad j = 1, 2, \ldots, m$

PCA factorization example

Top 50 eigenvalues of the correlation matrix of market returns computed on 5/1/2007 using a one-year window and a universe of 1417 stocks



Stock Selection and Residual Process Estimation

Stock selection

Rule: stocks with mean-reversion times less than 0.5 period (κ > 252/30 = 8.4)

Residual process estimation

Target OU process:

$$dX_i(t) = \kappa_i(m_i - X_i(t))dt + \sigma_i\, dW_i(t), \quad \kappa_i > 0$$

Regression estimation

$$R_n^S = \beta_0 + \beta R_n^I + \epsilon_n,$$
$$n = 1, 2, \ldots, 60$$

Auxiliary process

$$X_k = \sum_{j=1}^{k} \epsilon_j, \quad k = 1, 2, \ldots, 60$$

AR(1) estimation

$$X_{n+1} = a + bX_n + \zeta_{n+1},$$
$$n = 1, \ldots, 59$$

Parameter estimation

$$a = m(1 - e^{-\kappa\Delta t})$$
$$b = e^{-\kappa\Delta t}$$
$$\text{Variance}(\zeta) = \sigma^2 \frac{1 - e^{-2\kappa\Delta t}}{2\kappa}$$
$$\kappa = -\log(b) \times 252$$
$$m = \frac{a}{1-b}$$
$$\sigma = \sqrt{\frac{\text{Variance}(\zeta) \times 2\kappa}{1 - b^2}}$$
$$\sigma_{eq} = \sqrt{\frac{\text{Variance}(\zeta)}{1 - b^2}}.$$

***S*-score computation**

$$s = \frac{X(t) - m}{\sigma_{eq}}$$

Implementation of Statistical Arbitrage

Preset parameters

Stock pool: SP400 + SP500

Estimation window (corr.): 252 days

Estimation window (resid.): 60 days

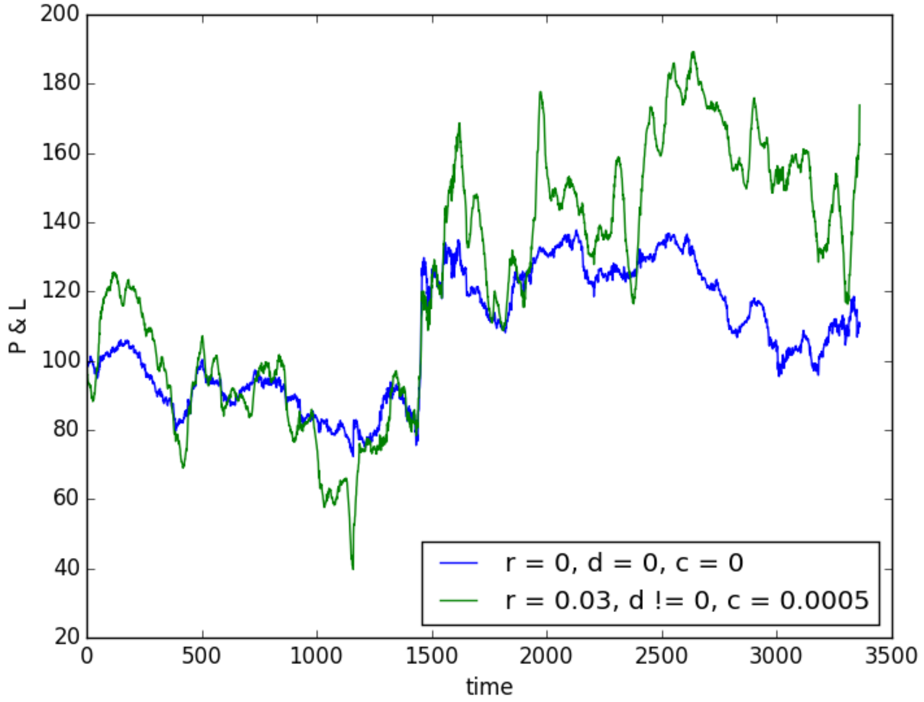
Trading signal cutoffs: $\bar{s}_{bo} = \bar{s}_{so} = 1.25$

$\bar{s}_{bc} = 0.75 \quad \text{and} \quad \bar{s}_{sc} = 0.50$

Turnover: 1 day

Open amount: 1 dollar

P&L results



Trading factors

Transaction cost	Sharpe ratio	Max trough down	Cum. return
—	0.0528	34.91%	9.84%
◯	0.0754	70.76%	73.82%

Conclusions and Improvements

- We have successfully implemented statistic arbitrage using PCA factorization approach.
- The very small turnover makes the transaction costs extremely high, which restricts the profits generating by the strategy.
- S&P 500 index hedge needs to be added to the portfolio.
- It might be worthy to readjust the PCA factors every one year.

Q & A
Thank you for listening to us!