Chapter 4: Vector Autoregressions.

Jesús Bueren

EUI

Introduction

- This chapter describes the dynamic interactions among a set of variables collected in an $(n \times 1)$ vector $\boldsymbol{y}_t$.
- A *p-th order vector autoregression*, VAR(p), is a vector generalization of an AR(p):

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t \tag{1}$$

- The $(n \times 1)$ vector $\boldsymbol{\epsilon}_t$ is a vector generalization of white noise:

$$E(\boldsymbol{\epsilon}_t) = 0$$

$$E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_\tau') = \begin{cases} \boldsymbol{\Omega} \text{ for } t = \tau \\ \boldsymbol{0} \text{ otherwise} \end{cases}$$

Introduction

- The first row of the vector system specifies that:

$$
\begin{aligned}
y_{1t} = c_1 &+ \phi_{1,1}^{(1)} y_{1,t-1} + \cdots + \phi_{1,n}^{(1)} y_{n,t-1} \\
&+ \phi_{1,1}^{(2)} y_{1,t-2} + \cdots + \phi_{1,n}^{(2)} y_{n,t-2} \\
&+ \quad \vdots \quad + \cdots + \quad \vdots \\
&+ \phi_{1,1}^{(p)} y_{1,t-p} + \cdots + \phi_{1,n}^{(p)} y_{n,t-p} \\
&+ \epsilon_{1,t}
\end{aligned}
$$

Thus a vector autoregression is a system in which each variable is regressed on a constant and p of its own lags as well as on p lags of each other variables.

Stationarity

- As we did in the univariate case, we can rewrite the VAR(p) system as a VAR(1):

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t,$$

where,

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

- If the eigenvalues of $\mathbf{F}$ all lie inside the unit circle, then the VAR turns out to be covariance stationary

Stationarity

- A vector $\mathbf{y}_t$ is said to be covariance-stationary if its first and second moments are independent of date t.

- Assuming covariance- stationarity, we can take expectations of both sides of equation (1) to find:

$$\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{c}$$

We can thus rewrote equation (1) as:

$$\mathbf{y}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \boldsymbol{\Phi}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t$$

The Conditional Likelihood Function

- The likelihood function is calculated in the same way as for a scalar autoregression.
- Conditional on the values of $\boldsymbol{y}$ observed from date $t-p$ to $t-1$, the value of $\boldsymbol{y}_t$ follows:

$$\boldsymbol{y}_t|\boldsymbol{y}_{t-1},\dots,\boldsymbol{y}_{t-p} \sim N(\mathbf{c}+\boldsymbol{\Phi}_1\mathbf{y}_{t-1}+\dots+\boldsymbol{\Phi}_p\mathbf{y}_{t-p},\boldsymbol{\Omega})$$

- The conditional MLE of $\boldsymbol{\Phi}$ coincides with n OLS regressions (prove it!).
- The conditional MLE of $\boldsymbol{\Omega}$ coincides with sample variance-covariance matrix of the OLS residuals (prove it!).

Granger Causality

- Very bad name: Granger predictability would be much better.
- One of the key questions that can be addressed with vector autoregression is how useful some variables are for **forecasting** others.
- In a bivariate VAR describing x and y, y does not Granger-cause x in case if it cannot help forecast x.
  - The coefficient matrices $\mathbf{\Phi}_j$ are lower triangular for all j

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{11}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \phi_{11}^{(p)} & 0 \\ \phi_{21}^{(p)} & \phi_{22}^{(p)} \end{bmatrix} \begin{bmatrix} x_{t-p} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Granger Causality

F-test

- A simple approach would be to consider the regression:

$$x_t = c_1 + \phi_{11}^{(1)} x_{t-1} + \cdots + \phi_{11}^{(p)} x_{t-p} + \phi_{12}^{(1)} y_{t-1} + \cdots + \phi_{12}^{(p)} y_{t-p} \quad (2)$$

- Then, you could conduct an F-test for the null hypothesis (no granger causality):

$$\mathrm{H}_0 : \phi_{12}^{(1)} = \cdots = \phi_{12}^{(p)} = 0$$

Granger Causality

Test for Granger Causality

- Estimate eq (2) and compute the sum of squared residuals:

$$RSS_1 = \sum_{t=1}^{T} \hat{u}_t^2$$

- Re-estimate eq (2) by imposing the null and compute the sum of squared residuals:

$$RSS_2 = \sum_{t=1}^{T} \hat{e}_t^2$$

- Compute:

$$S = \frac{T(RSS_2 - RSS_1)}{RSS_1}$$

- Under the null, reject if S greater than the 5% critical values for a $\chi^2(p)$

Granger Causality

Relation between 'causality' and 'Granger causality'

- Granger causality and causality are very different concepts.
- In fact, they can run in the opposite direction as we will see in the following example:
  - The price as a stock represent the expected discounted present value of future dividends: $P_t = E\left[\sum_{j=1}^{\infty} \frac{D_{t+j}}{(1+r)^j}\right]$
  - Imagine $D_t = d + u_t + \delta u_{t-1} + v_t, u_t$ and $v_t \sim WN$ and observable.
  - Then $E[D_{t+j}] = \begin{cases} d + \delta u_t \text{ if } j = 1 \\ d \text{ if } j > 1 \end{cases}$,
  - We can write:

$$P_t = \frac{d}{r} + \frac{\delta u_t}{1+r}$$

$$D_t = -\frac{d}{r} + (1+r)P_{t-1} + u_t + v_t$$

- Hence, in this model, Granger causation runs in the opposite direction than true causation.

