SVAR's

May 18, 2020

Structural Models

- Original meaning: Structural relative to an **intervention**.

Structural Models

- Original meaning: Structural relative to an **intervention**.
- There should be some element or elements of the model that we can imagine changing by a policy intervention, or an interpretable “shock” whose effects we want to predict (e.g. monetary policy tightens, crop yields decline because of global warming).

Structural Models

- Original meaning: Structural relative to an **intervention**.

- There should be some element or elements of the model that we can imagine changing by a policy intervention, or an interpretable “shock” whose effects we want to predict (e.g. monetary policy tightens, crop yields decline because of global warming).

- We need not only to be able to map the actual policy change or interpretable disturbance into an element of the model, we want to use the remaining, unchanged parts of the model to trace out the action or shock’s effects.

Structural Models

- Original meaning: Structural relative to an **intervention**.

- There should be some element or elements of the model that we can imagine changing by a policy intervention, or an interpretable "shock" whose effects we want to predict (e.g. monetary policy tightens, crop yields decline because of global warming).

- We need not only to be able to map the actual policy change or interpretable disturbance into an element of the model, we want to use the remaining, unchanged parts of the model to trace out the action or shock's effects.

- In a stochastic model, the intervention usually is mapped into a change in a random “disturbance term”. Since the rest of the model is supposed not to change, the disturbances we can change should be independent of other sources of randomness in the model.

Generic dynamic structural model

$$g(y_t, y_{t-1}, \varepsilon_t, \varepsilon_{t-1}) = 0\,.$$

- Elements of ε_t vector independent of each other, but not in general across time.
- We expect serial correlation of ε_t terms to be greater the finer the time unit.
- Completeness: we should be able to solve for y_t: $y_t = h(y_{t-1}, \varepsilon_t, \varepsilon_{t-1})$. Otherwise the model cannot be used to simulate a time path for y from given initial conditions on y, ε.

- With completeness, we can, from knowledge of the joint distribution of $\{\varepsilon_s, s = -\infty, \ldots, T\}$, find the joint distribution of a sample $\{y_1, \ldots, y_T\}$, assuming stationarity (so that the effects of initial y's die away.).

Invertibility

- The model is much more manageable if it can be solved for ε_t: $\varepsilon_t = f(y_{t-s}, s \geq 0)$.

Invertibility

- The model is much more manageable if it can be solved for ε_t: $\varepsilon_t = f(y_{t-s}, s \geq 0)$.

- In this case, but not otherwise, an assumed form for the distribution of $\varepsilon_t \mid \{\varepsilon_{t-s}, s \geq 1\}$ translates to a distribution for $y_t \mid \{y_{t-s}, s \geq 1\}$ by plugging in $f(y_{t-s}, y_{t-s-1})$ for ε_{t-s} in the pdf for $\varepsilon_t \mid \{\varepsilon_{t-s}, s \geq 1\}$, accounting for the Jacobian $|\partial f / \partial y_t|$ if it's non-constant.

Invertibility II

- Invertibility fails whenever ε_t is longer than y_t, which seems likely to be always, in principle.
- It is easy to construct theoretical examples where invertibility fails.
- This is not as serious a problem as it seems: We need only approximate invertibility.
- Approximate invertibility holds when the projection of the shock we are interested in (e.g. the monetary policy behavior shock) on current and past y produces a high R^2.

- We can usually get good approximate invertibility if we are sure to include in y variables that respond promptly to the structural shock we are interested in (e.g., interest rates for the monetary policy shock).

Checking approximate invertibility

A straightforward method: Usually a linearized dynamic structural model has the form

$$w_t = Gw_{t-1} + H\varepsilon_t$$
$$y_t = Fw_t$$

Also usually H is full column rank, so that if we know w_t and w_{t-1} we can recover ε_t exactly.

