Revisiting Identification and Estimation in Structural VARMA Models

C., Gourieroux, and A., Monfort

(March 2015)

Summary

The aim of the paper :

Revisit the standard practices for defining the shocks in structural VAR models and deriving the impulse response functions, that are the dynamic consequences of shocks on the behaviour of macroeconomic or financial time series.

In particular we explain why the identification problems encountered in SVAR models, and often solved by ad-hoc approaches, disappear in a non Gaussian framework.

These results are especially important if economic policies or stress-tests are based on such impulse response functions.

1. THE STANDARD PRACTICE

Step 1 : Specification of a dynamic simultaneous equation model

$$\Phi_0 Y_t = \Phi_1 Y_{t-1} + \varepsilon_t,$$

where the $\varepsilon_t' s$ are serially uncorrelated, $E\varepsilon_t = 0$, $V\varepsilon_t = \Sigma$, Y_t are observed macrovariables, dim $Y_t =$ dim $\varepsilon_t = n$.

Constraints are introduced on Φ_0, Φ_1, Σ, such as the diagonal elements of Φ_0 are equal to 1.

Example : $n = 3$

$$\begin{cases} \text{inflation rate :} & \pi_t = 0.9\pi_{t-1} + 0.2g_{t-1} + \varepsilon_{1,t}, \\ \text{GDP growth :} & g_t = -0.3(i_{t-1} - \pi_t) + \varepsilon_{2,t}, \\ \text{nominal interest rate :} & i_t = 0.9i_{t-1} + 1.5\pi_t + \varepsilon_{3,t}. \end{cases}$$

Reduced form

The model is rewritten as :

$$Y_t = \Phi_0^{-1}\Phi_1 Y_{t-1} + \Phi_0^{-1}\varepsilon_t,$$

which is more appropriate for statistical inference.

Example : $n = 3$

$$\begin{cases} \pi_t &= 0.9\,\pi_{t-1} + 0.2g_{t-1} + \varepsilon_{1,t}, \\ g_t &= 0.06\,g_{t-1} - 0.3i_{t-1} + 0.27\,\pi_{t-1} + 0.3\varepsilon_{1,t} + \varepsilon_{2,t}, \\ i_t &= 0.9\,i_{t-1} + 1.35\pi_{t-1} + 0.3\,g_{t-1} + 1.5\varepsilon_{1,t} + \varepsilon_{3,t}. \end{cases}$$

Step 3 : Estimation by Ordinary Least Squares

The reduced form is usually estimated by O.L.S. equation by equation. This provides estimators of :

$$\Phi_0^{-1}\Phi_1 \text{ and } \Phi_0^{-1}\Sigma(\Phi_0^{-1})'.$$

But this approach creates two identification issues :

i) It is not possible to identify Φ_0, Φ_1, Σ, separately,
ii) The OLS approach provides estimators such that the eigenvalues of $\Phi_0^{-1}\Phi_1$ have their modulus strictly smaller than 1.

Until recently the second identification issue (noncausal process), or its moving-average counterpart (nonfundamentalness in MA) has been neglected.

The first identification issue is partially "solved" by triangularization [Sims (1980) : Macroeconomics and Reality, Econometrica]

More precisely, for determining the impulse response function, we do not need the identification of Φ_0, Φ_1, Σ. Instead we have to consider an autoregressive specification such as :

$$Y_t = \psi Y_{t-1} + C\eta_t,$$

where the error terms η_t are cross-sectionally uncorrelated : $V(\eta_t) = Id$.
The introduction of the no correlation condition on the η is needed to perform shocks on $\eta_{1,t}$, say, with no impact on the other components $\eta_{2,t}, \dots, \eta_{n,t}$.

Step 4 : Triangularization

Since ψ is identifiable, the only identification problem concerns matrix C.

The (static) identification issue :

$$CC' = V_{t-1}(Y_t) \text{ is identifiable, but not } C \text{ itself.}$$

What is proposed by Sims : Assume that C is lower triangular.