Reduced-form IRFs

- Assuming stationarity, we can rewrite a the reduced form VAR(p) as a VMA (∞):

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \sum_{i=1}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{\epsilon}_{t-i}$$

- We could simply simulate the system to compute the IRFs.
- The IRF ($\frac{\partial y_{i,t+s}}{\partial \epsilon_{j,t}}$) describes the response of $y_{i,t+s}$ to a one-time unit change in $y_{j,t}$ holding all other variables at data t or earlier held constant.

Reduced-form IRFs

Interpretation

- Can we interpret the IRF as the causal effect of $y_{j,t}$ on $y_{i,t+s}$?
- Imagine that we knew $\{y_{t-1}, \dots, y_{t-p}\}$
- Suppose we were told at date t that $y_{1,t}$ was larger than expected, how would this cause us to revise our forecast about variable $y_{i,t}$? Is this $\frac{\partial y_{i,t+s}}{\partial \epsilon_{j,t}}$?
- No, unless Ω is a diagonal matrix.

Reduced-form IRFs

Error Bands

1. Estimate VAR and save $\hat{\mathbf{\Phi}}$, and residuals $\hat{\epsilon} = \{\hat{\epsilon}_1, \ldots, \hat{\epsilon}_T\}$
2. Draw uniformly and with replacement from these residuals and use $\hat{\mathbf{\Phi}}$ to construct a new simulated serie of $\mathbf{Y}^s$ (take $\mathbf{Y}_1^s$ from the data).
3. Estimate a new $\hat{\mathbf{\Phi}}$ from this new sample and its associated impulse response.
4. Go back to 2 until you generate M impulse response functions.

From the Structural to the Reduced-form VAR

- The impulse responses in terms of ϵ_t have a difficult economic interpretation.
- We are shocking one element in ϵ leaving the others unchanged but Ω is a non-diagonal matrix.
- As such, we cannot interpret them as the causal effect of one variable on another one.

The Structural Model

- Therefore let's think about writing the structural model (the data generating process):

$$\mathbf{B}_0 y_t = \mathbf{k} + \mathbf{B}_1 y_{t-1} + \cdots + \mathbf{B}_p y_{t-p} + \mathbf{u}_t, u_t \sim N(0, \mathbf{I}_n) \qquad (3)$$

  where $\mathbf{D}$ is a diagonal matrix.

- If we knew the data-generating process, we could understand contemporaneous and future *causal effects* of one variable over the other.

- In this VAR, shocks have a well defined economic interpretation.

- **Problem:** We cannot estimate the system (3) by a series of n OLS equations because of reverse causality.

The Structural Model

Example: Bivariate VAR(1)

- Imagine that we want to study the effect of changes in the interest rate (r_t) on output growth (y_t).

- A structural VAR can help us answering what is the effect of a change in the interest rate on output.

$$\begin{bmatrix} b^0_{1,1} & b^0_{1,2} \\ b^0_{2,2} & b^0_{2,2} \end{bmatrix} \begin{bmatrix} y_t \\ r_t \end{bmatrix} = \begin{bmatrix} b^1_{1,1} & b^1_{1,2} \\ b^1_{2,2} & b^1_{2,2} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u^y_t \\ u^r_t \end{bmatrix}$$

with

$$\begin{bmatrix} u^y_t \\ u^r_t \end{bmatrix} \sim N(0, \boldsymbol{I}_2), \boldsymbol{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- $b^0_{1,2}$ captures the contemporaneous causal effect of an change in the interest rate on output.

The Structural Model

Example: Bivariate VAR(1)

- We can rewrite the structural VAR as a system of two equations:

$$y_t = \alpha_1 r_t + \alpha_2 y_{t-1} + \alpha_3 r_{t-1} + u_t^y \quad (4)$$

$$r_t = \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 r_{t-1} + u_t^r \quad (5)$$

- We cannot estimate these two equations by OLS:

  $u_t^y \rightarrow y_t \rightarrow r_t$ so r_t is endogenous 4

  $u_t^r \rightarrow r_t \rightarrow y_t$ so y_t is endogenous 5

The Structural Model

Identification Problem

- We can premultiply the structural VAR(1) by $\boldsymbol{B}_0^{-1}$:

$$\boldsymbol{y}_t = \boldsymbol{B}_0^{-1}\mathbf{k} + \boldsymbol{B}_0^{-1}\mathbf{B}_1 y_{t-1} + \cdots + \boldsymbol{B}_0^{-1}\mathbf{B}_p y_{t-p} + \boldsymbol{B}_0^{-1}\mathbf{u}_t$$
$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t$$

  here we see that the reduced form VAR shocks ($\boldsymbol{\epsilon}_t$) are linear combination of the structural shocks ($\boldsymbol{u}_t$).

