Macroeconomic Forecasting (MFx)

IMFx

Structural Vector Autoregressive Models (SVAR)

Part 1

Session 1. Introduction to VARs

Introduction

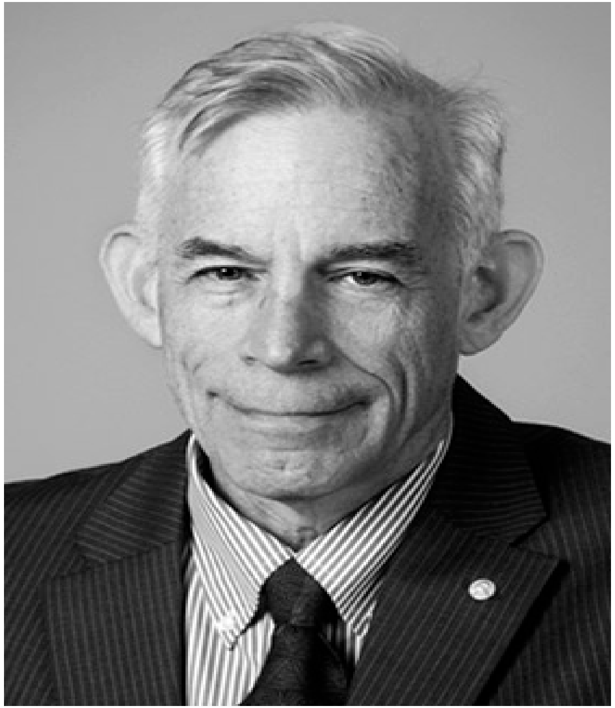
Del Negro and Schorfheide (2011):

> *"At first glance, VARs appear to be straightforward multivariate generalizations of univariate autoregressive models. At second sight, they turn out to be one of the key empirical tools in modern macroeconomics."*

What are VARs ?

- multivariate *linear* time-series models
- endogenous variable in the system are functions ***of lagged values of all endogenous*** variables
- simple and flexible alternative to the traditional multiple-equations models

Historical Overview: Sims’ Critique



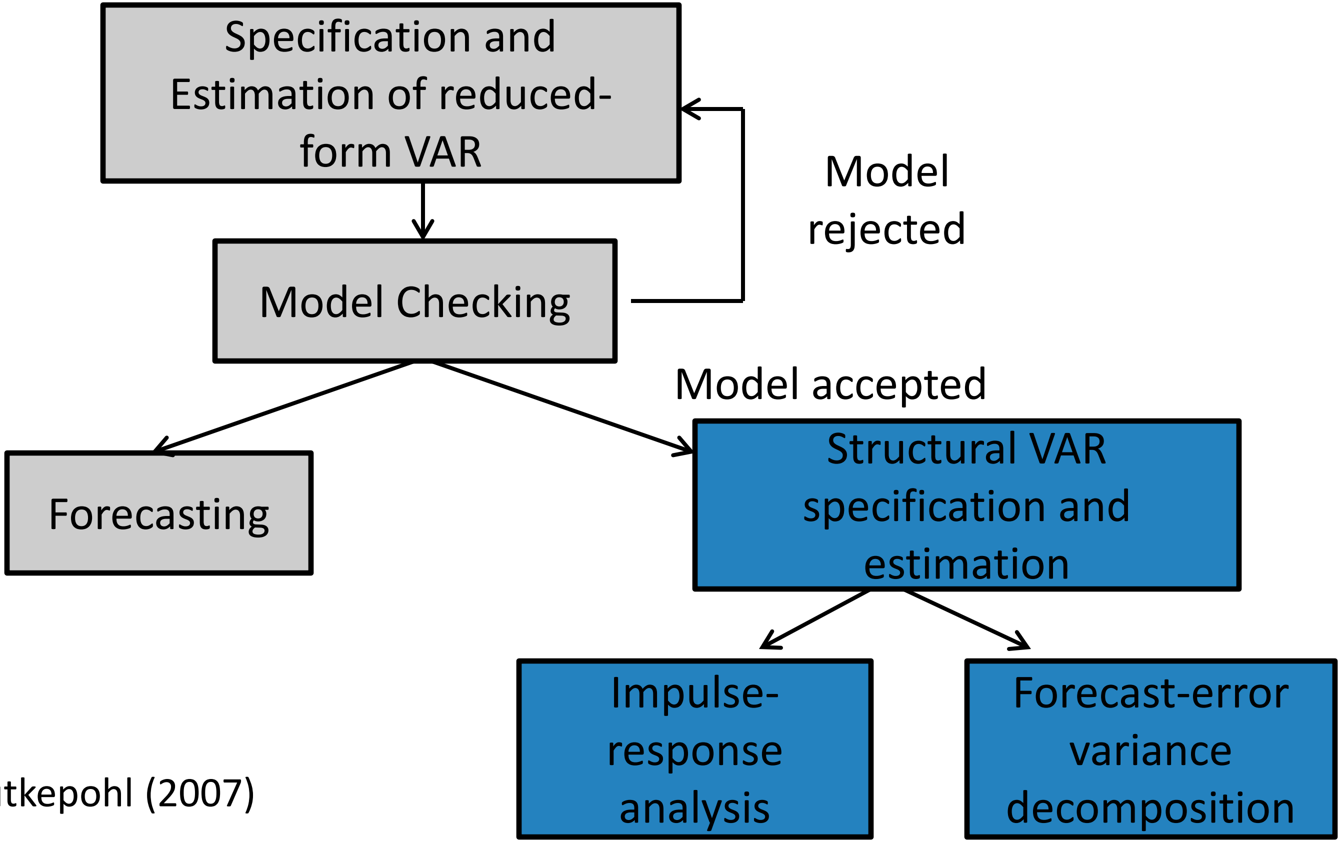
- In the 1980s criticized the large-scale macro-econometric models of the time
- Proposed VARs as an alternative that allowed one to model macroeconomic data informatively

