Large Precision Matrix Estimation for Time Series Data With Latent Factor Model

Clifford Lam

Department of Statistics
London School of Economics and Political Science

Sparsity in Machine Learning and Statistics
April 1, 2009

Outline of the Talk

- Part I
  - Introduction
    - The need for $p \to \infty$ as $n \to \infty$ asymptotics.
  - Factor Models and Scientific Applications.
  - Some Literatures.
- Part II
  - Estimation and Identifiability.
  - Numerical Examples.
  - Results for Estimation of Factors and Precision Matrix.
  - Summary and Further Researches.

Outline of the Talk

- Part I
  - Introduction
    - The need for $p \to \infty$ as $n \to \infty$ asymptotics.
  - Factor Models and Scientific Applications.
  - Some Literatures.
- Part II
  - Estimation and Identifiability.
  - Numerical Examples.
  - Results for Estimation of Factors and Precision Matrix.
  - Summary and Further Researches.

Part I
Introduction

Why $p \to \infty$ as $n \to \infty$?

- When dimension p is large compare to sample size n, $n \to \infty$ alone not realistic.
  - e.g. $p = 10000, n = 100 \to$ should model $p = O(n^2)$.
  - e.g. $p = 200, n = 300 \to$ should model $p = O(n)$.

- Various scientific applications have this scenario.
  - **Finance**:
    $\to$ Stocks (p) in thousands, time points data observed (n) in hundreds or thousands. $p = O(n)$.
    $\to$ For volatility matrix or the inverse, # parameters $= O(p^2)$.
  - **Bioinformatics**:
    $\to$ Gene expressions data, genes $p \sim 10^4$ to 10^6. Number of patients n in hundreds.
  - **Image Analysis**:
    $\to$ Number of pixels in an image $p \sim 10^4$ to 10^6 depends on resolution, number of images n in hundreds.

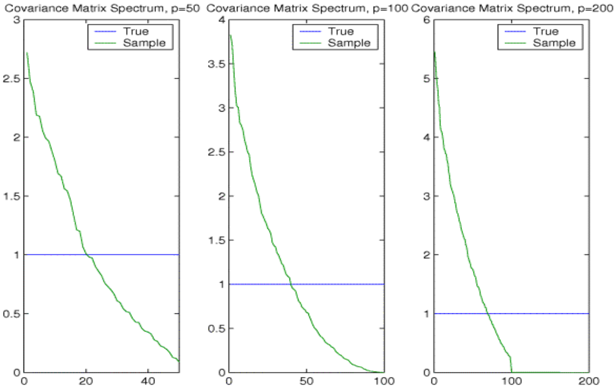
Why $p \to \infty$ as $n \to \infty$?

- When dimension p is large compare to sample size n, $n \to \infty$ alone not realistic.
  - e.g. $p = 10000, n = 100$ $\to$ should model $p = O(n^2)$.
  - e.g. $p = 200, n = 300$ $\to$ should model $p = O(n)$.

- Various scientific applications have this scenario.
  - **Finance**:
    $\to$ Stocks (p) in thousands, time points data observed (n) in hundreds or thousands. $p = O(n)$.
    $\to$ For volatility matrix or the inverse, # parameters $= O(p^2)$.
  - **Bioinformatics**:
    $\to$ Gene expressions data, genes $p \sim 10^4$ to 10^6. Number of patients n in hundreds.
  - **Image Analysis**:
    $\to$ Number of pixels in an image $p \sim 10^4$ to 10^6 depends on resolution, number of images n in hundreds.

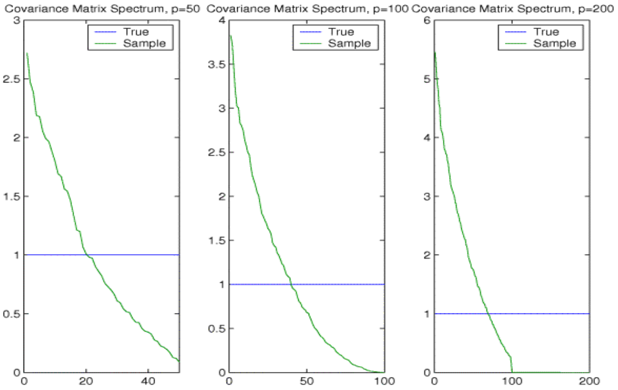
Effect of High-dimensionality

- A random sample of size 100 from $N(\mathbf{0}, \mathbf{I})$. We estimate the sample covariance matrix.



Effect of High-dimensionality

- A random sample of size 100 from $N(\mathbf{0}, \mathbf{I})$. We estimate the sample covariance matrix.



Sample covariance matrix can be extremely noisy in L_2 norm, even if element-wise consistency always holds.

Dimension Reduction - Factor Model

- A factor model has the form

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t, \ t = 1, \cdots, n,$$

where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors with (hopefully) $r \ll p$, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise.

- Multi-factor model originated from study of general intelligence factor in psychometrics. Has wide economic and finance applications.

- Essentially reduce the variation of a p-dimensional vector to the variation of r factors.

- Estimation errors not studied well for covariance or precision matrix estimation when p grows with n.

