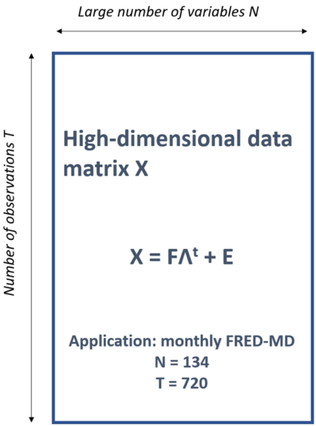
Estimation of factors using higher-order multi-cumulants in weak factor models

Kris Boudt
(joint with Guanglin Huang and Wanbo Lu)

R/Finance 2022

Summarizing high-dimensional data



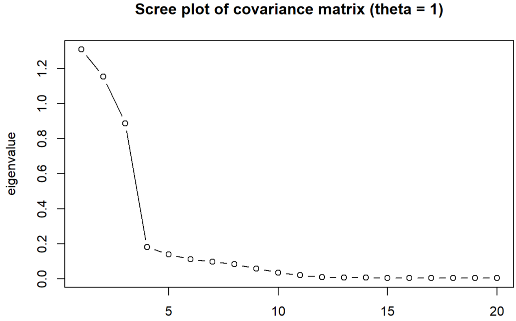
- Financial time series data can be overwhelming: the number of variables to anayze jointly (N) is large
- Reduce dimension by using a factor model:
  $$x_{it} = \lambda_i' f_t + e_{it},$$
  $R << N$ is the number of factors both f_t and λ_i are unobserved.
- Standard approach: estimate factors using PCA on the second moment covariation of the N variables: $X^t X$
- Question: Can we do better using PCA on the third moment covariation of the N variables: $X^t((XX^t) \circ (XX^t))X$

PCA-based factor analysis to decompose $X = F\Lambda^t + E$

- Reliable when E has low explanatory power for X^tX such that

$$X^tX \approx \Lambda F F^t \Lambda^t.$$

$\rightarrow$ Clear separation of eigenvalues of X^tX into a group of large eigenvalues representing factor-related variation and a group of small eigenvalues representing idiosyncratic variation



PCA-based factor analysis to decompose $X = F\Lambda^t + E$

- Following Ahn and Horenstein (2013): Select the number of factors based on maximizing the ratio of two adjacent eigenvalues arranged in descending order

$$\widehat{R}^{(k)} = \text{argmax}_{1 \leq r \leq R_{max}} \frac{r - th \text{ largest eigenvalue of } X^t X}{(r+1) - th \text{ largest eigenvalue of } X^t X}.$$

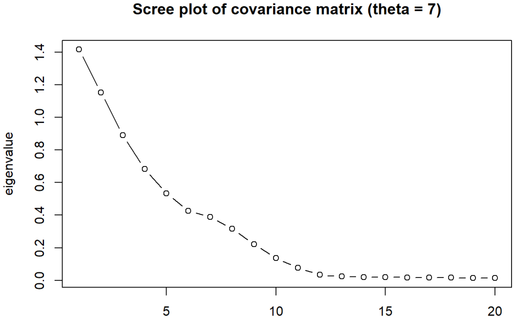
- Set the loading $\widehat{\Lambda}$ to $\sqrt{N}$ times the eigenvectors of $X^t X$
- Compute the factors as the linear fit $\widehat{F} = X\widehat{\Lambda}/N$

Limits of covariance analysis in case of weak factors

- Often, the variance of E is large such that it has a substantial finite sample contribution to

