Vector Autoregressions, Identification, and Causality

Alessio Moneta

Scuola Superiore Sant'Anna
a.moneta@santannapisa.it

Causality Club - University of Pisa

8 July 2024

Introduction

Vector Autoregressive (VAR) models are quite simple time series models and — in their *structural* form — the main tools for macroeconomic analysis (e.g. policy, sources of fluctuations, etc.).

Are they equipped to warrant causal claims?

Outline

- VAR (and SVAR) models
- Problem of identification
- Structural causal models and causal discovery

The Vector Autoregressive (VAR) model

Given a vector $\mathbf{y}_t$ of k variables:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t$$

where $\mathbf{A}_i$ $(i = 1, \ldots, p)$ are $(k \times k)$ matrices;

$\mathbf{u}_t$ is a $(k \times 1)$ vector of white-noise error terms (residuals or *forecast errors*), and $E(\mathbf{u}_t \mathbf{u}_t') = \boldsymbol{\Sigma}_u$;

$\boldsymbol{\mu}$ is a $(k \times 1)$ vector of constants (possibly including a deterministic trend).

The Structural Vector Autoregressive (SVAR) model

The SVAR model provides a more precise description of a data generating process:

$$\mathbf{B}_0\mathbf{y}_t = \boldsymbol{\eta} + \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \ldots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

where $\mathbf{B}_i$ $(i = 0, \ldots, p)$ are $(k \times k)$ matrices of structural coefficients;

$\boldsymbol{\varepsilon}_t$ is a $(k \times 1)$ vector of white-noise structural shocks, and $E(\mathbf{u}_t\mathbf{u}_t') = \boldsymbol{\Sigma}_u$;

$\boldsymbol{\eta}$ is a $(k \times 1)$ vector of constants (possibly including a deterministic trend).

Identifying a SVAR from a VAR

From VAR to SVAR model:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \tag{1}$$

$$\mathbf{B}_0 \mathbf{y}_t = \mathbf{B}_0 (\boldsymbol{\mu} + \mathbf{A}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t) \tag{2}$$

$$\mathbf{B}_0 \mathbf{y}_t = \boldsymbol{\eta} + \mathbf{B}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{B}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{3}$$

Relation between reduced-form residuals and structural shocks:

$$\mathbf{u}_t = \mathbf{B}_0^{-1} \boldsymbol{\varepsilon}_t \tag{4}$$

From estimation of (1) one can get (3) only by knowing $\mathbf{B}_0^{-1}$ (quite difficult) or under specific assumptions.

There is indeed a problem of identification.

But what does a SVAR model tell us about causality?

Note on identification

Note that a problem of causal inference is a problem of identification, but that not all the problems of identification are problems of causal inference.

This depends very much on the nature of the structure to be identified.

A structure is *causal* if it allows to predict the effects of interventions.

Identification and causality

Peters's et al. (2017) definition of intervention distribution: given $\mathbf{X} = X_1, \ldots, X_p$ and a Directed Acyclic Graph (DAG) over $\mathbf{X}$, let

$$P(x_1, \ldots, x_p | do(X_j = \tilde{p}(x_j))) := \prod_{i \neq j}^{p} P(x_i | x_{\mathbf{PA}_i}) \tilde{p}(x_j)$$

Is the SVAR model a structure that allows to predict the effects of interventions?

To answer this question, let us first check how such structures are formalized in the literature on causal discovery.

Structural Causal Model

SCMs are key to formalize causal structures.

Definition by Peters et al. (2017):

A SCM $\mathfrak{C} := (\mathbf{S}, P_{\mathbf{N}})$ consists of:

(i) a set A of k assignments

$$X_j := f_j(\mathbf{PA}_j, N_j) \quad j = 1, \ldots, k$$

where $\mathbf{PA}_j \subseteq \{X_1, \ldots, X_k\} \backslash \{X_j\}$ (*parents* of X_j);

(ii) a joint distribution $P_{\mathbf{N}} = P(N_1) \cdot \ldots \cdot P(N_k)$

A causal graph $\mathcal{G}$ is obtained by creating one vertex for each X_j and drawing $X_i \longrightarrow X_j$ if $X_i \in \mathbf{PA}_j$.