Then, by Cholevsky decomposition, there is a unique solution C to the equation $CC' = V_{t-1} Y_t$ (up to the signs of the diagonal elements).

Step 5 : Impulse Response Function (IRF)

Then derive the (deterministic) IRF due to a shock at date T on η_{1T}, say, by computing recursively the effects δY_{T+h} on Y_{T+h} via the recursive equation :

$$\delta Y_{T+h} = \hat{\psi}\delta Y_{T+h-1} + \hat{C}\delta\eta_{T+h-1}, h \geq 0,$$

where : $\delta Y_{T-1} = 0, \delta\eta_{T-1} = 0$,
$\delta\eta_T = \begin{pmatrix} \delta\eta_{1T} \\ 0 \end{pmatrix}, \delta\eta_{T+h} = 0, h \geq 1$, for a transitory shock.

The limits of the approach above

The IRF depends :

on the order of variables Y_t,

on preliminary linear transformations on these variables.

Not the same pattern of IRF, if we consider the order (g_t, π_t, i_t), or the order (π_t, i_t, g_t).

Not the same pattern of IRF, if the triangularization is applied to $(g_t - \pi_t, i_t - \pi_t, \pi_t)$, or to (g_t, i_t, π_t) (real vs nominal).

How to choose among these IRF ?

i) Select the IRF with a "realistic" pattern, or corresponding to the "expectations" of the policy maker.

ii) Introduce additional restrictions (often more structural) :

• short term causality restrictions (on Φ_0, Σ) ;

• long run restrictions (jointly on Φ_0, Φ_1, Σ) [Blanchard, Quah (1989), American Economic Review]

• sign restrictions on either coefficients, or IRF [Uhlig (2005), Journal of Monetary Economics]

in order to identify the IRF.

2. THE NEW APPROACH

Be more careful when defining an Impulse Response Function.

Message 1 : **The shocks have to be made on independent errors, not on uncorrelated errors only.**

Indeed uncorrelated variables may be strongly dependent, even in a deterministic relationship.

Example : $X \sim N(0,1)$, $Y = X^2$,

$$Cov(X, Y) = E(X^3) - E(X)E(X^2) = 0.$$

Message 2 : **To get reliable impulse response, we need "prediction bounds" by means of stochastic shocks, not deterministic ones.**

The shocks concern the distribution of the error, not only its location parameter.

Step 1 : A VARMA model, with independent errors

$$\Phi_0 Y_t = \Phi_1 Y_{t-1} + \Theta_0 \eta_t + \Theta_1 \eta_{t-1},$$

where the $\eta_t' s$ are serially independent, identically distributed, and also cross-sectionally independent ($E\eta_t = 0$, $V\eta_t = Id$).

We assume :

i) $\dim \eta_t = \dim Y_t = n$

ii) the eigenvalues of $\Phi_0^{-1}\Phi_1$ a modulus strictly smaller than 1.

iii) no restriction on the roots of the MA polynomials.

An associated moving average form :

$$Y_t = A_0 \eta_t + A_1 \eta_{t-1} + A_2 \eta_{t-2} + \dots$$

Step 2 : Condition for identification

In general there is no identification problem of the moving average representation.

Message 3 : **The MA coefficients $A_0, A_1, A_2 \ldots$, and the distributions of $\eta_{1t}, \ldots, \eta_{nt}$ are identifiable, if at most one η_{jt} is Gaussian.**

Chan, Ho, Tong (2006), Biometrika,
when η_t has second-order moments ;

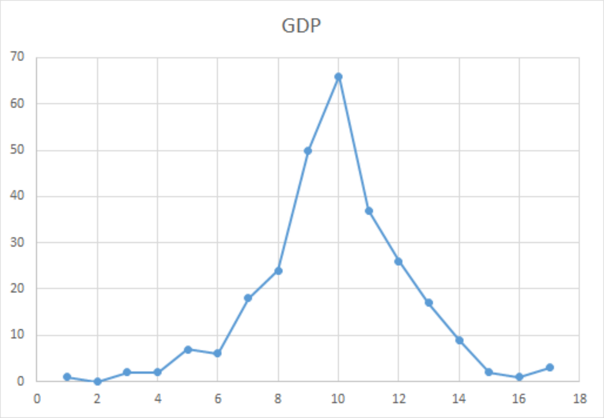
extended to $\eta' s$ with fat tails in Gourieroux, Zakoian (2014), Journal of Time Series Analysis.

