UNIVERSITÄT BIELEFELD
Fakultät für Wirtschaftswissenschaften

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

Dietmar Bauer

Bielefeld University

17th International Conference Computational and Financial Econometrics,

Berlin, 16-18 December 2023

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

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.

Approximate (generalized) Dynamic Factor Models: GDFMs

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

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

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

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

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

This assumption is a high level assumption that differs from the literature. Often assumptions on the underlying processes are stated such that the covariance estimates fulfill similar assumptions.

1. Introduction

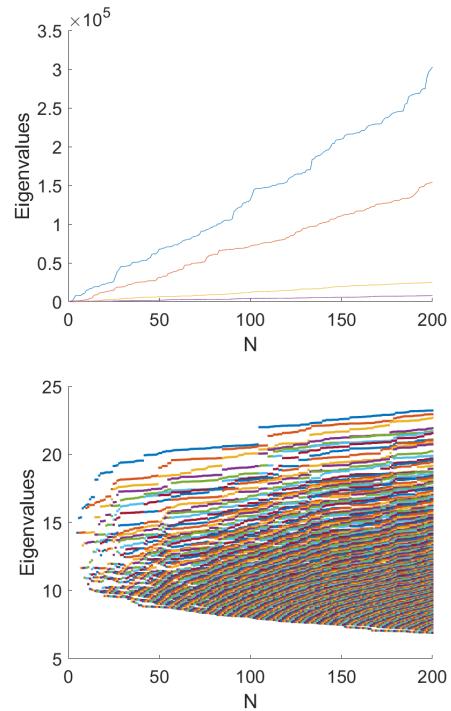
Assumption (Identification)

- $\Gamma_{\Xi,N} = \mathbb{E}\,\Xi_t^N(\Xi_t^N)'$: $\lambda_{max}(\Gamma_{\Xi,N}) \leq M_\Xi$.
- $\Gamma_F = \mathbb{E}\,F_t F_t' = I_r$.
- Λ_N is lower positive 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.

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



1. Introduction

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

$$F_t = \tilde{C}x_t, \quad x_{t+1} = \tilde{A}x_t + \tilde{B}u_{t+1}, \quad u_t \in \mathbb{R}^q .$$

where $\tilde{A} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$, $\tilde{B} \in \mathbb{R}^{\tilde{n}\times q}$, $\tilde{C} \in \mathbb{R}^{r\times\tilde{n}}$ and where $(u_t)_{t\in\mathbb{Z}}$ is a martingale difference sequence and has expectation zero and diagonal variance matrix Ω. The transfer function $b(z) = \tilde{C}(I_{\tilde{n}} - z\tilde{A})^{-1}\tilde{B}$ has no zeros or poles inside the unit circle.

- These assumptions imply that the spectral density:

  $$\Phi_\chi^N(z) = \Lambda_N b(z)\Omega b(z)^* \Lambda_N', .$$

  is rational and has rank q almost everywhere.
- r equals the rank of $\tilde{C} \Rightarrow q \leq r \leq \tilde{n}$.

$(u_t)_{t\in\mathbb{Z}}$: dynamic factors.

Tall transfer functions

$b(z) = \tilde{C}(I_{\tilde{n}} - z\tilde{A})^{-1}\tilde{B} \in \mathbb{C}^{r\times q}$ where $q < r$ typically: 'tall'.

- Tall transfer functions can be 'zero-less': $\text{rank}(b(z)) = q, \quad \forall |z| \leq 1$.

Deistler, Anderson, Filler, Zinner (2010):

- This case is generic. For almost every (in a certain sense) system $(\tilde{A}, \tilde{B}, \tilde{C})$ the corresponding $b(z)$ is zero-less.
- Zero-less transfer functions have autoregressive, stable pseudo-inverses:

$$F_t = b(L)u_t \Rightarrow v_t = Du_t = b^{\dagger}(L)F_t = \sum_{j=0}^{p} b_j^{\dagger} F_{t-j}.$$

- **But:** the pseudo-inverse $b^{\dagger}(z)$ is not unique. Not even the lag order. This complicates estimation.

