

Notes on Identification in Vector Autoregressions

1. Reduced Form Vector Autoregression

We want to estimate a simple VAR which can be written as:

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t$$

or

$$Y_t = [\nu \ A_1 \ A_2 \ \dots] X_{t-1} + U_t$$

where Y_t and U_t of the dimension $N \times T$ and X_{t-1} is $[1 \ y_{t-1} \ y_{t-2} \ \dots]'$ and of the dimension $(Np + 1) \times T$. In R, you additionally need to take care of the rows that are not included in the estimation, so the actual dataframe are smaller for rows where NA is present. This happens when you create lagged variables (for $p = 1$ (2) first (second) row is taken away etc.). When you have the two dataframes - Y and X , you estimate the coefficient with the least squares (henceforth: LS) estimator. The LS minimizes the squared error of the model. In stylized terms, this means:

$$\begin{aligned} U_t U_t' &= (Y - AX)(Y - AX)' \\ &= YY' - YX'A' + AXY' + AXX'A' \\ \frac{\partial U_t U_t'}{\partial A} &= -2XY' + 2AXX' = 0 \\ &= YX' + AXX' = 0 \\ YX' &= AXX' \\ \hat{A} &= YX'(XX')^{-1} = [\hat{\nu} \ \hat{A}_1 \ \hat{A}_2 \ \dots] \end{aligned}$$

where \hat{A} is of the dimension $N \times (Np + 1)$. This gives us our reduced form coefficient estimates. While our residual matrix is naturally:

$$\hat{U}_t = Y_t - \hat{A}X_{t-1}$$

The residual variance-covariance matrix $\hat{\Sigma}_u$ is calculated as:

$$\hat{\Sigma}_u = \frac{\hat{U}_t \hat{U}_t'}{T - Np - 1}$$

and is of the dimension $N \times N$, while the coefficient variance covariance is calculated as $\hat{\Sigma}_u \times (X'X)^{-1}$, a Kronecker product:

$$\hat{\Sigma}_A = \hat{\Sigma}_u \times (XX')^{-1}$$

which is of the $N(Np + 1) \times N(Np + 1)$ dimension. This follows from the derivation for the variance covariance estimator $Var[b|X] = \sigma^2(XX')^{-1}$. The standard errors of the

variances of the coefficients are calculated by taking the diagonal terms of $\hat{\Sigma}_A$ (which are the variances) and taking the square root. **Note that all of these values can also be estimated equation by equation of the system.** The problem is only that you don't get the whole U_t matrix that is needed for further analysis. To get the impulse responses, we rely on the **moving average representation (MA)**, to which we can arrive using lag operators L :

$$\begin{aligned}
y_t - (A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p}) &= \nu + u_t \\
(I_K - A_1 L - A_2 L^2 - \dots - A_p L^p) y_t &= \nu + u_t \\
A(L) y_t &= \nu + u_t \\
y_t &= A(L)^{-1} (\nu + u_t) \\
y_t &= \Phi(L) (\nu + u_t) \\
y_t &= \left(\sum_{i=0}^{\infty} \Phi_i \right) \nu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}
\end{aligned}$$

where $\Phi(L)$ is an operator that reflects the assumption that $A(L)^{-1}$ is equal to an infinite sum, where the common ratio is $A_1 L + A_2 L^2 + \dots + A_p L^p$ and we know that the first term in the sum is $\Phi_0 = I_K$. To find the other terms of this infinite sum, we use the fact that $A(L)\Phi(L) = I_K$, or written differently:

$$(I_K - A_1 L - A_2 L^2 - \dots - A_p L^p)(\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \dots) = I_K$$

We can see that since Φ_0 stands as the sole term, it is equal to I_K , while the other terms need to be zero to match the right-hand side. After some multiplication, we get:

$$\begin{aligned}
\Phi_0 + \Phi_1 L - \Phi_0 A_1 L + \Phi_2 L^2 - \Phi_1 A_1 L^2 - A_2 L^2 \Phi_0 + \dots &= I_K \\
\Phi_0 + (\Phi_1 - \Phi_0 A_1) L + (\Phi_2 - \Phi_1 A_1 - A_2 \Phi_0) L^2 + \dots &= I_K
\end{aligned}$$

where the conditions are:

$$\begin{aligned}
\Phi_0 &= I_K \\
\Phi_1 &= \Phi_0 A_1 \\
\Phi_2 &= \Phi_1 A_1 + \Phi_0 A_2 \\
&\vdots \\
\Phi_i &= \sum_{j=1}^i \Phi_{i-j} A_j
\end{aligned}$$

