Inference in Non-Stationary High-Dimensional VARs

Hecq, A., Margaritella, L.[(a)], Smeekes, S. (Maastricht University)

EC^2 Conference, Dec 11-12, 2020, CREST-ENSAE & ESSEC

Objectives

Design a method allowing to **test for Granger causality in High-Dimensional VAR models in levels, irrespectively of the integration and cointegration orders** of the time series.

Stylized fact

- Hecq et al.(2019) developed an LM test for Granger causality in **high-dimensional stationary VAR** models where the number of variables K is large and potentially larger than the sample T.
- Such test extends the **post-double-selection** (PDS) algorithm designed in Belloni et al. (2014) to the time series setting and allows for block-Granger causality as well.
- The **lasso** is used within the PDS to partial-out the effect of nuisance variables and **uniform asymptotic validity** of the post-selection (generalized) least squares estimator is attained.

- Hecq et al.(2019) relies on the stationarity assumption over the time series entering the VAR but **unit root and cointegration tests are biased**, we would like to avoid them!
- **Toda & Yamamoto (1995)**: simple d **lags-augmentation** of the system reconducts the asymptotics to standard stationary arguments
- d = max order of integration which we suspect in the model (we argue 2 is enough).

Problems with this procedure:

I. Especially in High-dimensions with **K > T**, e.g. $T = 100$, $K = 200$. Say the lasso reduces to $K = 80$. Now, even a one lag augmentation of all the K variables could be **OLS-infeasible**.

II. Purposely over-specifying the VAR model with extra lags comes at a cost of **power loss**.

Solution:

We show it is **not necessary to d-augment all K variables** to have standard inference when we are interested in testing bivariate (or blocks-) Granger causality: it is **enough to augment** the lag of the 2 (or block-size) **variables involved** to obtain **Asymptotic Normality** of the **OLS estimator**.

References

- Belloni, A., Chernozhukov, V., Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies, 81(2), 608-650.
- Hecq, A., Margaritella, L., Smeekes, S. (2019). Granger Causality Testing in High-Dimensional VARs: a Post-Double-Selection Procedure. arXiv preprint arXiv:1902.10991.
- Toda, H. Y., Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. Journal of econometrics, 66(1-2), 225-250.

[(a)] l.margaritella@maastrichtuniversity.nl

A glimpse of the algebra behind

- Let $(x_t, y_t, z_{1,t}, \dots, z_{K-2,t})'$ be a K-dimensional multiple time series process, generated by a VAR(p) and we want to test conditional bivariate Granger non-causality between x_t and y_t.
- We can express the equation of interest in compact form as:

$$y_t = \boldsymbol{\phi}'\boldsymbol{w}_{t,p} + \boldsymbol{\psi}'\boldsymbol{z}_{t,p} + \epsilon_{2,t}. \quad (1)$$

for: $\boldsymbol{w}_{t,p} = (x_{t-1}, y_{t-1}, \dots, x_{t-p}, y_{t-p})'$, $\boldsymbol{z}_{t,p} = (\boldsymbol{z}'_{t-1}, \dots, \boldsymbol{z}'_{t-p})'$,
$\boldsymbol{\phi} := (a_{21}^{(1)}, a_{22}^{(1)}, \dots, a_{21}^{(p)}, a_{22}^{(p)})'$, $\boldsymbol{\psi} := (a_{23}^{(1)}, \dots, a_{2K}^{(1)}, \dots, a_{23}^{(p)}, \dots, a_{2K}^{(p)})'$,
hence: $H_0 : a_{21}^{(1)} = \dots = a_{21}^{(p)} = 0$

- In order to test the null hypothesis, we augment only lags of x_t and y_t in the levels of the cross-sectional equation of interest:

$$y_t = \sum_{j=1}^{p}\left[a_{21}^{(j)} x_{t-j} + a_{22}^{(j)} y_{t-j} + a_{23}^{(j)} z_{1,t-j} + \cdots + a_{2K}^{(j)} z_{K-2,t-j}\right] + a_{21}^{(p+1)} x_{t-p-1} + a_{22}^{(p+1)} y_{t-p-1} + \cdots + a_{21}^{(p+d)} x_{t-p-d} + a_{22}^{(p+d)} y_{t-p-d} + \epsilon_{2,t},$$