The differentiation issue

- Often one works with integrated variables that are differenced to obtain stationary variables.
- For $y_{i,t} = \lambda_i' F_t + \chi_{i,t} \Rightarrow \Delta y_{i,t} = \lambda_i' \Delta F_t + \Delta \chi_{i,t}$.
- If F_t is I(1), then $\Delta F_t = b(L)u_t$ is stationary.
- Zero-lessness would imply that $b(1)$ has full column rank.
- Consequently, when working with first differences for all variables, zerolessness implies that all dynamic factors have long-run impact on $y_{i,t}$.
- If some dynamic factors only have short-run impact, we obtain w.r.o.g. $b(z) = b_{red}(z)\text{diag}(I_{r-d}, \Delta I_d)$ such that $b(1) = [b_{nz}(1), 0]$.
- Then after integration the last d dynamic factors only have short-run impact.
- In that situation ΔF_t does not possess an AR representation.

1. Introduction

Assumption (factor dynamics)

$b(z) = b_{red}(z) \begin{pmatrix} I_{q-d} & 0 \\ 0 & \Delta(z) I_d \end{pmatrix}, 0 \leq d \leq q$ where

- The matrix $D = b(0) \in \mathbb{R}^{r \times q}$ has full column rank q.
- There exists a wide rational and stable transfer function $c(z) = \sum_{j=0}^{\infty} c_j z^j \in \mathbb{R}^{q \times r}$, $c_j = \underline{C}\underline{A}^{j-1}\underline{B}$, such that $c(z) b_{red}(z) = I_q$.
- 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 $(F_t)_{t \in \mathbb{Z}}$ and the past spaces for $(u_t)_{t \in \mathbb{Z}}$ and $(F_t)_{t \in \mathbb{Z}}$ coincide.
- For $d > 0$ the tall transfer function $b(z)$ is not zeroless and there is no stable, rational left pseudo-inverse for $b(z)$.
- However, u_t can still be approximated arbitrarily well (in the mean square sense) by past F_t's.

Message of the talk

- Use the fact, that projections of the future space onto the past space of F_t are unique (as suggested in Forni, Lippi (2023))
- Embed this in canonical variate (CVA) subspace procedure to obtain consistent estimate of the transfer function $b(z)$.
- Show that this holds also for spectral zeros on the unit circle.

Note: This is different from Kapetanios and Marcellino (2009), who consider applying CVA on y_t^N (thus including a model for the idiosyncratics) and do not allow for spectral zeros.

Canonical Variate Analysis (CVA)

Innovations state space representation

$$b(z) = \tilde{C}(I_n - z\tilde{A})^{-1}\tilde{B} = D + zC(I - zA)^{-1}B \Rightarrow$$
$$F_t = Cx_t + Du_t, \qquad x_{t+1} = Ax_t + Bu_t.$$

CVA is based on two facts:

- The state x_t can be approximated by past observations:

$$x_t(p) = \sum_{j=1}^{p} K_j(p) F_{t-j} \overset{p\to\infty}{\to} x_t$$

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

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

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

2. Canonical Variate Analysis (CVA)

Jointly this implies

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

$$\mathcal{O}_f = \begin{pmatrix} C \\ \vdots \\ CA^{f-1} \end{pmatrix}, \mathcal{K}_p = \begin{pmatrix} K_1(p) & K_2(p) & \dots & K_p(p) \end{pmatrix},$$

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

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

CVA algorithm in the singular (tall transfer function) case

1. Choose f, p.
2. Perform a rank restricted regression 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

$$\hat{\Sigma}_T = T^{-1} \sum_{t=p+1}^{T} \hat{v}_t \hat{v}_t' = \hat{D}\hat{\Omega}_T \hat{D}' + \hat{R}_q$$

   using the q largest singular values.
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.

Note: CVA has been proposed by Larimore (1990).

