Quantitative Methods: Factor Analysis and PCA

Master

Financial Economics

ANNÉE SCOLAIRE / ACADEMIC YEAR 2011-2012

Intervenant/Lecturer: Sergio FOCARDI

Regressions and ARCH/GARCH

	Regression	ARCH	GARCH
Model	$Y = b_0 + b_1 X + \varepsilon$ $Y = b_0 + b_1 X_1 + \cdots + b_1 X_N + \varepsilon$	$Y_t = \sigma_t \varepsilon_t$ $\sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$	$Y_t = \sigma_t \varepsilon_t$ $\sigma_t^2 = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$
Estimation	OLS	MLE	MLE
Conditions and tests	Linear relationship between variables Significance test Residuals white noise	Dynamic model ARCH test	Dynamic model ARCH test

State-of-the-art Theory of Linear Factor Models

References

- Amengual, Dante and Mark W. Watson, 2006, “Consistent Estimation of The Number Of Dynamic Factors In A Large N And T Panel”, July 3, 2006
- Anderson, T. W. (1963). “Asymptotic theory for principal component analysis”. Annals of Mathematical Statistic 34, 122–148.
- Anderson, T. W. ,2003, *An introduction to multivariate statistical analysis*, Third ed. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- Bai, Jushan, 2003, “Inferential theory for factor models of large dimensions”, Econometrica, 71, 135–171, 2003.
- Bai, Jushan and Serena Ng, 2002, “Determining the number of factors in approximate factor models.” Econometrica, 70, 191–221. (Errata, Determining the number of factors in approximate factor models, by Jushan Bai and Serena Ng.)
- Biroli, Giulio, Jean-Philippe Bouchaud, Marc Potters, 2006, “On the top eigenvalue of heavy-tailed random matrices”, DSM/SPhT-T06/216 http://www-spht.cea.fr/articles/T06/216/
- Bossaerts, Peter, 1988, “Common Nonstationary Components Of Asset Prices”, Journal of Economic Dynamics and Contrcl 12 (1988) 347-364. North-Holland

- Bouchaud, Jean-Philippe, Laurent Laloux, M. Augusta Miceli, and Marc Potters, 2005. "Large dimension forecasting models and random singular value spectra", Science & Finance (CFM) working paper archive 500066, Science & Finance, Capital Fund Management.
- Breitung, Jorg and Uta Kretschmer, 2005, "Dynamic factor models", Deutsche Bundesbank Discussion Paper Series 1: Economic Studies, No 38/2005
- Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance.", Journal of Finance 52:1, 57-82.
- Chamberlain, Gary and Michael Rothschild, 1983, "Arbitrage, factor structure and mean-variance analysis in large asset markets", Econometrica 51, 1305–1324.
- Chamberlain, Gary, 1983, "Funds, factors, and diversification in arbitrage pricing models", Econometrica, 51, 1281–1304.
- Connor, Gregory and Robert Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: a new framework for analysis, Journal of Financial Economics, 15, 373-394
- Connor, Gregory and Robert Korajczyk, 1988, "Risk and Return in an Equilibrium APT: Application of a New Test Methodology", Journal of Financial Economics, 15, 373-394
- Connor, Gregory and Robert Korajczyk, 1993, "A Test for the Number of Factors in an Approximate Factor Model", Journal of Finance 48 (September 1993): 1263-1291.

- Cox, D.R. and H.D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall/CRC , 1977
- Fabozzi, Frank J., Sergio M. Focardi, and Petter N. Kolm, 2006 *Financial Modeling of the Equity Market: from CAPM to Cointegration*, 2006, Wiley, Hoboken, New Jersey
- Fabozzi, Frank J., Sergio Focardi, and Caroline Jonas, 2008. *Challenges in Quantitative Equity Management* (CFA Institute Research Foundation, Charlottesville, VA).
-
- Fama, Eugene F. and Kenneth R. French, 1993, "Common risk factors in the returns on stocks and bonds." Journal of Financial Economics 33, 3-56.
- Guttman, "The Determinacy of Factor Score Matrices With Implications for Five Other Basic Problems of Common-Factor Theory.", British Journal of Statistical Psychology, 1955,8, 65-81.
- Harding, Matthew C., 2008a, "Explaining the Single Factor Bias of Arbitrage Pricing Models in Finite Samples", Economics Letters, 99(1), 85-88, 2008
-

- Harding, Matthew C., 2008b, “Structural estimation of high-dimensional factor models”, MIT, Department of Economics, Job Market Paper
- Heaton, Chris and Victor Solo, 2003, “Asymptotic Principal Components Estimation of Large Factor Models”
- Heaton, Chris and Victor Solo, 2006, “Estimation Of Approximate Factor Models: Is It Important To Have A Large Number Of Variables?” Presented at the North American Summer Meeting of the Econometric Society at the University of Minnesota in June 2006
- Jolliffe, Iain T., *Principal Components Analysis*, Springer; 2nd edition, October 1, 2002
- Jolliffe, I. (2003). “A modified principal component technique based on the LASSO.” Journal of Computational and Graphical Statistics, 12, 531-547.
- Mehta, M. L. *Random Matrices and the Statistical Theory of Energy Levels*. New York, Academic Press, 1967. 1991, Second Edition, Academic Press, Boston.

- Plerou, Vasiliki , Parameswaran Gopikrishnan, Bernd Rosenow, Luıs A. Nunes Amaral, Thomas Guhr and H. Eugene Stanley, 2002, “Random matrix approach to cross correlations in financial data”, Physical Review E, Volume 65, 066126
- Scherrer, Wolfgang and Christiaan Heij, 1998, “Estimation of factor models by realization-based and approximation methods”, Econometric Institute Report 9831
- Schneeweiss, Hans, 1997, “Factors and principal components in the near spherical case”, Multivariate Behavioral Research, 32, 375-401.
- Schneeweiss, Hans and Hans Mathes, 1995, “Factor analysis and principal components.”, Journal of Multivariate Analysis, 55,105-124.
- Schonemann, P. H. “The Minimum Average Correlation Between Equivalent Sets of Uncorrelated Factors.”, Psychometrika, 197 1, 36.2 1-30.
- Spearman, C. (1904), “General intelligence, objectively determined and measured.”, American Journal of Psychology, 15, 201-293.