Dimension Reduction - Factor Model

- A factor model has the form

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t,\ t = 1, \cdots, n,$$

  where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors with (hopefully) $r \ll p$, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise.

- Multi-factor model originated from study of general intelligence factor in psychometrics. Has wide economic and finance applications.

- Essentially reduce the variation of a p-dimensional vector to the variation of r factors.

- Estimation errors not studied well for covariance or precision matrix estimation when p grows with n.

Dimension Reduction - Factor Model

- A factor model has the form

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t, \; t = 1, \cdots, n,$$

where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors with (hopefully) $r \ll p$, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise.

- Multi-factor model originated from study of general intelligence factor in psychometrics. Has wide economic and finance applications.

- Essentially reduce the variation of a p-dimensional vector to the variation of r factors.

- Estimation errors not studied well for covariance or precision matrix estimation when p grows with n.

Dimension Reduction - Factor Model

- A factor model has the form

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t,\ t = 1, \cdots, n,$$

where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors with (hopefully) $r \ll p$, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise.

- Multi-factor model originated from study of general intelligence factor in psychometrics. Has wide economic and finance applications.

- Essentially reduce the variation of a p-dimensional vector to the variation of r factors.

- Estimation errors not studied well for covariance or precision matrix estimation when p grows with n.

Literatures on Factor Models

- Classics:
  - APT and CAPM (Ross 1976, 1977).
  - Three-factor model (Fama and French 1993).
  - Many many others.

Literatures on Factor Models

- Classics:
  - APT and CAPM (Ross 1976, 1977).
  - Three-factor model (Fama and French 1993).
  - Many many others.
- Recent papers:
  - Element-wise consistency in Volatility matrices through factor modelling (Wang *et al.* 2008).
  - Non-stationary unknown factors, p fixed (Pan and Yao 2008).
  - Non-stationary unknown factors, p fixed (Pẽna and Poncela 2006).
  - Estimating number of factors under weak assumptions (Bai and Ng 2002).
  - Many others.

Literatures on Factor Models

- Classics:
  - APT and CAPM (Ross 1976, 1977).
  - Three-factor model (Fama and French 1993).
  - Many many others.

- Recent papers:
  - Element-wise consistency in Volatility matrices through factor modelling (Wang *et al.* 2008).
  - Non-stationary unknown factors, p fixed (Pan and Yao 2008).
  - Non-stationary unknown factors, p fixed (Pẽna and Poncela 2006).
  - Estimating number of factors under weak assumptions (Bai and Ng 2002).
  - Many others.

- Concerned with $p \to \infty$ as $n \to \infty$:
  - Johnstone and Lu 2006. Essentially simple one-factor model, done with sparse PCA.
  - Fan *et al.* 2008. They assumed factors are known in advance.

Part II
Estimation and Identifiability

The Factor Model - Assumptions

The model is

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t,\ t = 1, \cdots, n,$$

where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise. Furthermore, assume $\text{Var}(\boldsymbol{\epsilon}_t) = \sigma^2\mathbf{I}$, and

$$\text{Cov}(\boldsymbol{\epsilon}_t, \mathbf{X}_s) = \mathbf{0}\ \forall\ t \text{ and } s.$$

The Factor Model - Assumptions

The model is

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\epsilon}_t,\ t = 1, \cdots, n,$$

where $\mathbf{Y}_t \in \mathbb{R}^p$, $\mathbf{A}$ is a constant $p \times r$ matrix of factor loadings, $\mathbf{X}_t \in \mathbb{R}^r$ is a vector of r factors, and $\boldsymbol{\epsilon}$ is a p-vector of idiosyncratic noise. Furthermore, assume $\text{Var}(\boldsymbol{\epsilon}_t) = \sigma^2\mathbf{I}$, and

$$\text{Cov}(\boldsymbol{\epsilon}_t, \mathbf{X}_s) = \mathbf{0}\ \forall\ t \text{ and } s.$$

- Assumption $\text{Var}(\boldsymbol{\epsilon}_t) = \sigma^2\mathbf{I}$ allows us to have unbiased estimation of error variance. It can be relaxed to

$$\text{Var}(\boldsymbol{\epsilon}_t) = \text{diag}(\sigma_1^2\mathbf{I}_1 \cdots \sigma_k^2\mathbf{I}_k),$$

  with rate of k specified later.
- Assume $\mathbf{X}_t$ is stationary. See further researches.

The Factor Model - Assumptions

- One key assumption is, for all i,

$$\|\mathbf{a}_i\| = O(p^{1/2}),$$

where $\mathbf{A} = (\mathbf{a}_1 \cdots \mathbf{a}_r)$.

- Each factor is shared by the majority of the cross-sectional data points.