Issues

- F_t^- typically is a singular vector with rank $n + pq \leq rp$.
- We only use the projection $\hat{x}_t(p)$, which is well defined, even if the left pseudo-inverse is not.
- Spectral zeros (differentiation!) may imply that many lags p are needed to achieve a suitable approximation.
- In the zero-less case there exists an autoregressive left pseudo-inverse such that the approximation of x_t is perfect for $p \geq p_0$.

Asymptotic Results

Extraction of the static factors F_t

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

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

- 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 ($H_T \to I > 0, \hat{\Lambda}_N^\dagger = (\hat{\Lambda}_N' \hat{\Lambda}_N)^{-1} \hat{\Lambda}_N'$)

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

- We obtain

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

- 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.

3. Asymptotic Results

Under the assumption $T/N^2 \to 0$ we have $O(Q_T + 1/N) = O(Q_T)$ and hence the estimation of Λ_N does not change the order of the estimation accuracy:

Theorem (Differenced Process)

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

Then there exists a state space representation $(A_p, B_p, C_p, D_p) \overset{p\to\infty}{\to} (A, B, C, D)$ (a realization of $b(z)$) and matrices $\hat{G}_T \in \mathbb{R}^{q\times q}$ such that

$$\max\{\|\hat{A} - A_p\|, \|\hat{B}\hat{G}_T - B_p\|, \|\hat{C} - C_p\|, \|\hat{D}\hat{G}_T - D_p\|\} = O(\sqrt{p^5 Q_T}).$$

Hence the transfer function is estimated consistently.

Result based on Poskitt (2006), see Bauer (2023) for the square, non-singular case.

3. Asymptotic Results

Theorem (No differentiation)

Under the same assumptions as on the last slide, but where $d = 0$ and where $f = p = p(T) = -\lfloor \frac{e \log T}{2 \log \rho_0} \rfloor$, $e > 1$, for $\rho_0 > 0$ and $f = p \geq p_0$ for $\rho_0 = 0$ we have that $(A_p, B_p, C_p, D_p) = (A, B, C, D)$ and

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

This shows that there are two sources for the worse performance in the over-differenced case:

- The estimation error for the covariance sequence is amplified due to the smallest eigenvalues of $T^{-1} \sum_{t=p+1}^{T-f} F_t^- (F_t^-)'$ tending to zero as $1/p^2$.
- Slow approximation due to the fact that no AR(∞) representation exists where the approximation error between (A_p, B_p, C_p, D_p) and (A, B, C, D) typically is of order $1/p$, see Bauer (2023).

Specification of Integers

Illustration

Simulation system:

- $N = 200$, $T = 800$
- $r = 5, q = 2, n = 3$.
- Factor dynamics: $A = \text{diag}(0.8, -0.8, 0.4)$, all other matrices chosen randomly.
- Idiosyncratic terms: each individual series follows an AR(1) with randomly chosen $|\rho_i| \leq 0.7, i = 1, ..., N$ and noise with variance $\sigma^2 = 0.25$.
- $M = 1000$ replications.

We use three versions of the data:

- Original process: stationary.
- differenced process: all variables are differenced.
- integrated process: the static factors are integrated, the stationary idiosyncratic terms added afterwards.

Dimension of the static factor r

- There are many methods to choose r.
- To mention just one example, Bai and Ng (2002) suggest to use an information type criterion to select the number of static factors:

$$\widehat{IC}_2(k) = \log SSR_k + k\frac{N+T}{NT}\log(NT/(N+T)), \quad SSR_k = \| Y - \hat{\Lambda}_{N,k}\hat{F}_k' \|_{Fr}^2$$

- This penalizes the fit (in terms of variation not explained by the common component) against complexity (modelled as linear in number of factors).
- $\hat{r}$ minimizes the criterion.

