Algorithmic Trading & Quantitative Strategies

Lecture 4 (3/12/2024)

Giuseppe Paleologo (gardener)

Today's Session

- N.B.: a student suggested that I upload the slides in advance to get acquainted with material. I will try to do this from now on.
- Especially appropriate today. This is the second most mathematical lecture in the course.
- There will be a short multiple-answer quiz at the end of the lecture.
- Please send answers in the form
  - 1-a
  - 2-c
  - ...

  To gp2642@nyu.edu. You will have 5 minutes

Topics

- Statistical Models estimation
    - Basics. Geometry. SVD/PCA. Indeterminacy
    - Relationship with statistical estimation
    - Some practical considerations
- Plan for the rest of the course
    - Lecture 5: Finishing Statistical Models. Fundamental models, linking models
    - Lecture 6: advanced portfolio construction
    - Lecture 7: backtesting

Statistical Models

Why?

1. Minimal data requirements
2. Good performance
3. Complementarity
4. Theoretical foundations
5. Basis for alpha research

Why not?

1. Not using all the data
2. Interpretability

The Basics

As usual, returns matrix R (n x m), loadings matrix B (n x m), factor returns matrix (m x T). $||\cdot||_F$ is the Frobenius matrix.

Solve

$$\min_{\mathbf{B},\mathbf{F}} \|\mathbf{R} - \mathbf{BF}\|_F$$

The matrix has rank m. Every matrix with rank m can be written as BF. So rewrite

$$\min_{\text{rank}(\hat{\mathbf{R}}) \leq m} \left\|\mathbf{R} - \hat{\mathbf{R}}\right\|^2$$

Where I have not specified the norm type. Any unitarily invariant norm will do.

Solution

- Singular Value Decomposition $\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}'$.
- Truncate Singular Values $\hat{\mathbf{R}} = \mathbf{U}\mathbf{S}_m\mathbf{V}'$
- And set $\mathbf{B} = \mathbf{U}_m$

$$\mathbf{F} = \mathbf{S}_m\mathbf{V}'_m$$

- Related problem: PCA and eigenvalue problem $\max \mathbf{w}'\hat{\boldsymbol{\Sigma}}\mathbf{w}$

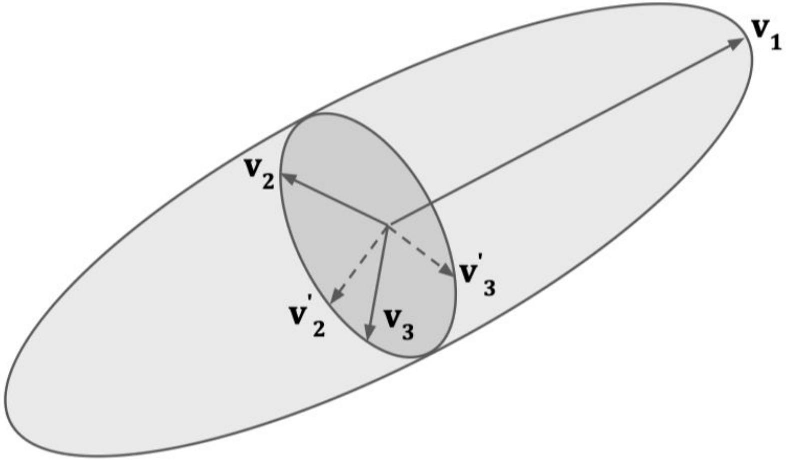
$$\text{s.t. } ||\mathbf{w}|| \leq 1$$

Relationship PCA/SVD

- Empirical Covariance Matrix $\hat{\mathbf{\Sigma}} = \frac{1}{T}\mathbf{R}\mathbf{R}' = \frac{1}{T}\mathbf{U}\mathbf{S}^2\mathbf{U}'$

- Rewrite problem
$$\max \mathbf{w}'\hat{\mathbf{\Sigma}}\mathbf{w} \quad \text{s.t. } ||\mathbf{w}|| \leq 1$$
$$\max \mathbf{v}'\mathbf{S}^2\mathbf{v} \quad \text{s.t. } ||\mathbf{w}|| \leq 1, \quad \mathbf{w} = \mathbf{U}\mathbf{v}, \quad \mathbf{v} \in \mathbb{R}^n$$

