UNIVERSITÄT
BIELEFELD
Fakultät für
Wirtschaftswissenschaften

Using subspace algorithms to estimate the factor dynamics in approximate dynamic factor models

ZeSt Kolloquium,
15.4.2025

Financing from the DFG (project GLASS No. 469278259) is gratefully acknowledged.

Dietmar Bauer

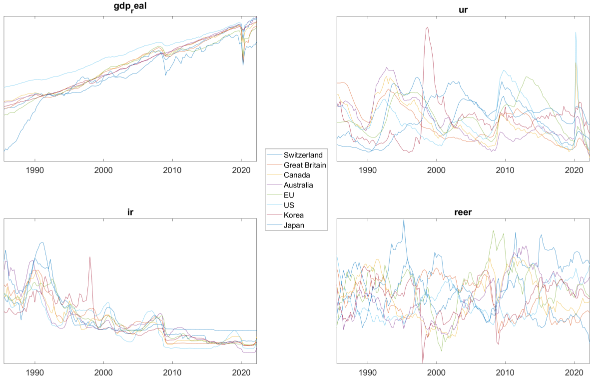
Introduction

1. Introduction

- Often in the analysis of, for example, macro-data or financial data one models a large number of variables jointly.
- Examples include the vectorization of matrix valued time series (MaTS), wherein a number of variables is observed in a number of regions.
- The number of variables typically is smaller but similar to the number of time points.
- To mention just one example, Bai and Ng (2019) use an example with $N = 128$ variables and $T = 676$ (monthly) observations from the FRED-MD data base.
- Joint models for such number of series necessarily contain many parameters and pose problems for specification and estimation.
- In order to cope with the high dimensionality, factor models and generalisations thereof are considered to reduce the dimensionality.

Typical data set

Quarterly data from 8 countries/regions on a number of variables:



Outline

- **Factor models (FM)** reduce the dimensionality by focusing on *common* features that appear in many time series first.
- Common factors typically are **dynamic**, that is not independent over time: state space models.
- **Subspace** procedures like CVA are used to estimate such models.
- **Consistency results** confirm that they provide useful information in the aDFM situation.

This is a methodological talk, I will not show real world data or modelling!

Specification of Factor Models

Approximate (generalized) Dynamic Factor Models: aDFMs

$$y_{it} = \chi_{it} + \xi_{it} = \underbrace{\lambda_i' F_t}_{\text{factor component}} + \underbrace{\xi_{it}}_{\text{idiosyncratic component}}, i = 1, .., N.$$

$$Y_t^N = \Lambda_N F_t + \Xi_t^N \in \mathbb{R}^N$$

- $Y_t^N = [y_{1t}, ..., y_{Nt}] \in \mathbb{R}^N, t = 1, ..., T$
- $F_t \in \mathbb{R}^r$... static factors, where $r \ll N$.

Assumption (Independence)

The factors F_t and the idiosyncratic component ξ_{is} are independent for all variables i and all times t, s.

Assumption (Factor Loadings)

The factor loadings λ_i are assumed deterministic such that
$N^{-1} \sum_{i=1}^N \lambda_i \lambda_i' = N^{-1} \Lambda_N' \Lambda_N \to M_\Lambda > 0$.
$\sup_N \max_i \|\lambda_i\| \leq M_\lambda$.

Stationarity

- In this talk we restrict attention to stationary processes.
- This may involve the need to transform some variables for example via taking temporal differences.

Assumption (Stationarity)

The processes $(F_t)_{t\in\mathbb{Z}}, (\Xi_t^N)_{t\in\mathbb{Z}}$ are jointly wide sense stationary with zero expected value for all N and possess spectral densities.
For each of the processes F_t, Ξ_t^N, y_t^N we have

$$\max_{0\leq k\leq H_T} \max_{i,j} \| T^{-1} \sum_{t=1+k}^{T} x_{t,i}x_{t-k,j} - \mathbb{E}\, x_{t,i}x_{t-k,j} \| = O(Q_T)$$

where $Q_T := \sqrt{(\log\log T/T)}$ and $H_T = (\log T)^a$ for some integer $a > 1$.