What Are VARs Used For?

- **Forecasting**
  - *Reduced-Form VARs*

- **Structural Analysis**
  - *Structural VARs*

Unit Plan/Roadmap



Source:Lutkepohl (2007)

Session 2. Estimation of VARs

Introduction to VARs

- Let $\mathbf{y}_t$ be a vector with the value of *n* variables at time *t*:

$$\mathbf{y_t} = [y_{1,t}\, y_{2,t} \cdots y_{n,t}]'$$

- A ***p-order vector autoregressive process*** generalizes a one-variable AR(*p*) process to *n* variables:

$$\mathbf{y_t} = \mathbf{G}_0 + \mathbf{G_1 y_{t-1}} + \mathbf{G_2 y_{t\text{-}2}} + \ldots + \mathbf{G_p y_{t\text{-}p}} + \mathbf{e_t}$$

Reduced form VAR

$\mathbf{G_0}$ = (n×1) vector of constants

$\mathbf{G_j}$ = (n×n) matrix of coefficients

$\mathbf{e_t}$ = (n×1) vector of white noise innovations

$$E[\mathbf{e_t}] = 0$$

$$E[\mathbf{e_t e_\tau}'] = \begin{cases} \Omega, & \text{if } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

NOT diagonal

Example: A VAR(1) in 2 Variables

$$y_{1,t} = g_{11} y_{1,t-1} + g_{12} y_{2,t-1} + e_{1,t}$$

$$y_{2,t} = g_{21} y_{1,t-1} + g_{22} y_{2,t-1} + e_{2,t}$$

In matrix notation:

$$\mathbf{y_t} = \mathbf{G_1} \mathbf{y_{t-1}} + \mathbf{e_t}$$

where

$$\mathbf{y_t} = \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix}, \text{ for example: } \mathbf{y_t} = \begin{bmatrix} \pi_t \\ gdp_t \end{bmatrix}$$

$$\mathbf{G_1} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}, \mathbf{e_t} = \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$$

Assumptions about the error terms:

$$\mathrm{E}[\mathbf{e_t e_t'}] = \begin{pmatrix} \sigma^2_{\mathbf{e}_1} & \sigma_{\mathbf{e}_1\mathbf{e}_2} \\ \sigma_{\mathbf{e}_1\mathbf{e}_2} & \sigma^2_{\mathbf{e}_2} \end{pmatrix} = \Omega$$

$$\mathrm{E}[\mathbf{e_t e_\tau'}] = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{for t} \neq \tau$$

Estimation: by OLS

- Performed with OLS applied equation by equation

- Estimates are:
  - consistent
  - efficient
  - equivalent to GLS

General Specification Choices

- *Selection of variables* to be included: in accordance with economic theory, empirical evidence and/or experience
- *Exogenous variables* can be included: constant, time trends, other additional explanators
- Non-stationary level data is often *transformed* (log levels, log differences, growth rates, etc.)
- The model should be *parsimonious*

Session 4. Stationary VARs

Stationarity of a VAR: Definition

A p-th order VAR is said to be *covariance-stationary* if:

1. The expected value of $\mathbf{y_t}$ does not depend on time

$$E[\mathbf{y}_t] = E[\mathbf{y}_{t+j}] = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix}$$

Finite and time-invariant first and second order moments

2. The covariance matrix of $\mathbf{y_t}$ and $\mathbf{y_{t+j}}$ depends on the time lapsed j and not on the reference period t

$$E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+j} - \boldsymbol{\mu})'] = E[(\mathbf{y}_s - \boldsymbol{\mu})(\mathbf{y}_{s+j} - \boldsymbol{\mu})'] = \boldsymbol{\Gamma}_j$$

Conditions for Stationarity

- The conditions for a VAR to be stationary are similar to the conditions for a univariate AR process to be stationary:

$$\mathbf{y}_t = \mathbf{G}_0 + \mathbf{G}_1\mathbf{y}_{t-1} + \mathbf{G}_2\mathbf{y}_{t-2} + \ldots + \mathbf{G}_p\mathbf{y}_{t-p} + \mathbf{e}_t$$