- Note that eigenvectors are not uniquely identified when eigenvalues are identical


$\mathbf{v}_1$
$\mathbf{v}_2$
$\mathbf{v}'_3$
$\mathbf{v}'_2$
$\mathbf{v}_3$

Maximum Likelihood and PCA

Recall the standard equation $\mathbf{r} = \mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}$

Assume $\mathbf{f} \sim N(0, \mathbf{I}_m)$ $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma}^2\mathbf{I}_n) \Rightarrow \mathbf{B}\mathbf{B}' + \sigma^2\mathbf{I}_n$

Compute the log likelihood

$$\mathcal{L}(\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}) = -\frac{T}{2}\left[\log\left|\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}\right| + \langle \hat{\boldsymbol{\Sigma}}_{\mathbf{r}}^{-1}, \boldsymbol{\Sigma}_{\mathbf{r}}\rangle + n\log(2\pi)\right]$$

Solve

$$\begin{aligned} \max \quad & -\log\left|\hat{\boldsymbol{\Sigma}}_{\mathbf{r}}\right| - \langle \hat{\boldsymbol{\Sigma}}_{\mathbf{r}}^{-1}, \boldsymbol{\Sigma}_{\mathbf{r}}\rangle \\ \text{s.t.} \quad & \hat{\boldsymbol{\Sigma}}_{\mathbf{r}} = \hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\sigma}^2\mathbf{I}_n \end{aligned} \quad \Rightarrow \quad \begin{aligned} \hat{\mathbf{B}} &= \mathbf{U}_m(\mathbf{S}_m^2 - \hat{\sigma}^2\mathbf{I}_n)^{1/2} \\ \hat{\sigma}^2 &= \bar{\lambda} \end{aligned}$$

Where lambda bar is the average of the last n-m eigenvalues of Sigma hat

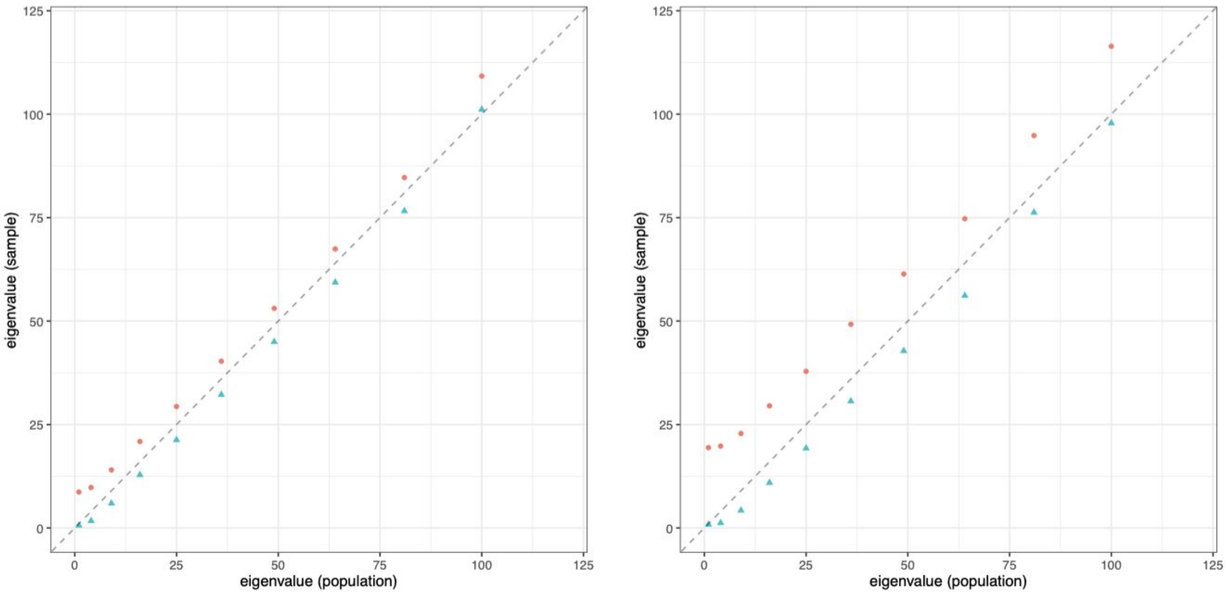
Alternative Solution

- Diagonal Factor Covariance matrix
- Shrinkage of the volatilities
- Asymptotically equal to PCA
- Eigenfactors are SVD

