

Handout 9

1 Solving for the identification matrix directly

While the Cholesky decomposition is a convenient way to impose a recursive identification scheme, we might want to impose a more complex set of restrictions that do not conform with a recursive scheme. In those cases we are left with inspecting the non-linear system of equations that links the structural- and the reduced-form representation of a VAR.

To fix ideas, let's stick with a very simple VAR(1) process for two variables and let's use, again, a recursive scheme, but this time, let's not take advantage of the Cholesky decomposition. Let the reduced-form representation for y_t take the form:

$$y_t = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} y_{t-1} + \epsilon_t$$

and let ϵ_t be white noise with variance Ω . Also, let the structural-form representation of y_t take the form:

$$\begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{21} & \tilde{a}_{22} \end{pmatrix} y_t = \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{21} & \tilde{b}_{22} \end{pmatrix} y_{t-1} + \tilde{\epsilon}_t,$$

where $\tilde{\epsilon}_t$ is white noise with diagonal variance Σ .

Let the short run restriction be that $a_{12} = 0$. From usual normalizations, we have that $a_{11} = 1$ and $a_{22} = 1$. Thus identification entails finding values of the parameters a_{21} and for Σ_1 and Σ_2 , the diagonal entries of Σ .

From the relationship between the structural form and the reduced form, we have that:

$$\tilde{a}^{-1} \Sigma \tilde{a}^{-1'} = \Omega.$$

Notice that our restrictions imply:

$$\tilde{a} = \begin{pmatrix} 1 & 0 \\ a_{21} & 1 \end{pmatrix}.$$

But then the determinant of a equals 1 and we can write:

$$\begin{pmatrix} 1 & 0 \\ -a_{21} & 1 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & -a_{21} \\ 0 & 1 \end{pmatrix} = \Omega.$$

Multiplying through yields:

$$\begin{pmatrix} \Sigma_1 & 0 \\ -a_{21}\Sigma_1 & \Sigma_2 \end{pmatrix} \begin{pmatrix} 1 & -a_{21} \\ 0 & 1 \end{pmatrix} = \Omega.$$

And finally:

$$\begin{pmatrix} \Sigma_1 & -a_{21}\Sigma_1 \\ -a_{21}\Sigma_1 & \Sigma_2 \end{pmatrix} = \Omega.$$

which yields three independent equations in the unknowns a_{21} , Σ_1 , Σ_2 , and the reduced-form coefficients in Ω .

2 Long-run restrictions

In this section we shall consider how to adapt the recursive scheme to achieve identification based on long-run restrictions.

To fix ideas, let's go back to a univariate process. Let's start from a simple AR(1) process:

$$y_t = \rho y_t + \epsilon_t$$

where ϵ_t is white noise. What is the long-run cumulative effect of ϵ_t ? That would be given by:

$$\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \epsilon_t} \epsilon_t = \sum_{k=0}^j \rho^k \epsilon_t = \frac{1}{1 - \rho} \epsilon_t.$$

Notice that the formula for the long-run effect of an innovation takes a very similar form in the case of an AR(n) process:

$$\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \epsilon_t} \epsilon_t = \frac{1}{1 - \sum_i^n \rho_i} \epsilon_t,$$

and can be extended to a VAR(n) process as:

$$\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \epsilon_t} = (I - \sum_i^n \rho_i)^{-1} \epsilon_t.$$

From this formula, we see that the variance for the long-run cumulative effect of an innovation for a VAR(n) process is:

$$\text{Var}(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \epsilon_t} \epsilon_t) = (I - \sum_i^n \rho_i)^{-1} \Omega (I - \sum_i^p \rho_i)^{-1'}.$$