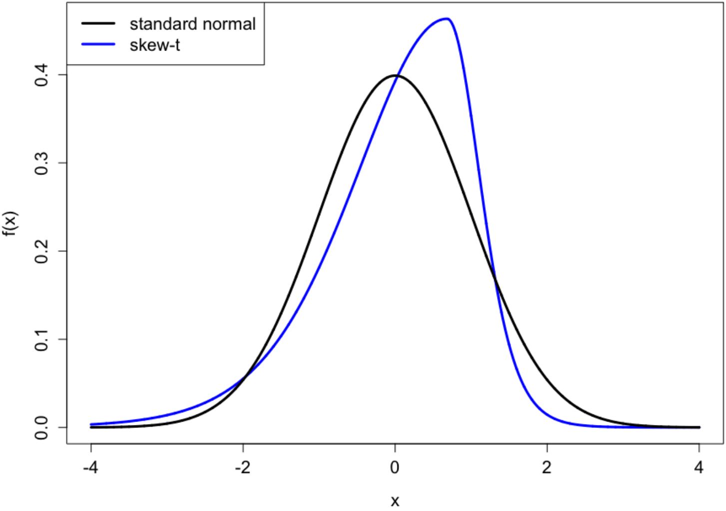
$$X^t X \approx \Lambda F F^t \Lambda^t + E^t E$$

- Due to the large explanatory power of the idiosyncratic factors, there is no clear separation of the eigenvalues



That's bad news

Besides being big or fat, data is often non-normal



Non-normality is summarized in coskewness and cokurtosis matrix

Case of three assets:

- Coskewness between 3 assets: $C^{(3)}_{ijk} = \mathbb{E}[X_i X_j X_k]$
- Coskewness matrix:

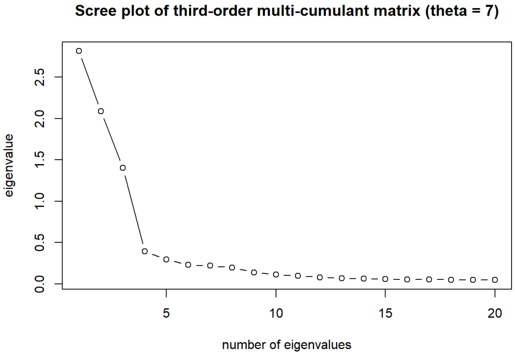
$$C^{(3)}_x = \begin{pmatrix} C^{(3)}_{111} & \mathbf{C^{(3)}_{112}} & C^{(3)}_{211} & C^{(3)}_{212} \\ C^{(3)}_{121} & C^{(3)}_{122} & \underline{C^{(3)}_{221}} & C^{(3)}_{222} \end{pmatrix}$$
$$= \begin{pmatrix} -0.458 & \mathbf{-1.107} & -1.107 & -1.950 \\ -1.107 & -1.950 & \underline{-1.950} & -4.522 \end{pmatrix} \times 10^{-6}$$

We transform this to a square matrix: $C^{(3)}_x C^{(3)t}_x$
A general estimator of the third-order covariation is:

$$\widetilde{C}^{(3)}_x \widetilde{C}^{(3)t}_x = \frac{1}{T^2} X^t((XX^t) \circ (XX^t))X$$

Higher order factor analysis to decompose $X = F\Lambda^t + E$

- Consider that non-normality is mostly driven by exposure to non-normal factors: $\widetilde{C}_x^{(3)} \widetilde{C}_x^{(3)t} \approx \Lambda C_f^{(3)} C_f^{(3)t} \Lambda^t$
- Sucess in estimating factors by doing eigenanalysis on $\widetilde{C}_x^{(3)} \widetilde{C}_x^{(3)t}$ instead of $X^t X$



Higher order factor analysis to decompose $X = F\Lambda^t + E$

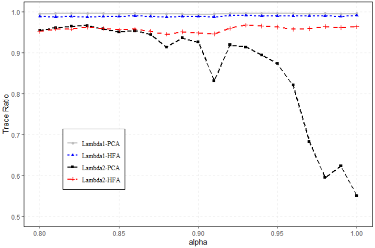
- Unlike PCA on covariance, we do PCA on $\widetilde{C}_x^{(3)}\widetilde{C}_x^{(3)t}$
  - Select the number of factors based on maximizing the ratio of adjacent eigenvalues of $\widetilde{C}_x^{(3)}\widetilde{C}_x^{(3)t}$ arranged in descending order
  - Set the loading $\widehat{\Lambda}$ to $\sqrt{N}$ times the eigenvectors of $\widetilde{C}_x^{(3)}\widetilde{C}_x^{(3)t}$
  - Compute the factors as the linear fit $\widehat{F} = X\widehat{\Lambda}/N$
- Backed by theory:
  - Asympotic properties: Consistency and asymptotic normality of HFA for $N, T \to \infty$
  - When to use it? Efficiency gain in case of weak factors: largest eigenvalue of E^tE grows at rate N^α with $\alpha \in (0, 1]$ (DeMol et al.(2008)) ($\leftrightarrow$ strong factors have $\alpha = 0$)

