Applied Econometrics for Macro and Finance

Vector Autoregression Models

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Historical Background

Traditionally, two types of models are used in macroeconomics:

- Large-scale macroeconometric model: large, simultaneous equations models (eg. Fair Model)
- Transfer-function-like model (see Chapter 5.1—5.4 in Enders)

Critics

- For Large-scale macroeconometric model:
  - It is *ad hoc*
  - The identification (model specification) for existing large-scale models is questionable
- For Transfer-function-like model: fail to account for possible feedback effects (not exactly exogenous).

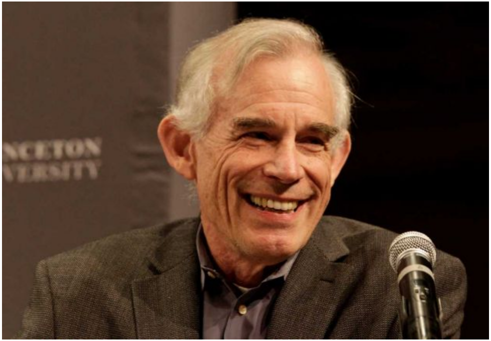
Sims (1980)'s solution

- Build a model in a style which does not tend to impose restrictions so arbitrarily
- Estimate models as unrestricted reduced forms, treating all variables as endogenous

VAR Models and Macroeconometrics

- VAR Models are now used as a central tool in empirical macroeconomics.
  - Macroeconomists now spend a lot of time examining the shocks in VAR models and their effects.
  - Examples:
    - (a) Monetary Policy Shocks
    - (b) Oil Price Shocks
    - (c) Confidence Shocks
- Christopher Sims (1980, *Econometrica*) provided a new macro-econometric framework that held great promise.
  - Christiano (2012) " Christopher A. Sims and Vector Autoregressions " *Scandinavian Journal of Economics*

Christopher Sims



- Christopher A. Sims (1942–), Princeton University. Nobel Laureate (2011)

Four Tasks of VAR for Macroeconometricians
– Stock and Watson (2001, *JEP*)

- Describe and summarize macroeconomic data
  - Granger casuality test
- Make macroeconomic forecasts
- Quantify what we do or do not know about the structure of the macroeconomy
  - backward-looking Taylor rule vs. forward-looking Taylor rule
  - monetary policy transmission
- Advise macroeconomic policymakers

Types of VAR:

1. Reduced-form VAR (simply, VAR)
2. Structural VAR (SVAR)
    - Structural inference: impulse response functions, variance decomposition, historical decomposition

Reduced-form VAR

a bivariate VAR(p) model is essentially an ARDL(p, p) model

- Consider a bivariate VAR(p) model:

$$x_t = c_x + \beta_1 x_{t-1} + \cdots + \beta_p x_{t-p} + \delta_1 z_{t-1} + \cdots + \delta_p z_{t-p} + \varepsilon_{xt}$$
$$z_t = c_z + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \theta_1 z_{t-1} + \cdots + \theta_p z_{t-p} + \varepsilon_{zt}$$

- The variance-covariance of ε_{xt} and ε_{zt} is

$$\Sigma_\varepsilon = \begin{bmatrix} Var(\varepsilon_{xt}) & Cov(\varepsilon_{xt}, \varepsilon_{zt}) \\ Cov(\varepsilon_{zt}, \varepsilon_{xt}) & Var(\varepsilon_{zt}) \end{bmatrix}$$

Reduced-form VAR

- Use VAR(1) as an example

$$x_t = c_x + \beta_1 x_{t-1} + \delta_1 z_{t-1} + \varepsilon_{xt}$$

$$z_t = c_z + \alpha_1 x_{t-1} + \theta_1 z_{t-1} + \varepsilon_{zt}$$

- In vector form

AR(1) in vector form

$$\underbrace{\begin{bmatrix} x_t \\ z_t \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} c_x \\ c_z \end{bmatrix}}_{c} + \underbrace{\begin{bmatrix} \beta_1 & \delta_1 \\ \alpha_1 & \theta_1 \end{bmatrix}}_{\Phi_1} \underbrace{\begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix}}_{y_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_{xt} \\ \varepsilon_{zt} \end{bmatrix}}_{\varepsilon_t}$$

$$y_t = c + \Phi_1 y_{t-1} + \varepsilon_t$$

Reduced-form VAR

- Given $y_t \in \mathbb{R}^k$,

$$\underbrace{y_t}_{k\times 1} = \underbrace{c}_{k\times 1} + \underbrace{\Phi_1}_{k\times k}\underbrace{y_{t-1}}_{k\times 1} + \cdots + \underbrace{\Phi_p}_{k\times k}\underbrace{y_{t-p}}_{k\times 1} + \underbrace{\varepsilon_t}_{k\times 1}$$

$$\varepsilon_t \sim^{i.i.d.} (0, \Sigma_\varepsilon) \quad \text{regression error}$$

- Estimation: equation by equation OLS

Choosing the Optimal Lag Length for a VAR

- Information criteria

$$AIC = T\log|\hat{\Sigma}_{\varepsilon}| + 2N$$
$$BIC = T\log|\hat{\Sigma}_{\varepsilon}| + N\log(T)$$

  where

  - $|\hat{\Sigma}_{\varepsilon}|$=determinant of the variance-covariance matrix of the residuals
  - $N = k^2p + k$ is the total number of regressors in all equations
  - T is the number of observations

- The values of the information criteria are constructed for $p = 0, 1, 2, \ldots, p_{\max}$
- The chosen number of lags is the one minimizing the value of the given information criterion

Applications of VAR Models

- Granger Causality Test (in a bivariate VAR model)
- Forecasts

Granger Causality Test

Definition (Granger Causality)

Given $y_t = (x_t \; z_t)'$. We say that z fails to Granger-cause x if for all $s > 0$,

$$\hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots, z_t, z_{t-1}, \ldots) = \hat{E}(x_{t+s}|x_t, x_{t-1}, \ldots)$$

where $\hat{E}$ denotes a linear projection along with a constant.

- That is, z is not linearly informative about future x.
- In practice, the Granger Causality Test is implemented by testing $H_0 : \delta_1 = \delta_2 = \cdots = \delta_p = 0$ in the following regression

$$x_t = c_x + \beta_1 x_{t-1} + \cdots + \beta_p x_{t-p} + \delta_1 z_{t-1} + \cdots + \delta_p z_{t-p} + \varepsilon_t$$

Granger Causality Test

- Caveat: Granger-causality relationships can be very different from causal relationships when economic variables respond to future expected values of other variables as in the rational expectations models.
- For instance,

$$p_t = \sum_{j=1}^{\infty} \beta^j E_t d_{t+j}$$

Predictions in VAR(p)

- To simplify our notation, we first introduce VAR(p) models in Demean Form
- Note that given $E(y_t) = \mu$

$$\mu = (I - \Phi_1 - \cdots - \Phi_p)^{-1} c$$

- Hence, we can obtain a demean VAR(p) model:

$$y_t - \mu = \Phi_1(y_{t-1} - \mu) + \cdots + \Phi_p(y_{t-p} - \mu) + \varepsilon_t$$

Predictions in VAR(p)

- Write the (demean) VAR(p) into a companion form:

$$Y_t = PY_{t-1} + \epsilon_t$$

$$Y_t = \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}, \quad \epsilon_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \quad P = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \dots & \Phi_p \\ I_k & 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \vdots \\ 0 & \dots & 0 & I_k & 0 \end{bmatrix}$$

- Hence, the dynamic (recursive) forecasts are

$$E_t(Y_{t+j}) = P^j Y_t$$