- Steiger, H., James , 1979, “Factor Indeterminacy In The 1930’s And The 1970’s Some Interesting Parallels”, Psychometrika, Vol 44. , No. 1. June, 1979
- Steiger, James H., 1996, “Coming Full Circle in the History of Factor Indeterminacy”, Multivariate Behavioral Research, 31 (4), 617-630
- Steiger, James H. and Peter H. Schonemann, 1978, “A History of Factor Indeterminacy”, in S. Shye Editor, Theory construction and data analysis in the behavioural sciences, 1978, Jossey-Bass, San Francisco.
- Stock, James H. and Mark W. Watson, 1998, “Diffusion Indexes”, NBER Working Paper 6702, august 1998
- Stock, James H. and Mark W. Watson, 2002a, “Forecasting using principal components from a large number of predictors”, Journal of the American Statistical Association, 97, 1167-79
- Stock, James H. and Mark W. Watson, 2002b, “Macroeconomic forecasting using diffusion indexes”, Journal of Business and Economics Statistics, 20:147–162, 2002b.
- Stock, James, H. and Mark W. Watson, 2004, “Forecasting With Many Predictors”, August 2004, Handbook of Economic Forecasting
- Thurstone, L. L., 1938, Primary mental abilities, Chicago, University of Chicago Press.
- Thurstone, L. L., 1947, *Multiple-Factor Analysis*. Chicago, University of Chicago Press.
- Zuur, A., F., I.D. Tuck, and N. Bailey, 2003, “Dynamic factor analysis to estimate common trends in fisheries time series”, Can. J. Fish. Aquat. Sci. 60: 542–552 (2003)

State-of-the-art theory of approximate linear factor models

- Factor models are a pillar of modern finance theory and practice
- The notion of absence of arbitrage (CAPM and ROSS)
- Alphas
- And Betas
- All depend on applying factor models

State-of-the-art theory of approximate linear factor models

- Classical factor model theory deals with static, strict factor models where residuals are white noise
- State-of-the-art theory of approximate linear factor models admits 1) correlated and autocorrelated residuals and 2) dynamic factors

Consequences of state-of-the-art theory of approximate linear factor models

- Factors are unique,
- Can be estimated with principal components
- We will first describe classical theory, then move to approximate factor models

Origin of factor models

- Classical factor models have a long history which goes back to the formalization of psychometric models
- At the beginning of the 20th century, psychologists needed to make sense of the psychometric tests that had been proposed
- Spearman (1904) introduced a one-factor model of mental abilities, the G factor
- Thurstone (1938, 1947) introduced the first multifactor model
- Hotelling (1933) described principal components analysis

Classical factor models

- Classical linear factor models as described, for example, in Anderson (2003), are strict factor models with a finite number of variables
- In a strict factor model, residuals are mutually uncorrelated and uncorrelated with factors
- This implies that all correlations are due to factors. A strict factor model is called a scalar strict factor model if all residuals have the same variance

- A linear factor model has the following form:

- where
- = the *i*-th variable to explain
- = the average of the *i*-th variable
- = the proportionality constant of the *i*–th variable *j*-th factor(the factor loading)
- = the *j-th* factor
- = the *i*-th residual term.

Static factor models

- Consider a multivariate time series of returns:

$$\mathbf{r}_t = (r_{it}), i = 1,2,\ldots,N, t = 1,2,\ldots,T$$

- Assume returns $\mathbf{r}_t$ are zero-mean, correlated stationary processes (or subtract the variables' means) …
- We do not consider integrated processes, e.g., prices, which require different theoretical considerations

Linear static factor models

- A classical linear static factor model has the following form:

$$\mathbf{r}_t = \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$$

- A factor model of this form is called static because it includes relationships between variables only at the same time
- It is a multiple regression model

Factor loadings

The matrix

$$\boldsymbol{\beta} = \begin{matrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{Nk} \end{matrix}$$

is the Nxk matrix of factor loadings

Loadings are constant

Factors, components and residuals

- $\mathbf{f}_t$ is a *k*-vector of common factors
- A factor model identifies:

- $\boldsymbol{\beta}\mathbf{f}_t$ which is an N-vector called the common component and
- $\boldsymbol{\varepsilon}_t$ which is a *N*-vector of residuals sometimes called idiosyncratic factors

Data representation

- A return sample can be represented by a matrix
- Each row is an observation and each column a variable:

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1t} & \cdots & r_{Nt} \\ \vdots & \ddots & \vdots \\ r_{1T} & \cdots & r_{NT} \end{bmatrix}$$

Matrix representation of factors and residuals

- Factors and residuals can be represented as:

$$\mathbf{F} = \begin{bmatrix} f_{11} & \cdots & f_{k1} \\ \vdots & \ddots & \vdots \\ f_{1t} & \cdots & f_{kt} \\ \vdots & \ddots & \vdots \\ f_{1T} & \cdots & f_{kT} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{N1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1t} & \cdots & \varepsilon_{Nt} \\ \vdots & \ddots & \vdots \\ \varepsilon_{1T} & \cdots & \varepsilon_{NT} \end{bmatrix}$$

- Factors and residuals are hidden, non-observed variables

Factor models in matrix form

- We can write a linear factor model as a relationship between the four matrices:

$$\mathbf{R} = \mathbf{F}\boldsymbol{\beta}' + \boldsymbol{\varepsilon}$$

- This is a matrix equation formed by *T*x*N* relationships

A hidden variables model

- Returns are observed variables
- Factors and residuals are hidden variables
- If we do not place restrictions on residuals…
- the model is empirically void of content
- It is only a definition of residuals given the factors

The strict factor model

- Zero mean variables: $E(\mathbf{r}_t) = 0,\ E(\mathbf{f}_t) = 0,\ E(\boldsymbol{\varepsilon}_t) = 0$
- Uncorrelated residuals: $\text{cov}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t) = \mathbf{D}$

  $\mathbf{D}$ diagonal matrix $\mathbf{D},\ \text{diag}(\mathbf{D}) = (\sigma_1^2, \ldots, \sigma_N^2)$
- Serially uncorrelated residuals at all lags:

  $E(\varepsilon_{it}\varepsilon_{js}) = 0,\ t \neq s, \forall i, j$
- Uncorrelated factors / residuals

  $E(\varepsilon_{it} f_{js}) = 0,\ \forall i, j, t, s$
- Serially uncorrelated factors at all lags

Indeterminacy of the strict factor model

- The factors of a strict factor model are determined up to an invertible linear transformation

$$\boldsymbol{\beta} = \boldsymbol{\beta}^* \mathbf{A}^{-1}, \mathbf{f}_t = \mathbf{A}\mathbf{f}_t^*$$

