Prediction of singular VARs and application to the generalized dynamic factor model

G.Nisol[1] and Siegfried Hörmann

[1]Université libre de Bruxelles

Introduction

- We observe $X_{it}, \; i = 1, \ldots, N, \; t = 1, \ldots, T$.
- Generalized Dynamic factor models :

$$\begin{aligned} X_{it} &= \chi_{it} + \xi_{it} \\ &= b_{i1}(L)u_{1t} + \ldots + b_{iq}(L)u_{qt} + \xi_{it}, \end{aligned}$$

- $u_t = (u_{1t}, \ldots, u_{qt})$ is an unobservable orthonormal white noise, orthogonal to ξ_{it}.
- Cross covariances among ξ_{it} are weak.
- All components (potentially $N > T$) are driven by a small number of factors q.

FHLZ

Classical methods assume that the span of the common components is finite dimensional.

$$X_{it} = \lambda_{i1} F_{1t} + \ldots + \lambda_{ir} F_{rt} + \xi_{it},$$

Estimation and prediction method without this assumption :
Forni, Hallin, Lippi and Zafaroni (2016) , Dynamic Factor Models with Infinite-Dimensional Factor Space: estimation.
Empirical success :
Forni, Giovannelli, Lippi and Soccorsi (2016), Dynamic Factor model with infinite dimensional factor space : forecasting.

VAR representation

Under the assumption of *rational spectral density* :

$$\chi_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \ldots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt},$$

it is shown in FHLZ that for almost all sets of parameters, any subvector $\chi_t^{(\boldsymbol{i})} = (\chi_{i_1 t}, \chi_{i_2 t}, \ldots, \chi_{i_{q+1} t})',\ 1 \leq i_1 < i_2 < \ldots < i_{q+1} \leq N$ has the following VAR representation :

$$\chi_t^{(\boldsymbol{i})} = A_1^{(\boldsymbol{i})} \chi_{t-1}^{(\boldsymbol{i})} + \ldots + A_S^{(\boldsymbol{i})} \chi_{t-S}^{(\boldsymbol{i})} + R^{(\boldsymbol{i})} u_t, \quad t = S+1, \ldots, T, \ (1)$$

- $A_j^{(\boldsymbol{i})}, j = 1, \ldots, S$ are $(q+1) \times (q+1)$ real matrices,
- $R^{(\boldsymbol{i})}$ is an $(q+1) \times q$ matrix
- S finite
- Covariance matrix of the noise $\Omega^{(\boldsymbol{i})}$ is of rank q.

Summary

- FHLZ uses the Yule-Walker equations to estimate the parameters of the VAR processes.
- Singular noise structure implies that

$$\Gamma_S := Cov\left((\chi_t^{(i)}, \chi_{t-1}^{(i)}, \dots, \chi_{t-S+1}^{(i)})'\right)$$

  might be close to singularity,
- which implies a bad accuracy of prediction.
- We therefore suggest a regularization method
- and show on simulation that we outperform.

Forecasting methodology : 1. Covariance estimation

1. Estimate the spectral density matrices of (X_t)

$$\hat{\mathcal{F}}_X(\theta) := \sum_{-B \leq h \leq B} \left(1 - \frac{|h|}{B}\right) \hat{\gamma}_X(h) e^{-\mathrm{i}h\theta}.$$

2. Derive estimators $\hat{\mathcal{F}}_\chi(\theta)$ based on a spectral expansion of $\hat{\mathcal{F}}_X(\theta)$
3. Estimate the auto-covariance operators of the common components through

$$\hat{\Gamma}_\chi(k) = \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} \hat{\mathcal{F}}_\chi(\theta_i) e^{\mathrm{i}k\theta_i}, \quad k = -S, \ldots, S.$$