Checking approximate invertibility

Let $z_t = w_t - Gw_{t-1}$. Using the projection matrix $\Theta = (H'H)^{-1}H'$,

$$\Theta(w_t - Gw_{t-1}) = \Theta z_t = \varepsilon_t \text{ and}$$

$$\mathrm{Var}(\varepsilon_t \mid t) = \Theta\,\mathrm{Var}(z_t \mid t)\Theta' .$$

Starting from any initial variance matrix for w, the Kalman filter calculations deliver as a byproduct a sequence of $\mathrm{Var}(z_t \mid t)$ matrices that do not depend on the y_t sequence and that usually converge. Check whether the above expression converges to zero for those elements of the ε_t vector that matter. (Sims and Zha, *Macroeconomic Dynamics* 2006).

SVAR identification

Complete reference: Rubio-Ramírez, Waggoner, and Zha (2010). Available on Rubio-Ramirez Duke website.

SVAR:

$$A(L)y_t = \varepsilon_t \,.$$

(ignoring the possibility of a constant or exogenous variables).

Reduced form:

$$(I - B(L))y_t = u_t \,, \qquad \mathrm{Var}(u_t) = \Sigma \,,$$

where $A_0 u_t = \varepsilon_t$, therefore $A_0^{-1}(A_0^{-1})' = \Sigma$, and $A_0(I - B(L)) = A(L)$.

The RF fully characterizes the probability model. The SVAR has more parameters than the RF, so there is an id problem. (There could be an id problem even if the parameter count matched; the SVAR might restrict the probability model for the data even if it had more parameters than the RF.)

Policy interventions as “shocks” vs interventions as “rule changes”

Long run restrictions: Blanchard and Quah

"Shock j should have no permanent effect on variable i."

$$y_t = A(L)\varepsilon_t \tag{1}$$

Restriction: $a_{ij}(s) \to 0$ as $t \to \infty$.

But in a stationary model, this is not in fact a restriction!

So this can only work if there is non-stationarity. Consider a model in Δy_t:

$$\Delta y_t = A_0\varepsilon_t + (A_1 - A_0)\varepsilon_{t-1} + \cdots = \Delta A(L)\varepsilon_t$$

$$\sum_{s=0}^{T} \Delta a_{ij}(s) = a_{ij}(T)$$

$$a_{ij}(s) \to 0 \equiv \sum_{s=0}^{\infty} \Delta a_{ij}(s) = 0\,.$$

B & Q continued

Restrictions on A_0: concentrated likelihood

If the SVAR restrictions are on A_0 alone and leave A_0 invertible, they leave $B(L) = -A_0^{-1}A^+$ unrestricted. The log likelihood as a function of A_0, maximized over $B(L)$, (sometimes called the **concentrated** likelihood) can be written as

$$\frac{T}{2}\log(2\pi) + T\log|A_0| - \tfrac{1}{2}\operatorname{trace}\Big(A_0'A_0\sum_{t=1}^{T}\hat{u}_t\hat{u}_t'\Big)\ ,$$

where $\hat{u}_t = (I - \hat{B}(L))y_t$ are the least-squares residuals.

Restrictions on A_0: integrated likelihood

If we are instead interested in the likelihood integrated over B (e.g. if we are calculating marginal data density or are doing MCMC sampling from the marginal density of A_0), we use the fact that, conditional on Σ the joint distribution of the coefficients in B is $N(\hat{B}_{OLS}, \Sigma \otimes (X'X)^{-1})$, where X is the $T \times (nk+1)$ matrix of right-hand side variables that appear in each equation of the reduced form (k lags of each of n variables, and a constant). Integrating the likelihood over this joint normal distribution gives us the log posterior

$$-\frac{(T-k)n}{2}\log(2\pi) + (T-nk)\log|A_0| - \frac{n}{2}\log(|X'X|) - \tfrac{1}{2}\,\mathrm{trace}\big(A_0'A_0\sum_1^T \hat{u}_t\hat{u}_t'\big)$$

Restrictions on A_0: integrated likelihood

$$-\frac{(T-k)n}{2}\log(2\pi)+(T-nk)\log|A_0|-\frac{n}{2}\log(|X'X|)-\tfrac{1}{2}\operatorname{trace}\big(A_0'A_0\sum_1^T\hat{u}_t\hat{u}_t'\big)$$

Since $\Sigma = A_0^{-1}(A_0^{-1})'$, we can maximize this expression over the free parameters in A_0 to get an estimate based on the maximum posterior marginal density of A_0. Usually that would be the start of a procedure that used this same expression to generate MCMC draws of A_0, accompanied by direct (non-MCMC) draws from the normal distribution of $B \mid \mathcal{A}_0$ or $A^+ \mid A_0$.