- The models:

$$\mathbf{r}_t = \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \ \mathbf{r}_t = \boldsymbol{\beta}^* \mathbf{A}^{-1}\mathbf{A}\mathbf{f}_t^* + \boldsymbol{\varepsilon}_t$$

  are observationally equivalent

- We assume

$$\operatorname{cov}\left(\mathbf{f}_t, \mathbf{f}_t\right) = \mathbf{I}_k$$

- But the model is still determined only up to an orthogonal rotation

Covariance matrix of the strict factor model

- The covariance matrix of a factor model can be decomposed as follows:

$$\mathbf{\Sigma} = \mathrm{cov}(\mathbf{r}_t, \mathbf{r}_t)(N \times N), \;\; \mathbf{\Lambda} = \mathrm{cov}(\mathbf{f}_t, \mathbf{f}_t)(k \times k)$$
$$\mathbf{\Psi} = \mathrm{cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t)(N \times N), \;\; \mathbf{\Sigma} = \boldsymbol{\beta}\mathbf{\Lambda}\boldsymbol{\beta}' + \mathbf{\Psi}$$

- If the model is a strict factor model, factors are orthonormal and residuals uncorrelated

$$\mathbf{\Sigma} = \mathrm{cov}(\mathbf{r}_t, \mathbf{r}_t)(N \times N), \;\; \mathbf{\Lambda} = \mathbf{I}_k$$
$$\mathbf{\Psi} = \mathbf{D}, \;\; \mathbf{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{D}$$

Estimation of loadings

- The factor loadings of a strict factor model can be estimated with ML techniques
- Under the assumption of joint normality, the joint distribution of returns is:

$$\mathbf{r}_t \sim N(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) = \left[(2\pi)^N |\boldsymbol{\Sigma}| \right]^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{r}_t' \boldsymbol{\Sigma}^{-1} \mathbf{r}_t \right\}$$

- and the likelihood

$$L(\boldsymbol{\Sigma}) = \left[(2\pi)^N |\boldsymbol{\Sigma}| \right]^{-\frac{K}{2}} \prod_{t=1}^{T} \exp\left\{ -\frac{1}{2} \mathbf{r}_t' \boldsymbol{\Sigma}^{-1} \mathbf{r}_t \right\}$$

With maximum likelihood

- Take the loglikelihood:

$$\log L(\boldsymbol{\Sigma}) = l(\boldsymbol{\Sigma}) = -\tfrac{NK}{2}\log(2\pi) - -\tfrac{K}{2}\log\left(|\boldsymbol{\Sigma}|\right) - \frac{1}{2}\sum_{t=1}^{T}\left\{\mathbf{r}_t\,'\boldsymbol{\Sigma}^{-1}\mathbf{r}_t\right\}$$

- and maximize with respect to β and σ with numerical methods under the restriction:

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}\,' + \mathbf{D},\ \mathbf{D} = \left(\sigma_1^2, \ldots, \sigma_N^2\right)\mathbf{I}_N$$

Estimation of factors

- Factors are not uniquely identified and cannot be estimated with MLE methods
- See Steiger and Schönemann *History of factor indeterminacy* with interesting cultural references
- A non unique estimate of factors can be obtained with cross sectional GLS regressions of returns on the betas

Or with principal components

- If residuals have the same variance

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \sigma^2 \mathbf{I}$$

- then the *k* factors can be estimated as the first *k* principal components

In summary

- Strict factor models are sets of simultaneous multiple regressions...
- Such that residuals are mutually uncorrelated
- And the covariance structure depends only on factors
- No local covariance is allowed

Principal Components Analysis
and
Singular Value Decomposition

Principal Components Analysis (PCA)

- PCA the key technique in modern factor analysis
- Wherever modern factor analysis can be applied,
- Factors are unique and can be estimated with PCA
- Hence the importance of PCA

The concept of principal components

- Consider the matrix of returns

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1t} & \cdots & r_{Nt} \\ \vdots & \ddots & \vdots \\ r_{1T} & \cdots & r_{NT} \end{bmatrix}$$

- Form the linear combination of returns (portfolio) of maximum variance
- s.t. the condition that the vector of coefficients has unitary length

PCA as a Karhunen-Loéve transform

- The Karhunen-Loéve transform finds the axis of maximum variance
- Consider a linear combination of returns

$$\sum_{j=1}^{N} \beta_{1j} r_{tj} = \boldsymbol{\beta}_1' \mathbf{r}_t, \ \mathbf{r}_t = (r_{t1}, \ldots, r_{tN})', \ \mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_T)'$$

  where the weights $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1N})'$ satisfy $\boldsymbol{\beta}_1' \boldsymbol{\beta}_1 = 1$

- And find:

$$\boldsymbol{\beta}_1 = \underset{\text{s.t. } \boldsymbol{\beta}_1' \boldsymbol{\beta}_1 = 1}{\operatorname{argmax}} \left(\sum_{t=1}^{T} \left(\sum_{j=1}^{N} \beta_{1j} r_{tj} \right)^2 \right) = \underset{\text{s.t. } \boldsymbol{\beta}_1' \boldsymbol{\beta}_1 = 1}{\operatorname{argmax}} \left(\sum_{t=1}^{T} \left(\boldsymbol{\beta}_1' \mathbf{r}_t \right)^2 \right)$$

Karhunen-Loéve transform

- As a second step: apply the same procedure to the residuals:

$$\mathbf{r}_t^{(1)} = \mathbf{r}_t - \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{'} \mathbf{r}$$

and write the condition:

$$\boldsymbol{\beta}_2 = \underset{\text{s.t. } \boldsymbol{\beta}_2^{'}\boldsymbol{\beta}_2 = 1,\ \boldsymbol{\beta}_1^{'}\boldsymbol{\beta}_2 = 0}{\operatorname{argmax}} \left(\sum_{t=1}^{T} \left(\boldsymbol{\beta}_2^{'} \mathbf{r}_t^{(1)} \right)^2 \right)$$

- Proceed in this way for *N* steps

- The matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)'$ coincides with the matrix of the eigenvectors of the covariance matrix

PCA

- We can write returns as the sum of *N* rank-1 matrices of the form:

$$\mathbf{r}_t = \boldsymbol{\beta}_i \boldsymbol{\beta}_i^{'} \mathbf{r}_t$$

- (Each term is a projection of returns on the *i-th* maximum variance direction)
- PCA is obtained by stopping the process after *k* terms

