Tutorial lecture 3
Reducing the dimension of the parameter space: Factor Models

Modeling of comovement or of relations between single time series in multivariate time series. Here we consider (static and dynamic)

- Principal component models
- Frisch or idiosyncratic noise model
- Reduced rank regression

3.1 The basic framework:

We restrict ourselves to the stationary case:

$$y_t = \Lambda(z)\xi_t + u_t, \quad \mathbb{E}\xi_t u_s' = 0 \tag{1}$$

where

y_t	...	observations (n–dim.)
ξ_t	...	factors (unobserved) ($r << n$–dim.)
$\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j$, $\Lambda_j \in \mathbb{R}^{n \times r}$	...	factor loadings
$\hat{y}_t = \Lambda(z)\xi_t$	...	latent variables
$\Lambda = \Lambda_0$	...	(quasi-static) case.

Spectral densities:

$$f_y = \Lambda f_\xi \Lambda^* + f_u \tag{2}$$

Ass.: $f_y(\lambda) > 0$, $f_\xi(\lambda) > 0$, $\mathrm{rk}\Lambda = r$

for the quasi-static case we obtain

$$\Sigma_y = \Lambda \Sigma_\xi \Lambda^* + \Sigma_u \quad \text{where e.g. } \Sigma_y = \mathbb{E} y_t y_t' \tag{3}$$

Identifiability questions:

- Identifiability of $f_{\hat{y}} = \Lambda f_\xi \Lambda^*$ and f_u
- Identifiability of Λ and f_ξ

Estimation of integers and real-valued parameters:

- Estimation of r
- Estimation of the free parameters in Λ, f_ξ, f_u
- Estimation of ξ_t

Forecasting model for factors:

$$\xi_{t+1} = a(z)\xi_t + d(z)x_t + \epsilon_{t+1}, \quad (\epsilon_t) \text{ white noise, } \mathbb{E}x_t\epsilon_s' = 0 \tag{4}$$

stability condition: $\det(I - za(z)) \neq 0 \quad |z| \leq 1$

3.2 Principal Component Analysis

1. The quasi-static case:

Eigenvalue decomposition of Σ_y:

$$\Sigma_y = O\Omega O' = \underbrace{O_1\Omega_1 O_1'}_{\Sigma_{\hat{y}}} + \underbrace{O_2\Omega_2 O_2'}_{\Sigma_u},$$

where Ω_1 is the $r \times r$–dim. diogonal matrix containing the r largest eigenvalues of Σ_y.
This decomposition is unique for $\omega_r > \omega_{r+1}$.
A special choice for the factor loading matrix is $\Lambda = O_1$, then

$$y_t = O_1\xi_t + u_t$$

$\xi_t = O_1' y_t$, $u_t = y_t - O_1 O_1' y_t = O_2 O_2' y_t$

Note: Factors are linear functions of y_t.

Estimation:

Determine r from $\omega_1, \ldots, \omega_n$
Estimate $\Lambda, \Sigma_\xi, \Sigma_u, \xi_t$ from the eigenvalue decomposition of $\hat{\Sigma}_y = \frac{1}{T}\sum_{t=1}^{T} y_t y_t'$

2. The dynamic case:
We commence from the spectral density f_y rather than from Σ_y

$$f_y(\lambda) = \underbrace{O_1(\lambda)\Omega_1(\lambda)O_1^*(\lambda)}_{f_{\hat{y}}(\lambda)} + \underbrace{O_2(\lambda)\Omega_2(\lambda)O_2^*(\lambda)}_{f_u(\lambda)},$$

then

$$y_t = O_1(z)\xi_t + u_t$$

$\xi_t = O_1^*(z)y_t$

Note: Here $\mathbb{E}u_t'u_t$ is minimal among all decompositions where $\mathrm{rk}(\Omega(z)) = r$ a.e.

Again $\xi_t = O_1^*(z)y_t$, i.e. factors are linear transformations of (y_t)
Problem: In general, the filter $O_1^*(z)$ will be non-causal and non-rational. Thus, naive forecasting may lead to infeasible forecasts for y_t. Restriction to causal filters is required.
In estimation, we commence from a spectral estimate.

3.3 The Frisch model

Here the additional assumption f_u is diagonal is imposed in (1).

Interpretation: Factors describe the common effects, the noise u_t takes into account the individual effects, e.g. factors describe markets and sector specific movements and the noise the firm specific movements of stock returns.

For given $\hat{y}_t$ the components of y_t are conditionally uncorrelated.