Forecasting methodology : 2. sub VAR estimation

4. Divide the cross-section into $N/(q+1)$ subsets of indexes of cardinal $q+1$.
5. Obtain the autocovariances $\hat{\Gamma}^{(\boldsymbol{i})}(k)$ of the subvectors $\chi_t^{(\boldsymbol{i})}$.
6. Estimate, by the Yule-Walker equations based on $\hat{\Gamma}^{(\boldsymbol{i})}(k)$, the VAR model (1) for all the subvectors.
   One obtains the estimators $\hat{A}_j^{(\boldsymbol{i})}$, $j = 1, \ldots, S$, .
7. Compute the residuals $\hat{\epsilon}_t^{(\boldsymbol{i})}$. described by

$$X_t^{(\boldsymbol{i})} = \hat{A}_1^{(\boldsymbol{i})} X_{t-1}^{(\boldsymbol{i})} + \ldots + \hat{A}_S^{(\boldsymbol{i})} X_{t-S}^{(\boldsymbol{i})} + \hat{\epsilon}_t^{(\boldsymbol{i})}, \quad t = S+1, \ldots, T,$$

Forecasting methodology 3 : combining

8. Combine the elements of the previous step in order to find the matrices $\hat{A} = (\hat{A}_1, \ldots, \hat{A}_S)$ and the residuals $\hat{\epsilon}_t$ of the entire cross-section
9. Estimate the matrix R and the factors u_t through a principal component analysis on the residuals $\hat{\epsilon}_t$.
10. Compute $\hat{C}(L) = \hat{C}_0 + \hat{C}_1 L + \ldots : \ = \hat{A}(L)^{-1}$ and deduce the h-step ahead prediction
$\hat{X}_{T+h} = \hat{\chi}_{T+h} = \hat{C}_h \hat{R} \hat{u}_T + \hat{C}_{h+1} \hat{R} \hat{u}_{T-1} + \ldots$.
11. Repeat the last algorithm (from step 4) several times permuting the indexes and average the predictions.

Notation

- Assume χ_t is a $(q+1)$-dimensional VAR process with matrix coefficients $A = (A_1, \ldots, A_S)$ and with noise ϵ_t.
- Let $C_t = (\chi_t', \ldots, \chi_{t-S+1}')'$.
- $m = (q+1)S$.
- $\Gamma_S = Cov(C_t)$, $\Omega = Cov(\epsilon_t)$, $\gamma = Cov(\chi_t)$.
- Spectral Decomposition : $\hat{\Gamma}_S = \sum_{k=1}^{m} \hat{\lambda}_k \hat{e}_k \hat{e}_k'$.
- $$\bar{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{S-1} & A_S \\ I & O & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \\ O & O & \cdots & I & O \end{pmatrix},$$

Ill-conditioning

Throughout, we assume that Ω is of rank q.

Proposition

Let $\|M\|$ be the spectral radius of $\|M\| := \sqrt{\lambda_{\max}(MM')}$. Then,

$$\frac{\lambda_{max}(\Gamma_S)}{\lambda_{min}(\Gamma_S)} \geq \frac{\|\Omega\|}{\|A_S\|^2\|\gamma\| \left(1 + \sum_{j=1}^{S-1} \|A_j\|^2\right)}.$$

Hence, overestimation of S implies a infinitely large condition number. Moreover, if $A_j = a_j I_{q+1}$, for $a_j \in \mathbb{R}$, one can show that $\lambda_{min}(\Gamma_S) = 0$.

Prediction Inaccuracy 1

Assume that we observe $\tilde{C}_t = C_t + \Xi_t$ where Ξ is the error term. We look at the prediction error, PE, between the actual forecast and the optimal forecast :

$$PE := \Pi_{q+1} \hat{\bar{A}} \tilde{C}_s - \Pi_{q+1} \bar{A} C_s.$$

It can be decomposed into

$$\begin{aligned} PE &= \Pi_{q+1} \Psi \hat{\Gamma}^{-1} C_s && + \Pi_{q+1} \bar{A} \Xi_s && + \Pi_{q+1} (\hat{\bar{A}} - \bar{A}) \Xi_s \\ &= H_{1s} && + H_{2s} && + H_{3s}, \end{aligned}$$

where Π_{q+1} is the projection onto the first $q+1$ components and Ψ is a $m \times m$ matrix that depends on χ_t and ϵ_t.