Eigenvectors and eigenvalues

Consider the returns and their covariance matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{N1} \\ \vdots & \ddots & \vdots \\ r_{1t} & \cdots & r_{Nt} \\ \vdots & \ddots & \vdots \\ r_{1T} & \cdots & r_{NT} \end{bmatrix}, \ \boldsymbol{\Sigma} = \frac{1}{T}\mathbf{R}'\mathbf{R}, \ \boldsymbol{\Sigma} = \left[N \times N\right]$$

and the matrices:

$$\mathbf{R}'\mathbf{R} \ \left[N \times N\right] \ \mathbf{R}\mathbf{R}' \ \left[T \times T\right]$$

Eigenvectors and eigenvalues

- Consider the equation: $\boldsymbol{\Sigma}\mathbf{v} = \lambda\mathbf{v}$
- Non trivial solutions of this equations are the roots of the equation:

$$\det\left(\boldsymbol{\Sigma} - \lambda\mathbf{I}\right) = 0$$

- The λ_i are called eigenvalues and the
- The corresponding $\mathbf{v}_i$ are called eigenvectors
- While $\mathbf{w}'\boldsymbol{\Sigma} = \lambda\mathbf{w}$ are the left eigenvectors

Eigenvectors and eigenvalues

- If a matrix is multiplied by a scalar α
- Its eigenvalues are multiplied by α
- But its eigenvectors do not change

$$\mathbf{Av} = \lambda\mathbf{v},\ (\alpha\mathbf{A})\mathbf{v} = (\alpha\lambda)\mathbf{v}$$

- Therefore the matrices

$$\mathbf{\Sigma} = \frac{1}{T}\mathbf{R'R} \text{ and } \mathbf{R'R}$$

- have the same eigenvectors but not the same eigenvalues

Diagonalization

- If the eigenvectors are all distinct and linearly independent,
- then the corresponding matrix can be diagonalized:

$$\mathbf{\Sigma V} = \mathbf{V\Lambda},\ \mathbf{V} = \left[\mathbf{v}_1, \ldots, \mathbf{v}_N\right],\ \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_N \end{bmatrix}$$

- and

$$\mathbf{\Sigma} = \mathbf{V\Lambda\Lambda}^{-1},\ \mathbf{\Lambda} = \mathbf{V}^{-1}\mathbf{\Sigma V}$$

Principal components

- Principal components are given by the products $\beta \mathbf{r}_t$
- Which are orthogonal time series
- We find again the same results of the Karhunen-Loéve transform

A reminder….

Eigenvalues and eigenvectors of covariance and correlation matrices, ctd...

Compute eigenvalues and eigenvectors

- [V D]=eig(C)

V =

0.2766	-0.1299	0.1127	-0.1561	0.9325
-0.2708	0.6659	0.2767	-0.6368	0.0330
-0.2973	0.5736	-0.4061	0.5660	0.3119
-0.5045	-0.1589	0.7520	0.3804	0.1003
-0.7122	-0.4306	-0.4246	-0.3241	0.1483

D =

0.5326	0	0	0	0
0	0.8487	0	0	0
0	0	0.9369	0	0
0	0	0	1.2857	0
0	0	0	0	1.4408

Principal components

- Multiply the matrix X by the matrix of eigenvectors V

$$Y = XV$$

$$\text{cov}(Y) = \frac{Y'Y}{T-1} = \frac{V'X'XV}{T-1} = V'CV = D$$

- Obtain a new matrix Y with a diagonal covariance matrix; terms on the diagonal are the eigenvalues of C equal to the variances of Y

Covariance matrix of principal components

Compute the covariance matrix of Y and verify it is equal to D

Y'*Y/(T-1) =

0.5326	-0.0000	-0.0000	-0.0000	-0.0000
-0.0000	0.8487	-0.0000	-0.0000	0.0000
-0.0000	-0.0000	0.9369	-0.0000	-0.0000
-0.0000	-0.0000	-0.0000	1.2857	-0.0000
-0.0000	0.0000	-0.0000	-0.0000	1.4408

Represent X as a linear combination of the Y

- Left-multiply Y=X*V by V’ and obtain Y*V’=X*V*V’=X
- X can be represented as a weighted sum of uncorrelated principal components
- Hence the variance of X is the weighted sum of the variances of the principal components each multiplied by the square of its weight

Represent X as a linear combination of the Y, ctd...

$$X = [X(:,1), \ldots, X(:,1)]$$
$$Y = [Y(:,1), \ldots, Y(:,1)]$$
$$V = \begin{bmatrix} V_{11} & \cdots & V_{1N} \\ \vdots & \ddots & \vdots \\ V_{N1} & \cdots & V_{NN} \end{bmatrix}, \; V' = \begin{bmatrix} V_{11} & \cdots & V_{N1} \\ \vdots & \ddots & \vdots \\ V_{1N} & \cdots & V_{NN} \end{bmatrix}$$

$$X(:,1) = V_{11}Y(:,1) + \cdots + V_{1N}Y(:,N)$$
$$\cdots\cdots\cdots$$
$$X(:,N) = V_{N1}Y(:,1) + \cdots + V_{NN}Y(:,N)$$