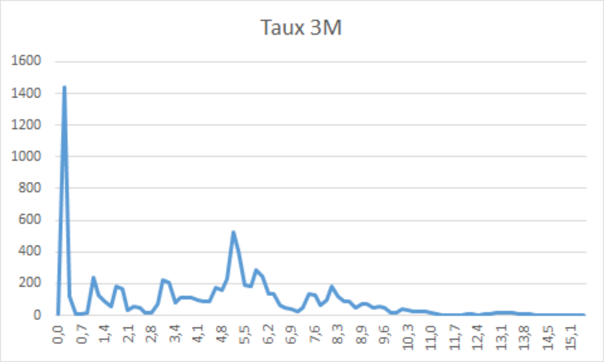
Are the observed macroseries or their associated VAR errors Gaussian ?

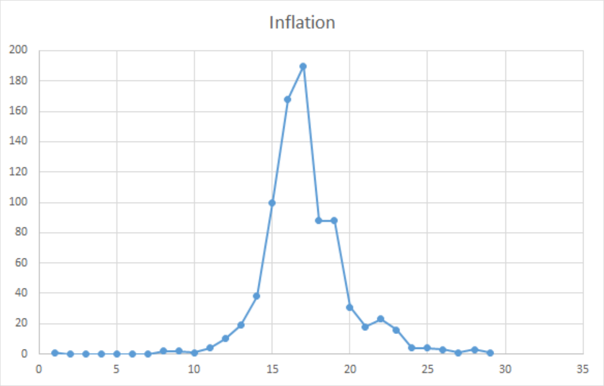
It has to be checked, but :

• the inflation rate is the example given for the first introduction of ARCH models [Engle (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation", Econometrica]

• the evolution of the interest rate is generally described by means of Cox, Ingersoll, Ross model with gamma type distribution, or even bimodal distribution, for taking into account the possibility of a regime of very low rates.


GDP
70
60
50
40
30
20
10
0
0
2
4
6
8
10
12
14
16
18


Taux 3M
1600
1400
1200
1000
800
600
400
200
0
0,0
0,7
1,4
2,1
2,8
3,4
4,1
4,8
5,5
6,2
6,9
7,6
8,3
8,9
9,6
10,3
11,0
11,7
12,4
13,1
13,8
14,5
15,1


Inflation
200
180
160
140
120
100
80
60
40
20
0
0
5
10
15
20
25
30
35

Steps 3-4 : Estimate (see next sections) and derive the stochastic impulse responses.

3. NONFUNDAMENTAL REPRESENTATIONS

Different economic arguments have been given in the literature for ill-located roots of the moving average polynomial.

Typically, for $n = 1$:

$$y_t = \varphi y_{t-1} + c\eta_t - c\theta\eta_{t-1}, |\varphi| < 1, |\theta| > 1.$$

i) A lag between the occurrence of a "shock" and its maximum effect on y_t.

a "shock" of productivity [Lippi, Reichlin (1994), Journal of Econometrics].

a "shock" of fiscal policy [Leeper, Walker, Yang (2013), Econometrica].

ii) A consequence of rational expectation

[Hansen, Sargent (1991) : "Two Difficulties in Interpreting Vector Autoregressions"]

$$y_t = E_t(\sum_{h=0}^{\infty} \beta^h w_{t+h}),$$

with : $w_t = \varepsilon_t - \theta\varepsilon_{t-1}, 0 < \beta < 1, |\theta| < 1.$

If the available information is : $I_t = (\varepsilon_t, \varepsilon_{t-1}, \dots).$, the solution of the RE model is :

$$y_t = (1 - \beta\theta)\varepsilon_t - \theta\varepsilon_{t-1}.$$

The root of the MA polynomial is : $(1 - \beta\theta)/\theta.$

This root can be larger, or smaller than 1.