Prediction Inaccuracy 2

Proposition

Assume that $\xi = \mathcal{N}_m(0, \sigma_\xi^2 I_m)$. *It holds that*

$$\frac{1}{T}\sum_{s=1}^{T-1} E\|H_{1s}\|^2 = \frac{1}{T}\sum_{\ell=1}^{m} E\|N_{\ell,T}\|^2;$$

$$\frac{1}{T}\sum_{s=1}^{T-1} E\|H_{2s}\|^2 = \sigma_\xi^2 \sum_{k=1}^{S} \operatorname{tr}(A_k A_k');$$

$$\frac{1}{T}\sum_{s=1}^{T-1} E\|H_{3s}\|^2 = \frac{\sigma_\xi^2}{T}\sum_{\ell=1}^{m} E\left(\frac{\|N_{\ell,T}\|^2}{\hat{\lambda}_\ell}\right),$$

where $(N_{1,T}', \ldots, N_{m,T}')' \xrightarrow{\mathcal{D}} \mathcal{N}_m(0, I_S \otimes \Omega)$

If $\hat{\lambda}_k$ is too small, MPSE explodes.

Solution proposed

- Estimate A by
$$\hat{\hat{A}}^{\mathrm{TRIM}} = \hat{\hat{A}} P_K$$
where P_K is the projection onto the K first eigenvectors of $\hat{\Gamma}_S$.
- It limits the previous exploding sum to its first K term(s).
- The value of K is chosen to minimize a proxy of the MSPE.

Simulation

We simulate from the data generating process (DGP) described by

$$X_{it} = \frac{c_{i1}(L)}{d_{i1}(L)} u_{1t} + \frac{c_{i2}(L)}{d_{i2}(L)} u_{2t} + \ldots + \frac{c_{iq}(L)}{d_{iq}(L)} u_{qt} + \xi_{it},$$

- $c_{ij}(L) = a_{ij}$ and $d_{ij}(L) = 1 + \alpha_{ij} L$ where $a_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{U}[-1, 1]$ and $\alpha_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{U}[-0.8, 0.8]$.
- $u_{jt} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1),\ j = 1, \ldots, q,\ t = 1, \ldots, T.$
- $\xi_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$
- Our quality measure : $\frac{\|\hat{X}^{(TRIM)}_{T+1} - X^{(opt)}_{T+1}\|}{\|\hat{X}^{(FHLZ)}_{T+1} - X^{(opt)}_{T+1}\|}$ at every repetition where $X^{(opt)}_{T+1} = E[X_T | A(L), \chi_{i,T}, \chi_{i,T-1}, \ldots]$

Results

	$q=1$			
N, T	60	120	240	480
30	0.37 (0.2)	0.43 (0.23)	0.56 (0.38)	0.54 (0.41)
60	0.42 (0.18)	0.55 (0.22)	0.65 (0.28)	0.63 (0.47)
120	0.47 (0.18)	0.57 (0.22)	0.66 (0.21)	0.61 (0.23)
240	0.5 (0.18)	0.55 (0.19)	0.63 (0.19)	0.61 (0.24)
	$q=2$			
N, T	60	120	240	480
30	0.96 (0.14)	0.97 (0.12)	0.93 (0.17)	0.84 (0.27)
60	0.97 (0.05)	0.96 (0.07)	0.94 (0.1)	0.86 (0.17)
120	0.96 (0.04)	0.96 (0.05)	0.93 (0.08)	0.86 (0.12)
240	0.96 (0.04)	0.96 (0.04)	0.93 (0.06)	0.86 (0.09)

Table 1: Average (standard deviation) of quality measure across repetitions.

Conclusion

Also in the paper:

- An adaptation of the information criterion to choose the VAR order,
- a simulation setting for singular VAR in general,
- a macro-economic application.