SVAR's

October 16, 2014

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

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

- 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: $y_t = h(y_{t-1}, \varepsilon_t)$. Otherwise the model cannot be used to simulate a time path for y from given initial conditions.

- 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: $\varepsilon_t = f(y_t, y_{t-1})$.

Invertibility

- The model is much more manageable if it can be solved for ε_t: $\varepsilon_t = f(y_t, y_{t-1})$.
- 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 get 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 —

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

Starting from any initial variance matrix for w, the Kalman filter delivers a sequence of $\mathrm{Var}(w_t \mid t)$ 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-Ramirez, 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.)

Long run restrictions: Blanchard and Quah

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

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

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

$$(2\pi)^{(n(T-nk-1)/2} \left|\Sigma\right|^{(nk+1)/2} \left|X'X\right|^{-n/2} \exp\left(-\tfrac{1}{2}\,\mathrm{trace}\big(\Sigma^{-1}\sum_{1}^{T}\hat{u}_t\hat{u}_t'\big)\right)$$

As a function of Σ, this is proportional to a Wishart pdf, the multivariate generalization of the chi-squared distribution. There are packaged functions to generate draws from it.

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

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

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

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

$$\begin{bmatrix} x & 0 & 0 \\ 0 & x & x \\ x & x & x \end{bmatrix} \Rightarrow$$ identified, but *adding* 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.

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

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

Identification through varying heteroskedasticity

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

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

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.

*

References

RUBIO-RAMIREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): “Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference,” *Review of Economic Studies*, 77(2), 665–696.