Causal Inference, Time Series and Economic History

4. Vector Autoregressions

Jason Lennard

Overview

- Vector autoregressions
  - Impulse response functions
  - Forecast error variance decompositions
  - Historical decompositions
- *Class discussion paper:* Nicolini, E. A., 'Was Malthus right? A VAR analysis of economic and demographic interactions in pre-industrial England', *European Review of Economic History*, 11 (2007), pp. 99-121

Vector Autoregressions (VARs)

- VARs are a popular method for understanding the dynamics between endogenous variables
- The term vector suggests more than 1 equation
- The term autoregression implies lagged values of a variable
- Commonly used in economic policy, macroeconomics and economic history

VARs

- Let's begin with the bivariate structural model (see Feinstein and Thomas, 2002, pp. 468-9, 491-4 for a primer on structural and reduced-form models):

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$

$$x_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$

- OLS estimation of these equations will lead to biased estimates of the parameters of interest because the zero conditional mean assumption is violated (reverse causality)

VARs

- A little manipulation seemingly overcomes the problem. Starting with the structural model:
$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$
$$x_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$
- Replacing x_t in the first equation with right-hand side of the second:
$$y_t = \beta_0 + \beta_1 \left(\gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x\right) + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$
- Expand out of the brackets:
$$y_t = \beta_0 + \beta_1 \gamma_0 + \beta_1 \gamma_1 y_t + \beta_1 \gamma_2 y_{t-1} + \beta_1 \gamma_3 x_{t-1} + \beta_1 u_t^x + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$
- Subtract $\beta_1 \gamma_1 y_t$ from both sides:
$$y_t - \beta_1 \gamma_1 y_t = \beta_0 + \beta_1 \gamma_0 + \beta_1 \gamma_2 y_{t-1} + \beta_1 \gamma_3 x_{t-1} + \beta_1 u_t^x + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$

VARs

- Collect terms on the left-hand side:
$$y_t\left(1-\beta_1\gamma_1\right)=\beta_0+\beta_1\gamma_0+\beta_1\gamma_2 y_{t-1}+\beta_1\gamma_3 x_{t-1}+\beta_1 u_t^x+\beta_2 y_{t-1}+\beta_3 x_{t-1}+u_t^y$$
- Rearrange and collect terms on the right-hand side:
$$y_t\left(1-\beta_1\gamma_1\right)=\beta_0+\beta_1\gamma_0+\beta_1\gamma_2 y_{t-1}+\beta_2 y_{t-1}+\beta_1\gamma_3 x_{t-1}+\beta_3 x_{t-1}+\beta_1 u_t^x+u_t^y$$
$$y_t\left(1-\beta_1\gamma_1\right)=\left(\beta_0+\beta_1\gamma_0\right)+\left(\beta_1\gamma_2+\beta_2\right)y_{t-1}+\left(\beta_1\gamma_3+\beta_3\right)x_{t-1}+\left(\beta_1 u_t^x+u_t^y\right)$$
- Divide both sides by $1-\beta_1\gamma_1$ to yield the first equation of the reduced form bivariate VAR:
$$y_t=\frac{\beta_0+\beta_1\gamma_0}{1-\beta_1\gamma_1}+\frac{\beta_1\gamma_2+\beta_2}{1-\beta_1\gamma_1}y_{t-1}+\frac{\beta_1\gamma_3+\beta_3}{1-\beta_1\gamma_1}x_{t-1}+\frac{\beta_1 u_t^x+u_t^y}{1-\beta_1\gamma_1}$$

VARs

- Moving on to the second equation of the reduced form bivariate VAR:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y$$

$$x_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$

- Replace y_t in the second equation with right-hand side of the first:

$$x_t = \gamma_0 + \gamma_1 \left(\beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t^y\right) + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$

- Expand out of the brackets:

$$x_t = \gamma_0 + \gamma_1 \beta_0 + \gamma_1 \beta_1 x_t + \gamma_1 \beta_2 y_{t-1} + \gamma_1 \beta_3 x_{t-1} + \gamma_1 u_t^y + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$