1. The quasi-static case:

Identifiability: More demanding compared to PCA

$$\Sigma_y = \underbrace{\Lambda \Sigma_\xi \Lambda'}_{\Sigma_{\hat{y}}} + \Sigma_u, \quad (\Sigma_u) \text{ digaonal} \tag{5}$$

Identifiability of $\Sigma_{\hat{y}}$:

Uniqueness of solution of (5)
for given n and r, the number of equations (i.e. the number of free elements in Σ_y) is $\frac{n(n+1)}{2}$. The number of free parameters on the r.h.s. is $nr - \frac{r(r-1)}{2} + n$. Now let

$$B(r) = \frac{n(n+1)}{2} - (nr - \frac{r(r-1)}{2} + n) = \frac{1}{2}((n-r)^2 - n - r)$$

then the following cases may occur:

$B(r) < 0$: In this case we might expect non-uniqueness of the decomposition
$B(r) \geq 0$: In this case we might expect uniqueness of the decomposition

The argument can be made more precise, in particular, for $B(r) > 0$ generic uniqueness can be shown.

Given Σ_y, if $\Sigma_\xi = I_r$ is assumed, then Λ is unique up to postmultiplication by orthogonal matrices (rotation).

Note that, as opposed to PCA, here the factors ξ_t, in general, cannot be obtained as a function of the observations y_t. Thus, the factors have to be approximated by a linear function of y_t. The following two approximations are used:

1. The regression method investigated by Thomson:
   The idea, here, is to estimate ξ_t by a linear function of y_t such that the variance of the estimation error, $\xi_t - \hat{\xi}_t$, is minimal. Therefore, $\hat{\xi}_t$ is given by the regression of ξ_t onto y_t,

$$\hat{\xi}_t^T = \Lambda' \Sigma_y^{-1} y_t, \tag{6}$$

   since by the above assumptions

$$\mathbb{E} y_t \xi_t' = \mathbb{E}[(\Lambda \xi_t + u_t)\xi_t'] = \Lambda. \tag{7}$$

   As can easily be seen, this estimator is biased in a certain sense, since $\mathbb{E}(\hat{\xi}_t^T | \xi_t) = \Lambda' \Sigma_y^{-1}(\Lambda \xi_t + \mathbb{E}(u_t | \xi_t)) \neq \xi_t$.

2. Bartlett's method:
   In his method Bartlett suggests to minimize the sum of the standardized residuals with respect to $\hat{\xi}_t$, i.e.,

$$\min_{\hat{\xi}_t}(y_t - \Lambda\hat{\xi}_t)'\Sigma_u^{-1}(y_t - \Lambda\hat{\xi}_t). \quad (8)$$

   Thus, the estimate for ξ_t is given by

$$\hat{\xi}_t^B = (\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}y_t. \quad (9)$$

   This estimate is unbiased in the same sense as above, if $\mathbb{E}(u_t|\xi_t) = 0$ holds true, since $\mathbb{E}(\hat{\xi}_t^B|\xi_t) = (\Lambda'\Sigma_u^{-1}\Lambda)^{-1}\Lambda'\Sigma_u^{-1}(\Lambda\xi_t + \mathbb{E}(u_t|\xi_t)) = \xi_t$.

Estimation:

If ξ_t and u_t were Gaussian white noise, then the (negative logarithm of the) likelihood function has the form

$$
\begin{aligned}
L_T(\Lambda, \Sigma_u) &= \frac{1}{2}T\log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2}\sum_{t=1}^{T} y_t'(\Lambda\Lambda' + \Sigma_u)^{-1}y_t = \\
&= \frac{1}{2}T\log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2}T\mathrm{tr}((\Lambda\Lambda' + \Sigma_u)^{-1}\hat{\Sigma}_y). \qquad (10)
\end{aligned}
$$

2. The dynamic case:

Here Equation (1) together with the assumption

$$f_u \text{ is diagonal.}$$

is considered. Again u_t represents the individual influences and ξ_t the comovements. The only difference to the previous section is that Λ is now a dynamic filter and the components of u_t are orthogonal to each other for all leads and lags.

There are still many unsolved problems.

3.4 Reduced Rank Regression model

Here we consider a regression model of the form

$$y_{t+1} = F \underbrace{G\tilde{x}_t}_{=\xi_{t+1}} + u_{t+1}, \qquad t \in \mathbb{Z}, \tag{11}$$

where the $\tilde{m}$–dimensional vector process $(\tilde{x}_t)$ of explanatory variables contains possibly lagged inputs x_t and lagged observed variables y_t and (u_t) denotes the n–dimensional noise process. In addition we assume:

(i) (x_t) and (u_t) are uncorrelated, i.e. $\mathbb{E}x_t u_s' = 0 \; \forall s, t$

(ii) (x_t) is stationary with a non-singular spectral density

(iii) (u_t) is white noise with $\mathbb{E}u_t u_t' > 0$

(iv) a stability assumption

Assumption: $\beta = FG$ is of rank $r < \min(n, \tilde{m})$.
Thus, $F \in \mathbb{R}^{n \times r}$ and $G \in \mathbb{R}^{r \times \tilde{m}}$ and $G\tilde{x}_t$ can be interpreted as the r–dimensional factor process (ξ_{t+1}), the matrix F can be interpreted as the corresponding factor loading matrix.