For such nonfundamental representations, the errors to be shocked are the errors with an economic interpretation.

Message 4 : **The errors to be shocked may differ from the innovations of the observable process.**

4. IMPULSE RESPONSE FUNCTIONS

Conditions on the errors to be shocked.

- The components of the errors have to be independent.
- The observable macrovariables have to be functions of the current and lagged values of the errors.
- These errors need clear economic interpretations.

Example : Is it sufficient to give to the errors an appealing name ?

$$\left\{ \begin{array}{lcl} g_t & = & -0.3(i_{t-1} - \pi_t) + \eta_{r,t} \leftarrow \text{ real shock} \\ i_t & = & 0.9 i_{t-1} + 1.5\pi_t + \eta_{mp,t} \leftarrow \text{ monetary policy shock} \\ \pi_t & = & 0.9\pi_{t-1} + 0.2 g_{t-1} + \eta_{cp,t} \leftarrow \text{cost push shock.} \end{array} \right.$$

Well chosen overidentifying restrictions are crucial for the economic interpretation.

The impulse response function

For a model $Y_t = \Phi Y_{t-1} + C_0\eta_t + C_1\eta_{t-1}$,
with well-located roots for Φ, ill or well-located roots for the MA polynomial.
Can be done in the standard way from data on $Y_1, \ldots, Y_T$, and estimated parameters $\hat{\eta}_T, \hat{\Phi}, \hat{C}_0, \hat{C}_1, \hat{f}_1, \ldots, \hat{f}_n$ (distributions)
Without shock :
Step 1 : Simulate $\eta^s_{T+1}, \ldots, \eta^s_{T+h-1}$ from $\hat{f}_1, \ldots, \hat{f}_n$.
Step 2 : Deduce simulated paths : $Y^s_{T+1}, \ldots, Y^s_{T+h}, \ldots$
by applying $Y^s_{T+h} = \hat{\Phi} Y^s_{T+h-1} + \hat{C}_0\hat{\eta}^s_{T+h} + \hat{C}_1\hat{\eta}^s_{T+h-1}$ (with $Y^s_T = Y_T, \hat{\eta}^s_T = \hat{\eta}_T$)
With shocks : The same principle, but with drawings in shocked distributions $\hat{f}^c_1, \ldots, \hat{f}^c_n$, for instance (for permanent shocks).

5. SEMI-PARAMETRIC ESTIMATION

Message 5 : **Don't estimate parameters Φ, C_0, C_1 by the standard approaches, such as Gaussian pseudo-maximum likelihood, or Kalman filter.**

These methods may provide misleading results, since they inherit the static and dynamic lack of identification existing in the Gaussian framework

In other words : first and second-order moments are not enough informative.

A three step semi-parametric approach

Easily explained for the one-dimensional model :

$$y_t = \varphi y_{t-1} + c\eta_t - c\theta\eta_{t-1}.$$

Step 1 : Estimation of autoregressive coefficient φ by instrumental variable

$$\hat{\varphi} = (\Sigma y_t y_{t-2})/(\Sigma y_{t-1} y_{t-2}).$$

Then deduce the instrumental variable residuals :

$$\hat{z}_t = y_t - \hat{\varphi} y_{t-1}.$$

Step 2 : Estimation of c, θ by a method of moment based on cross-moments of $\hat{z}_t$ up to order 3.

$$\hat{\theta} = -\Sigma(\hat{z}_t^2 \hat{z}_{t-1}) / \Sigma(\hat{z}_t \hat{z}_{t-1}^2),$$

and a corresponding expression for $\hat{c}$.