Simulation study

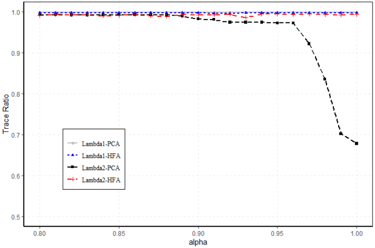
- A two-factor model $x_{it} = \lambda_{i1} f_{1t} + \lambda_{i2} f_{2t} + e_{it}$, $\lambda_i \sim \mathcal{N}(0, \mathbf{I})$, $e_t \sim \mathcal{N}(0, G_N)$
- Explanatory power of factors differs: $Var[f_{1t}] = 5$ (strong factor), $Var[f_{2t}] = 1$
- Largest eigenvalue of E'E is of the order N^{α}.
- When α increases, explanatory power of f_{2t} is only moderately larger than $\sigma_1(N^{-1} G_N)$ (Weakly influential factor).
- What you will see: Covariance-based approaches break down to estimate and select the weak factors as α increases.

Sensitivity of accuracy of loading estimates to explanatory power of idiosyncratic factors (α)

- Assume you know there are two factors, how accurately are the loadings estimated? Trace ratio which is 1 if perfect estimation.



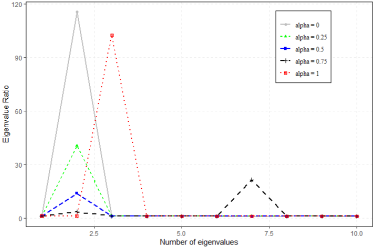
(a) $(N, T) = (100, 100)$



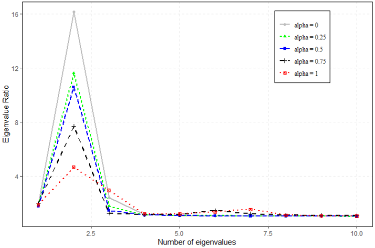
(b) $(N, T) = (500, 500)$

Sensitivity of eigenvalue ratio estimates to explanatory power of idiosyncratic factors (α)

- Factor selection: True number is 2.



(c) Eigenvalue Ratio test with $X^t X$



(d) Eigenvalue Ratio test with $X^t((XX^t) \circ (XX^t))X$

Quid kurtosis?

- Methodology is the same, but uses the following fourth order cumulant matrix:

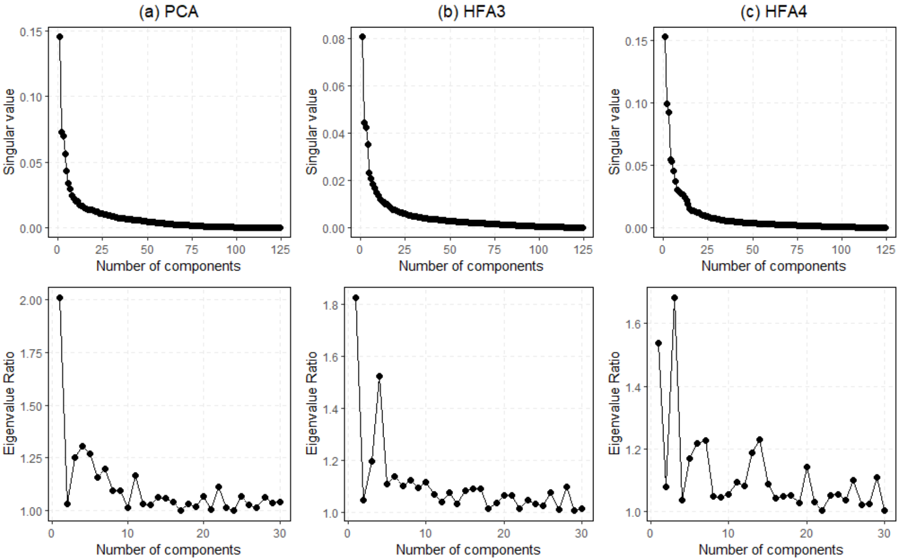
$$\widetilde{C}_x^{(4)}\widetilde{C}_x^{(4)t} = \frac{1}{T^2}X^t((XX^t)\circ(XX^t)\circ(XX^t))X + \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3,$$