2. Specification of Factor Models

Covariance of observations:

$$\Gamma_{y,N} = \mathbb{E}\, y_t^N (y_t^N)' = \Lambda_N (\mathbb{E}\, F_t F_t') \Lambda_N' + \Gamma_{\Xi,N}.$$

Two issues:

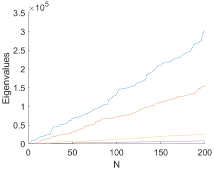
(I) separate common and idiosyncratic parts
(II) identify common factors and loadings from the product $\Lambda_N F_t$.

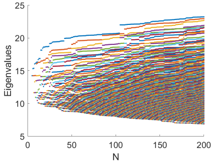
Assumption (Identification)

- $\Gamma_{\Xi,N} = \mathbb{E}\, \Xi_t^N (\Xi_t^N)'$: $\sup_N \lambda_{max}(\Gamma_{\Xi,N}) \leq M_\Xi$.
- $\Gamma_F = \mathbb{E}\, F_t F_t' = I_r$.
- Λ_N is positive lower triangular (for example, it starts with a lower triangular matrix with positive diagonal entries).

It follows that

- $\mathbb{E}\,\chi_t^N(\chi_t^N)' = \Lambda_N(\mathbb{E}\,F_t F_t')\Lambda_N' = \Lambda_N\Lambda_N'$
- $\Lambda_N'\Lambda_N/N \to M_\Lambda > 0$: all eigenvalues of $\Lambda_N\Lambda_N'$ grow essentially linearly as a function of N.
- In $\Gamma_{y,N} = \Lambda_N\Lambda_N' + \Gamma_{\Xi,N}$ the first r eigenvalues are asymptotically (in N) proportional to N.
- The remaining $N - r$ ones remain bounded.
- This suggest to estimate F_t using PCA.





Asymptotic identification in the sense of Chamberlain and Rothschild (1983) $\Rightarrow$ approximate DFM (aDFM).

State Space Modelling

Modelling Dynamics

- Factors typically are not uncorrelated in time.
- Typical models are autoregressive processes:

$$F_t = A_1 F_{t-1} + A_2 F_{t-2} + ... + A_p F_{t-p} + v_t,$$
$$a(z) F_t = v_t, \qquad a(z) = I - A_1 z - A_2 z^2 - ... - A_p z^p.$$

- To understand the dynamics we can rewrite the AR(p) into a higher dimensional AR(1) model:

$$\underbrace{\begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-p+1} \end{pmatrix}}_{x_{t+1}} = \underbrace{\begin{pmatrix} A_1 & A_2 & \dots & \dots & A_p \\ I & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & I & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} F_{t-1} \\ F_{t-2} \\ \vdots \\ F_{t-p} \end{pmatrix}}_{x_t} + \underbrace{\begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{B} v_t$$

$$x_{t+1} = A x_t + B v_t,$$
$$F_t = \underbrace{\begin{pmatrix} A_1 & A_2 & \dots & \dots & A_p \end{pmatrix}}_{C} x_t + v_t.$$

From MA(1) to state space

- Rewrite a MA(1) $F_t = v_t + B_1 v_{t-1}$ into a state space model:

$$F_t = \underbrace{B_1}_{C} \underbrace{v_{t-1}}_{x_t} + v_t$$

$$x_{t+1} = v_t = 0 * x_t + Bv_t$$

- This leads to general state space system of the form:

$$F_t = Cx_t + v_t$$
$$x_{t+1} = Ax_t + Bv_t$$

- One can show that each ARMA system can be rewritten into a state space system.
- The dynamics are contained in the state equation.
- If the matrix A is stable, stationary solutions are given by

$$F_t = v_t + \sum_{j=0}^{\infty} CA^j Bv_{t-j-1}.$$

3. State Space Modelling

Assumption (Rationality; cf. Lippi, Deistler, Anderson (2023))

The static factor process $(F_t)_{t\in\mathbb{Z}}, F_t \in \mathbb{R}^r$, does not depend on N and has a minimal state space representation as

$$F_t = Cx_t + Du_t, \quad x_{t+1} = Ax_t + Bu_t, \quad u_t \in \mathbb{R}^q .$$

where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times q}, C \in \mathbb{R}^{r\times n}, D \in \mathbb{R}^{r\times q}$ and where $(u_t)_{t\in\mathbb{Z}}$ is a stationary, ergodic martingale difference sequence and has expectation zero and variance matrix $\Omega = I_q$.
The transfer function $k(z) = D + zC(I_n - zA)^{-1}B$ has no zeros or poles inside the unit circle.

