Algorithmic Trading & Quantitative Strategies

Lecture 5 (4/2/2024)

Giuseppe Paleologo (gardener)

Today's Session

- I will send a project description on Thursday
- 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

Recap on Last Quiz

1. A drawback of statistical models is
   a. Low predictive power
   b. Computational burden
   c. Limited history
   d. **Low information use**
2. The standard PCA is based on an underlying factor model
   a. True
   b. **False**
   c. Uncertain
3. If two eigenvalues of a matrix are identical, its eigenvectors have
   a. Indefinite norm
   b. **Indefinite direction**
   c. The jargon name "Dick" and "Jane" in the academic literature
   d. Very important
4. Why can't you use Marchenko-Pastur for cleaning correlation matrices?
   a. Unit variance assumption does not hold
   b. **Zero correlation assumption does not hold**
   c. Sample is too small
5. PPCA recommends to
   a. Shrink the factor volatilities and correlations
   b. Shrink the factor correlations
   c. **Shrink the factor volatilities**
6. PPCA recommends to
   a. Shrink the idio volatilities
   b. **Shrink the big idio volatilities, increase the small ones**
   c. Leave the damn things unchanged

Topics

- Statistical Models
  - Practical considerations: z-scoring and reducing turnover
- Fundamental Factor Models
- Next Time:
  - Recap on Fundamental Models
  - Advanced Portfolio Construction

Why Fundamental Models?

- Interpretable
- Extensible
- Good performance
- Connects to academic research and to alpha research

Primitives

- Returns
  - Which returns? Closing auctions. Based on equal weighted/EWMA/volume weighted averages
- Characteristics
  - "Features", "Terms", "Signals"
  - Continuous variables
  - Discrete/categorical

Process

- Data ingestion and quality assurance
  - Missingness
  - Turnover
- Winsorization
- Estimation universe selection
- Loadings generation
- Cross-sectional regression
- Time-series estimation

Statistical Models

Reminder: positive features of the spectrum of a matrix are:

- The spectrum is high and separated from the bulk. The spiked assumptions hold better
- Spike eigenvalues are separated. The eigenvectors are identifiable

Whitening with idio vol proxy helps

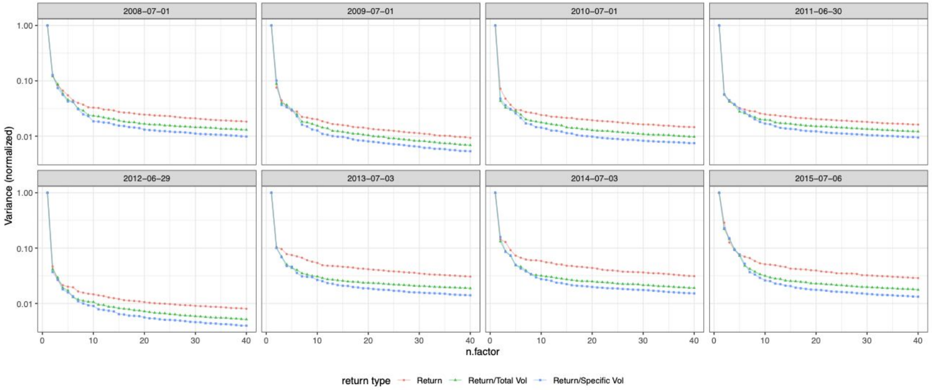


Figure 10.6: Variances of the eigenfactors (normalized to the variance of the first eigenfactor) for the first forty factors. Note that the scale of the y axis is logarithmic.

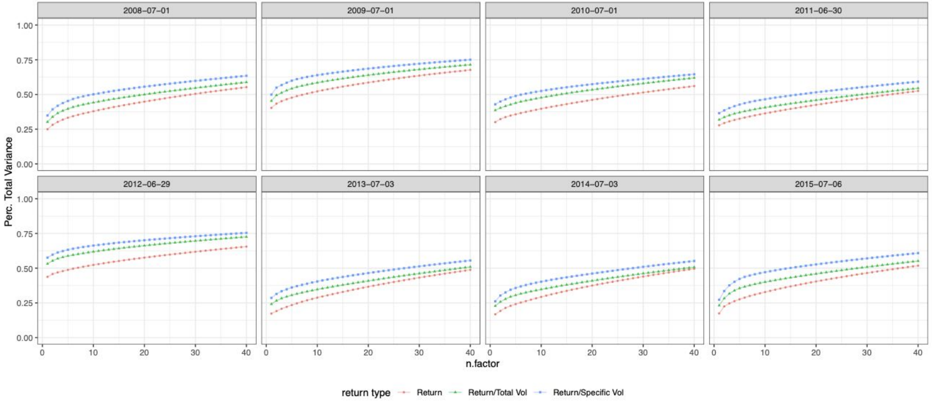


Figure 10.7: Cumulative percentage of variance described by the first n factors, for difference covariance matrices.