Compute the variance of X as a weighted sum of the variances of PCs

(var(Y))*(V'.*V‘) =(diag(D))'*(V'.*V')=

0.9221 0.9339 1.0792 0.9334 1.0357

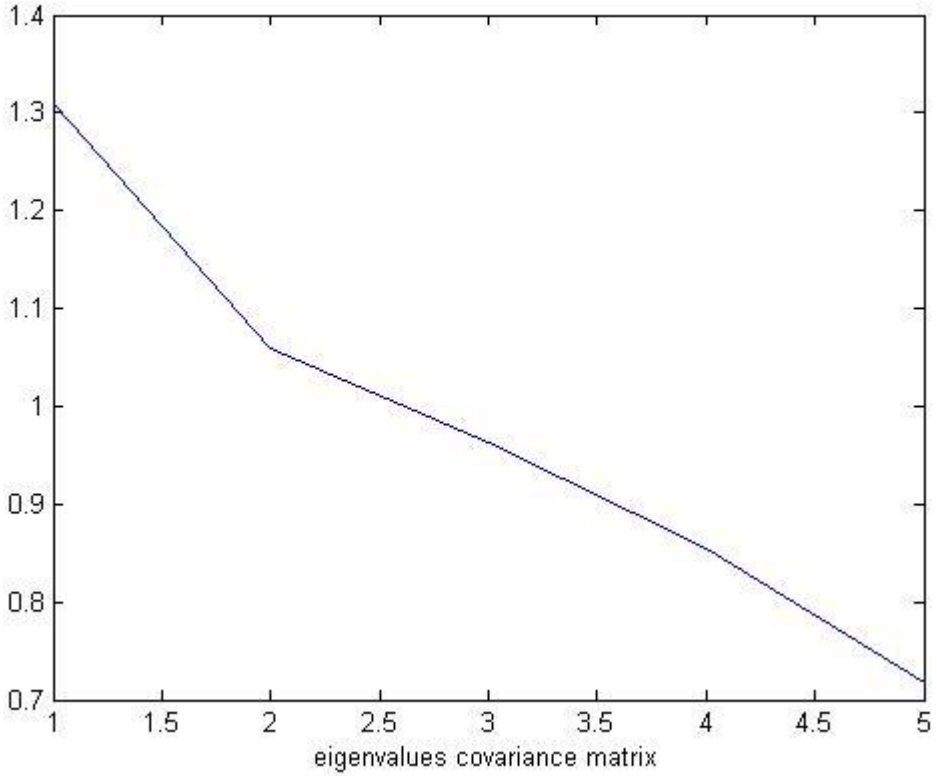
var(X) =

0.9221 0.9339 1.0792 0.9334 1.0357

What are the PCs

- The first PC associated to the largest eigenvalue is the portfolio of maximum variance
- The second PC is the portfolio of maximum variance orthogonal to the first PC
- The third PC is the portfolio of maximum variance orthogonal to the first two PCs
- And so on

Plot of eigenvalues



Truncate representation to the first N1 PCs

- Lets truncate the representation of X to the first N1 PCs, say N1=2

$$X(:,1) = V_{11}Y(:,1) + V_{12}Y(:,2) + (V_{13}Y(:,3) + V_{14}Y(:,4) + V_{15}Y(:,5))$$

$$\ldots\ldots\ldots$$

$$X(:,5) = V_{51}Y(:,1) + V_{52}Y(:,2) + (V_{53}Y(:,3) + V_{54}Y(:,4) + V_{55}Y(:,5))$$

$$X = Y(:,1:2)V1 + Y(:,3:5)V2$$

$$V1 = V'(1:2,:)$$

$$V2 = V'(3:5,:)$$

- X=Y(:,1:2)V1+err

Truncate representation to the first N1 PCs

- XPC=X-Y(:,1:2)*W(1:2,:)

var(X) =

0.9221 0.9339 1.0792 0.9334 1.0357

var(XPC) =

0.6248 0.2570 0.9316 0.6318 0.8879

mean(var(XPC)./var(X))=67%

Singular Value Decomposition SVD

- The SVD is a decomposition of a generic *MxN* matrix $\mathbf{X}$:

$$\mathbf{X} = \mathbf{USV}', \ \mathbf{X} = [M \times N], \ \mathbf{U} = [M \times M], \ \mathbf{V} = [M \times N], \ \mathbf{S} = [M \times N]$$

- The matrices $\mathbf{U,V}$ are orthonormal
- $\mathbf{U}$ is formed by the eigenvectors of $\mathbf{XX'}$
- $\mathbf{V}$ is formed by the eigenvectors of $\mathbf{X'X}$
- $\mathbf{S}$ is diagonal formed by the singular values

$$\mathbf{S} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_N \\ 0 & \cdots & 0 \end{bmatrix}$$

SVD, eigenvalues

- The singular values are equal to the square roots of the eigenvalues of the covariance matrix of $\mathbf{X}$
- The principal components are the $\mathbf{U}$ multiplied by the singular values:

$$\mathbf{X} = \mathbf{USV}', \ \mathbf{XV} = \mathbf{US}$$

Economy-size SVD

- If *M>N* then *M-N* columns of $\mathbf{U}$ are zero
- We can write the economy-size SVD as before:

$$\mathbf{X} = \mathbf{USV}', \ \mathbf{XV} = \mathbf{US}$$

- But $\mathbf{U}$ is *M*x*N* and $\mathbf{S}$ is *N*x*N*

PCA and SVD

- We can now restate PCA in terms of the SVD
- Consider the matrix of returns $\mathbf{R}$ and its economy-size SVD:

$$R = U\Sigma V'$$

- We can rewrite the SVD as the sum of *p=min(N,T)* rank-1 outer products:

$$\mathbf{R} = \mathbf{E}_1 + \cdots \mathbf{E}_p, \ \mathbf{E}_i = \sigma_i u_i v_i'$$

Limitations of strict factor models

- If the number of returns is large, no reasonable number of factors leaves uncorrelated residuals
- The asymptotic distribution work for large *T* small *N*
- The theory of factor models does not work for large *N* and large *T*

Approximate factor models

- Chamberlain and Rothshild (1983) introduced approximate factor models
- Approximate factor models allow residuals to be mutually correlated
- The covariance matrix of returns can be simplified as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}'\boldsymbol{\beta} + \boldsymbol{\Psi}$$

where $\boldsymbol{\Psi}$ is no longer diagonal

Local and global correlations

Chamberlain and Rothschild understood:

A rigorous theory of approximate factor models can be developed only in the asymptotic limit of an infinite market

where we can separate local correlations

from correlations due to common factors

Diversification

- The distinction between local and global correlations is fundamental from the point of view of diversification
- and thus of risk and arbitrage
- Local correlations can be diversified without limits;
- Global correlations cannot be diversified

Arbitrage

As stated by Ross (1976):

Excess returns due to global correlations compensate for risk while excess returns associated with local correlations lead to arbitrage.

Hence in an infinite economy alpha must be zero except for a finite number of securities

Correlations and eigenvalues

The size of the eigenvalues is proportional to:

1) the strength of correlations and…

2) the number of correlated entries

Case 1: in the absence of correlations, if all variances are equal eigenvalues are all equal to the common variance

Key examples

- Case 2: in the absence of correlations, if variances are different, eigenvalues are equal to the idiosyncratic variances ρ
- Case 3: if all correlations are equal to a constant then the largest eigenvalue is equal to $1+(N-1)\rho$
- while all other eigenvalues will be equal to

$$1-\rho$$

The intuition behind large eigenvalues

- A large eigenvalue signals the presence of widespread correlations,
- or of strong and diffused covariances
- A small eigenvalue is associated with correlations that interest only a small number of return processes
- In the limit, an eigenvalue 1 is associated with uncorrelated returns

The approximate factor model

Chamberlain and Rothschild define an approximate factor model for an infinite market if:

- When the number of processes *N* grows and becomes infinite,
- *K* eigenvalues of the covariance matrix of returns grow and become infinite,
- While all other eigenvalues remain bounded

The need of infinite markets

- This condition is rigorously defined for infinite markets
- as we can separate those eigenvalues that grows without limit
- from those that remain bounded
- In a finite market, we should separate large from small eigenvalues
- In arbitrary ways