The Factor Model - Estimation

The factor model implies

$$\begin{aligned}\mathbf{\Sigma}_y(k) &= \mathsf{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) \\ &= \mathbf{A}\mathsf{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^T \\ &= \mathbf{A}\mathbf{\Sigma}_x(k)\mathbf{A}^T,\ k \geq 1.\end{aligned}$$

The Factor Model - Estimation

The factor model implies

$$\begin{aligned}\mathbf{\Sigma}_y(k) &= \mathsf{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) \\ &= \mathbf{A}\mathsf{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^T \\ &= \mathbf{A}\mathbf{\Sigma}_x(k)\mathbf{A}^T,\ k \geq 1.\end{aligned}$$

If $\mathbf{B}$ of size $p \times (p-r)$ is an orthogonal complement of $\mathbf{A}$, then

$$\mathbf{B}^T\mathbf{\Sigma}_y(k) = \mathbf{0},\ \ k = 1, \cdots, k_0.$$

The Factor Model - Estimation

The factor model implies

$$\begin{aligned}\mathbf{\Sigma}_y(k) &= \mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) \\ &= \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^T \\ &= \mathbf{A}\mathbf{\Sigma}_x(k)\mathbf{A}^T,\ k \geq 1.\end{aligned}$$

If $\mathbf{B}$ of size $p \times (p-r)$ is an orthogonal complement of $\mathbf{A}$, then

$$\mathbf{B}^T\mathbf{\Sigma}_y(k) = \mathbf{0},\ \ k = 1, \cdots, k_0.$$

Hence for $\mathbf{L} = \sum_{k=1}^{k_0} \mathbf{\Sigma}_y(k)\mathbf{\Sigma}_y(k)^T$, we have

$$\mathbf{L}\mathbf{B} = \mathbf{0}.$$

This implies that r is the number of non-zero eigenvalues of $\mathbf{L}$.

The Factor Model - Estimation

We estimate a version of $\mathbf{A}$ by $\hat{\mathbf{A}}$, which is a matrix of r unit eigenvectors corresponding to the largest r eigenvalues of $\tilde{\mathbf{L}}$,

$$\tilde{\mathbf{L}} = \sum_{k=1}^{k_0} \tilde{\boldsymbol{\Sigma}}_y(k)\tilde{\boldsymbol{\Sigma}}_y(k)^T,$$

where $\tilde{\boldsymbol{\Sigma}}_y(k) = (n-k)^{-1}\sum_{t=1}^{n-k}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t+k} - \bar{\mathbf{Y}})^T$.

The Factor Model - Estimation

We estimate a version of $\mathbf{A}$ by $\hat{\mathbf{A}}$, which is a matrix of r unit eigenvectors corresponding to the largest r eigenvalues of $\tilde{\mathbf{L}}$,

$$\tilde{\mathbf{L}} = \sum_{k=1}^{k_0} \tilde{\boldsymbol{\Sigma}}_y(k)\tilde{\boldsymbol{\Sigma}}_y(k)^T,$$

where $\tilde{\boldsymbol{\Sigma}}_y(k) = (n-k)^{-1}\sum_{t=1}^{n-k}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t+k} - \bar{\mathbf{Y}})^T$.

The factor $\mathbf{X}_t$ is estimated by

$$\hat{\mathbf{X}}_t = \hat{\mathbf{A}}^T\mathbf{Y}_t,$$

and the error by

$$\mathbf{e}_t = (\mathbf{I} - \hat{\mathbf{A}}\hat{\mathbf{A}}^T)\mathbf{Y}_t.$$

The Factor Model - Identifiability of $\mathbf{A}$

Not identifiable since $\mathbf{AX}_t = \mathbf{AHH}^{-1}\mathbf{X}_t$ for any invertible $\mathbf{H}$.

- With the QR decomposition on $\mathbf{A}$, where $\mathbf{A} = \mathbf{QR}$ with $\mathbf{Q}$ a $p \times r$ matrix such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_r$, and $\mathbf{R}$ an upper triangular matrix,

$$\mathbf{Y}_t = \mathbf{QX}_t' + \boldsymbol{\epsilon}_t,$$

where $\mathbf{X}_t' = \mathbf{RX}_t$.

The Factor Model - Identifiability of $\mathbf{A}$

Not identifiable since $\mathbf{A}\mathbf{X}_t = \mathbf{A}\mathbf{H}\mathbf{H}^{-1}\mathbf{X}_t$ for any invertible $\mathbf{H}$.

- With the QR decomposition on $\mathbf{A}$, where $\mathbf{A} = \mathbf{Q}\mathbf{R}$ with $\mathbf{Q}$ a $p \times r$ matrix such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_r$, and $\mathbf{R}$ an upper triangular matrix,
$$\mathbf{Y}_t = \mathbf{Q}\mathbf{X}'_t + \boldsymbol{\epsilon}_t,$$
where $\mathbf{X}'_t = \mathbf{R}\mathbf{X}_t$.
- With the assumption $\|\mathbf{a}_i\| = O(p^{1/2})$ for all i, and noting that $\mathbf{R}$ has i-th diagonal elements $\|\mathbf{a}_i\|$, for $i = 1, \cdots, r$, $t = 1, \cdots, n$ we have
$$X'_{ti} = O_P(p^{1/2}), \; (\boldsymbol{\Sigma}_x(k))_{ls} = \mathrm{Cov}(X'_{tl}, X'_{(t+k)s}) = O(p).$$

The Factor Model - Identifiability of $\mathbf{A}$

Not identifiable since $\mathbf{A}\mathbf{X}_t = \mathbf{A}\mathbf{H}\mathbf{H}^{-1}\mathbf{X}_t$ for any invertible $\mathbf{H}$.