  where $\mathcal{N}_1 = -3(X^t((b+b')\circ(XX^t))X)/T^2$,
  $b = (a, a, \ldots, a) \in \mathbb{R}^{T\times T}$, $a = (a_1, a_2, \ldots, a_T)'$,
  $a_t = \sum_i \sum_j x_{it}x_{jt}\widetilde{\Sigma}_{x,ij}$ for $t = 1, 2, \ldots, T$;
  $\mathcal{N}_2 = 3\text{vec}(\widetilde{\Sigma}_x)^t\text{vec}(\widetilde{\Sigma}_x)\widetilde{\Sigma}_x\widetilde{\Sigma}_x$; $\mathcal{N}_3 = 6\widetilde{\Sigma}_x\widetilde{\Sigma}_x\widetilde{\Sigma}_x\widetilde{\Sigma}_x$ and $\widetilde{\Sigma}_x = X^tX/T$.
- Seems cumbersome, but computationally convenient.
- Increases power to detect also the factors that are symmetric.

US Equity Risk Premium forecasting

- A key variable in portfolio maangement is the expected equity risk premium:
$$ERP_t = \log(1 + r_t^m) - \log(1 + r_t^f),$$
- Many candidate predictors. We consider factors extracted from the 134 monthly macroeconomic time series in the FRED-MD ($N = 134$, $T = 720$)
$$ERP_{t+1} = \alpha + \beta(L)\widehat{f_t} + \gamma_h(L)ERP_t + \epsilon_{t+1},$$
- Period: 1959-2018. Rolling samples of 26 years (312 observations).

Scree plots of of FRED-MD database disagree



Out of sample performance for predicting the equity premium

- Our of sample period: 1985-2018.
- Accuracy evaluation in terms of Mean Squared Error

	1985-2007/10	2007/11-2018	1985-2018
Panel A: Select factors based on covariance			
PCA on covariance			2.316
PCA on 3rd order cumulant			2.308
PCA on 4th order cumulant			2.309
Panel B: Select factors based on covariance, coskewness and cokurtosis (the largest R)			
PCA on covariance			2.323
PCA on 3rd order cumulant			**2.280***
PCA on 4th order cumulant			2.284*

Out of sample performance for predicting the equity premium

- Conclusion of gains of using PCA on higher order moments is robust across subsamples.

	1985-2007/10	2007/11-2018	1985-2018
Panel A: Select factors based on covariance			
PCA on covariance	2.113	2.730	2.316
PCA on 3rd order cumulant	2.075*	2.786	2.308
PCA on 4th order cumulant	2.076*	2.787	2.309
Panel B: Select factors based on covariance, coskewness and cokurtosis (the largest R)			
PCA on covariance	2.111	2.758	2.323
PCA on 3rd order cumulant	2.081*	**2.687***	**2.280***
PCA on 4th order cumulant	**2.068***	2.727	2.284*

Conclusion

- Often data is high-dimensional and factors are used to summarize them
- Standard PCA fails in case of weak factors
- Solution is PCA on the higher order moments
- Complete framework: Factor selection and estimation. Computationally convenient.

- R package: `hofa` (https://github.com/GuanglinHuang/hofa)
- HFA in R: illustration using simulations (https://rpubs.com/guanglin/876536)
- Paper is available on SSRN. (https://ssrn.com/abstract=3599632)