- Then also $\chi_t^N = \Lambda_N F_t$ has a (singular) state space representation.
- $(u_t)_{t\in\mathbb{Z}}$: white noise process of *dynamic factors*.

Assumption (factor dynamics)

$k(z) = D + zC(I_n - zA)^{-1}B$ where

- The matrix $D = k(0) \in \mathbb{R}^{r \times q}$ has full column rank q such that the heading $q \times q$ submatrix is lower triangular with positive entries on the diagonal.
- There exists a pseudo inverse $D^\dagger$ such that $\underline{A} = A - BD^\dagger C$ is stable. Then the transfer function $c(z) = \sum_{j=0}^{\infty} c_j z^j \in \mathbb{R}^{q \times r}$, $c_j = C\underline{A}^{j-1}BD^\dagger$ is a pseudo left-inverse such that $c(z)k(z) = I_q$.
- There exists a real value $\rho_0 < 1$ such that $\|c_j\| \leq \rho_0^j \mu, \forall j \in \mathbb{N}$ for $0 < \mu < \infty$.
- If $c(z)$ is a polynomial $\rho_0 = 0$. The degree of $c(z)$ then is denoted as p_0.

Then Du_t are the innovations for the process $(F_t)_{t \in \mathbb{Z}}$.

Canonical Variate Analysis (CVA)

4. Canonical Variate Analysis (CVA)

$$F_t = Cx_t + Du_t, \qquad x_{t+1} = Ax_t + Bu_t.$$

$\Rightarrow$ estimation would be simple, if the state was known!

CVA is based on two facts:

1. The state x_t can be approximated by past observations:

$$x_t(p) = \sum_{j=1}^{p} \mathcal{K}_j(p) F_{t-j} = \mathcal{K}_p F_t^- \overset{p\to\infty}{\to} x_t$$

2. Predictions of $F_{t+h}, h \geq 0$, based on the past of F_t are a function of the state:

$$F_{t+h} = \underbrace{CA^h x_t}_{F_{t+h|t-1}} + \underbrace{Du_{t+h} + \sum_{j=0}^{h-1} CA^j Bu_{t+h-j-1}}_{v_{t+h|t-1}}$$

This holds for $h = 0, 1, ..., f$ and can be seen as a multi-step long VAR approximation.

4. Canonical Variate Analysis (CVA)

Jointly this implies using $x_t(p) \approx \mathcal{K}_p F_t^-$

$$\underbrace{\begin{pmatrix} F_t \\ \vdots \\ F_{t+f} \end{pmatrix}}_{F_t^+} = \mathcal{O}_f \mathcal{K}_p \underbrace{\begin{pmatrix} F_{t-1} \\ \vdots \\ F_{t-p} \end{pmatrix}}_{F_t^-} + V_t(p), \qquad (*)$$

$$\mathcal{O}_f = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^f \end{pmatrix}, \mathcal{K}_p = \begin{pmatrix} \mathcal{K}_1(p) & \mathcal{K}_2(p) & \dots & \mathcal{K}_p(p) \end{pmatrix},$$

$$V_t(p) = \begin{pmatrix} v_{t|t-1} \\ \vdots \\ v_{t+f|t-1} \end{pmatrix} + \mathcal{O}_f(x_t - x_t(p)).$$

$(*)$ is a regression equation, where the matrix $\mathcal{O}_f \mathcal{K}_p$ has low rank $n \ll (f+1)r$.

Rank restricted regression (RRR)

How to estimate $\beta \in \mathbb{R}^{f \times p}$ in $Y_t = \beta X_t + U_t, t = 1, ..., T$ subject to $\text{rank}(\beta) = n < \min(f, p)$?