$$
\begin{aligned}
\mathbf{B} &= \mathbf{U}_m \\
\boldsymbol{\Sigma}_{\mathbf{f}} &= (\mathbf{S}_m^2 - \bar{\lambda}\mathbf{I}_n) \\
\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} &= \bar{\lambda}\mathbf{I}_n
\end{aligned}
$$

Performance: Shrinkage of the largest eigenvectors works

Left 1000 assets, Right 3000. Same 10 factors. 250 days
Circle: non-shrunk (PCA), Triangle Shrunk (PPCA)



The relationship between PCA, X-sectional regression and time series regression

1. Take the first m eigenvectors as loadings
2. Do a cross-sectional regression (like in fundamental models)
3. You will recover the factor returns from the PCA

1. Take the the factor returns
2. Do a time-series regression on the asset returns
3. You will recover the first m eigenvectors as loadings

PCA loadings and returns are the only ones that have this consistency property.

The Consistency Property, Cont.

Why does this matter in practice?

You do cross-sectional performance attribution in fundamental models, and get $0 PnL attributed to a factor

Then, you take time series PnL of that strategy and time-regress it against factor returns, and estimate a non-zero slope and non-zero PnL

Reconciling the two is just hard

Theory: The Spiked Covariance Model

Asymptotic behavior of empirical covariance matrix as T $\to\infty$: there is a fixed m such that

$$\lambda_i := \lim_{T\to\infty} \lambda_{T,i} \begin{cases} = 1 & \text{for all } i > m & \text{"bulk"} \\ > Cn & \text{for all } i \leq m & \text{"spike"} \end{cases}$$

So we have *pervasive risk* and *diversifiable risk*.

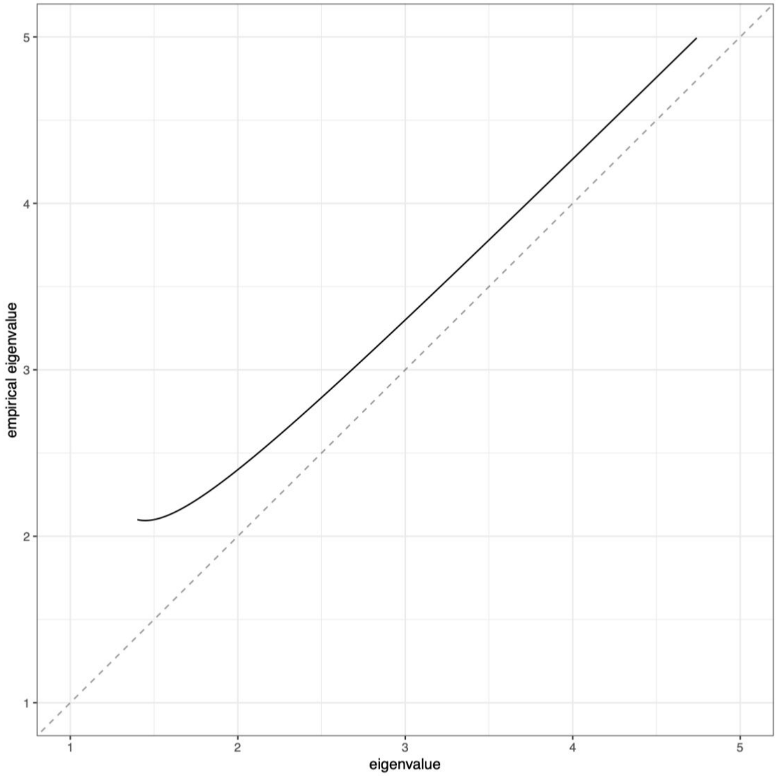
Asymptotic Behavior

Characterized 2000-onward. This is not Marchenko-Pastur (if you have heard of it), and M-P *does not* apply to covariance matrix estimation

