Financial Data Project: Notes on Factor Analysis

Dom Owens

18/11/2019

Factor Analysis

Within this project, we will use **Factor Analysis** as a dimension reduction technique for our multivariate time series

$$\mathbf{x}_t$$

The problem is treated as follows:

We wish to reduce the dimension of an observable random vector

$$\mathbf{x} \in \mathbb{R}^p$$

to a smaller vector

$$\mathbf{x} \in \mathbb{R}^m$$

of latent variabes, where $m << p$.

We do this by expressing each entry x_i as a linear combination of the factors:

$$x_i = \lambda_{i1} f_1 + ... + \lambda_{im} f_m + \epsilon_i \qquad \mathbf{x} = \Lambda \mathbf{f} + \boldsymbol{\epsilon}$$

Here, Λ are the **factor loadings**, and

$$\boldsymbol{\epsilon}$$

are errors.

We make the following assumptions:

- $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $E(\boldsymbol{f}) = \mathbf{0}$, $E(\boldsymbol{x}) = \mathbf{0}$ (WLOG)
- $E(\boldsymbol{\epsilon\epsilon^T}) = \boldsymbol{\Psi}$ is a diagonal matrix
- $E(\boldsymbol{ff^T}) = \boldsymbol{I}_m$, so that the factors are independent
- For inferential purposes, we make distributional assumptions on $\boldsymbol{f}$ or $\boldsymbol{x}$ (often multivariate normality)

We will be working with **time series** data and models, meaning our observations

$$\boldsymbol{x}_t$$

are indexed in time by $t \in \{0, 1, ...T\}$. We further assume that

- The **Covariance** matrix Σ is constant with respect to t (this is the **static** model, as opposed to the more complicated **dynamic** model)
- The differenced logarithm of the series is second-order stationary

Indeed, underlying the whole idea of forecasting financial markets is the *Big Assumption*, which is that economic activity in the near future will closely resemble economic activity in the past.

Estimation

As opposed to the similar-looking regression problem

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

in which the desgin matrix X is known, we know neither $\boldsymbol{f}$ nor Λ; hence, any "best-fit" solutions will not be unique.

For conducting estimation in practice, we often find $\hat{\Lambda}$ and $\hat{\Phi}$, then find $\hat{\boldsymbol{f}}$. In disciplines such as finance and economics, the factors can be pre-specified according to theoretical justifications (see the Fama-French factor model; we will use a mathematical approach instead.

One way of finding estimates is through Principal Components Analysis (PCA), called **Principal Factor Analysis (PFA)**. is what we will use in this project.

We work with the covariance and sample covariances matrices

$$\Sigma = \Lambda\Lambda^T + \Psi \text{ and } S = \frac{1}{T}XX^T$$

Suppose that we have the principal components decomposition

$$\boldsymbol{x}_t = A^T\boldsymbol{z}_t$$

where $A \in \mathbb{R}^{p \times n}$ is a matrix consisting of eigenvectors $\boldsymbol{\alpha_i}$, each corresponding to an eigenvalue l_i in decreasing order.

$$\boldsymbol{z} \in \mathbb{R}^p$$

where \$ p \$ is the number of different series being measured.

By partitioning the decomposition into the principal m and minor $p - m$ components, we obtain a factor analysis and error accordingly:

$$\begin{aligned}\mathbf{x}_t &= (A_m|A^*_{p-m})^T \left(\frac{\mathbf{z}_{m,t}}{\mathbf{z}^*_{p-m,t}}\right) \\ &= A_m^T\mathbf{z}_{m,t} + (A^*_{p-m})^T\mathbf{z}^*_{p-m,t} \\ &= \Lambda\boldsymbol{f}_t + \boldsymbol{\epsilon}_t\end{aligned}$$

Substituting in the sample covariance and normalising gives us our estimates (for a model with $\Psi = \sigma^2 I$ this maximises the log-likelihood function (PRML p.548); here we do not assume this)

$$\hat{\Lambda} = \sqrt{p}A_m^T, \hat{\Phi} = pA^{*T}_{p-m}A^*_{p-m}$$

Factor Analysis: Selecting the number of factors m

We can select the number of factors to use, m, using the knowledge that each factor explains a decreasing amount of the total variance, given our static model estimated with PCA.

We might plot the Scree plot (with `screeplot`) and identify a drop-off in the variance explained, or pick an amount of factors sufficient to explain, say, 70% of the variance. Alternatively, information criteria can give a systematic means of selecting m, though these are slightly more complicated.

Forecasting

We opt for the **direct forecast** procedure, where data at $t + h$ is projected onto factors $\mathbf{f}_t$.

Forecasting can be conducted for a single series y_t, which is dependent on the factors, or for the whole vector of series $\mathbf{x}_t$.

References

Pattern Recognition and Machine Learning, Christopher Bishop

Dynamic Factor Models Matteo Barigozzi

Forecasting Using Principal Components From a Large Number of Predictors Stock, Watson