- Unrestricted estimate: $\hat{\beta} = Y'X(X'X)^{-1}$, typically of full rank $\min(f, p)$.
- Perform SVD on weighted version:

$$W_f^+ \hat{\beta} W_p^- = \begin{pmatrix} \hat{U}_n & \hat{U}_R \end{pmatrix} \begin{pmatrix} \hat{S}_n & 0 \\ 0 & \hat{S}_R \end{pmatrix} \begin{pmatrix} \hat{V}_n' \\ \hat{V}_R' \end{pmatrix} = \underbrace{\hat{U}_n \hat{S}_n \hat{V}_n'}_{\text{rank n}} + \hat{R}$$

- Estimate
$$\tilde{\beta} = (W_f^+)^{-1} \hat{U}_n \hat{S}_n \hat{V}_n' (W_p^-)^{-1}.$$
- MLE: $W_p^- = (T^{-1}X'X)^{1/2}, W_f^+ = (T^{-1}Y'Y)^{-1/2}$.
- OLS: $W_p^- = (T^{-1}X'X)^{1/2}, W_f^+ = I_f$.

See Reinsel and Velu (1998).

CVA algorithm in the singular (tall transfer function) case

1. Choose f, p.
2. Perform a rank restricted regression (using OLS weighting) of F_t^+ onto F_t^-. In this step the order n needs to be specified.
3. Use the estimate $\hat{\mathcal{K}}_p$ from the last step to estimate the state $\hat{x}_t(p) = \hat{\mathcal{K}}_p F_t^-$.
4. Estimate C by regressing F_t onto $\hat{x}_t(p)$. This step provides $\hat{v}_t = F_t - \hat{C}\hat{x}_t(p)$.
5. Obtain $\hat{D}$ from a truncated SVD of
$$\hat{\Sigma}_T = T^{-1} \sum_{t=p+1}^{T} \hat{v}_t \hat{v}_t' = \hat{D}\hat{D}' + \hat{R}_q$$
using the q largest singular values, where $\hat{D}$ is p.l.t.
6. Regress $\hat{x}_{t+1}(p)$ onto $\hat{x}_t(p)$ and $\hat{u}_t = \hat{D}^\dagger \hat{v}_t$ to obtain the estimates $\hat{A}$ and $\hat{B}$.
7. Convert $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ to an overlapping form, for example echelon forms.

CVA has been proposed by W. Larimore (1983).

For an R implementation see here.

Asymptotic Results

Extraction of the static factors F_t

- First step: static PCA on $Y_t^N \in \mathbb{R}^N$ using $\frac{\hat{\Gamma}_Y}{N} = (NT)^{-1} \sum_{t=1}^T Y_t^N (Y_t^N)'$.

$$\frac{\hat{\Gamma}_Y}{N} = \frac{\Lambda_N}{\sqrt{N}} \hat{\Gamma}_F \frac{\Lambda_N}{\sqrt{N}}' + \frac{\hat{\Gamma}_\Xi}{N} + \text{cross terms} = \hat{U}_{N,r} \hat{S}_{N,r} \hat{U}_{N,r}' + \hat{R}_N$$

- We get $\hat{\Lambda}_N = \hat{U}_{N,r} \hat{S}_{N,r}^{1/2} \hat{L}_N$ ($\hat{L}_N$ is introduced to fulfill the identification restrictions)
- Then use $(\hat{\Lambda}_N^\dagger = (\hat{\Lambda}_N' \hat{\Lambda}_N)^{-1} \hat{\Lambda}_N')$

$$\hat{F}_t = \hat{\Lambda}_N^\dagger y_t^N = \underbrace{(\hat{\Lambda}_N^\dagger \Lambda_N)}_{\Delta_T} F_t + \hat{\Lambda}_N^\dagger \Xi_t^N.$$

- We obtain

$$T^{-1} \sum_{t=1}^T (\hat{F}_t \hat{F}_t' - \Delta_T F_t F_t' \Delta_T') = O(Q_T + 1/N).$$

- Furthermore $\|\hat{\Lambda}_N/\sqrt{N} - \Lambda_N/\sqrt{N}\| = O(Q_T + 1/N)$.
- In this step the dimension r of the static factor needs to be specified.

