On-line Appendix to Nonlinear Forecasts and Impulse Responses for Causal-Noncausal (S)VAR Models

C. Gourieroux*, and J. Jasiak†

August 28, 2022

*University of Toronto, Toulouse School of Economics and CREST, e-mail: *Christian.Gourieroux @ensae.fr*

†York University, e-mail: *jasiakj@yorku.ca*

On-Line APPENDIX B

Identification Conditions

B.1 Independent Component Analysis

Let us consider the independent component model:

$$Y = P\epsilon, \tag{b.1}$$

where the observed vector Y is of dimension n and the components $\epsilon_1, ..., \epsilon_n$ are independent.

Proposition B.1. [Eriksson, Koivunen (2004), Th. 3, and Comon (1994), Th 11]

Under the following conditions:

i) P is invertible,

ii) the components $\epsilon_1, ..., \epsilon_n$ are independent and at most one of them has a Gaussian distribution,

the matrix P is identifiable up to the post multiplication by DQ, where Q is a permutation matrix and D a diagonal matrix with non-zero diagonal elements.

The matrix P is identifiable up to a permutation of indexes and to signed scaling $\epsilon_i \to \pm\sigma_i\epsilon_i$,with $\sigma_i > 0,\ i = 1, ..., n$. The only local identification issue is the positive scaling and can only be solved by introducing identifying restrictions.

Proposition B.2. [Hyvarinen et al. (2001)]

Under the assumptions of Proposition B.1. the local identification issue is solved if P is an orthogonal matrix: $P'P = Id$.

B.2 Two-Sided Multivariate Moving Averages

Proposition B.1 has been extended by Chan, Ho (2004), Chan, Ho, Tong (2006) to two sided moving averages. We give a version of their result for structural mixed models:

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + Pu_t.$$

Proposition B.3

Let us assume that:

i) The roots of $\det(Id - \Phi_1 z - \cdots - \Phi_p z^p) = 0$ are not on the unit circle,

ii) Matrix P is invertible,

iii) (u_t) is i.i.d. with independent components,

iv) Each component admits a finite even moment of order k larger than 3, and at least one non-zero cumulant of order larger than 3.

Then i) $\Phi_1, ..., \Phi_p$ are identifiable, ii) P is identifiable up to the identification issues given in Proposition B.1.

This result corresponds to Condition 4 in Chan, Ho, Tong (2006). Assumption iv) implies that all distributions of the components are non-Gaussian.

The conditions of Proposition B.2. are sufficient for identification. Other sufficient conditions based on the cross-moments of 3rd and 4th order have been considered in the literature to weaken the assumption of cross-sectional independence [see. e.g. Velasco (2022)].

References to On-Line Appendix B

Axler, S., Bourdon, P. and W. Ramey (2001): "Harmonic Function Theory" 2nd Ed., Graduate Texts in Mathematics, Springer.

Chan, K. and L. Ho (2004): "On the Unique Representation of Non-Gaussian Multivariate Linear Processes", Technical Report 341, University of Iowa.

Chan, K., Ho, L. and H. Tong (2006): "A Note on Time Irreversibility of Multivariate Linear Processes", Biometrika, 93, 221-227.

Comon, P. (1994): " Independent Component Analysis: A New Concept", Signal Processing, 36, 287-314

Eriksson, J. and V. Koivunen (2004): "Identifiability, Separability and Uniqueness of Linear ICA Models", IEEE Signal Processing Letters, 11, 601-604.

Hyvarinen, A., Karhunen, J. and E. Oja (2001): "Independent Component Analysis", Wiley.

On-Line Appendix C: Predictive Density Estimation

This Appendix describes the kernel-based estimation of the predictive density given in Proposition 1 from the following time series:

$$\hat{\epsilon}_t = Y_t - \hat{\Phi}_1 Y_{t-1} - \cdots - \hat{\Phi}_p Y_{t-p+1}, \ \ t = 1, ..., T,$$
$$\hat{Z}_{2,t} = \hat{A}^2 \begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix}, \ \ t = 1, ..., T.$$