Intuition: the extra lags of x_t and y_t have true coefficient = 0 and via algebraic trick allow us to rewrite the original p lags of x_t and y_t in first differences $\longrightarrow$ Asymptotic Normality

Post-Double-Selection augmented LM

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\phi} + \boldsymbol{Z}\boldsymbol{\psi} + \boldsymbol{e} = \boldsymbol{W}_{GC}\boldsymbol{\phi} + \boldsymbol{W}_y\boldsymbol{\varphi} + \boldsymbol{Z}\boldsymbol{\psi} + \boldsymbol{e} = \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \boldsymbol{e}^{(0)}$$

	y_{t-1}	x_{t-1}	$z_{1,t-1}$	$z_{2,t-1}$	y_{t-2}	x_{t-2}	$z_{1,t-2}$	$z_{2,t-2}$
$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}^{(0)} + \boldsymbol{e}^{(0)}$	✓	✗	✗	✓	✓	✗	✗	✗
$\boldsymbol{X}_{GC}^{(1)} = \boldsymbol{X}_{-GC}^{(1)}\boldsymbol{\beta}^{(1)} + \boldsymbol{e}^{(1)}$	✓	O	✗	✗	✓	✓	✓	✗
$\boldsymbol{X}_{GC}^{(2)} = \boldsymbol{X}_{-GC}^{(2)}\boldsymbol{\beta}^{(2)} + \boldsymbol{e}^{(2)}$	✓	✓	✗	✗	✓	O	✓	✗
$\boldsymbol{X}_{\hat{S}}$	✓	✗	✗	✓	✓	✗	✓	✗

- Employ the **restricted augmentation** such that

$$\boldsymbol{y} = \boldsymbol{W}^*_{d,-GC}\boldsymbol{\phi}^*_{d,-GC} + \boldsymbol{Z}_{\hat{S}}\boldsymbol{\psi} + \boldsymbol{\xi}, \quad = \boldsymbol{X}^*_{\hat{S}}\boldsymbol{\beta}^{*\dagger} + \boldsymbol{\xi}, \quad (2)$$

where $\boldsymbol{W}^*_{d,-GC}$ contains the original p lags of y_t plus the d augmented lags of both y_t and x_t but NOT the original p lags of x_t

- Regress the OLS residuals $\hat{\boldsymbol{\xi}} = \boldsymbol{y} - \boldsymbol{X}^*_{\hat{S}}\boldsymbol{\beta}^{*\dagger}$ of (2) onto the variables retained by the previous regularization steps plus the original p lags of the Granger causality variable: $\hat{\boldsymbol{\xi}} = \boldsymbol{X}^*_{\hat{S}\cup GC}\boldsymbol{\beta}^{*\ddagger} + \boldsymbol{\nu}$
Given the residuals $\hat{\boldsymbol{\nu}} = \hat{\boldsymbol{\xi}} - \boldsymbol{X}^*_{\hat{S}\cup GC}\boldsymbol{\beta}^{*\ddagger}$, obtain $R^2 = 1 - \hat{\boldsymbol{\nu}}'\hat{\boldsymbol{\nu}}/\hat{\boldsymbol{\xi}}'\hat{\boldsymbol{\xi}}$.
- Reject H_0 if $TR^2 > q_{\chi^2_p}(1-\alpha)$, where $q_{\chi^2_p}(1-\alpha)$ is the $1-\alpha$ quantile of the χ^2 distribution with p degrees of freedom.