Additional Benefits of Idio Whitening: Returns

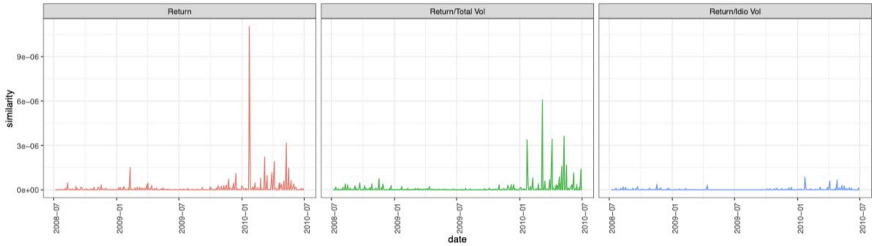
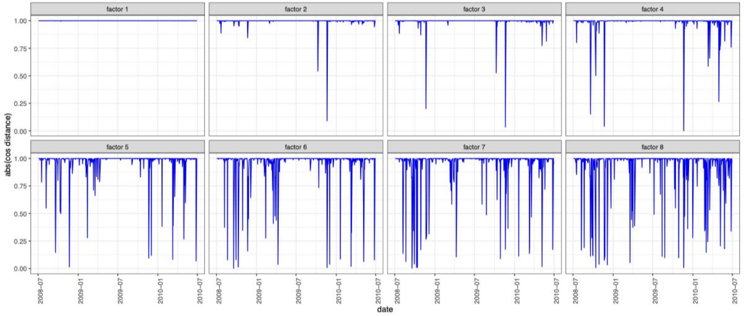
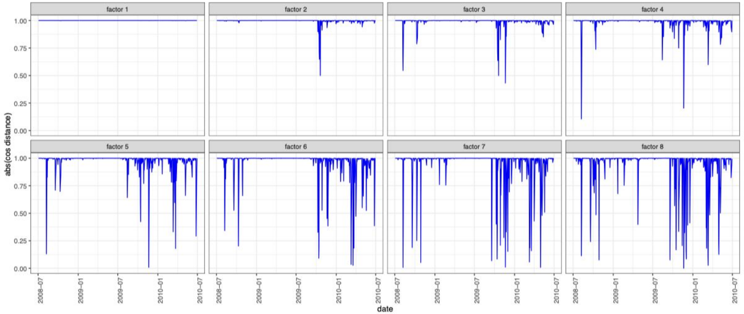


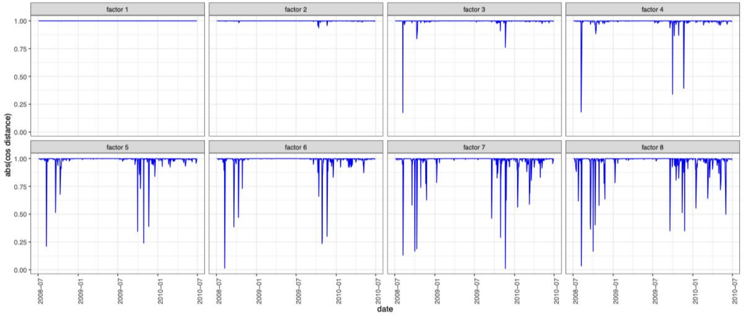
Figure 9.9: Distance between column subspaces of the first eight eigenfactors in consecutive periods. The eigenfactors are generated by PCAs on total returns, total returns/total vol, and total return/idio vol.



(a)



(b)



(c)

Additional Benefits of Idio Whitening: Returns

Figure 9.10: Factor returns for the first four eigenvactors. The eigenfactors are generated by PCAs on total returns, total returns/total vol, and total return/idio vol.

Reducing Turnover

Factors are turn over a lot. They can also change sign. They can be stabilized by rotating them from one period to the next.

$$
\begin{aligned}
\mathbf{B}_0 :=& \mathbf{B}_0 \\
\tilde{\mathbf{B}}_{t+1} =& \arg\min \; \|\mathbf{B}_t - \mathbf{Y}\|_F^2 \\
& \text{s.t. } \mathbf{Y} = \mathbf{B}_{t+1}\mathbf{X} \\
& \mathbf{X}'\mathbf{X} = \mathbf{I}_m \\
& \mathbf{X} \in \mathbb{R}^{m\times m}
\end{aligned}
$$

There is a clever solution for this problem.