Assume

1. Finite 4th moments
2. $\lambda_1 \geq \lambda_2 = \lambda_3 = \ldots > \lambda_n = 1$

Empirical Behavior



$$\gamma := n/T \in [0, \infty)$$

If $\lambda_i > 1 + \sqrt{\gamma}$ then $\hat{\lambda}_i \to \mu_i := \lambda_i \left(1 + \frac{\gamma}{\lambda_i - 1}\right)$

If $\lambda_i \leq 1 + \sqrt{\gamma}$ then $\hat{\lambda}_i \to (1 + \sqrt{\gamma})^2$

First case: eigenvectors converge to true eigenvectors

Second case: indeterminacy

More general results are in the notes

Optimal Shrinkage of Eigenvalues

This depends on the loss function! Simple inversion:

$$\ell(\lambda) = \frac{(\lambda + 1 - \gamma) + \sqrt{(\lambda + 1 - \gamma)^2 - 4\lambda}}{2}, \quad \lambda \geq 1 + \sqrt{\gamma}$$

But a simplified form, inspired by this formula is

$$\begin{aligned} \ell(\lambda) &= \kappa_1 \lambda - \kappa_2 \\ \kappa_2 &\geq \lambda_{\min} \\ \kappa_1 &\in (0, 1) \end{aligned}$$

Performance

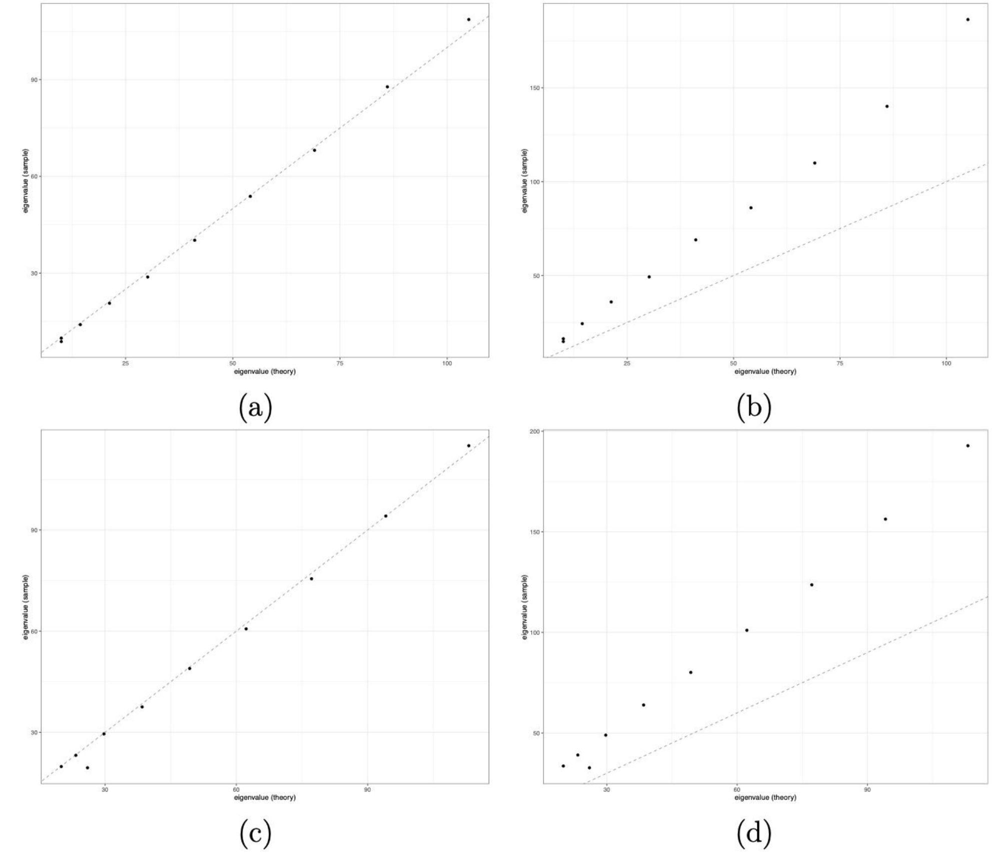


Figure 10.5: (a): 1000 assets, normally distributed returns; (b) 1000 assets, t-distributed returns; (c): 3000 assets, normally distributed returns; (d) 3000 assets, t-distributed returns. The x-axis denotes the population eigenvalues, while the y-axis denotes the shrinked empirical eigenvalues. The dashed line is the line $y = x$.