- Subtract $\gamma_1 \beta_1 x_t$ from both sides:

$$x_t - \gamma_1 \beta_1 x_t = \gamma_0 + \gamma_1 \beta_0 + \gamma_1 \beta_2 y_{t-1} + \gamma_1 \beta_3 x_{t-1} + \gamma_1 u_t^y + \gamma_2 y_{t-1} + \gamma_3 x_{t-1} + u_t^x$$

VARs

- Collect terms on the left-hand side:

$$x_t\,(1-\gamma_1\beta_1)=\gamma_0+\gamma_1\beta_0+\gamma_1\beta_2 y_{t-1}+\gamma_1\beta_3 x_{t-1}+\gamma_1 u_t^y+\gamma_2 y_{t-1}+\gamma_3 x_{t-1}+u_t^x$$

- Rearrange and collect terms on the right-hand side:

$$x_t\,(1-\gamma_1\beta_1)=\gamma_0+\gamma_1\beta_0+\gamma_1\beta_2 y_{t-1}+\gamma_2 y_{t-1}+\gamma_1\beta_3 x_{t-1}+\gamma_3 x_{t-1}+\gamma_1 u_t^y+u_t^x$$

$$x_t\,(1-\gamma_1\beta_1)=(\gamma_0+\gamma_1\beta_0)+(\gamma_1\beta_2+\gamma_2)\,y_{t-1}+(\gamma_1\beta_3+\gamma_3)\,x_{t-1}+\left(\gamma_1 u_t^y+u_t^x\right)$$

- Divide both sides by $1-\gamma_1\beta_1$ to yield the second equation of the reduced form bivariate VAR:

$$x_t=\frac{\gamma_0+\gamma_1\beta_0}{1-\gamma_1\beta_1}+\frac{\gamma_1\beta_2+\gamma_2}{1-\gamma_1\beta_1}y_{t-1}+\frac{\gamma_1\beta_3+\gamma_3}{1-\gamma_1\beta_1}x_{t-1}+\frac{\gamma_1 u_t^y+u_t^x}{1-\gamma_1\beta_1}$$

VARs

- Simplifying gives the reduced-form VAR, which omits the contemporaneous terms:

$$y_t = \frac{\beta_0 + \beta_1\gamma_0}{1 - \beta_1\gamma_1} + \frac{\beta_1\gamma_2 + \beta_2}{1 - \beta_1\gamma_1} y_{t-1} + \frac{\beta_1\gamma_3 + \beta_3}{1 - \beta_1\gamma_1} x_{t-1} + \frac{\beta_1 u_t^x + u_t^y}{1 - \beta_1\gamma_1}$$

$$x_t = \frac{\gamma_0 + \gamma_1\beta_0}{1 - \gamma_1\beta_1} + \frac{\gamma_1\beta_2 + \gamma_2}{1 - \gamma_1\beta_1} y_{t-1} + \frac{\gamma_1\beta_3 + \gamma_3}{1 - \gamma_1\beta_1} x_{t-1} + \frac{\gamma_1 u_t^y + u_t^x}{1 - \gamma_1\beta_1}$$

$$y_t = \alpha + \phi_1 y_{t-1} + \psi_1 x_{t-1} + \varepsilon_t^y$$

$$x_t = \mu + \theta_1 y_{t-1} + \eta_1 x_{t-1} + \varepsilon_t^x$$

- The reduced-form parameters are composites of the structural parameters ($\alpha = \frac{\beta_0+\beta_1\gamma_0}{1-\beta_1\gamma_1}$, $\phi_1 = \frac{\beta_1\gamma_2+\beta_2}{1-\beta_1\gamma_1}$, $\psi_1 = \frac{\beta_1\gamma_3+\beta_3}{1-\beta_1\gamma_1}$, $\varepsilon_t^y = \frac{\beta_1 u_t^x+u_t^y}{1-\beta_1\gamma_1}$, etc.)