Looking at the MA representation, we can see that the elements of $\Phi(L)$ are already the impulse responses to a unit shock. Using the representation without the constant ν :

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \dots$$

where the response of variable i in y_t after s periods when variable j is hit by a unit shock in t :

$$\frac{\partial y_{i,t+s}}{\partial u_{j,t}} = \Phi_s(i, j) = \phi_{ij,s}$$

Note that these responses are transitory if $y_t \sim I(0)$. To get to the structural impulse responses, we can show that reduced form shocks u_t can be mapped on to more primitive **structural shocks** ε_t . To show this, the reduced form VAR:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t$$

can be rewritten as:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \mathcal{A}^{-1} \mathcal{B} \varepsilon_t$$

or:

$$\mathcal{A} y_t = A_1^* y_{t-1} + A_2^* y_{t-2} + \dots + A_p^* y_{t-p} + \mathcal{B} \varepsilon_t$$

which is the **structural VAR** representation, where we used the fact that there might also be contemporaneous relationships between the variables governed by \mathcal{A} matrix. Since we normally assume orthogonality of shocks, \mathcal{B} is an identity matrix by assumption. We have a mapping between the two types of shocks:

$$u_t = \mathcal{A}^{-1} \mathcal{B} \varepsilon_t$$

where u_t and ε_t are $T \times N$ and the structural matrices are $N \times N$. Thus, structural shocks can be computed as:

$$\varepsilon_t = \mathcal{B}^{-1} \mathcal{A} u_t$$

2. Cholesky Decomposition/Triangularization

The whole point of Cholesky decomposition is that the reduced form VAR estimates arise only from a statistical exercise - there are not baked in assumption with regard to the relationship between the variables in the system. Triangularization is a simple way to quantitatively represent a causal chain the system itself. While some variables affect others contemporaneously, the others only impact it with a lag. A classic example is the trivariate VAR, where the interest rate is ordered last - it affect everything with a lag, but other variables do impact it through a policy function. The higher (more left) in the dataframe we go, the more rigid are the variables. This assumption can be easily implemented using Cholesky decomposition.

To do this, we take the the residual variance covariance matrix $\hat{\Sigma}_u$ and do a Cholesky decomposition to get:

$$\hat{\Sigma}_u = PP'$$

where P is lower triangular. Since we also have the mapping between the reduced form and structural shocks, we know that:

$$Var(u_t) = \mathbb{E}(u_t u_t') = \mathcal{A}^{-1} \mathcal{B} \mathbb{E}(\varepsilon_t \varepsilon_t') \mathcal{B}' \mathcal{A}'^{-1}$$

Assuming that a reduced form shock is only driven by one more primitive shock, $\mathcal{B} = I_K$, and we can thus rewrite the expression as:

$$Var(u_t) = \mathbb{E}(u_t u_t') = \mathcal{A}^{-1} \mathbb{E}(\varepsilon_t \varepsilon_t') \mathcal{A}'^{-1} \quad (1)$$

where \mathcal{A}^{-1} is actually equal to the Cholesky P matrix. Note that we could also write the restrictions as \mathcal{B} being a diagonal matrix, with non-zero diagonal terms, and \mathcal{A} being a lower triangular matrix with non-zero diagonal elements. This would yield the same mapping, but the separate matrices would be a bit different. We can do this because the system is still identified. Note that:

- reduced form has $k + pk^2 + \frac{k(k+1)}{2}$ parameters (constants, coefficients, unique values of Σ_u)
- structural form has $k + (p+2)k^2$ parameters

which gives us the difference of $k + (p+2)k^2 - k - pk^2 - \frac{k(k+1)}{2} = k^2 + \frac{k(k-1)}{2}$, giving us the number of the additional restriction necessary to identify all the structural coefficients. In the accompanying Cholesky example, our $k = 5$, meaning that we need 35 additional restrictions. We get 20 with imposing zero restrictions on off-diagonal elements of \mathcal{B} and 15 restrictions by imposing zero restrictions on the lower triangular elements of \mathcal{A} and 1s on its diagonal. An equivalent representation is by making $\mathcal{B} = I_K$, and only have zero restrictions in \mathcal{A} as already mentioned. By performing the Cholesky decomposition, we in effect get $\mathcal{A}^{-1} \mathcal{B}$, because the matrix is exactly identified if it is lower triangular. Thus, we can get ε_t :

$$\varepsilon_t = P^{-1} u_t$$

Note that \mathcal{B} is actually a diagonal matrix where the diagonal terms are equal to the diagonal terms of P (if \mathcal{A} is assumed to have 1 on the diagonal).