Restrictions on A_0: Conclusions

Thus if the restrictions are on A_0 alone,

- Likelihood maximization is OLS, followed by nonlinear maximization on A_0 alone.
- Posterior simulation can be done in blocks, with the B block a simple draw from a multivariate normal.

Extensions by RWZ

- They show a straightforward method for checking global identification. (Hamilton had shown a local id check.)
- They show that certain kinds of nonlinear restrictions (e.g. on impulse responses) can also be handled with their approach.
- They claim that the nonlinear maximization can be done faster in identified cases by searching explicitly for the rotation of the Choleski decomposition of the RF Σ that satisfies the restrictions.

The cases for exact id 0-restrictions in a 3d system

$$\begin{bmatrix} x & x & x \\ 0 & 0 & 0 \\ x & x & x \end{bmatrix} \text{ or } \begin{bmatrix} 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix} \Rightarrow \text{incomplete}$$

$$\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \end{bmatrix} \Rightarrow \text{identified}$$

$$\begin{bmatrix} x & x & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix} \Rightarrow \text{not identified, but first equation is overid'd}$$

$$\begin{bmatrix} x & 0 & 0 \\ 0 & x & x \\ x & x & x \end{bmatrix} \Rightarrow \text{identified, but } \textit{adding} \text{ a restriction can undo id}$$

The most paradoxical case

$$\begin{bmatrix} x & x & 0 \\ x & 0 & x \\ 0 & x & x \end{bmatrix} \qquad \Rightarrow \text{ local exact id, global overid, } \textit{and} \text{ unid}$$

This case is "globally overidentified" in the sense that there are Σ matrices such that no A_0 matrix satisfying the zero restrictions generates that Σ matrix. It is locally identified in the sense that except on a measure zero set of values of the A_0 matrix coefficients, there is a unique one-one mapping between A_0 and Σ in the neighborhood of every A_0. But it is also globally unidentified, in the sense that there are pairs of A_0 matrices that are not the same, but generate the same Σ.

The most paradoxical case, numerical example

Here are two A_0's that generate the same Σ:

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} 0.000000 & 0.4082483 & 0.0000000 \\ 2.236068 & 0.0000000 & -0.8944272 \\ 0.000000 & 0.9128709 & 1.0954451 \end{bmatrix}$$

both of which have the cross product

$$\begin{bmatrix} 5 & 0 & -2 \\ 0 & 1 & 1 \\ -2 & 1 & 2 \end{bmatrix}$$

Typical contemporaneous ID for money

r, fast block y, slow block z:

$$\begin{bmatrix} x & ? & 0 \\ x & x & x \\ 0 & 0 & x \end{bmatrix}$$

Block triangular normalization

Thm: Linear transformations of the equations of a system can always make it triangular with an identity covariance matrix.

Identification through varying heteroskedasticity

We have two or more Σ_j's from different time periods or different groups, generated by variation in the variances of the structural shocks, not in the form of A_0. A normalization is needed, for example that the diagonal of A_0 is all ones, or that the variances of structural shocks for the $j = 1$ period or group are all one.

$$\Sigma_1 = A_0^{-1}\Lambda_1(A_0^{-1})' \qquad \Sigma_2 = A_0^{-1}\Lambda_2(A_0^{-1})' \tag{2}$$

$$\therefore \Sigma_1^{-1}\Sigma_2 = A_0'\Lambda_1^{-1}\Lambda_2(A_0^{-1})' \tag{3}$$

Identification through varying heteroskedasticity

$$\Sigma_1 = A_0^{-1}\Lambda_1(A_0^{-1})' \qquad \Sigma_2 = A_0^{-1}\Lambda_2(A_0^{-1})' \tag{4}$$

$$\therefore \Sigma_1^{-1}\Sigma_2 = A_0'\Lambda_1^{-1}\Lambda_2(A_0^{-1})' \tag{5}$$

This last expression is in the form of the usual eigenvector decomposition of a matrix. It asserts that the columns of A_0' are the right eigenvectors of $\Sigma_1^{-1}\Sigma_2$. So if we can do an eigenvector decomposition, and the roots we find are all distinct (meaning every variance has changed) we can calculate A_0.