To fix notation, consider again the reduced-form VAR(n) representation of a process y_t ,

$$y_t = c + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_n y_{t-n} + \epsilon_t, \quad (1)$$

where ϵ_t has variance Ω . Again, let the structural representation of the process y_t take the form

$$\tilde{\rho}_0 y_t = \tilde{c} + \tilde{\rho}_1 y_{t-1} + \tilde{\rho}_2 y_{t-2} + \dots + \tilde{\rho}_n y_{t-n} + \tilde{\epsilon}_t, \quad (2)$$

where $\tilde{\epsilon}_t$ is IID with variance Σ , which is related to Ω above by $\Omega = (\tilde{\phi}_0^{-1}) \Sigma (\tilde{\phi}_0^{-1})'$. Furthermore $\phi_j = \frac{\tilde{\phi}_j}{\tilde{\phi}_0}$ for $j \in \{1, \dots, n\}$ and $c = \frac{\tilde{c}}{\tilde{\phi}_0}$. We could identify ϕ_0 using restrictions on the long-run effects of shocks. Akin to the recursive short-run identification scheme, let only the first shock have a nonzero long-run cumulative effect on the first variable; let only the first and second shock have a nonzero long-run cumulative effect on the second variable; and so on.

Now, consider what the structural representation implies for the long-run cumulative effect of a shock and its variance:

$$\begin{aligned} \lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t &= (\tilde{\rho}_0 - \sum_i^n \tilde{\rho}_i)^{-1} \tilde{\epsilon}_t \\ \text{Var}(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t) &= (\tilde{\rho}_0 - \sum_i^n \tilde{\rho}_i)^{-1} \Sigma (\tilde{\rho}_0 - \sum_i^p \tilde{\rho}_i)^{-1'}. \end{aligned}$$

Collecting the term $\tilde{\rho}_0$, one obtains:

$$\text{Var}(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t) = (\tilde{\rho}_0 (I - \sum_i^n \tilde{\rho}_0^{-1} \tilde{\rho}_i))^{-1} \Sigma (\tilde{\rho}_0 (I - \sum_i^n \tilde{\rho}_0^{-1} \tilde{\rho}_i))^{-1'}.$$

But notice that by simplifying further:

$$Var\left(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t\right) = \left(I - \sum_i^n \tilde{\rho}_0^{-1} \tilde{\rho}_i\right)^{-1} \tilde{\rho}_0^{-1} \Sigma \tilde{\rho}_0^{-1'} \left(I - \sum_i^n \tilde{\rho}_0^{-1} \tilde{\rho}_i\right)^{-1'}$$

But this suggests that we can estimate $Var\left(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t\right)$ as

$$Var\left(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t\right) = \left(I - \sum_i^n \hat{\rho}_i\right)^{-1} \hat{\Omega} \left(I - \sum_i^n \hat{\rho}_i\right)^{-1'}$$

Notice that the long-run restriction scheme described above, together with the normalization that Σ is diagonal, implies that $(\tilde{\rho}_0 - \sum_i^n \tilde{\rho}_i)^{-1}$ should be triangular and equal to R , the Cholesky factor of $Var\left(\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t}\right)$. As a result, we can estimate $\tilde{\rho}_0^{-1}$ by considering the following set of restrictions:

$$\left(\hat{\rho}_0 - \sum_i^n \hat{\rho}_i\right)^{-1} = R$$

which leads to

$$\left(\hat{\rho}_0 - \hat{\rho}_0 \sum_i^n \hat{\rho}_i\right)^{-1} = R.$$

Collecting terms:

$$\left(I - \sum_i^n \hat{\rho}_i\right)^{-1} \hat{\rho}_0^{-1} = R,$$

and premultiplying by $\left(I - \sum_i^n \hat{\rho}_i\right)^{-1}$

$$\hat{\rho}_0^{-1} = \left(I - \sum_i^n \hat{\rho}_i\right) R.$$

2.1 Using Instrumental Variables to Implement Long-Run Restrictions

See Shapiro and Watson “Sources of Business Cycle Fluctuations”, NBER Macroeconomic Annual, 1988.

To fix ideas consider a VAR(1) process with only two variables.

$$\begin{pmatrix} 1 & \tilde{a}_{12} \\ \tilde{a}_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{21} & \tilde{b}_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \tilde{\epsilon}_{1t} \\ \tilde{\epsilon}_{2t} \end{pmatrix}, \quad (3)$$

where $\tilde{\epsilon}_t$ is white noise with diagonal variance Σ .

Notice that OLS equation by equation would produce inconsistent estimates because of endogeneity. We know that y_{2t} in the first equation is correlated with y_{1t} , as can be seen from the second equation, but then it must also be correlated with ϵ_{1t} . A similar reasoning applies to the term y_{1t} in the second equation.