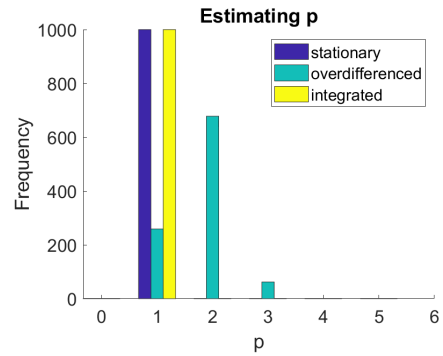
4. Specification of Integers

f, p

- f needs to be large enough such that the observability matrix $\mathcal{O}_f$ is of full rank.
- In the GDFM case, a small value typically is sufficient, as $r \geq n$ is reasonable, where hence $f = 1$ would suffice.
- p should be chosen such that the approximation of x_t by $x_t(p)$ is accurate.
- This is related to the lag selection in a long VAR representation for $\hat{F}_t$.
- Since this in the GDFM setting is a singular process, the information criterion needs to be adapted.
- A simple fix is: $\hat{\Sigma}_T(p)$ denoting the innovation variance estimate for a lag p AR approximation of F_t:

$$IC(p; C_T) = \underbrace{\text{tr}}_{\log\det}\left[\hat{\Sigma}_T(p)\right] + \frac{2r^2 p C_T}{T}.$$

$T = 800, N = 200$



T	200	400	800	1600
stat.	1.00	1.00	1.00	1.00
overdiff.	1.00	1.04	1.80	3.39
integr.	1.00	1.00	1.00	1.00

4. Specification of Integers

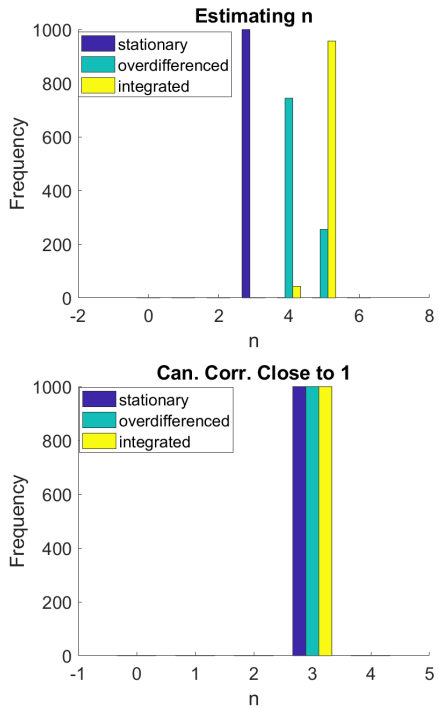
System order n

- Within CVA the system order is often estimated using the singular values.
- Selection can be done using

$$SVC(n) = \sum_{j=1}^{n} \log \hat{\sigma}_j^2 + \frac{2rnC_T}{T}$$

- For sufficiently large penalty term C_T we achieve consistent estimation of the order.
- If $N^2/T \to \infty$ the estimation error in $\hat{F}_t$ is dominated by the sampling error Q_T. Thus the same rates as in the square non-singular case (see Bauer, 2005) arise and $C_T/T \to 0$, $C_T/(fp \log\log T) \to \infty$ is sufficient for consistency.
- Breitung and Pigorsch (2013) show, the canonical correlations are close to one in the singular case.

$T = 800, N = 200$



$N = 50$, percent $\hat{n} = n$

T	200	400	800	1600
stat.	1.00	1.00	1.00	1.00
overdiff.	0.00	0.01	0.28	0.81
integr.	0.06	0.36	0.95	1.00

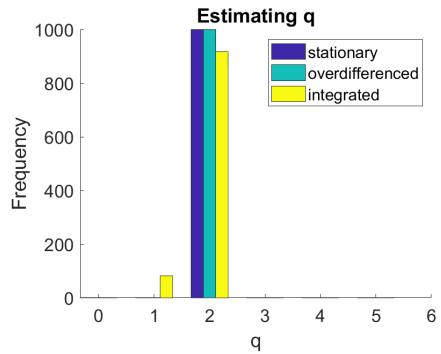
Dynamic factor dimension q

- Typically this would be determined based on dynamic PCA.
- An alternative is to use the estimated innovation variance $\hat{\Sigma}_T \in \mathbb{R}^{r\times r}$.
- The largest q eigenvalues will tend to their non-zero limits, the remaining ones tending to zero.
- Consequently also here information criteria can be used:

$$IC(q; C_T) = \sum_{j=1}^{q} \mu_j(\hat{\Sigma}_T) + \frac{rqC_T}{T}.$$

where $\mu_j(\hat{\Omega}_T)$ denotes the j-th largest eigenvalue.