- With the QR decomposition on $\mathbf{A}$, where $\mathbf{A} = \mathbf{Q}\mathbf{R}$ with $\mathbf{Q}$ a $p \times r$ matrix such that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_r$, and $\mathbf{R}$ an upper triangular matrix,

$$\mathbf{Y}_t = \mathbf{Q}\mathbf{X}'_t + \boldsymbol{\epsilon}_t,$$

  where $\mathbf{X}'_t = \mathbf{R}\mathbf{X}_t$.

- With the assumption $\|\mathbf{a}_i\| = O(p^{1/2})$ for all i, and noting that $\mathbf{R}$ has i-th diagonal elements $\|\mathbf{a}_i\|$, for $i = 1, \cdots, r$, $t = 1, \cdots, n$ we have

$$X'_{ti} = O_P(p^{1/2}),\ (\boldsymbol{\Sigma}_x(k))_{ls} = \mathrm{Cov}(X'_{tl}, X'_{(t+k)s}) = O(p).$$

- WLOG with $\|\mathbf{a}_i\| = O(p^{1/2})$ for the original factor, can assume the factor model has $\mathbf{A}$ and $\mathbf{X}_t$ satisfying

$$\mathbf{A}^T\mathbf{A} = \mathbf{I}_r,\ X_{ti} = O_P(p^{1/2}).$$

The Factor Model - Identifiability of $\mathbf{A}$

The estimation procedure actually identifies a version of $\mathbf{A}$ as follows. We have

$$\mathbf{L} = \sum_{k=1}^{k_0} \mathbf{\Sigma}_y(k)\mathbf{\Sigma}_y(k)^T, \text{ with } \mathbf{\Sigma}_y(k) = \mathbf{A}\mathbf{\Sigma}_x(k)\mathbf{A}^T \text{ for } k = 1, \cdots, k_0.$$

Then $\mathbf{L}\mathbf{A} = \mathbf{A}\sum_{k=1}^{k_0} \mathbf{\Sigma}_x(k)\mathbf{\Sigma}_x(k)^T = \mathbf{A}\mathbf{Q}_x\mathbf{D}_x\mathbf{Q}_x^T$ for $\mathbf{D}_x > \mathbf{0}$ (i.e. $\mathbf{A}$ are the r unit eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{L}$). Hence

$$\mathbf{L}\mathbf{A}(\pm\mathbf{Q}_x) = \mathbf{A}(\pm\mathbf{Q}_x)\mathbf{D}_x.$$

The factor model we want to estimate is then

$$\mathbf{Y}_t = \underbrace{(\pm\mathbf{A}\mathbf{Q}_x)}_{\text{new } \mathbf{A}}\underbrace{(\pm\mathbf{Q}_x^T\mathbf{X}_t)}_{\text{new } \mathbf{X}_t} + \boldsymbol{\epsilon}_t.$$

The Factor Model - Identifiability of $\mathbf{A}$

The estimation procedure actually identifies a version of $\mathbf{A}$ as follows. We have

$$\mathbf{L} = \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)^T, \text{ with } \boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}^T \text{ for } k = 1, \cdots, k_0.$$

Then $\mathbf{LA} = \mathbf{A}\sum_{k=1}^{k_0} \boldsymbol{\Sigma}_x(k)\boldsymbol{\Sigma}_x(k)^T = \mathbf{AQ}_x\mathbf{D}_x\mathbf{Q}_x^T$ for $\mathbf{D}_x > \mathbf{0}$ (i.e. $\mathbf{A}$ are the r unit eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{L}$). Hence

$$\mathbf{LA}(\pm\mathbf{Q}_x) = \mathbf{A}(\pm\mathbf{Q}_x)\mathbf{D}_x.$$

The factor model we want to estimate is then

$$\mathbf{Y}_t = \underbrace{(\pm\mathbf{AQ}_x)}_{\text{new } \mathbf{A}}\underbrace{(\pm\mathbf{Q}_x^T\mathbf{X}_t)}_{\text{new } \mathbf{X}_t} + \boldsymbol{\epsilon}_t.$$

- Can always find a version of $\mathbf{A}$ close to our estimate (arranging eigenvalues in $\mathbf{D}_x$, or multiplied by $\pm$ sign).

Simulations and Theoretical Results

Toy Example

We simulate a one-factor model with factor $(\mathbf{A})_i = 2\cos(2\pi i/p)$, $X_t = 0.9X_{t-1} + \eta_t$, $i = 1, \cdots, p$, $t = 1, \cdots, n$ where $\eta_t \sim N(0, 2^2)$ and is independent of all other variables. The model is

$$\mathbf{Y}_t = \mathbf{A}X_t + \boldsymbol{\epsilon}_t,\ \ \epsilon_{tj} \sim \text{ i.i.d. } N(0, 2^2).$$

- Number of simulations = 50. Sample size $n = 200, 500$.
- Results for $p = 20, 180, 400, 1000$.