$$
\begin{aligned}
X_1 &:= f_1(X_3, N_1)\\
X_2 &:= f_2(X_1, N_2)\\
X_3 &:= f_3(N_3)\\
X_4 &:= f_4(X_2, X_3, N_4)
\end{aligned}
$$

$$
\begin{tikzcd}
X_1 \arrow[r] & X_2 \arrow[d] \\
X_3 \arrow[u] \arrow[r] & X_4
\end{tikzcd}
$$

```
# R-code snippet 1
# generate a sample from the SCM distribution
set.seed(7)
X3<-runif(1000)-0.5
X1<-2*X3 + rnorm(1000)
X2<- (0.5*X1)^2 + rnorm(1000)^2
X4<-X2 + 2*sin(X3 + rnorm(1000))
```

Causal discovery

The literature on causal discovery (e.g. Spirtes et al. 2000, Pearl 2009, Peters et al. 2017) has developed several algorithms to learn a causal graph $\mathcal{G}$ (or a set of observational equivalent graphs) from observational data:

- constraint-based causal discovery (i.e. based on conditional independence tests)

  e.g. PC algorithm, FCI algorithm (Spirtes et al. 2000)

- causal discovery based on specific assumptions about the SEM

  e.g. LiNGAM (Shimizu et al. 2006), which is based on ICA/Non-Gaussianity; RESIT (Peters et al. 2014), which is based on nonlinearity.

SCM for time series

Is a SVAR a SCM?

- Yes: a SVAR with independent shocks ε_t can be formalized as a linear SCM and one can associate to it a *full-time graph*.

Full time graph (example)

$$
\begin{tikzcd}
y_{1t} \arrow[r] & y_{2t} \arrow[r] & y_{3t} \\
y_{1,t-1} \arrow[u] \arrow[ur] \arrow[r] & y_{2,t-1} \arrow[u] \arrow[ur] \arrow[r] & y_{3,t-1} \arrow[u] \\
\cdots \arrow[u] \arrow[ur] & \cdots \arrow[u] \arrow[ur] & \cdots \arrow[u]
\end{tikzcd}
$$

(*Note*: there are algorithms — cf. Runge et al. 2017; Entner-Hoyer 2010 — that try to learn a full-time graph skipping the SVAR)

SCM for time series (cont'd)

Is a SVAR a SCM?

- Yes: a SVAR with independent shocks ε_t can be formalized as a linear SCM and one can associate to it a *full-time graph*.
- No: it is not guaranteed to remain stable only under types of intervention (macroeconomist's view — cf. so-called Lucas's critique)

Impulse Response Function Analysis

Wold decomposition (inverting the autoregressive part):

$$(\mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{u}_t$$

$$\mathbf{y}_t = (\mathbf{I} - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)^{-1}\mathbf{u}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{u}_{t-j}$$

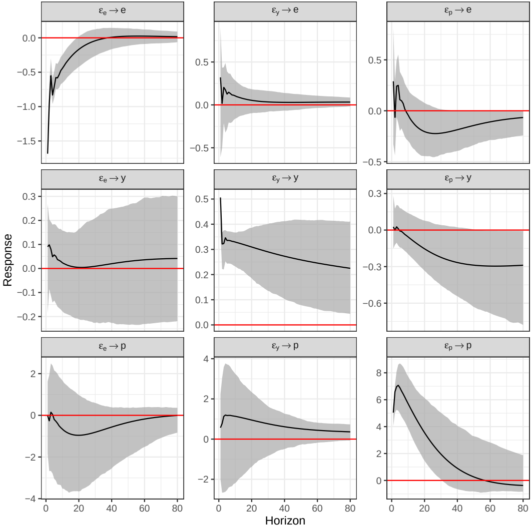
where $\mathbf{\Phi}_0 = \mathbf{I}$, $\mathbf{\Phi}_i = \sum_{j=1}^{i} \mathbf{A}_j \mathbf{\Phi}_{i-j}$ for $i = 1, 2, \ldots$

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{u}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}_j \mathbf{B}_0^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_j \varepsilon_t$$