Maximum likelihood estimate is obtained by an OLS estimation of β followed by a weighted singular value decomposition, where only the largest r singular values are kept.

Identifiability: F is unique only up to postmultiplication by a nonsingular matrix and an analogous statement holds for G and ξ_{t+1}.

Singular value decomposition of $\beta = U\Sigma V'$, where U and V are orthogonal matrices of dimensions n and $\tilde{m}$, resp., and $\Sigma \in \mathbb{R}^{n \times \tilde{m}}$ is the matrix of singular values, σ_i, $i = 1, \ldots, \min(n, \tilde{m})$, arranged in decreasing order. The strictly positive singular values are assumed to be different and the singular vectors, corresponding to these positive singular values, are unique up to sign change and suitably normalized in order to obtain uniqueness.

Direct procedure: Let $\hat{\beta}$ denote the OLS estimator of β and let $\hat{\beta} = \hat{U}\hat{\Sigma}\hat{V}'$ denote its singular value decomposition. The reduced rank estimator of β, denoted as *direct* estimator, then is given by

$$\hat{\hat{\beta}}_D = \hat{U}_1\hat{\Sigma}_1\hat{V}_1' \tag{12}$$

where $\hat{\Sigma}_1 \in \mathbb{R}^{r\times r}$ is the matrix formed from the r largest singular values of $\hat{\Sigma}$ and $\hat{U}_1$ and $\hat{V}_1$, resp., are formed from the first r columns of $\hat{U}$ and $\hat{V}$, resp.

Indirect procedure: SVD for a suitably weighted matrix. For a canonical correlations analysis one would consider

$$\Sigma_y^{-\frac{1}{2}} y_{t+1} = \Sigma_y^{-\frac{1}{2}} \beta \Sigma_{\tilde{x}}^{\frac{1}{2}} \Sigma_{\tilde{x}}^{-\frac{1}{2}} \tilde{x}_t + \Sigma_y^{-\frac{1}{2}} u_{t+1}. \qquad (13)$$

Replacing the population second moments by their sample counterparts, consider the SVD

$$\hat{\Sigma}_y^{-\frac{1}{2}} \hat{\beta} \hat{\Sigma}_{\tilde{x}}^{\frac{1}{2}} = \hat{U} \hat{\Sigma} \hat{V}' \qquad (14)$$

where $\hat{\beta}$ is the least squares estimator. Note, $\hat{U}$, $\hat{\Sigma}$ and $\hat{V}$ are different from $\hat{U}$, $\hat{\Sigma}$ and $\hat{V}$ mentioned above. Retaining only the r largest singular values one obtains (using an obvious notation)

$$\hat{\hat{\beta}}_I = \hat{\Sigma}_y^{\frac{1}{2}} \hat{U}_1 \hat{\Sigma}_1 \hat{V}_1' \hat{\Sigma}_{\tilde{x}}^{-\frac{1}{2}}, \qquad (15)$$

where again $\hat{U}_1$, $\hat{\Sigma}_1$ and $\hat{V}_1$ are different from $\hat{U}_1$, $\hat{\Sigma}_1$ and $\hat{V}_1$ in Equation (12). Furthermore, note that (15) is the ML estimate if there are no lagged variables of y_t contained in $\tilde{x}_t$.

Model specification: Selection of input variables out of a possibly large set of candidate inputs, specification of the dynamics of the inputs and outputs and the number of factors. AIC or BIC-type criterion of the form

$$AIC(\tilde{m}, r) = \log\det\hat{\Sigma}_{u(\tilde{m},r)} + d(\tilde{m}, r)\frac{2}{T}$$

$$BIC(\tilde{m}, r) = \log\det\hat{\Sigma}_{u(\tilde{m},r)} + d(\tilde{m}, r)\frac{\log T}{T},$$

where $d(\tilde{m}, r) = nr + r\tilde{m} - r^2$ is the number of free parameters in β for a given specification and $\hat{\Sigma}_{u(\tilde{m},r)}$ is the one step ahead (in sample) prediction error variance covariance matrix corresponding to the specification indicated and to one of the estimation procedures described above.