Then consistent approximations of errors η_t are deduced by inverting appropriately the MA polynomial (backward if $|\hat{\theta}| < 1$, forward, if $|\hat{\theta}| > 1$) :

$$\hat{\eta}_t = \frac{1 - \hat{\varphi} L}{\hat{c}(1 - \hat{\theta} L)} y_t.$$

Step 3 : Estimation of the distribution of $(\eta_{j,t}), j = 1, \ldots, n$

By kernel smoothing of the historical distribution of the $\hat{\eta}_{j,t}, t = 1, \ldots, T$.

Test for overidentifying restrictions

In the non Gaussian case, the parameters of the MA reduced form are identified, and the additional structural restrictions corresponding to causality restrictions, long run restrictions, nonfundamentalness hypothesis are generally overidentifying restrictions.

Message 6 : **The structural restrictions are overidentifying restrictions and can be tested.**

They can be rejected, or accepted.

6. NONCAUSAL REPRESENTATIONS

What is a mixed causal/noncausal process ?

A stationary process satisfying :

$$Y_t = \Phi Y_{t-1} + C\eta_t,$$

where the eigenvalues of Φ can have a modulus strictly smaller, or stricly larger than 1, and (η_t) satisfies the serial and cross-sectional independence assumptions, with $E(\eta_t) = 0$, $V(\eta_t) = Id$.

The unique stationary solution of the autoregressive system is a two-sided moving average in η_t :

$$Y_t = \sum_{j=-\infty}^{+\infty} A_j \eta_{t-j}.$$

For non Gaussian errors, the two-sided MA reduced form is identifiable, i.e. the sequence of A_j, and the distributions of the $\eta_{i,t}$ are identifiable.

Remark : It is important to distinguish the model :
$y_t = 2y_{t-1} + \varepsilon_t$, where ε_t is independent of $y_{t-1}, y_{t-2}, \ldots$

and the model :

$y_t = 2y_{t-1} + \varepsilon_t$, where y_t is strongly stationary.

In the first case :

y_t is explosive (non stationary), ε_t is its innovation.

in the second case :

y_t is stationary, ε_t is not the (linear) innovation of y_t.

Are such mixed autoregressive processes encountered in practice ?

• When estimating VAR model by maximum likelihood with Student errors (i.e. non Gaussian errors), ill-located autoregressive roots systematically appear.
[Lanne, Saikkonen (2013) : Econometric Theory]

• Why ?
In fact a stationary process satisfying : $y_t = 2y_{t-1} + \varepsilon_t$, with for instance $\varepsilon_t \sim$ Cauchy, is a stationary martingale :
$E_{t-1}(y_t) = y_{t-1}$, with trajectories showing recurrent speculative bubbles [Gourieroux, Zakoian (2014)]

A speculative bubble is a local explosion (followed by a crash)

to be distinguished from a global explosion (a trend)

speculative bubble : stationary martingale.

trend : nonstationary martingale $\leftrightarrow$ random walk.

- **As solution of rational expectation equilibrium models**

A RE equilibrium model [Diba, Grossman (1988) : "Explosive Rational Bubbles in Stock Prices ?, AER].

$$y_t = aE_t(y_{t+1}) + \varepsilon_t, a > 0.$$

Standard result valid if $Ey_t^2 < \infty$.

There exists a unique stationary solution $y_t^0 = \varepsilon_t$, if $a < 1$;

There exists an infinite number of stationary solutions, if $a > 1$:

combination of the forward solution : $y_t^0 = \varepsilon_t$

and of the perfect foresight solution : $y_t^1 = \dfrac{L}{L-a}\varepsilon_t$.