VARs

- Moving from the reduced-from bivariate VAR(1) to a reduced-form bivariate VAR(p):

$$y_t = \alpha + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{p} \psi_j x_{t-j} + \varepsilon_t^y$$

$$x_t = \mu + \sum_{j=1}^{p} \theta_j y_{t-j} + \sum_{j=1}^{p} \eta_j x_{t-j} + \varepsilon_t^x$$

- These equations can be estimated separately using OLS

Steps

- Step 1: Determine lag length (p) using information criteria
- Step 2: Estimate VAR(p)
- Step 3: Check the residuals
- Step 4: Display key results: Impulse response function (IRF), forecast error variance decomposition (FEVD), historical decomposition (HD)

IRF

- VARs have multiple variables and can have more than 1 lag
- It can be difficult to interpret many coefficients
- Sometimes different lags for the same variable may switch sign
- It is therefore common to report an IRF instead
- An IRF measures the effect of a shock to an endogenous variable on itself or on another endogenous variable at a particular horizon

IRF

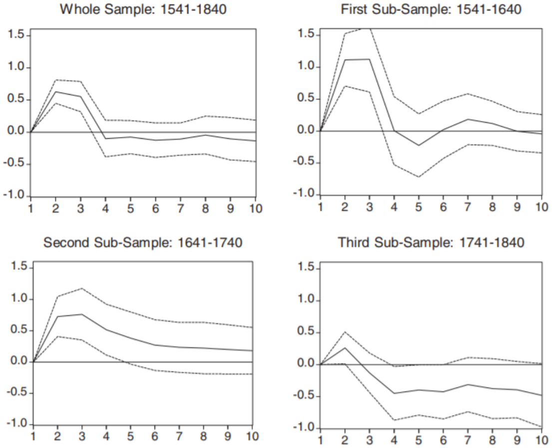


Figure 4. *Evolution of preventive check: responses of CBR to a shock in LRW.*

IRF and CIRF

- Shock the system with an impulse ($\varepsilon_t^i \neq 0$) (common to use σ_{ε^i} or 1)
- All past observations, past and future shocks set to 0
- The impulse response h steps ahead is defined as:

$$IRF_h = \frac{\partial y_{t+h}}{\partial \varepsilon_t^i}$$

- The cumulative impulse response function (CIRF) is calculated as:

$$CIRF_h = \sum_{i=0}^{h} IRF_i$$

- The CIRF is typically used when the time series are in differences or growth rates
- Easier in practice, let's look at an example in Excel

Pros and Cons of VARs

Pros

- Simple. VARs can be estimated equation by equation with OLS
- Produce good forecasts

Cons

- Make strong assumptions about the underlying data generating process (Jordà, 2005)
- Can consume many degrees of freedom, i.e., $df = (k \times p) + 1$
- Difficult to interpret coefficients
- Identification!

Identification

- In VAR models, we analyse the effect of a shock on a set of variables
- But what really is this shock?
- Remember our reduced-form errors are composites of the structural shocks:

$$\varepsilon_t^y = \frac{1}{1 - \beta_1 \gamma_1} \left(u_t^y + \beta_1 u_t^x \right)$$

$$\varepsilon_t^x = \frac{1}{1 - \beta_1 \gamma_1} \left(\gamma_1 u_t^y + u_t^x \right)$$

Identification

- Short-run restrictions/ Cholesky decomposition
  - Bernanke and Blinder (1992)
  - Nicolini (2007)
- Uses recursive zero restrictions on the contemporaneous coefficients
- In the bivariate case, set either β_1 or γ_1 to zero and estimate the other
- The restriction(s) are motivated by theory or institutional knowledge
- In practice, this can be implemented using the Cholesky decomposition

