

Handout 8

1 VARs

Let y_t represent an $(n \times 1)$ vector containing the observations for n distinct variables at time t . Consider the process:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_n y_{t-n} + \epsilon_t,$$

where ϵ_t is identically and independently distributed with mean $\bar{0}$ and variance Ω , $\phi_1, \phi_2, \dots, \phi_n$ are $n \times n$ matrices of coefficients and c is an $n \times 1$ vector of coefficients.

Why is it interesting to consider VARs? Could we not simply stick with univariate processes? Notice that if the true data-generating process for the vector y_t is indeed a VAR, a univariate representation for any of the individual variables, if it exists, would take an ARMA form of possibly infinite order. The direct estimation of the VAR form might then involve a more parsimonious specification even if we are really interested in only one of the entries of the vector y_t .

1.1 Basic properties

Stationarity. Following the same reasoning as for univariate processes, rewrite the VAR process above in companion form:

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ y_{t-n} \end{pmatrix} = \begin{pmatrix} c \\ \bar{0} \\ \cdot \\ \cdot \\ \cdot \\ \bar{0} \end{pmatrix} + \Phi \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \cdot \\ \cdot \\ \cdot \\ y_{t-n-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \bar{0} \\ \cdot \\ \cdot \\ \cdot \\ \bar{0} \end{pmatrix}$$

where

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_n \\ I & \bar{0} & \dots & \bar{0} \\ \bar{0} & I & \dots & \bar{0} \\ \bar{0} & \bar{0} & \dots & I \end{pmatrix}$$

If all of the eigenvalues of Φ are within the unit circle, the VAR process will be covariance stationary.

Unconditional Mean

If the process is covariance stationary, than we have:

$$E[y_t] = c + \phi_1 E[y_t] + \phi_2 E[y_t] + \dots + \phi_n E[y_t].$$

Denoting the unconditional mean by μ

$$\mu = (I - \phi_1 - \phi_2 - \dots - \phi_n)^{-1} c$$

1.2 MA representation

If the VAR process is covariance stationary, then it also has a an $MA(\infty)$ representation. To fix notation, let this representation take the form:

$$y_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots$$

where μ is the mean of the process and the remaining matrices are square and conformable with the innovation vectors. These matrices of coefficients can be found easily by proceeding numerically as follows:

1. Momentarily discard the c vector and pretend the mean of the process is zero.
2. For each of n innovations of the process at time 0 consider forming the (unidentified) responses of y starting from the pretend mean to unit increases in the innovations in period 0 only. Collect as the responses of y for as many periods as the desired coefficients $\psi_1, \psi_2, \dots, \psi_n$ above.
3. Collect the responses of y through time in history matrices H_i , where $H_i = [y_{0,i}, y_{1,i}, y_{2,i}, \dots, y_{n,i}]$ where $y_{t,i}$ denotes the response of y to the i^{th} at time t .
4. Form ψ_j by collecting the relevant columns of the history matrices $\psi_j = [y_{j,1}, \dots, y_{j,n}]$.

1.3 Estimation

Just as we have seen for an AR process, if the ϵ_t term is normally distributed, and independent through time, OLS estimates equation by equation will be equivalent to maximum likelihood estimates for the whole system, conditioning on the first n observations being non-stochastic. This implies that for an unrestricted VAR, we have license to ignore the information contained in the correlation in the residual across different equations.

The essential ingredient needed for this result to go through is that the lag structure of the VAR be the same in each equation. If instead, one is interested in estimating a system with parametric restrictions on the coefficients, one can obtain more efficient estimates by considering the correlation of the residuals across equations. Instead of OLS equation by equation, in that case, one can deploy a SUR estimator or form the conditional maximum likelihood and optimize it numerically.

If you cannot stomach the assumption that the first n observations be non-stochastic, then you can proceed numerically using unconditionally maximum likelihood.