Toy Example

$n = 200$	$\|\hat{\mathbf{A}} - \mathbf{A}_x\|$	$\|\mathbf{S}^{-1} - \boldsymbol{\Sigma}_y(0)^{-1}\|$	$\|\hat{\boldsymbol{\Sigma}}_y(0)^{-1} - \boldsymbol{\Sigma}_y(0)^{-1}\|$
$p = 20$	.022(.005)	.24(.03)	.009(.002)
$p = 180$	.023(.004)	79.8(29.8)	.007(.001)
$p = 400$	.022(.004)	-	.007(.001)
$p = 1000$	.023(.004)	-	.007(.001)

$n = 200$	$\|\mathbf{S} - \boldsymbol{\Sigma}_y(0)\|$	$\|\hat{\boldsymbol{\Sigma}}_y(0) - \boldsymbol{\Sigma}_y(0)\|$
$p = 20$	218(165)	218(165)
$p = 180$	1962(1500)	1963(1500)
$p = 400$	4102(3472)	4103(3471)
$p = 1000$	10797(6820)	10800(6818)

Toy Example

$n=500$	$\|\hat{\mathbf{A}}-\mathbf{A}_x\|$	$\|\mathbf{S}^{-1}-\mathbf{\Sigma}_y(0)^{-1}\|$	$\|\hat{\mathbf{\Sigma}}_y(0)^{-1}-\mathbf{\Sigma}_y(0)^{-1}\|$
$p=20$	.014(.003)	.12(.02)	.005(.002)
$p=180$	.014(.002)	1.23(.06)	.004(.001)
$p=400$	.014(.002)	20.0(1.9)	.004(.001)
$p=1000$	.014(.002)	-	.004(.001)

$n=500$	$\|\mathbf{S}-\mathbf{\Sigma}_y(0)\|$	$\|\hat{\mathbf{\Sigma}}_y(0)-\mathbf{\Sigma}_y(0)\|$
$p=20$	136(95)	136(95)
$p=180$	1344(1108)	1344(1108)
$p=400$	3134(2561)	3134(2561)
$p=1000$	7696(5528)	7696(5528)

Some Results - Factor Loadings Matrix

Weak convergence in L_2 norm for $\hat{\mathbf{A}}$

With all assumptions valid, there is a version of $\mathbf{A}$ (say $\mathbf{A}_x$) such that

$$\|\hat{\mathbf{A}} - \mathbf{A}_x\| = O_P(rn^{-l_{xx}} + r^{1/2}n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}),$$

where $0 < l_{xx}, l_{x\epsilon}, l_{\epsilon\epsilon} \leq 1/2$.

Some Results - Factor Loadings Matrix

Weak convergence in L_2 norm for $\hat{\mathbf{A}}$

With all assumptions valid, there is a version of $\mathbf{A}$ (say $\mathbf{A}_x$) such that

$$\|\hat{\mathbf{A}} - \mathbf{A}_x\| = O_P(rn^{-l_{xx}} + r^{1/2}n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}),$$

where $0 < l_{xx}, l_{x\epsilon}, l_{\epsilon\epsilon} \leq 1/2$.

- Since we assumed the original factor has $\|\mathbf{a}_i\| = O(p^{1/2})$, all cross-sectional observations share the factors in a non-trivial way.
- Var$(\boldsymbol{\epsilon}_t)$ can be a general covariance matrix here.
- Higher dimension means more data around a particular factor level in the re-defined factor space (confined in $[-1,1]^r$), hence more accurate estimation.

Some Results - Precision Matrix for $\mathbf{Y}_t$

- We estimate $\hat{\sigma}^2 = (np)^{-1}\|\hat{\mathbf{E}}\|_F^2$, where $\mathbf{E} = (\mathbf{e}_1 \cdots \mathbf{e}_n)$.
- We estimate $\hat{\boldsymbol{\Sigma}}_x(0) = \hat{\mathbf{A}}^T\tilde{\boldsymbol{\Sigma}}_y(0)\hat{\mathbf{A}} - \hat{\sigma}^2\mathbf{I}$.

Weak convergence in L_2 norm for $\hat{\boldsymbol{\Sigma}}_y(0)^{-1}$

With all assumptions valid, by estimating

$$\hat{\boldsymbol{\Sigma}}_y(0) = \hat{\mathbf{A}}\hat{\boldsymbol{\Sigma}}_x(0)\hat{\mathbf{A}}^T + \hat{\sigma}^2\mathbf{I},$$

we have

$$\|\hat{\boldsymbol{\Sigma}}_y(0)^{-1} - \boldsymbol{\Sigma}_y(0)^{-1}\| = O_P(r^{3/2}n^{-l_{xx}} + rn^{-l_{x\epsilon}} + r^{1/2}n^{-l_{\epsilon\epsilon}}),$$

where $0 < l_{xx}, l_{x\epsilon}, l_{\epsilon\epsilon} \leq 1/2$.