[Gourieroux, Laffont, Monfort (1982), Econometrica]

This result is modified if we allow for stationary solutions without first-order moment. The RE equilibrium model is obtained by matching demand and supply and involves two economic shocks : ε_t, w_t, say. It can be shown that, even if $a < 1$, there exists an infinite number of stationary solutions. For instance if (ε_t) and (w_t) are independent, we get solutions :

$$y_t = \varepsilon_t + y_t^*,$$

with $y_t^* = \rho y_{t+1}^* + \varepsilon_t^*$,
where (ε_t^*) and (ε_t) are independent white noises, ε_t^* an appropriate nonlinear function of past current and future values of w_t, and $|\rho| < 1$.
stationary noncausal process $\leftrightarrow$ bubble $\leftrightarrow$ selffulfilling prophetics

How to estimate semi-parametrically noncausal processes ?

By covariance estimators based on the sample counterpart of covariance restrictions :

$$Cov[a(Y_t - \Phi Y_{t-1}), b(Y_{t-h} - \Phi Y_{t-h-1})] = 0, \forall h, \forall a, b.$$

Select a set of functions a, b including nonlinear functions to avoid the identification issue of methods based on first and second-order moments only

$a(y) = y, b(y) = y^2$, for capturing the leverage effects.

$a(y) = y^2, b(y) = y^2$, for capturing the volatility persistence.

Impulse Response Function

The VAR model :

$$\left\{ \begin{array}{lcl} y_t & = & \varepsilon_t + y_t^*, \\ z_t & = & \varepsilon_t, \end{array} \right.$$

where $y_t^* = \rho y_{t+1}^* + \varepsilon_t^*, |\rho| < 1$.

cannot be used directly for deriving the IRF.

Indeed ε_t^* is function of the future values $Y_t, Y_{t+1}, \ldots,$ and has no economic interpretation.

The errors with structural interpretations are :

$$\varepsilon_t \text{ and } w_t,$$

(directly deduced from the errors in demand and supply)

The dynamic system can be rewritten in terms of the structural shocks as :

$$\begin{cases} y_t &= \varepsilon_t + y_t^*, \\ y_t^* &= g(y_{t-1}^*, w_t; s, \rho) \\ z_t &= \varepsilon_t, \end{cases}$$

with a nonlinear function g (s :stability parameter).

The noncausal linear model has also a causal nonlinear representation.

The causal nonlinear representation has to be used for deriving the IRF :

Nonlinear Impulse Response Function

7. CONCLUDING REMARKS

Message 1 : The errors to be shocked have to be independent, not only uncorrelated.

Message 2 : The impulse responses have to be derived from stochastic shocks.

Message 3 : There is no problem of identification of the moving average representation, if the errors are not Gaussian.

Message 4 : The errors to be shocked can differ from the (linear) innovations, in case of nonfundamentalness.

Message 5 : The standard "Box-Jenkins" type estimation methods can be misleading, since they inherit the identification issues existing in the Gaussian framework.

Message 6 : In the non Gaussian case, the structural (causality, or long run) restrictions are overidentifying restrictions. They can be tested and possibly rejected.

Message 7 : When the process has noncausal components, it can be necessary to derive the nonlinear innovations and the nonlinear impulse response functions.

Some of these messages have been given a long time ago and well received in other fields than Economics.

M1-M2 : usual in operational research

M3 : known since 10 years. The static version of M3 is known since 40 years and is the basis of Independent Component Analysis.

M4 : nonfundamentalness in moving average dynamics, mentioned since 20 years.

M7 : noncausal process, known since 40 years.

Some are much more recents : The drawback of Box-Jenkins method M5,

The consistent semi-parametric estimation methods (This paper).

The possibility to test the over-identifying structural restrictions M6 (This paper).

The interpretation of noncausal AR process in terms of speculative bubbles M7. (GZ 2014).

How to understand the time necessary to transfer these messages to the practice of macroeconomists (macroeconometricians), even if some messages have been initially given by either well known economists : Hansen, Sargent, Reichlin..., or statisticians : Rosenblatt, Davis, Tong... ?