1.4 Identification

Realizing that the VAR process for y_t is simply a dynamic system of equations, one can see that it can be thought as the reduced-form of the structural system:

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

where $\tilde{\epsilon}_t$ is IID with variance Σ , which is related to Ω above by $\Omega = (\phi_0^{-1})\Sigma(\phi_0^{-1})'$. Furthermore $\phi_j = \frac{\tilde{\phi}_j}{\tilde{\phi}_0}$ for $j \in \{1, \dots, n\}$ and $c = \frac{\tilde{c}}{\tilde{\phi}_0}$.

It is also apparent from the equation above, that unless particularly restrictive conditions hold, the estimates of the innovations for the reduced form of the system will co-mingle the structural innovations and will have no easy economic interpretation.

It is apparent from the structural form of the VAR that there are N^2 more structural parameters than reduced form parameters. For identification of the structural parameters through the reduced form estimates, there needs to be a one-to-one mapping between the coefficients of the two forms. Normalizing the diagonal elements of ω to be 1, implies that that for identification, we need to place at least $N^2 - N$ restrictions on the structural parameters of the model. This is where economic theory is supposed to come to the rescue.

While a typical presentation of the methods relating to the estimation and identification of dynamic systems focuses only parametric restrictions on $\tilde{\phi}_0$ to $\tilde{\phi}_n$, the literature on identified vector auto-regressions typically relies on restrictions on $\tilde{\phi}_0$ and Σ .

2 A Recursive Identification Scheme

As a first example of identifying the innovations of a VAR, consider the recursive identification scheme. The assumptions underlying this scheme are that $\tilde{\phi}_0$ is lower triangular and that Σ is diagonal. What is the meaning of these restrictions? With Σ diagonal, the structural innovations are taken to be uncorrelated with each other. The fact that the matrix $\tilde{\phi}_0$ is taken to be lower triangular implies that only the first shock can affect the first variable in the VAR contemporaneously; only the first and second shocks can affect the second variable contemporaneously; and so on.

Notice that the identification scheme described above involves $N^2 - N$ restrictions. $\frac{N^2 - N}{2}$ restrictions come from the assumption of a triangular $\tilde{\phi}_0$, and an additional $\frac{N^2 - N}{2}$ come from a diagonal Σ (taking into account that variance covariance matrices are symmetric).

One way to retrieve the structural coefficients would be to estimate them directly by maximum likelihood, but we can follow a route that is less numerically taxing.

For this consider the Cholesky decomposition. Given a positive definite matrix A the Cholesky decomposition yields a triangular matrix R such that $A = RR'$. Remember also that the inverse of a triangular matrix is itself triangular. These two facts suggest the following estimation strategy:

1. Estimate the variance covariance matrix of the unidentified residuals by $\hat{\Omega}$, where the i, j entry of $\hat{\Omega}$ is given by $\sum_{k=1+n}^T \frac{e_{i,k}e_{j,k}}{T}$ where n is the order of the VAR, T is the number of observations and $e_{i,k}, e_{j,k}$ are the OLS residuals from the i^{th} and j^{th} equation of the VAR for period k , respectively.
2. By construction, $\hat{\Omega}$ is positive definite, so we can entertain computing its Cholesky decomposition to construct $\hat{\Omega} = RR'$
3. Extract the diagonal elements of R , place them in the diagonal matrix *Sigma*
4. Form a new matrix Q by copying the non-diagonal elements of R . Place elements equal to 1 along the diagonal of Q (Notice that, by construction, $\hat{\Omega} = Q\hat{\Sigma}Q'$.)
5. Form $\hat{\phi}_0$ the estimate of $\tilde{\phi}_0$ by inverting the matrix Q .

3 Variance decompositions

Because the identified innovations produced by the recursive identification scheme are uncorrelated with each other, we can easily decompose the variance of the in-sample predicted values into the individual contributions of the various identified innovations.

For variable i in the VAR, the contributions to its predicted variance by the k innovation can be sizes as $var(\hat{y}_i|k)/var(\hat{y}_i)$. where $\hat{y}_i|k$, is the in-sample predicted series for \hat{y}_i conditional on only turning on the k innovations only.