Identification

- Nicolini (2007), for example, orders the crude birth rate (CBR) before the crude death rate (CDR), real wages (W) and the crude marriage rate (CMR), which assumes that CBR is not affected by CDR, W or CMR *contemporaneously*
- This is based on the idea that a change in CDR, W or CMR in year t is unlikely to result in a change in CBR in year t because of the lags involved in conception and pregnancy

Identification

- Long-run restrictions
  - Blanchard and Quah (1989)
- Sign restrictions
  - Uhlig (2005)
- Other restrictions
  - Blanchard and Perotti (2002)
- External instruments
  - Stock and Watson (2008, 2012) and Mertens and Ravn (2013)

FEVD

- A measure of each shock's relative importance in explaining the variability of a variable at a particular horizon
- Based on the (orthogonal) impulse response function (following the Cholesky decomposition, for example)
- Let's look at an example in Excel

Table 3. *Variance decomposition after ten years.*

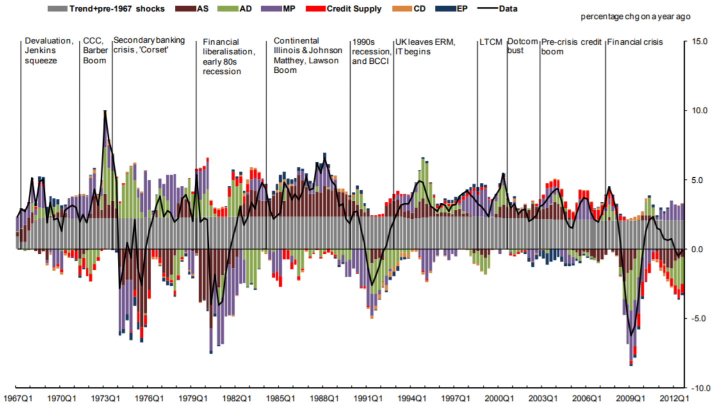
	CBR	CDR	LRW
Percentage of variance of CBR due to:			
Whole sample	69.79	16.35	13.86
First sub-sample	69.58	8.09	22.33
Second sub-sample	44.54	24.26	31.20
Third sub-sample	67.26	2.07	30.67
Percentage of variance of CDR due to:			
Whole sample	11.57	81.15	7.29
First sub-sample	14.18	62.53	23.29
Second sub-sample	4.54	90.48	4.98
Third sub-sample	13.70	78.49	7.81
Percentage of variance of LRW due to:			
Whole sample	0.80	5.04	94.16
First sub-sample	5.49	3.54	90.97
Second sub-sample	3.49	11.20	85.31
Third sub-sample	11.09	1.49	87.42

Historical Decomposition

- What is the contribution of a specific shock to the data?
- This can be shown with the historical decomposition
- It shows the contribution of each shock to a variable over time

Historical Decomposition

Chart 8: Historical decomposition of GDP growth (with policy response)



Class Discussion Paper: Nicolini (2007)

Research Question

- To test the Malthusian hypothesis in England using a long time series

Class Discussion Paper: Nicolini (2007)

- Crude birth rate (CBR): Wrigley and Schofield (1981)
- Crude death rate (CDR): Wrigley and Schofield (1981)
- Log real wages of London labourers (W): Allen (2001)
- Crude marriage rate (CMR): Wrigley and Schofield (1981)
- Dummy variables for epidemics (1557, 1558, 1559, 1563, 1603, 1625, 1658, 1659, 1665, 1681, 1728, 1729)
- Sample: 1541-1840

Class Discussion Paper: Nicolini (2007)

Model

- VAR(4)
- Identification: Cholesky decomposition
- Ordering: *CBR*, *CDR*, *W*, *CMR*

Next Class

Class discussion paper: Cloyne, J., 'Discretionary tax changes and the macroeconomy: New narrative evidence from the United Kingdom', *American Economic Review*, 103 (2013), pp. 1507-28.

Further Reading

- Stock and Watson, *Introduction to Econometrics*, chapter 17 (pp. 649-55)