$$\hat{\Sigma}_u = PP' = P\mathcal{B}^{-1}\mathcal{B}\mathcal{B}'\mathcal{B}'^{-1}P' \quad (2)$$

In essence, (1) and (2) are the same - one is theoretical and one empirical, where $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathcal{B}\mathcal{B}'$. What's more, they allow us to distinguish between unit and one standard deviation impulse responses. Writing the MA representation, we can see that:

$$y_t = \Phi(L)u_t = \Phi(L)P\mathcal{B}^{-1}\varepsilon_t = \Phi(L)\mathcal{A}^{-1}\varepsilon_t$$

$$\frac{\partial y_{t+h-1}}{\partial \varepsilon_t} = \Phi(L)\mathcal{A}^{-1}$$

which corresponds to the the impulse responses to a **unit structural shock**, which has $\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathcal{B}\mathcal{B}'$. Whereas, the following:

$$y_t = \Phi(L)u_t = \Phi(L)P\mathcal{B}^{-1}\varepsilon_t = \Phi(L)Pz_t$$

$$\frac{\partial y_{t+h-1}}{\partial z_t} = \Phi_{h-1}P$$

corresponds to the impulse responses to a **one standard deviation structural shocks**. This is because \mathcal{B}^{-1} is multiplied with ε_t to yield the standardized structural shock z_t , which has $\mathbb{E}(z_t z_t') = I_K$. We know that \mathcal{B} is the standard deviation matrix of ε_t because the product needs to overall lead back to the Cholesky representation PP' . This is only possible if $\text{Var}(\varepsilon_t) = \mathcal{B}\mathcal{B}'$. In his book, Lütkepohl implies that P can be decomposed into W and D matrices, where W is equal to P , but with 1s on the diagonal. Furthermore, D is a diagonal matrix with the elements of P . This is exactly the same decomposition that is done by directly assuming an **AB model**. One could also assume an **A model** (with $\mathcal{B} = I_K$) and follow the same steps.

2. Long Run Restrictions (alla Blanchard Quah)

The motivation is that economic theory sometimes gives us little guidance about what will happen in the short run. But we know from the standpoint of nominal rigidities for example, that not output, but only prices should be impacted by aggregate demand shocks in the long run (the so-called **neutrality of money**). Empirically, this assumption can be incorporated in the structural VAR framework by applying restrictions to the long run impulse response matrix. Going back to our Wold representation:

$$y_t = A(L)^{-1}u_t = \Phi(L)u_t = \Phi(L)\mathcal{A}^{-1}\mathcal{B}\varepsilon_t$$

as already shown before. We are interested in calculating $\Phi(1)\mathcal{A}^{-1}\mathcal{B}$, where $\Phi(1)$ is the infinite sum of all MA coefficients. We know that we can get this easily by calculating $A(1)^{-1}$. But in this case, we do not use Cholesky decomposition on $\hat{\Sigma}_u$. This would correspond to short-run zero restrictions. We wish to use it on the long-run response matrix. This is why we do not know what is \mathcal{C} :

$$\mathcal{C} = A(1)^{-1}\mathcal{A}^{-1}\mathcal{B}$$

However, note that we can calculate:

$$\mathcal{C}\mathcal{C}' = A(1)^{-1}\mathcal{A}^{-1}\mathcal{B}\mathcal{B}'(\mathcal{A}^{-1})'(A(1)^{-1})'$$

where the inner product is known:

$$\mathcal{C}\mathcal{C}' = A(1)^{-1}\Sigma_u(A(1)^{-1})'$$

Thus we can calculate \mathcal{C} , because every terms is observable and we can also get the short-run structural matrix $\mathcal{A}^{-1}\mathcal{B}$. We calculate \mathcal{C} by using the Cholesky matrix, which assumes zero restrictions. For example:

$$\begin{bmatrix} y_{t,t+\infty} \\ u_{t,t+\infty} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{\text{Supply}} \\ \varepsilon_t^{\text{Demand}} \end{bmatrix}$$

meaning that the supply shock has a long run effect on output (c_{11}) as well as unemployment (c_{21}), whereas the non-technological shock only has a long run effect on unemployment (c_{22}). We can then get the short run structural matrix as:

$$A(1)^{-1}\mathcal{C} = \mathcal{A}^{-1}\mathcal{B}$$

where \mathcal{B} is still assumed to be diagonal or identity as before. Thus, we can calculate the whole sequence of responses.

where the **2. Proxy-VAR**