Estimation of factors

- Under the previous assumption,
- Chamberlain and Rothschild demonstrate that in a *k*-factor approximate factor model
- 1) Factors are unique, and…
- 2) Factors can be estimated with the first *k*-principal components

Implications

- This discovery is very important
- First it states that factor structures are unique,
- making it reasonable to try to find the exact number of factors
- Second it gives a computationally simple procedure – PCA or equivalently SVD - to discover factors

Factors and PCs

- Prior to this work,
- The equivalence of factors and principal components has been conjectured but proved only in restricted cases
- This work makes this equivalence theoretically sound

Practical implications

- First we have to ascertain if the paradigm of infinite models can be applied in practice
- That is, if large *N* models can be considered sufficiently close to infinite models,
- so that the same conclusions apply

Practical implications

- If the conclusion is positive, then we have a universal procedure for factor modeling
- And portfolios based on factors should be substantially similar
- Given that portfolios are constructed with optimization techniques that do not allow any customization
- Except for the risk appetite

Introducing a more general model

- Before we begin to answer these questions, we need to introduce new factor models
- That can be used not only for assessing risk exposures but also for forecasting

The time dimension

- Chamberlain and Rotschild assume that returns, factors, and residuals are stationary processes without any serial correlation
- However, if factor models have to be used in forecasting returns,
- Some autocorrelation or autocross correlation must exist

Adding autocorrelations and heteroscedasticity

- The problem of inclusion of correlation along the time dimension has been tackled by several researchers
- In particular, in a series of papers, Jushan Bai and Serena Ng introduced an extension of approximate fractor models,
- which has become the state-of-the-art of static factor models

- Bai and Ng extend approximate factor models,
- Introducing a time dimension,
- And allowing for heteroscedasticity,
- And autocorrelations

The Bai and Ng model

$$E(\boldsymbol{\varepsilon}_t) = 0,\ E\left(\left|\boldsymbol{\varepsilon}_t\right|^8\right) < \infty$$

$$E\left(\frac{\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_t}{N}\right) = E\left(N^{-1}\left(\sum_{i=1}^{N} \varepsilon_{is}\varepsilon_{it}\right)\right) = \gamma_N(s,t)$$

$$\left|\gamma_N(s,s)\right| \le M,\ T^{-1}\left(\sum_{s=11}^{T}\sum_{t=1}^{T}\left|\gamma_N(s,t)\right|\right) \le M,\ \forall t, \forall s$$

$$E\left(\varepsilon_{it}\varepsilon_{ji}\right) = \tau_{ij,t},\ \left|\tau_{ij}\right| \le \tau_{ij},\ N^{-1}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tau_{ij}\right|\right) \le M$$

$$E\left(\varepsilon_{is}\varepsilon_{jt}\right) = \tau_{ij,st},\ (NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|\tau_{ij,st}\right| \le M$$

$$E\left|N^{-\frac{1}{2}}\sum_{i=1}^{N}\left[\varepsilon_{is}\varepsilon_{it} - E\left(\varepsilon_{is}\varepsilon_{it}\right)\right]\right|^4 \le M$$

The Bai and Ng model

- The model has lost the simplicity of the original approximate factor model,
- Introducing a number of conditions directly on individual covariances
- But it retains the basic idea of applying factorization to an infinite market
- By letting both *T* and *N* grow to infinity

Determining the Optimal Number of Factors

The number of factors

- In every factor model, it is assumed that the number of factors is a well determined quantity
- This is true for all strict factor models, where factors are responsible for all correlations,
- And for the infinite approximate factor models we introduced

The Cattell Scree-test

- A nearly heuristic method for determining the number of factors
- Based on plotting eigenvalues in descending order
- The number of factors corresponds to the point where the plot changes steepness

Growing eigenvalues

- It would be natural to gauge the number of factors
- by observing how many eigenvalues grow with growing dimensions of the market
- However, this does not work for two reasons

Growing eigenvalues

- First, it is practically impossible to let both the number of returns and the length of return series grow
- Because if we lengthen the time window, we find fewer series
- Second, because theoretically, in a large market, all eigenvalues grow
- Though only some eigenvalues grow without bounds

Formal methods

Two basic methods have been proposed to determine the number of factors:

1) A method based on Information Theory proposed by Bai and Ng

2) A method based on Random Matrix Theory proposed by Bouchaud et al.

Method based on Information Theory

- It is a model selection method
- It works in the same way as criteria for selecting the optimal model
- Introducing a penalty function,
- The number of factors minimize the following function

$$\mathbf{Q} = \arg\min_{Q,NT} \left(\frac{1}{NT} \left\| \mathbf{R}_t - \boldsymbol{\beta} \mathbf{F}_{\mathbf{Q},t} \right\|^2 + k\sigma^2 \left(\frac{N+T}{NT} \right) \log \left(\frac{NT}{N+T} \right) \right)$$

Method based on Information Theory

- Bai and Ng demonstrate that their criteria work asymptotically
- That is, asymptotically the criteria pick the correct number of factors
- When applied to finite markets, however, they pick the optimal number of factors
- Given the size of markets and the length of time series

Methods based on Random Matrix Theory

- These methods are based on the asymptotic distribution of eigenvalues in a random matrix
- A random matrix is a matrix with random entries
- True factors correspond to those eigenvalues that fall outside the distribution of random eigenvalues

Estimation and Uniqueness of Factors

Estimation and uniqueness

- In any of the approximate factor models introduced above,
- Factors are unique up to an invertible linear transformation,
- and can be estimated with principal components (belong to the space spanned by principal components)

Estimation and uniqueness

What has been demonstrated:

If we compute the first *k* principal components

$$\arg\min_{f,\beta} \frac{1}{NT}\left(\sum_{t=1}^{T}\left\|\mathbf{R}_{Nt} - \boldsymbol{\beta}\mathbf{F}_{Nt}\right\|^2\right)$$

There is a linear transformation $\mathbf{H}$

of the true factors such that

the first *k* principal components tend to this transformation when *N,T* grow

$$\lim_p \left\| \overline{\mathbf{F}} - \mathbf{HF} \right\| = 0$$

Determining the Sample Distribution of Factors in Large Markets

Can we apply the paradigm of large *N,T* factor models?

- In the previous section, we outlined the state-of-the-art theory of factor models
- This theory is widely used in practice in macroeconomics
- For example, some Central Banks use the Stock and Watson indexes based on factor models
- and the FAVAR (Factor Augmented Vector Auto Regression) introduced by Bernanke

Can we apply the paradigm of large *N,T* factor models?