- For known factors, Fan *et al.* 2006 has established

$$\|\hat{\boldsymbol{\Sigma}}_y^{-1} - \boldsymbol{\Sigma}_y^{-1}\|_F = o_P((p^2r^4\log n/n)^{1/2}).$$

If the error variance is

$$\mathsf{Var}(\boldsymbol{\epsilon}_t) = \mathsf{diag}(\sigma_1^2 \mathbf{I}_1 \cdots \sigma_k^2 \mathbf{I}_k)$$

with the size of each identity matrix $\mathbf{I}_i$ being $O(s)$, then:

Weak convergence in L_2 norm for $\hat{\boldsymbol{\Sigma}}_y(0)^{-1}$

With all assumptions hold, and $\hat{\boldsymbol{\Sigma}}_x(0) = \hat{\mathbf{A}}^T \tilde{\boldsymbol{\Sigma}}_y(0) \hat{\mathbf{A}} - \hat{\mathbf{A}}^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} \hat{\mathbf{A}}$,

$$\|\hat{\boldsymbol{\Sigma}}_y(0)^{-1} - \boldsymbol{\Sigma}_y(0)^{-1}\| = O_P\Big(\sqrt{\frac{pr}{s}} h_n\Big),$$

where $h_n = rn^{-l_{xx}} + r^{1/2} n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}$.

- The order actually does not depend on p even when $s = 1$, but it may not be going to zero either.

Result for the Covariance Matrix

From the simulation is is seen that factor model does not lead to better estimate than sample covariance matrix. Indeed:

Rate of Convergence for $\hat{\boldsymbol{\Sigma}}_y(0)$

With previous assumptions hold,

$$\|\hat{\boldsymbol{\Sigma}}_y(0) - \boldsymbol{\Sigma}_y(0)\| = O_P(ph_n) = \|\mathbf{S} - \boldsymbol{\Sigma}_y(0)\|,$$

where $h_n = rn^{-l_{xx}} + r^{1/2}n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}$.

Convergence of Markowitz Optimal Portfolio

The Markowitz optimal portfolio has the form

$$\boldsymbol{\xi} = a_n \boldsymbol{\Sigma}_y(0)^{-1}\mathbf{1} + b_n \boldsymbol{\Sigma}_y(0)^{-1}\boldsymbol{\mu}_n,$$

where $\boldsymbol{\mu} = E(\mathbf{Y}_t)$, a_n and b_n are constants depending on $\boldsymbol{\mu}_n$ and the target return.

Weak convergence for $\hat{\boldsymbol{\xi}}$

With all assumptions valid, by estimating $\boldsymbol{\mu}_n$ by $\bar{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_y(0)^{-1}$ by $\hat{\boldsymbol{\Sigma}}_y(0)^{-1}$, we have

$$p^{-1/2}\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\| = O_P\Big(\sqrt{\frac{pr}{s}}h_n\Big),$$

where $h_n = rn^{-l_{xx}} + r^{1/2}n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}$.

Convergence of the Variance of the Optimal Portfolio

The Markowitz optimal portfolio has variance

$$\mathbf{V}_{\boldsymbol{\xi}} = \boldsymbol{\xi}^T \boldsymbol{\Sigma}_y(0) \boldsymbol{\xi}.$$

Weak convergence for $\hat{\mathbf{V}}_{\boldsymbol{\xi}}$

With all assumptions valid, by estimating $\boldsymbol{\xi}$ by $\hat{\boldsymbol{\xi}}$ and $\boldsymbol{\Sigma}_y(0)^{-1}$ by $\hat{\boldsymbol{\Sigma}}_y(0)^{-1}$, we have

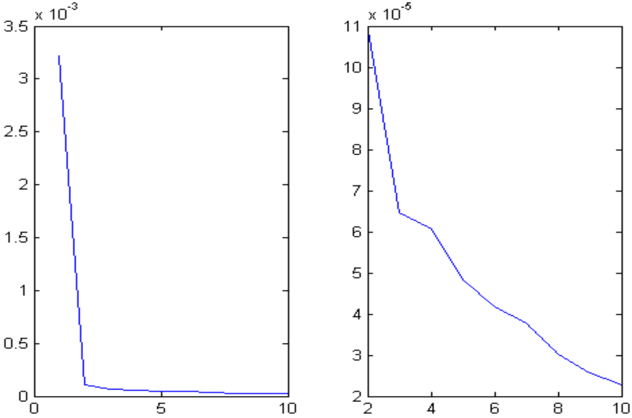
$$|\hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\Sigma}}_y(0) \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^T \boldsymbol{\Sigma}_y(0) \boldsymbol{\xi}| = O_P\left(p\sqrt{\frac{pr}{s}} h_n\right),$$

where $h_n = rn^{-l_{xx}} + r^{1/2} n^{-l_{x\epsilon}} + n^{-l_{\epsilon\epsilon}}$.

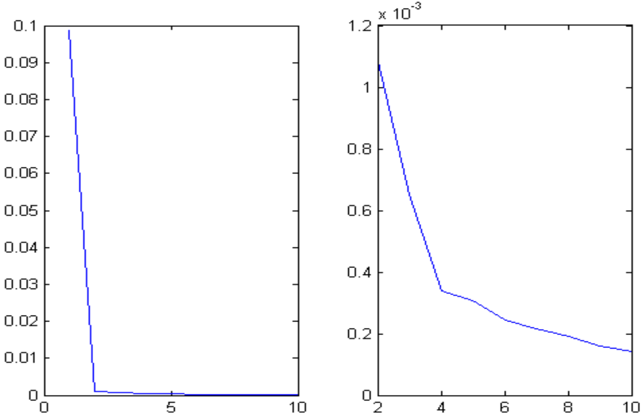
- $\mathbf{V}_{\boldsymbol{\xi}}$ actually depends on $\boldsymbol{\Sigma}_y(0)^{-1}$ only functionally, but not $\boldsymbol{\Sigma}_y(0)$ itself.