Choosing the Number of Factors

- Threshold method: volatility *level* exceeding a threshold

$$m = \max\{k|\hat{\lambda}_k \geq 1 + \sqrt{\gamma}\}$$

- Change point method: *change* in volatility exceeding a threshold. Basically scree plot

$$m = \max_{2 \leq k \leq k_{\max}} \left(\hat{\lambda}_k - \hat{\lambda}_{k-1}\right)$$

$$m = \max_{2 \leq k \leq k_{\max}} \left(\log \hat{\lambda}_k - \log \hat{\lambda}_{k-1}\right)$$

- Penalty-Based

$$\min_{k, \mathrm{rank}(\hat{\mathbf{R}}) \leq k} \left\|\mathbf{R} - \hat{\mathbf{R}}\right\|^2 + k f(n, T)$$

$$f(n, T) = \frac{n+T}{nT} \log\left(\frac{nT}{n+T}\right)$$

What's left? Whitening

- We assumed that the bulk had unit variance
- So, we need to studentize returns, either by total or specific volatility
- Short story: studentize by specific volatility
- But we don't have specific vol. Remedy: use a proxy from a first-stage PCA.

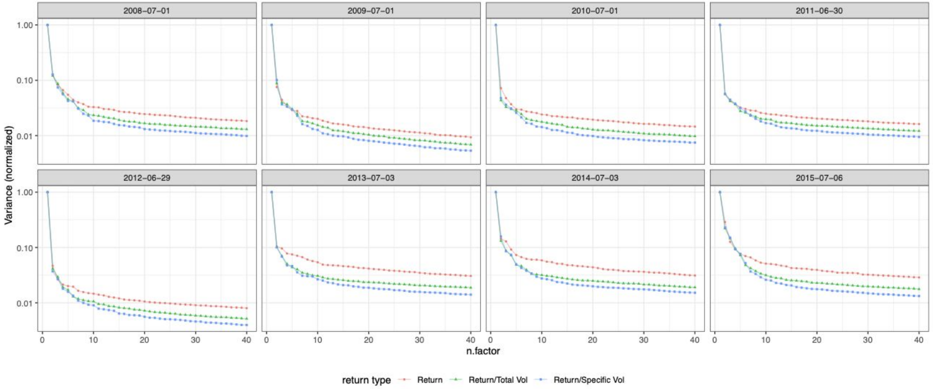


Figure 10.6: Variances of the eigenfactors (normalized to the variance of the first eigenfactor) for the first forty factors. Note that the scale of the y axis is logarithmic.

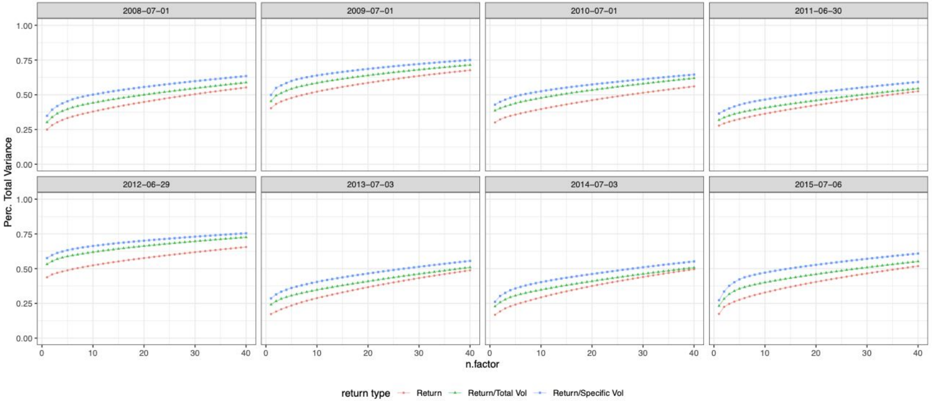


Figure 10.7: Cumulative percentage of variance described by the first n factors, for difference covariance matrices.

Preview of What's Left for Statistical Models

- Integrating all the steps
- Interpreting PCA
- Reducing turnover