- Can we apply the same paradigm to finance? In particular to returns?
- If yes, consequences are momentous as we have to conclude that all users of factor models (most quant firms are users of factor models) ultimately use the same models,
- Or marginally different versions of the same PCA-based models

Can we apply the paradigm of large *N,T* factor models?

If not, we need some theory to replace factor model theory to:

- Discriminate factor vs non-factor correlations
- Gauge the optimality of our factor models
- Gauge the uniqueness of our factor models

The sample distribution of factors

- Our first task is to determine the theoretical sample distribution of factors
- This problem was solved by Bai and Ng
- They found: in the asymptotic limit, the deviations of principal components from true factors are normally distributed
- They computed the covariance matrix of this distribution

The theoretical sample distribution of common componnets

- Recall that given a factor model:

$$\mathbf{r}_t = \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$$

- The product $\boldsymbol{\beta}\mathbf{f}_t$ is called the common component *C*
- Bai and Ng demonstrate that the theoretical asymptotic distribution of the common components is normal
- And determine its covariance matrix as follows

Theoretical ditribution of common components:

$$\frac{\left(\tilde{C}_{it} - C_{it}\right)}{\left(\frac{V_{it}}{N} + \frac{W_{it}}{T}\right)^{\frac{1}{2}}} \xrightarrow{d} N(0,1)$$

$$V_{it} = \boldsymbol{\lambda}_i' \boldsymbol{\Sigma}_{\Lambda}^{-1} \boldsymbol{\Gamma}_t \boldsymbol{\Sigma}_{\Lambda}^{-1} \boldsymbol{\lambda}_i$$

$$W_{it} = \mathbf{F}_t' \boldsymbol{\Sigma}_{F}^{-1} \boldsymbol{\Phi}_i \boldsymbol{\Sigma}_{F}^{-1} \mathbf{F}_t$$

Theoretical ditribution of common components:

Where:

The vector $\boldsymbol{\lambda}_i$ is the vector of weights of the *i*-th return

The matrix $$\lim \left\| \frac{\boldsymbol{\Lambda}' \boldsymbol{\Lambda}}{N} - \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}} \right\| = 0$$

The matrix $$\lim \frac{\mathbf{F}_t \mathbf{F}_t^{'}}{N} = \boldsymbol{\Sigma}_{\mathbf{F}}$$

The matrix $$\boldsymbol{\Gamma}_t = \lim \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_j E\left(e_{it} e_{jt}\right)$$

The matrix $$\boldsymbol{\Phi}_i = \lim \sum_{s=1}^{T} \sum_{t=1}^{T} E\left(\mathbf{F}_t \mathbf{F}_s e_{it} e_{is}\right)$$

Estimation of covariances

The terms in the covariances are estimated by:

$$\tilde{V}_{it} = \tilde{\boldsymbol{\lambda}}_i' \left(\tilde{\boldsymbol{\Lambda}}'\tilde{\boldsymbol{\Lambda}}\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{e}_{it}^2 \tilde{\boldsymbol{\lambda}}_{\mathbf{i}}' \tilde{\boldsymbol{\lambda}}_{\mathbf{i}}'\right) \left(\tilde{\boldsymbol{\Lambda}}'\tilde{\boldsymbol{\Lambda}}\right)^{-1} \tilde{\boldsymbol{\lambda}}_i$$

$$\tilde{W}_{it} = \tilde{\mathbf{F}}_t' \tilde{\Theta}_i \tilde{\mathbf{F}}_t'$$

where $\tilde{\Theta}_i$ is the heteroscedasticity and autocorrelation consistent (**HAC**) estimator of

$$\tilde{F}_t e_{it}$$

Theoretical ditribution of common factors:

Factors are normally distributed with variances:

$$\sqrt{N}\left(\tilde{\mathbf{F}}_t - \mathbf{H}\mathbf{F}_t\right) \xrightarrow{d} N\left(0, \mathbf{V}^{-1}\mathbf{Q}\boldsymbol{\Gamma}_t\mathbf{Q}'\mathbf{V}^{-1}\right)$$

where $\mathbf{V}$ is the limit of the diagonal matrix formed with the first *k*-largest eigenvalues of the matrix $\frac{\mathbf{R}'\mathbf{R}}{T}$ and

$$\underset{N,T\to\infty}{p\lim}\frac{\tilde{\mathbf{F}}'\mathbf{F}}{T} = \mathbf{Q}$$

Computing the Confidence Bands for Factors and Common Components

Computing the confidence-bands for factors and common components

- The distributions of common components allow to compute confidence bands

- Confidence bands at 95% are

$$\left(\tilde{\mathbf{C}}_{it} - 1.96\left(\frac{V_{it}}{N} + \frac{W_{it}}{T} \right)^{\frac{1}{2}} N^{\frac{1}{2}}, \tilde{\mathbf{C}}_{t} + 1.96\left(\frac{V_{it}}{N} + \frac{W_{it}}{T} \right)^{\frac{1}{2}} N^{\frac{1}{2}} \right) \quad t = 1,2,\ldots,T$$

A sector-based factor model

- Let's construct a sector-fundamental model
- Assume that each stock return is exposed to a market factor and to a factor typical of the industry segment to which it belongs
- The construction of factors is based on Hamelink et al (2001)
- We represent returns using the following model:

A sector-based factor model of the Russell 1000

Consider the following model:

$$r_{it} = f_{Mt} + \sum_{j=1}^{Q} \delta_{ij} f_{jt} + \varepsilon_{it}$$

where f_{Mt} is the market factor

and the factor loadings δ_{ij} are 1 if the stock *i* belongs to industry *j*, zero otherwise

A sector-based factor model of the Russell 1000

- In this form, the model cannot be estimated because factors are perfectly collinear
- We apply the restriction $$\sum_{j=1}^{Q} \delta_{ij} f_{jt} = 0$$
- so that the market factor has no exposure to the sectors
- Using this restriction, we can express one factor in function of the other factors
- Factors can be estimated through a cross sectional regression

Performing Procrustes Analysis of PCA-based and Sector-based Factors

Performing Procrustes analysis of PCA-based and sector-based factors

The principle of Procrustes analysis applied to factor analysis is the following:

Suppose $\mathbf{F}$, $\mathbf{G}$ are matrices representing two distinct *Q*-factors:

$$\mathbf{F} = \begin{bmatrix} f_{11} & \cdots & f_{1Q} \\ \vdots & \ddots & \vdots \\ f_{T1} & \cdots & f_{TQ} \end{bmatrix}, \mathbf{G} = \begin{bmatrix} g_{11} & \cdots & g_{1Q} \\ \vdots & \ddots & \vdots \\ g_{T1} & \cdots & g_{TQ} \end{bmatrix}$$