Stocks Data Illustration

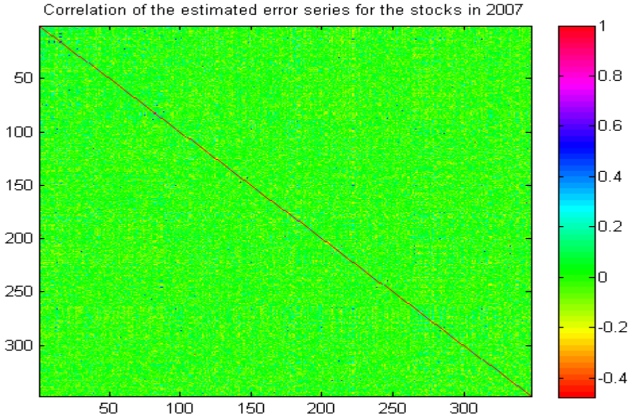
- NASDAQ 347 most traded stocks at the beginning of 2007.
- Application of the factor model for the year 2007, and 2008.



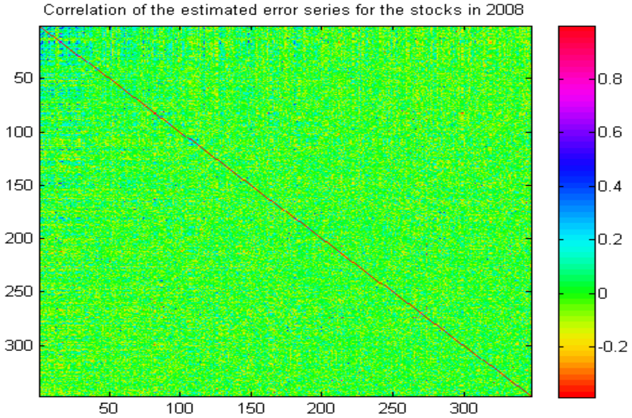
Left : 10 largest eigenvalues of the matrix $\tilde{L}$ for 2007
Right : Second to ninth largest eigenvalues.



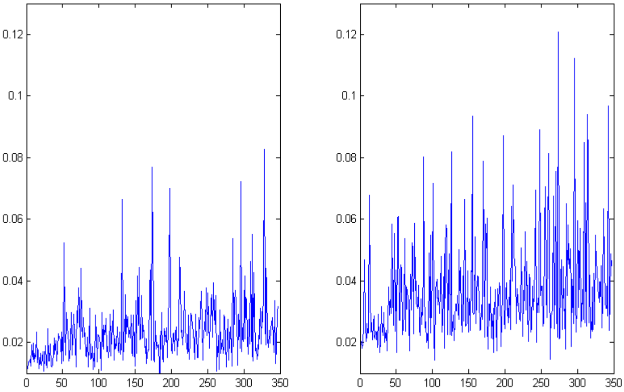
Left : 10 largest eigenvalues of the matrix $\tilde{L}$ for 2008
Right : Second to ninth largest eigenvalues.



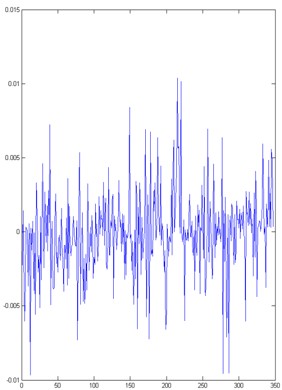
Heatmap of correlations for the estimated error series during 2007 for all 347 stocks.

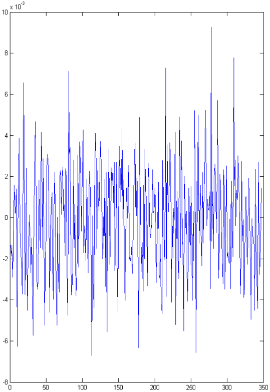


Heatmap of correlations for the estimated error series during 2008 for all 347 stocks.



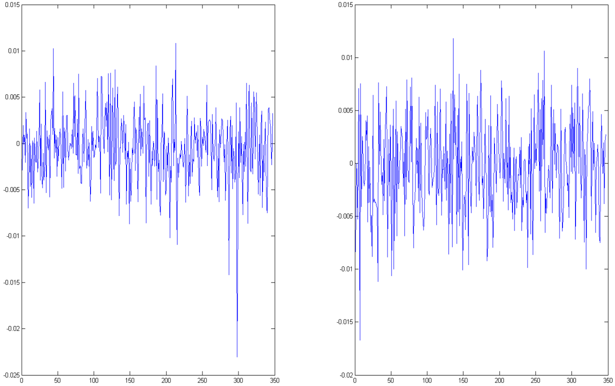
Left : SD for the estimated error of each stock in the factor model for 2007.
Right : SD for the estimated error of each stock in the factor model for 2008.