- The model implies that:

$$\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}(\boldsymbol{B}_0^{-1})' \tag{6}$$

- We can estimate consistently $\boldsymbol{\Omega}$ but we cannot we cannot identify $\boldsymbol{B}_0^{-1}$ as there are infinite $\boldsymbol{B}_0$ that satisfy 6.

The Structural Model

Example

- Think of $\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}(\boldsymbol{B}_0^{-1})'$ as a system of equations:

$$\begin{bmatrix} \sigma_y^2 & \sigma_{y,r}^2 \\ \sigma_{y,r}^2 & \sigma_r^2 \end{bmatrix} = \begin{bmatrix} b_{11}^{-1} & b_{12}^{-1} \\ b_{21}^{-1} & b_{22}^{-1} \end{bmatrix} \begin{bmatrix} b_{11}^{-1} & b_{21}^{-1} \\ b_{12}^{-1} & b_{22}^{-1} \end{bmatrix}$$

- We can write it as:

$$\begin{aligned} \sigma_y^2 &= (b_{11}^{-1})^2 + b_{12}^{-1} b_{12}^{-1} \\ \sigma_{y,r}^2 &= b_{11}^{-1} b_{21}^{-1} + b_{12}^{-1} b_{22}^{-1} \\ \sigma_{y,r}^2 &= b_{21}^{-1} b_{11}^{-1} + b_{22}^{-1} b_{12}^{-1} \\ \sigma_r^2 &= (b_{21}^{-1})^2 + (b_{22}^{-1})^2 \end{aligned}$$

$\Rightarrow$ 3 equations and 4 unknowns

Recursive VARs

- A common solution is to impose restrictions on the structural model (based in economic theory) so that the can recover the structural parameters.
- Imagine that we are willing to restrict the contemporaneous relation of the different variables:
  - $\boldsymbol{B}_0$ is lower triangular/upper triangular.
- Procedure:
  1. Estimate the reduce form VAR.
  2. Based on $\hat{\boldsymbol{\Omega}}$ compute B_0 using the Cholesky decomposition.
  3. Compute IRF and confidence bands using the structural shocks

Recursive VAR

- Cholesky Decomposition decomposes $\boldsymbol{\Omega}$ into

$$\boldsymbol{\Omega} = \boldsymbol{B}_0^{-1}(\boldsymbol{B}_0^{-1})'$$

where B_0^{-1} is lower triangular.

- Once we have an estimate of $\boldsymbol{B}_0^{-1}$, we can shock one element of $\boldsymbol{u}_t$ (the structural shock) and compute the contemporaneous and future causal impact of the innovation of a particular variable on another one.

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{B}_0^{-1} \boldsymbol{u}_t$$

Recursive VAR

- In our bivariate VAR example, we could assume that monetary policy works with a lag and thus has no contemporaneous effect on output.

- This would entail:

$$\begin{bmatrix} b^0_{1,1} & \cancel{b^0_{1,2}} \\ b^0_{2,2} & b^0_{2,2} \end{bmatrix} \begin{bmatrix} y_t \\ r_t \end{bmatrix} = \begin{bmatrix} b^1_{1,1} & b^1_{1,2} \\ b^1_{2,2} & b^1_{2,2} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} u^y_t \\ u^r_t \end{bmatrix}$$

- And thus the system of equations from $\boldsymbol{B}^1_0(\boldsymbol{B}^1_0)' = \boldsymbol{\Omega}$ would become:

$$\begin{aligned} \sigma^2_y &= (b^{-1}_{11})^2 \\ \sigma^2_{y,r} &= b^{-1}_{11} b^{-1}_{21} \\ \sigma^2_r &= (b^{-1}_{21})^2 + (b^{-1}_{22})^2 \end{aligned}$$

This is now a system of 3 equation and 3 unknows solved by the Cholesky decomposition.

Structural IRFs

- Assuming stationarity, we can rewrite a structural VAR(p) as a VMA (∞):

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{u}_t + \sum_{i=1}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{u}_{t-i}$$

Structural IRFs

Error Bands

1. Estimate VAR and save $\hat{\mathbf{\Phi}}$, and residuals $\hat{\epsilon} = \{\hat{\epsilon}_1, \ldots, \hat{\epsilon}_T\}$
2. Draw uniformly and with replacement from these residuals and use $\hat{\mathbf{\Phi}}$ to construct a new simulated serie of $\mathbf{Y}^s$ (take $\mathbf{Y}_1^s$ from the data).
3. Estimate $\tilde{\mathbf{\Phi}}^s$, $\tilde{\mathbf{\Omega}}^s$ and a new $(\tilde{\boldsymbol{B}}_0^s)^{-1}$ from this new sample and its associated impulse response.
4. Go back to 2 until you generate S impulse response functions.