Simulations

DGP2=non-sparse toeplitz-like VAR(2) in levels, Identity error-covariance

DGP	Size/Power	K\T	50	100	200	500	1000
		10	7.6	7.7	4.9	5.6	5.1
2	Size	20	6.9	8.3	7.9	4.9	7.3
		50	7.1	6.4	5.7	7.0	6.3
		100	6.9	6.4	6.7	5.9	5.4
		10	15.5	27.4	50.2	94.8	99.9
2	Power	20	12.9	24.5	50.6	92.5	99.9
		50	11.2	24.1	44.7	90.3	99.9
		100	9.4	20.7	48.5	90.4	99.8

Choice of lag-length p

- Consider a range of $p = 1, \dots, 10$ and $\forall p$ fit a diagonal VAR(p), i.e. for each equation run 10 AR(p)
- Call $\hat{\boldsymbol{e}}$ the $T \times K$ matrix of estimated residuals, then $\hat{\boldsymbol{\Omega}} := T^{-1}(\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}})$ will be the $K \times K$ empirical covariance matrix
- Select the model which minimise an information criterion as

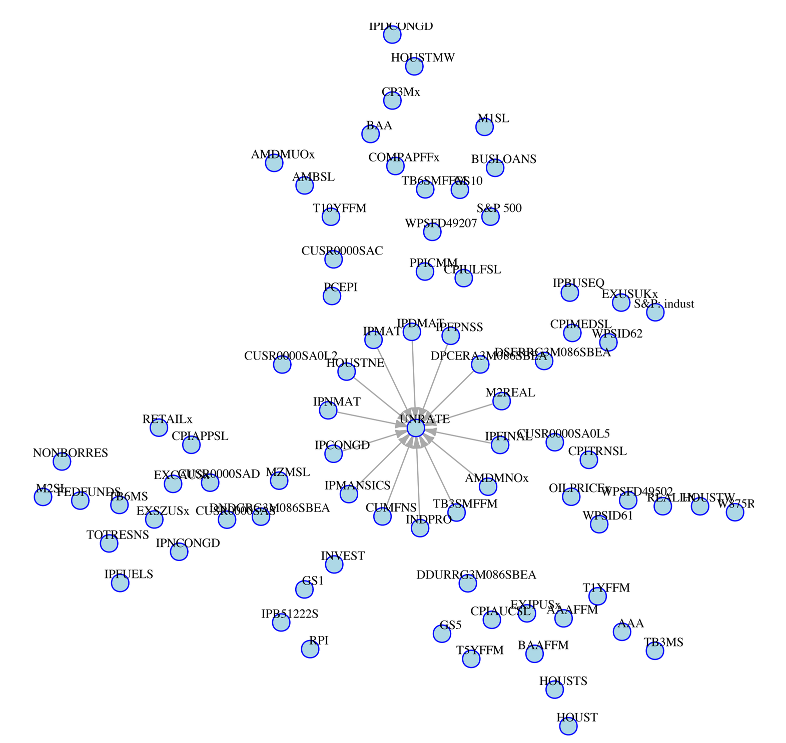
$$\text{AIC:} \quad \log(det(\hat{\boldsymbol{\Omega}})) + \frac{2pK}{T}; \qquad \text{BIC:} \quad \log(det(\hat{\boldsymbol{\Omega}})) + \frac{\log(T)}{T}pK$$

- Problem: if $K > T$ we bypass the dimensionality issue with the diagonal VAR(p) but $\hat{\boldsymbol{\Omega}}$ will be singular
- Solution: we approximate the determinant of $\hat{\boldsymbol{\Omega}}$ by using the product of its diagonal elements, such that:

$$\text{BIC}^*: \quad \log\left(\prod_{i=1}^{K}(\hat{\boldsymbol{\Omega}}_{ii})\right) + \frac{\log(T)}{T}pK \equiv \text{tr}\left(\log(\hat{\boldsymbol{\Omega}})\right) + \frac{\log(T)}{T}pK.$$

Application: FRED-MD without transformations

- **Idea**: analyze the main driving factors of inflation and unemployment in the US, 1959-2019, $K = 79$, $T = 729$, $p = 4$
- **Why**: economic debate over the *flattening of the Phillips curve.*
- **How**: building *Granger-causality networks* of the raw FRED-MD series, representing the direction of the predicting connections among macroeconomic variables.



- Latest version at https://sites.google.com/view/luca-margaritella
- R pkg **HDGCvar** at https://github.com/Marga8/HDGCvar, soon on CRAN!