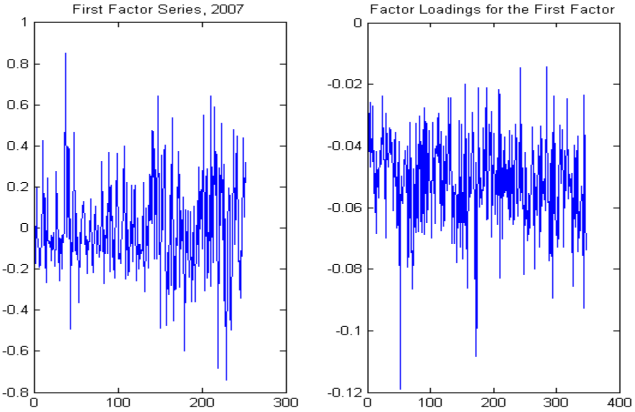
Left : Correlations between the first factor and the estimated error for each stock, 2007.

Right : Correlations between the second factor and the estimated error for each stock, 2007.

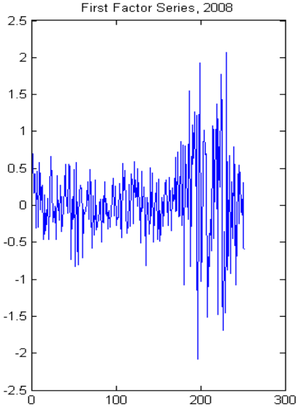


Left : Correlations between the first factor and the estimated error for each stock, 2008.

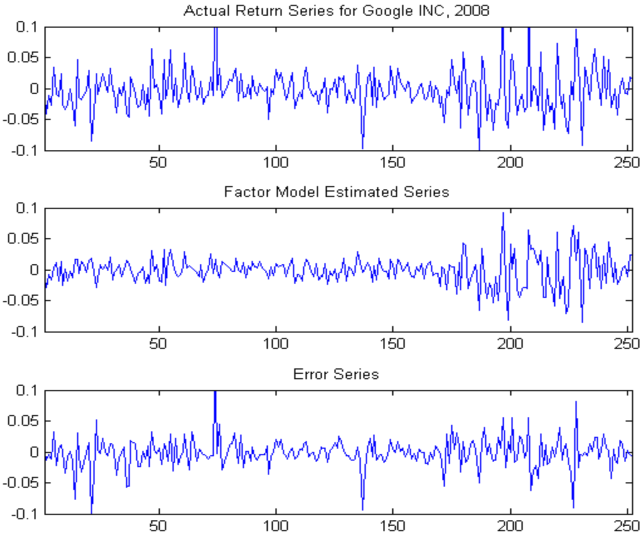
Right : Correlations between the second factor and the estimated error for each stock, 2008.



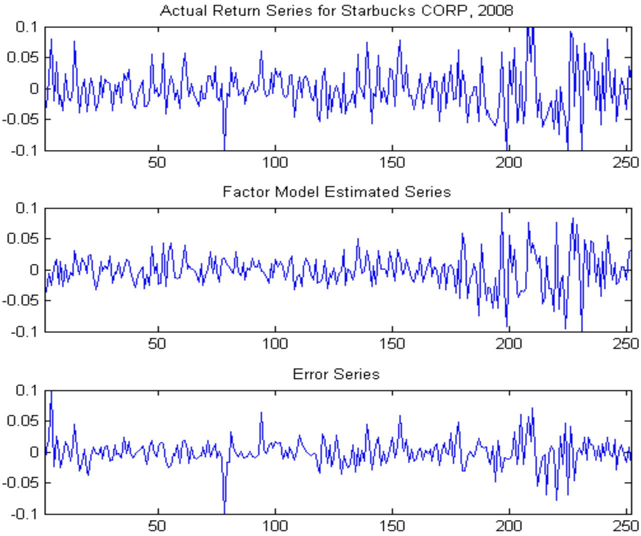
Left : First factor series, 2007.
Right : Corresponding factor loadings for each stock.



Left : First factor series, 2008.
Right : Corresponding factor loadings for each stock.



Actual Return Series for Google INC, 2008
Factor Model Estimated Series
Error Series
0.1
0.05
0
-0.05
-0.1
50
100
150
200
250


Actual Return Series for Starbucks CORP, 2008
0.1
0.05
0
-0.05
-0.1
50
100
150
200
250
Factor Model Estimated Series
0.1
0.05
0
-0.05
-0.1
50
100
150
200
250
Error Series
0.1
0.05
0
-0.05
-0.1
50
100
150
200
250

Remarks and Further Researches

- The covariance matrix $\mathbf{\Sigma}_y(0)$ is estimated no better than the sample covariance matrix, which behaves badly when p is large compare to n. Prediction, or risk assessment in finance, need this to be estimated accurately $\longrightarrow$ the need for further regularisation.

- The general result for estimation of $\mathbf{A}$ provides insights. Wavelet method can be used.

- Develop results for high dimensional non-stationary time series. Need in particular effect of convergence of generalised sample covariance matrices (Chen and Wei (1988) or Tanaka (1996) considered fixed dimension).

- Constant factor loadings matrix may not be adaptive enough to complex financial data. Time-varying $\mathbf{A}$ can potentially improve.

Thank You!