Procrustes analysis

- Suppose that the mean has been subtracted by each column,
- and that the two matrices have been rescaled by dividing them for the respective Euclidean/Frobenius norms
- The Euclidean/Frobenius norm of a *mxn* matrix is a generalization of the Euclidean norm of a vector

The Euclidean/Frobenius norm

- The Euclidean/Frobenius norm is defined as the square root of the sum of the squares of the matrix elements:

$$\|\mathbf{A}\| = \sqrt{trace(\mathbf{A}'\mathbf{A})} = \sqrt{\sum_{i}^{m}\sum_{j}^{n} a_{ij}^{2}}$$

- We use the Euclidean/Frobenius norm as a measure of the distance of two models

Procrustes analysis

- Procrustes analysis tries to find an orthogonal matrix that minimizes the Euclidean/Frobenius norm of the matrix

$$\mathbf{F} - \mathbf{UG}$$

$$\mathbf{U} = \arg\min_{\mathbf{X}} \left(\left\| \mathbf{F} - \mathbf{XG} \right\| \right)$$

- Procrustes analysis is implemented in major packages

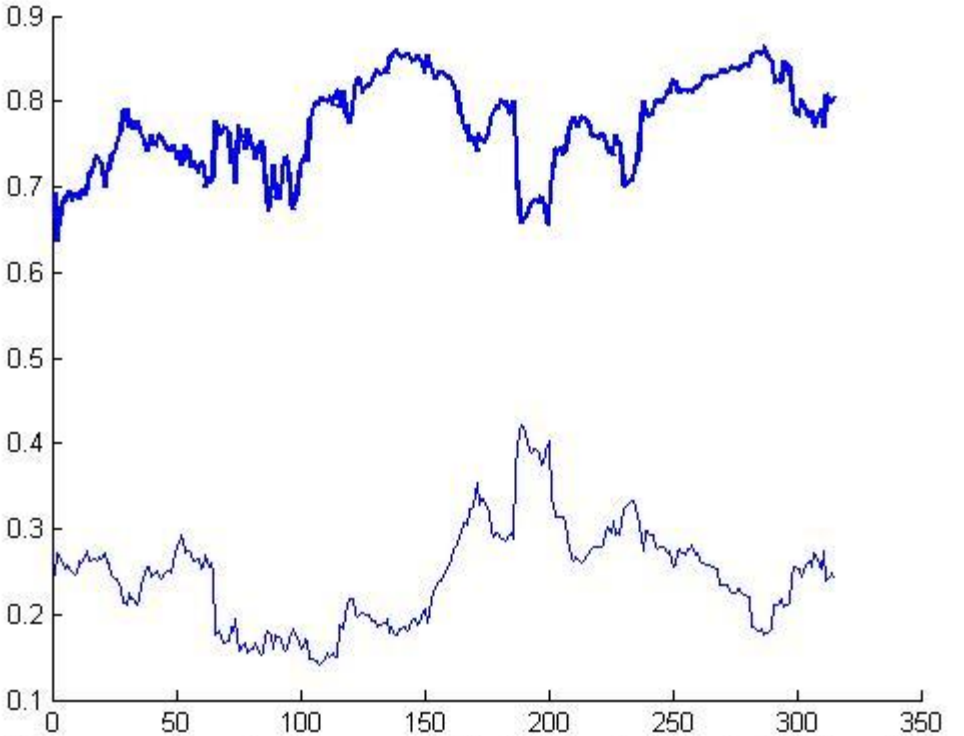
Diversity of factors

- In order to ascertain the degree of similarity between factors,
- we report both the relative mean squared error (MSE) of Procrustes analysis and the correlation coefficient between target factors
- and factors rotated after Procrustes analysis.

MSE of Procrustes analysis and the correlation coefficient for each pair of factors

	Industry- based/ PCA- based
Factor 1	0.9895
Factor 2	0.6548
Factor 3	0.6840
Factor 4	0.8304
Factor 5	0.7019
Average Factor Correlation	0.7721
Relative MSE	1.0180

Correlation coefficient PCA- and industry-based factors



Mean corr. coeff. and mean square residuals sector factors and rotated pca factors

Forecasting

	Industry- based Factors	PCA- based Factors
Average Forecast Correlation	0.0775	0.0732
Relative MSE	1.0180	1.0318

Dynamic factor models

- Dynamic factor models are models that allow to specify a dynamics for factors and for the processes themselves.
- Dynamic factor models now have important applications outside the area of financial econometrics, for example in ecological studies (see, for example, Zuur, Tuck and Bailey, 2003).
- The development of dynamic factor models is recent in comparison with static factor models.
- While modern static multi-factor models were proposed by Thurstone and Hotelling in the 1930s, the first dynamic factor models were proposed in econometrics only in 1977 by Geweke (1977) and by Sargent and Sims (1977).

- The subsequent development of dynamic factor models followed three lines
- dynamic factor models of stationary processes in the "finite *N*, large (infinite) *T*" case
- dynamic factor models of stationary processes in the "large (infinite) *N*, large (infinite) *T*" case, and
- dynamic factor models of integrated processes. The literature on dynamic factor models of integrated processes overlaps with the large literature on cointegration.

- Dynamics enter factor models in three different ways:
- specifying a dynamics for the factors,
- specifying a dynamics for the residuals, and
- allowing regression on lagged factors
- Dynamics is typically specified as an autoregressive process.

- Sargent and Sims (1977) and Geweke (1977) proposed a dynamic factor model of the type:

$$r_t = \sum_{i=0}^{\infty} \beta_i f_{t-i} + \varepsilon_t$$

- where returns are an N vector, the betas are NxQ matrices, f is a vector for each *t* and epsilon is a vector.
- It is assumed that *N* is finite, *K*<<*N* and *T* tend to infinity.
- It is also assumed that factors and residuals are uncorrelated and that residuals are mutually uncorrelated though possibly autocorrelated.
- This model is the dynamic equivalent of the strict factor model.

A general dynamic factor model

Peña and Box (1987) studied the following more general model:

$$r_t = \beta f_t + \varepsilon_t$$
$$\Phi(L) f_t = \Theta(L) \eta_t$$
$$\Phi(L) = I - \Phi_1 L - \cdots - \Phi_p L^p$$
$$\Theta(L) = I - \Theta_1 L - \cdots - \Theta_q L^q$$