Local projections

If we have a (non-structural) first-order VAR

$$y_t = Ay_{t-1} + \varepsilon_t \,,$$

we know that this implies

$$y_t = \sum_{s=0}^{t-1} A^s \varepsilon_{t-s} + A^t y_0 \,.$$

(This discussion generalizes to higher-order VAR's stacked to become first-order.)

Local projections

A^s is the matrix of sth order terms in the impulse responses of the system. (These are responses to the innovations, not orthogonalized.) But as is easily seen, A^s is also the matrix with typical element

$$\frac{\partial E_t[y_{i,t+s}]}{\partial y_{j,t}} .$$

Since in a VAR $E_t[y_{t+s}]$ is by construction a linear function of y_t, We can estimate A^s as the coeffcient matrix B_s in the regression equation system

$$y_{t+s} = B_s y_t + \xi_t .$$

Advantages of using $\hat{B}_s$

Estimating A^s as $\hat{B}_s$ is an alternative to estimating A in the usual one-step-ahead system $y_t = Ay_{t-1} + \varepsilon_t$ and estimating A^s as $(\hat{A})^s$.

If the data are actually from a non-linear process, or are from a VAR with more lags than have been specified in the VAR model, $\hat{B}_s$ nonetheless estimates the best (in RMSE sense) linear predictor of y_{t+s} based on y_t. If the model is mis-specified in these ways, $(\hat{A})^s$ is not a best linear predictor, and indeed can be quite a bad predictor. (E.g, $s = 2$, data are from $y_t = .5y_{t-2} + \varepsilon_t$, we mis-specify the model as first order and estimate $y_t = \rho y_{t-1} + \varepsilon_t$.)

If one needs only impulse responses of one variable to one shock, k steps ahead, local projection requires estimating only k linear regressions, instead of the n (number of variables) regressions required to estimate a VAR, and no matrix powers A^s need be calculated.

Disadvantages of $\hat{B}_S$

- If the VAR is correctly specified, local projection is inefficient, possibly extremely inefficient.
- To get asymptotically correct error bands requires estimating the autocorrelation pattern of the residuals, which is as computationally demanding as estimating a VAR.
- If local projection differs strongly from what emerges from a Bayesian posterior, the VAR is probably mis-specified. But in that case there is no “impulse response function”.

Local projection does not estimate a model of the data

- Nothing like the simulated draws from the posterior of the ir's is available, so answering questions like "what is the probability of a hump-shaped impulse response" is impossible.

- Of course the posterior expectation of y_{t+s} given the whole sample up to time t is not $(\hat{A})^s y_t$, Instead it is the mean of the forecast paths obtained by making posterior draws from the joint distribution of A and σ_ε^2.

Orthogonalized irf's from local projection

- What we've described so far delivers raw irf's to innovations, which are the same as what we would get by triangularizing the system and looking at responses to the shock last in the ordering — so when it changes, none of the other shocks change.

- If instead, in the local projection regression we include $y_{i,t}$, but not $y_{j,t}$ for $j \neq i$, plus the full y_{t-s} vector for $s = 1, \ldots, m$ (where m is the number of lags in the VAR), the $\hat{B}_s$ coefficients trace out irf's for the case where variable i is first in the ordering, so all other shocks respond to it contemporaneously.

SVAR ID through "external instruments"

- Suppose we are interested in an SVAR, and in the response to a particular structural shock ε_i.
- Suppose further that we somehow acquire data on ε_i itself. Then of course we could estimate the impulse responses directly by a regressions of the y_t vector on many lagged values of ε_{it}. (One could also estimate a bunch of "local projection" regressions.)
- Getting data on ε_{it} itself seems unlikely. But finding data on an error-ridden proxy ε^*_{it} for ε_{it} seems possible.
- What about just regressing y_t on many lagged values of ε^*_{it}?