Are there convenient instruments that might take care of the endogeneity problem? Fortunately, yes.

Consider again the long-run cumulative effect of an innovation ϵ_t .

$$\lim_{j \rightarrow \infty} \sum_{k=0}^j \frac{\partial y_{t+k}}{\partial \tilde{\epsilon}_t} \tilde{\epsilon}_t = (a - b)^{-1} \tilde{\epsilon}_t.$$

The long-run restriction scheme described in the previous section implies that $(a - b)^{-1}$ is equal to a lower-triangular matrix c . We then have

$$c = (a - b)^{-1},$$

but this also implies that:

$$c(a - b) = I.$$

For our example we have that:

$$\begin{pmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} 1 - \tilde{b}_{11} & \tilde{a}_{12} - \tilde{b}_{12} \\ \tilde{a}_{21} - \tilde{b}_{21} & 1 - \tilde{b}_{22} \end{pmatrix} = I$$

Notice that the 1, 2 entry of the system implies that:

$$c_{11}(\tilde{a}_{12} - \tilde{b}_{12}) = 0$$

As long as c_{11} is not 0, then it must be that $\tilde{a}_{12} = \tilde{b}_{12}$. Notice that exploiting this restriction we can rewrite the first equation in (3) as:

$$y_{1t} = \tilde{b}_1 y_{1t-1} - \tilde{b}_{12}(y_{2t} - y_{2t-1}) + \epsilon_{1t}.$$

We can then estimate the equation above using y_{t-1} and y_{t-2} as instruments.

Using the IV estimates of the equation above, we can construct the residual $\tilde{\epsilon}_{1t}$. This can then be used as an instrument in the estimation of the second equation of (3).

3 Sign Restrictions

See Jon Faust “On the Robustness of Identified VAR Conclusions about Money”, Carnegie Rochester Conference Series on Public Policy, 1998.

Going back the identification problem in equations (1) and (2), in the case of the recursive ordering restrictions, we used a Cholesky decomposition for the variance matrix of the unidentified residuals, Ω , so that $\Omega = RR'$. Normalizing the matrix Σ to have unit variances, we could then take R as the inverse of $\tilde{\phi}_0$. However, for some applications it might be hard to justify any particular ordering scheme. In those cases we might still achieve identification using restrictions on the identified impulse responses.

Think about augmenting the Cholesky decomposition of the matrix Ω with a matrix Q , such that $QQ' = I$. In that case $\Omega = RQQ'R'$. The new identifying matrix RQ (our candidate for the inverse of $\tilde{\phi}_0$) would then give rise to a new set of identified shocks that share the property that their variance covariance matrix is the same as those of the unidentified shocks. If we could systematically consider all the Q matrices that have the property that $QQ' = I$, then we could discriminate among possible candidate identifying matrices based on the properties of the identified impulse response functions.

To fix ideas take a simple bivariate VAR. A candidate for Q is then:

$$Q = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \quad (4)$$

This particular choice for the matrix Q is called a Givens rotation. Varying θ between 0 and π we could then systematically consider all the matrices such that $QQ' = I$. To obviate the problem that θ can take a continuum of values, we could make a fine grid.

For each value on the grid, we could calculate a new set of impulse responses and store them away, only if the responses satisfy particular properties chosen in accordance with economic theory. As an example, consider the case of a VAR including petroleum production levels and petroleum prices. We could require that supply shocks drive the price up and the quantity down (or vice versa), and that demand shocks drive up (or down) both the quantity and the price.

In practice, extending this scheme to VARs that encompass more variables runs into the curse of dimensionality. One of the fixes that is used in practice involves abandoning the idea that we can systematically consider all Q matrices such that $QQ' = I$. Instead, we could draw candidate Q matrices at random. One way to generate random draws of Q is this:

1. Draw each entry of a square matrix A , conformable with Ω from the Standard Normal distribution.
2. Take the QR decomposition of the matrix A , which produces matrices Q and R such that $A = QR$, where Q is unitary and R is triangular.

The identification scheme involves generating enough random draws of Q to ensure that we find at least n matrices that satisfy the desired set of properties for the impulse responses. The programs accompanying this handout implement this kind of identification scheme for a 3-variable VAR.