The elements of $\mathbf{\Psi}_j$ are the impulse response functions:

$$\frac{\partial \mathbf{y}_{t+j}}{\partial \varepsilon_t} = \mathbf{\Psi}_j$$

Example



Note on the Wold decomposition:

The Wold decomposition

$$\mathbf{y}_t = (I - \mathbf{A}_1 L - \ldots - \mathbf{A}_p L^p)^{-1} \mathbf{u}_t$$

is possible only under stability, that is if

$$\det \mathbf{A}(z) = \det(I - \mathbf{A}_1 z - \ldots - \mathbf{A}_p z^p) \neq 0 \text{ for } z \in \mathbb{C}, |z| \leq 1.$$

But in general (even with non-stationary variables), the forecast error associated with an *h*-step forecast is:

$$\mathbf{y}_{t+h} - \mathbf{y}_{t+h|t} = \mathbf{u}_{t+h} + \mathbf{\Phi}_1 \mathbf{u}_{t+h-1} + \ldots + \mathbf{\Phi}_{h-1} \mathbf{u}_{t+1}.$$

Thus we have:

$$\frac{\partial \mathbf{y}_{t+j}}{\partial \mathbf{u}_t} = \mathbf{\Phi}_j; \qquad \frac{\partial \mathbf{y}_{t+j}}{\partial \boldsymbol{\varepsilon}_t} = \mathbf{\Phi}_j \mathbf{B} = \mathbf{\Psi}_j \quad \text{(IRF)}$$

Back to the identification problem...

Identification of a SVAR model reduces to the problem of finding the "right" mixture of $\mathbf{u}_t$:

$$\mathbf{B}_0\mathbf{u}_t = \boldsymbol{\varepsilon}_t \quad \text{or} \quad \mathbf{u}_t = \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t$$

Main assumption:

$E[\varepsilon_t\varepsilon_t'] = I$, which implies that $E[\mathbf{u}_t\mathbf{u}_t'] = \mathbf{B}_0^{-1}(\mathbf{B}_0^{-1})'$

SVAR Identification methods

1. Choleski decomposition of $E[\mathbf{u}_t \mathbf{u}_t']$.
2. A priori zero restrictions on $\mathbf{B}_0$
3. Long-run restrictions.
4. Sign restrictions.
5. Graphical causal discovery applied on $\mathbf{u}_t$.
6. Independent component analysis applied to $\mathbf{u}_t$
7. External instruments
8. Heteroskedasticity
9. ...

Structural VAR

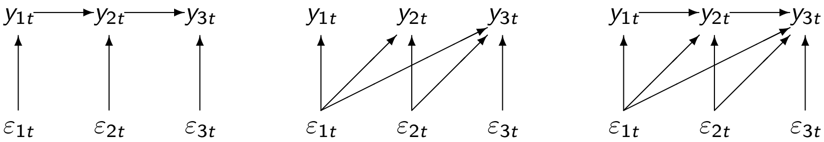
Alternative SVAR formulations...

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{B}_1\mathbf{y}_{t-1} + \mathbf{B}_2\mathbf{y}_{t-2} + \ldots + \mathbf{B}_p\mathbf{y}_{t-p} + \varepsilon_t \tag{1}$$

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0^{-1}\varepsilon_t \tag{2}$$

$$\mathbf{G}_0\mathbf{y}_t = \mathbf{G}_1\mathbf{y}_{t-1} + \mathbf{G}_2\mathbf{y}_{t-2} + \ldots + \mathbf{G}_p\mathbf{y}_{t-p} + \mathbf{F}\varepsilon_t \tag{3}$$

...corresponding to different causal structures, for example:



Some concluding remarks /queries

- SVAR identification for causal discovery or causal discovery for SVAR identification?
- A linear SVAR is not credible that remains invariant to the types of interventions studied in the causal discovery literature (e.g. systematic interventions).
- More work to be done to enrich the possibility of representing causal structures + interventions in a SVAR: not only nonlinearity, but also more complex relations between shocks and structures, etc.

References

Kilian, L., & Lütkepohl, H. (2017). Structural vector autoregressive analysis. Cambridge University Press.

Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms (p. 288). The MIT Press.