$T = 800, N = 200$



integr., percent $\hat{q} = q$

T	200	400	800	1600
$N = 50$	0.09	0.34	0.95	1.00
$N = 100$	0.08	0.29	0.93	1.00
$N = 150$	0.08	0.28	0.92	1.00
$N = 200$	0.09	0.31	0.92	1.00

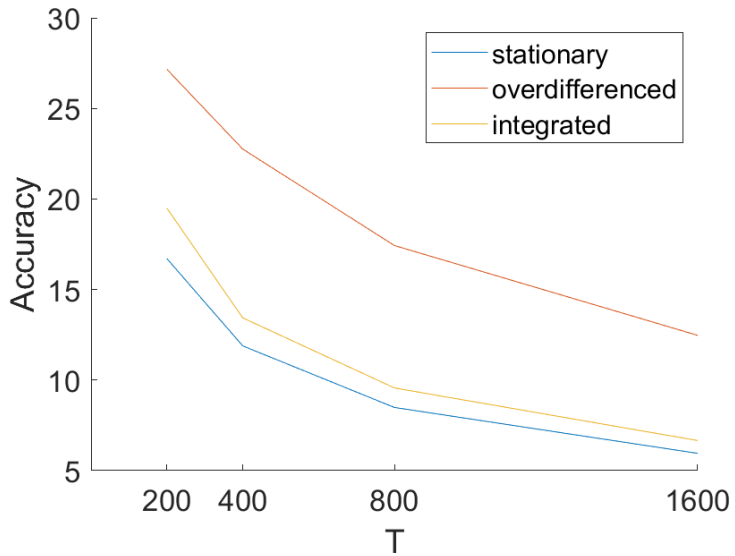
4. Specification of Integers

Estimation Accuracy

- Accuracy measured as

$$\|\hat{D}\hat{\Omega}\hat{D}' - D\Omega D'\| + \sum_{j=0}^{10} \|\hat{C}\hat{A}^j\hat{K}\hat{\Omega}(\hat{C}\hat{A}^j\hat{K})' - CA^jK\Omega(CA^jK)'\|$$

- Averaged over 1000 replications for $N = 50$.
- Convergence is clearly visible.
- Overdifferenced case leads to worst results.



Conclusions

5. Conclusions

- CVA can be used to obtain consistent estimates in the GDFM setting.
- The main assumption beside the GDFM model structure is that $T/N^2 \to 0$ such that the cross-sectional dimension grows faster than the time dimension.
- The procedure is simple being based on static PCA and regression methods.
- The method provides information on all required integer parameters.
- The method also works, if the transfer function relating the dynamic factors to the static factors is rank deficient at zero frequency, as would be expected when working with (over)differenced or seasonally adjusted data.

6. Literature

- Bai, J.; Ng, S. (2002): Determining the Number of Factors in Approximate Factor Models. In: Econometrica 70 (1), S. 191â€“221.
- Bai, J.; Ng, S. (2004): A PANIC Attack on Unit Roots and Cointegration. In: Econometrica (42), S. 1127â€“1177.
- Bai, J.; Ng, S. (2019): Rank regularized estimation of approximate factor models. In: Journal of Econometrics 212 (1), S. 78â€“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), 1281â€“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. 211â€“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. 74â€“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. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In 29th IEEE Conference on Decision and Control (pp. 596-604). 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. 3â€“16.
- Poskitt, D.S., 2007. Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible case. Annals of Institute of Statistical Mathematics 59, 697â€“725.

Thank you for your attention!