5. Asymptotic Results

Under the assumption $T/N^2 \to 0$ we have $O(Q_T + 1/N) = O(Q_T)$.

Theorem (Consistency)

- Let the data be generated according to the before mentioned Assumptions.
- Assume that $T/N^2 \to 0, T \to \infty$.
- Let $p(T) \leq H_T, f \geq n, p = p(T) \geq -\frac{e \log T}{2 \log \rho_0}, e > 1$, for $\rho_0 > 0$ and $p \geq p_0$ for $\rho_0 = 0$.
- Further let $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ denote the CVA estimators (with the OLS weighting) based on the PCA estimates $\hat{F}_t$ for given order n converted to an appropriate overlapping form.

Then

$$\max\{\|\hat{A} - A\|, \|\hat{B} - B\|, \|\hat{C} - C\|, \|\hat{D} - D\|\} = O(Q_T).$$

Hence the transfer function is estimated consistently.

Extensions

Different degrees of persistence:

- We can allow for (co-)integrated factor processes, but then need to use different weights in RRR step.
- In that case consistency still holds for a slightly larger lower bound for the increase of p.
- No knowledge on the number of common trends = integrated state components is needed.
- Also over-differentiation is possible such that the process F_t has simple (i.e. not higher order) spectral zeros at $z = 1$. In that case the rate of convergence drops from $O(Q_T)$ to $O(p^{2.5}Q_T + p^{-1})$.

Conclusions

6. Conclusions

- CVA can be used to obtain consistent estimates in the aDFM setting.
- The main assumption beside the aDFM model structure is that $T/N^2 \to 0$ such that the cross-sectional dimension grows fast enough compared to the time dimension.
- The procedure is simple being based on static PCA and regression methods.
- The method provides information on all required integer parameters, which in all cases lead to consistent estimation.

7. Literature

- Bai, J.; Ng, S. (2002): Determining the Number of Factors in Approximate Factor Models. In: Econometrica 70 (1), S. 191–221.
- Bai, J.; Ng, S. (2004): A PANIC Attack on Unit Roots and Cointegration. In: Econometrica (42), S. 1127–1177.
- Bai, J.; Ng, S. (2019): Rank regularized estimation of approximate factor models. In: Journal of Econometrics 212 (1), S. 78–96.
- Bauer, D. (2005): Estimating linear dynamical systems using subspace methods. In: Econ. Theory 21 (01).
- Bauer, D. (2023): Using Subspace Algorithms for the Estimation of Linear State Space Models for Over-Differenced Processes. Paper submitted to Economics Letters, 2023. Available at Github.
- Breitung, J.; Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors. Oxford Bulletin of Economics and Statistics, 75(1), 23-36.
- Chamberlain, G., Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 51 (5), 1281–1304.
- Deistler, M.; Anderson, B. D.O.; Filler, A.; Zinner, Ch.; Chen, W. (2010): Generalized Linear Dynamic Factor Models: An Approach via Singular Autoregressions. In: European Journal of Control 16 (3), S. 211–224. DOI:
- Forni, M.; Hallin, M.; Lippi, M.; Zaffaroni, P. (2017): Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. In: Journal of Econometrics 199 (1), S. 74–92.
- Forni, M., Lippi, M. (2023) Approximating Singular by Means of NonSingular Structural VARs. EIEF Working Paper 23/1, October 2023.
- Kapetanios, G.; Marcellino, M. (2009). A parametric estimation method for dynamic factor models of large dimensions. Journal of Time Series Analysis, 30(2), 208-238.
- Larimore, W. E. (1983).System identification, reduced-order filtering and modeling via canonical variate analysis. 1983 American Control Conference (pp. 445-451). IEEE.
- Lippi, M.; Deistler, M.; Anderson, B.D.O. (2023): High-Dimensional Dynamic Factor Models: A Selective Survey and Lines of Future Research. In: Econometrics and Statistics 26, S. 3–16.
- Reinsel, G., Velu, R. (1998) Multivariate Reduced-Rank Regression. Lecture Notes in Statistics (LNS, volume 136), Springer.

Thank you for your attention!