Read the notes for details

$$
\begin{aligned}
\mathbf{A} &:= \mathbf{B}_t'\mathbf{B}_{t+1} \\
\mathbf{A} &= \mathbf{USV}' \\
&\boxed{\mathbf{X} = \mathbf{VU}'}
\end{aligned}
$$

Putting it all Together

Procedure 9.1: *Statistical Model Estimation*

1. **Inputs**: $\mathbf{R} \in \mathbb{R}^{n\times T}$, $\tau_s \geq \tau_f > 0$, $p \in \mathbb{N}$, $m > 0$.
2. **Time-Series Reweighting**:

$$\begin{aligned}\mathbf{W}_{\tau_f} &:= \kappa\,\mathrm{diag}\left(\exp(-T/\tau_f), \ldots, \exp(-1/\tau_f)\right)\\ \tilde{\mathbf{R}} &= \mathbf{R}\mathbf{W}_{\tau_f}\end{aligned}$$

3. **First stage PCA**: $\tilde{\mathbf{R}} := \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{V}}'$
4. **Idio Proxy Estimation**:

$$\begin{aligned}\mathbf{E} &= \tilde{\mathbf{R}} - \tilde{\mathbf{U}}_p\tilde{\mathbf{S}}_p\tilde{\mathbf{V}}_p' && \textit{(truncated SVD)}\\ \sigma_i^2 &= \sum_t [\mathbf{E}]_{i,t}^2 && \textit{(idio vol proxies)}\\ \mathbf{W}_\sigma &:= \mathrm{diag}\left(\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right)\end{aligned}$$

5. **Idio Reweighting**:

$$\begin{aligned}\mathbf{W}_{\tau_s} &:= \kappa\,\mathrm{diag}\left(\exp(-T/\tau_s), \ldots, \exp(-1/\tau_s)\right)\\ \hat{\mathbf{R}} &:= \mathbf{W}_\sigma\mathbf{R}\mathbf{W}_{\tau_s}\end{aligned}$$

6. **Second Stage PCA**: $\hat{\mathbf{R}} := \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}$
7. **Second-Stage Factor Model**: $\hat{\mathbf{r}} = \hat{\mathbf{U}}_m\mathbf{f} + \hat{\boldsymbol{\epsilon}}$

$$\begin{aligned}\text{where: } \mathbf{f} &\sim N\left(0, \mathrm{diag}\left(\ell(s_1^2), \ldots, \ell(s_m^2)\right)\right)\\ \boldsymbol{\epsilon} &\sim N(0, \bar{\lambda}\mathbf{I}_n)\\ \bar{\lambda} &= \frac{1}{n-m}\sum_{i=m+1}^{n} s_i^2\end{aligned}$$

8. **Output: Final Factor Model**: $\mathbf{r} := \mathbf{B}\mathbf{f} + \boldsymbol{\epsilon}$

$$\begin{aligned}\text{where: } \mathbf{B} &= \mathbf{W}_\sigma^{-1}\hat{\mathbf{U}}_m\\ \mathbf{f} &\sim N\left(0, \mathrm{diag}\left(\ell(s_1^2), \ldots, \ell(s_m^2)\right)\right)\\ \boldsymbol{\epsilon} &\sim N(0, \bar{\lambda}\hat{\mathbf{W}}_\sigma^{-2})\end{aligned}$$

Fundamental Models

Takeaways for Fundamental Models

- At their core, very simple: ***at their core***, alpha and risk modeling are very simple
- There are many details
  - Data
  - Shrinkage
  - Heteroskedastic and correlation correction
  - Regime Change

Steps

Steps:

1. Data ingestion: integrity, temporal continuity, missingness, outliers
2. Estimation universe: data quality, continuity, liquidity, relevance
3. Winsorization
4. Loadings Generation
5. Cross-Sectional Regression
6. Covariance Generation

Cross-Sectional Regression $\mathbf{r}_t = \mathbf{B}_t \mathbf{f}_t + \boldsymbol{\epsilon}_t, \qquad t \in \mathbb{N}$

Assumptions:

1. Loadings are full rank
2. Homoskedastic idio returns are homoskedastic (jkg)
3. Factor returns and idio returns are independent
4. At least fourth moments are finite

Basics of X-sec regression

Gaussian Likelihood

$$\prod_{t=1}^{T} \frac{1}{(2\pi)v|\boldsymbol{\Omega}_\epsilon|} \exp\left(-\frac{1}{2}(\mathbf{r}_t - \mathbf{B}\mathbf{f}_t)'\boldsymbol{\Omega}_\epsilon^{-1}(\mathbf{r}_t - \mathbf{B}\mathbf{f}_t)\right)$$

$$\min \left\|\boldsymbol{\Omega}_\epsilon^{-1/2}(\mathbf{R} - \mathbf{B}\mathbf{F})\right\|^2$$
$$\text{s.t. } \mathbf{F} \in \mathbb{R}^{m \times T}$$

$$\hat{\mathbf{f}}_t = (\mathbf{B}'\boldsymbol{\Omega}_\epsilon^{-1}\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\Omega}_\epsilon^{-1}\mathbf{r}_t$$