Restrictions on the error in the external instrument

- This is an "errors in variables" problem. If the errors are uncorrelated with the residual (i.e. with all the other structural shocks), there is only a downward bias in the irf estimates, by a uniform factor across all lags.
- But the requirement that ε_i^* be uncorrelated with all other structural shocks is very strong.

External instruments from restrictions on A_0

Suppose our external instrument is z_t and we have an SVAR model for y_t the we write as

$$y_t = C(L)\varepsilon_t \,,$$

where y_t and ε_t are $n \times 1$. We suppose z to be an instrument for ε_{1t}, the first structural shock, which means that

- Adding z_t to the VAR system would not improve forecasts, i.e. y is Granger Causally Prior (GCP) to z;
- The innovation in z_t (its forecast error when forecast from all past values of y and z), which we will call $\tilde{z}_t$, is correlated with ε_{1t}, but not with ε_{it} for $i \neq 1$.

Constructing an expanded SVAR

Let $\tilde{z}^* = \tilde{z}_t - \rho\varepsilon_{1t}$ be the residual in a regression of $\tilde{z}_t$ on ε_{1t}. It is by construction serially uncorrelated and uncorrelated with current and past y and ε.

Adding z_t at the top of the variable list and $\tilde{z}_t^*$, scaled to have unit variance, to the top of the structural shock list to form a new SVAR:

$$\begin{bmatrix} z_t \\ y_t \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ 0 & C_{0,\cdot,1} & C_{0,\cdot,2} \end{bmatrix} \begin{bmatrix} \tilde{z}_t^* \\ \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} f(L) & F(L) \\ 0 & C^+(L) \end{bmatrix} \begin{bmatrix} \tilde{z}_t^* \\ \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} ,$$

where ε_{2t} now represents a vector of all the elements of ε_t except the first, $C_{0,\cdot,1}$ is the first column of C_0, $C_{0,\cdot,2}$ is the remaining columns of C_0, and $C^+(L)$ is $C(L) - C_0$, i.e. the part of $C(L)$ involving positive powers of L.

The zero restrictions

Obviously this expanded SVAR has known zero restrictions on its coefficients. Some are on the matrix mapping structural shocks to observables, some are on the lag coefficients. This SVAR can be estimated subject to these restrictions, which is a fully efficient way to exploit the restrictions. The restrictions imply (via quite a bit of matrix algebra we won't go through here) that in the *reduced form* VAR for this variable list, the responses of y to $\tilde{z}^*$ are proportional to the responses of y to ε_1.

Two recent references that explore the connection among identification methods for SVAR's are Plagborg-Møller and Wolf (2019); Plagborg-Møller and Wolf (2019).

References on monetary SVAR's

Baumeister and Hamilton (2013)

Uhlig (2001)

Sims (1992)

Sims (1980)

Sims (1986)

*

References

BAUMEISTER, C., AND J. D. HAMILTON (2013): "Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information," Discussion paper, University of Notre Dame and University of California at San Diego, http://econweb.ucsd.edu/~jhamilto/bh1.pdf.

PLAGBORG-MØLLER, M., AND C. K. WOLF (2019): "Local Projections and VARs Estimate the Same Impulse Responses," Discussion paper, Princeton University, https://scholar.princeton.edu/sites/default/files/mikkelpm/files/lp_var.pdf.

Plagborg-Møller, M., and C. K. Wolf (2019): "Instrumental Variable Identification of Dynamic Variance Decompositions Princeton University This version: June 9, 2019," Discussion paper, Princeton University, https://scholar.princeton.edu/sites/default/files/mikkelpm/files/decomp_iv.pdf.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: The Theory for Identification and Algorithms for Inference," *Review of Economic Studies*, 77, 665–696.

Sims, C. A. (1980): "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review*, 70, 250–57.

——— (1986): "Are Forecasting Models Usable for Policy Analysis?," *Quarterly Review of the Minneapolis Federal Reserve Bank*, 10, 2–16.

——— (1992): “Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy,” *European Economic Review*, 36, 975–1011.

UHLIG, H. (2001): “What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure,” Discussion paper, Humboldt University, Berlin.