The above time series are used to approximate the density g of ϵ_t and density l_2 of $Z_{2,t}$ as follows:

$$\hat{g}_T(\epsilon) = \frac{1}{T}\frac{1}{h^m}\sum_{t=1}^{T} K_m\left(\frac{\epsilon - \hat{\epsilon}_t}{h}\right),$$

and

$$\hat{l}_{2,T}(z_2) = \frac{1}{T}\frac{1}{h^{n_2}}\sum_{t=1}^{T} K_{n_2}\left(\frac{z_2 - \hat{Z}_{2,t}}{h}\right),$$

where h_1, h_2 are bandwidths and K_m, K_{n_2} are multivariate kernels of dimensions m and n_2 respectively. Then, the estimated predictive density is:

$$\hat{l}_T(y|\underline{Y}_T) = \frac{\hat{l}_{2,T}\left[\hat{A}^2 \begin{pmatrix} y \\ \tilde{Y}_T \end{pmatrix}\right]}{\hat{l}_{2,T}\left[\hat{A}^2 \begin{pmatrix} Y_T \\ \tilde{Y}_{T-1} \end{pmatrix}\right]} \, |\det \hat{J}_2| \, \hat{g}_T(y - \hat{\Phi}_1 Y_T - \cdots - \hat{\Phi}_p Y_{T-p+1}),$$

This formula is easily extended to bandwidths adjusted for each component, by replacing for example $\frac{1}{h^m} K_m\left(\frac{\epsilon - \hat{\epsilon}_t}{h}\right)$ by $\prod_{j=1}^{m} \frac{1}{h_j} K\left(\frac{\epsilon_j - \hat{\epsilon}_{j,t}}{h_j}\right)$, where K is a univariate kernel. Such an adjustment can account for different component variances.

Let us consider the example of a bivariate VAR(1) process with one noncausal component and a scalar noncausal eigenvalue λ_2 (see Section 6). The estimated coefficients of the inverse of A are denoted by

$$\hat{A}^{-1} = \begin{pmatrix} \hat{a}^{11} & \hat{a}^{12} \\ \hat{a}^{21} & \hat{a}^{22} \end{pmatrix}.$$

The predictive density depends on unknown scalar parameters λ_1, λ_2 and functional parameters l_2, g that can be estimated. The marginal density $l_2[A^2 y]$ can be approximated by a kernel estimator:

$$\hat{l}_{2,T}(\hat{A}^2 y) = \frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{\hat{a}^{21}(y_1 - y_{1,t}) + \hat{a}^{22}(y_2 - y_{2,t})}{h_2}\right),$$

while the density $l_2[A^2 Y_T]$, can be approximated by a kernel estimator:

$$\hat{l}_{2,T}(y_T) = \frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{\hat{a}^{21}(y_{1,T} - y_{1,t}) + \hat{a}^{22}(y_{2,T} - y_{2,t})}{h_2}\right),$$

where h_2 is a bandwidth. The joint density $g(y - \Phi y_T)$ can be approximated by

$$\hat{g}_T(y - \hat{\Phi} y_T) = \frac{1}{T}\frac{1}{h_{11}h_{12}}\sum_{t=1}^{T} K\left(\frac{y_1 - \hat{\phi}_{1,1}y_{1,T} - \hat{\phi}_{1,2}y_{2,T} - \hat{\epsilon}_{1,t}}{h_{11}}\right) K\left(\frac{y_2 - \phi_{2,1}y_{1,T} - \phi_{2,2}y_{2,T} - \hat{\epsilon}_{2,t}}{h_{12}}\right).$$

where $\hat{\epsilon}_{1,t}$ and $\hat{\epsilon}_{2,t}$ are residuals $\hat{\epsilon}_t = y_t - \hat{\Phi} y_{t-1}$ and h_{11}, h_{12} are two bandwidths adjusted for the variation of $\hat{\epsilon}_{1,t}$ and $\hat{\epsilon}_{2,t}$, respectively. We get:

$$\begin{aligned}
\hat{l}_T(y_1, y_2 | Y_T) \quad = \quad & \frac{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{\hat{a}^{21}(y_1 - y_{1,t}) + \hat{a}^{22}(y_2 - y_{2,t})}{h_2}\right)}{\frac{1}{T}\frac{1}{h_2}\sum_{t=1}^{T} K\left(\frac{\hat{a}^{21}(y_{1,T} - y_{1,t}) + \hat{a}^{22}(y_{2,T} - y_{2,t})}{h_2}\right)} \\
& |\hat{\lambda}_2| \frac{1}{T}\frac{1}{h_{11}h_{12}}\sum_{t=1}^{T} K\left(\frac{y_1 - \hat{\phi}_{1,1}y_{1,T} - \hat{\phi}_{1,2}y_{2,T} - \hat{\epsilon}_{1,t})}{h_{11}}\right) \\
& K\left(\frac{y_2 - \hat{\phi}_{2,1}y_{1,T} - \hat{\phi}_{2,2}y_{2,T} - \hat{\epsilon}_{2,t})}{h_{12}}\right)
\end{aligned}$$