Note the similarity with PCA

Note that the estimation is exactly the same as factor mimicking portfolios

And this is why you should use squared loss

Shrinkage

- Factor volatilities are biased upwards because they are regression coefficients with errors, so they should be shrinked:

$$\mathrm{var}(\hat{\mathbf{f}}_t) = \mathbf{\Omega}_\mathbf{f} + \frac{1}{T}(\mathbf{B}'\mathbf{\Omega}_\epsilon^{-1}\mathbf{B})^{-1}$$

- Anderson's asymptotics for $T \to \infty$ doesn't apply. Recommended: simple Ledoit-Wolf. Let's not overkill.

$$\mathbf{\Omega}_{\text{shrink}}(\rho) = (1 - \rho_1)\hat{\mathbf{\Omega}}_\mathbf{f} + \rho_2 \frac{\text{trace}\left(\hat{\mathbf{\Omega}}_\mathbf{f}\right)}{m} \mathbf{I}_m$$

Separation Volatility-Correlation

Correlations change at a slower time-scale than volatilities

When people say, "in a crisis all correlations go to 1"... they should think twice [discuss!]

$$\boldsymbol{\Omega}_{\mathbf{f}} = \mathbf{VCV}$$

$$\mathrm{diag}\left(\mathbf{V}_t^2\right) = \kappa_V \sum_{s=0}^{T} e^{-s/\tau_V} \hat{\mathbf{f}}_{t-s} \circ \hat{\mathbf{f}}_{t-s}$$

$$\mathbf{C} := \kappa_C \sum_{s=0}^{T} e^{-s/\tau_C} \mathbf{V}_{t-s}^{-1} \hat{\mathbf{f}}_{t-s} \hat{\mathbf{f}}_{t-s}' \mathbf{V}_{t-s}^{-1}$$

Dynamic Factor Volatility

Add a "modulating term".

The intuition is that cross-sectional factor returns are pretty large, and responsive.

If you get them wrong, adjust.

$$\begin{aligned}\mathbf{f}_t &= e^{x_t/2}\mathbf{C}_t^{1/2}\mathbf{V}_t\boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t &\sim N(\mathbf{0}, \mathbf{I}_n) \\ x_{t+1} &= \phi\xi_t + \sigma\gamma_t\end{aligned}$$

$$\begin{aligned}\mathbf{u}_t &:= \mathbf{V}_t^{-1}\mathbf{C}_t^{-1/2}\mathbf{f}_t \\ \kappa &=: E(\log\|\boldsymbol{\eta}_t\|^2) \\ \epsilon_t &=: \kappa - \log\|\boldsymbol{\eta}_t\|^2 \\ x_t &:= \xi_t \\ y_t &:= \log\|\mathbf{u}_t\|^2 - \kappa\end{aligned}$$

$$e^{\hat{x}_t/2} = \kappa_0 \exp\frac{1}{2}\left(\sum_{s=0}^{\infty} e^{-s/\tau_0}(\log\|\mathbf{u}_{t-s}\|^2 - \kappa)\right)$$

$$\boxed{\simeq \kappa_0 \sum_{s=0}^{\infty} e^{-s/\tau_0}\frac{\|\mathbf{u}_{t-s}\|}{\sqrt{m}}}$$

Autocorrelation Correction

You need this. Example: Asynchronous Returns.

$$(\mathbf{C}_l)_{i,j} = \mathrm{cov}(\mathbf{f}_{t,i}, \mathbf{f}_{t-l,j})$$

Scholes correction:

$$\mathbf{\Omega}_{\mathbf{f}} = \hat{\mathbf{\Omega}}_{\mathbf{f}} + \frac{1}{2}\sum_{l=1}^{l_{\max}}(\mathbf{C}_l + \mathbf{C}_l')$$

Newey-West correction:

$$\mathbf{\Omega}_{\mathbf{f}} = \hat{\mathbf{\Omega}}_{\mathbf{f}} + \frac{1}{2}\sum_{l=1}^{l_{\max}}\left(1 - \frac{l}{l_{\max}}\right)(\mathbf{C}_l + \mathbf{C}_l')$$

Winsorization

- Ideally, choose the right estimation/investment universe and **never** winsorize
- In practice, the investment universe selection is imperfect
  - Bad data sources
  - Illiquid assets
  - Bankruptcies relistings
- So use a simple measure

$$d_{i,t} = \frac{|\log(1 + r_{i,t})|}{\text{median}(\log(1 + r_{i,t-1}), \ldots, \log(1 + r_{i,t-T}))}$$