Lag operator

$$\underbrace{(\mathbf{I}_n - \mathbf{G}_1\mathbf{L} - \mathbf{G}_2\mathbf{L}^2 - \ldots - \mathbf{G}_p\mathbf{L}^p)}\mathbf{y}_t = \mathbf{G}_0 + \mathbf{e}_t$$

$$\mathbf{G}(\mathbf{L})\mathbf{y}_t = \mathbf{G}_0 + \mathbf{e}_t$$

Lag polynomial

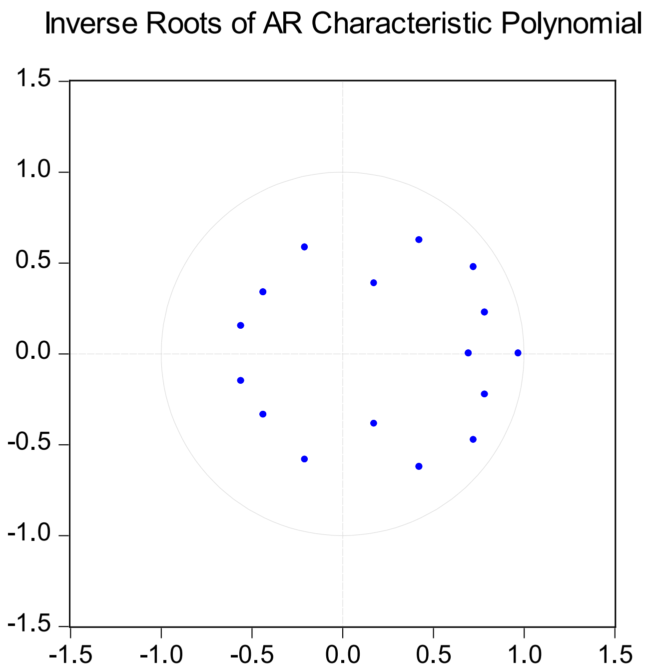
- For $\mathbf{y}_t$ to be stationary, the matrix polynomial in the lag operator $\mathbf{G}(\mathbf{L})$ must be invertible

Conditions for Stationarity

- A VAR(p) process is stationary (thus invertible) if all the *np* roots of the *characteristic polynomial* are (in modulus) outside the unit imaginary circle

$$\det(\mathbf{I_n} - \mathbf{G_1 L} - \mathbf{G_2 L^2} - \ldots - \mathbf{G_p L^p}) = 0$$

- EViews calculates the *inverse roots* of the characteristic AR polynomial, which should then lie *within* the unit imaginary circle



Vector Moving Average Representation of a VAR

- If a VAR is stationary, the $\mathbf{y_t}$ vector can be expressed as a sum of all of the past white noise shocks $\mathbf{e_t}$ (VMA(∞) representation)

$$\mathbf{y_t} = \boldsymbol{\mu} + \mathbf{G(L)^{-1} e_t}, \ \text{where } \boldsymbol{\mu} = \mathbf{G(L)^{-1} G_0}$$

$$\mathbf{y_t} = \boldsymbol{\mu} + (\mathbf{I_n} + \boldsymbol{\Psi}_1 \mathbf{L} + \boldsymbol{\Psi}_2 \mathbf{L}^2 + \ldots)\mathbf{e_t}$$

$$\mathbf{y_t} = \boldsymbol{\mu} + \mathbf{e_t} + \boldsymbol{\Psi}_1 \mathbf{e_{t-1}} + \boldsymbol{\Psi}_2 \mathbf{e_{t-2}} + \ldots$$

$$\mathbf{y_t} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\Psi}_\mathbf{i} \mathbf{e_{t-i}}$$

Wold theorem

- where $\boldsymbol{\Psi}_\mathbf{i}$ is a (n x n) matrix of coefficients, and $\boldsymbol{\Psi}_0$ is the identity matrix.

- From the VMA(∞) representation it is possible to obtain impulse response functions

Session 5. Lag Specification Criteria

Lags Needed for the VAR

What number is most appropriate? …

- If p is extremely short, the model may be poorly specified
- If p is extremely long, too many degrees of freedom will be lost
- The number of lags should be sufficient for the residuals from the estimation to constitute individual white noises

The Curse of Dimensionality

- VARs are very densely parametrized
  - In a VAR(*p*) we have *p* matrices of dimension *nxn*: G_1,.., G_p
  - Assume G_0 is an intercept vector (dimension: *nx1*)
  - The number of *total* coefficients/parameters to be estimated is:

$$n+nxnxp=n(1+nxp)$$

Overfitting versus Omitted Variable Bias

- *Over-fitting* problem

poor-quality estimates and bad forecasts

- *Omitted variable bias*

- Possible solutions:
  - Core VAR plus rotating variables
  - Bayesian analysis

Lag Length Criteria

- As for univariate models, one can use multidimensional versions of the:
  - AIC: Akaike information criterion
  - SC: Schwarz information criterion
  - HQ: Hanna-Quinn information criterion
- *Information-based criteria*: trade-off between parsimony and reduction in sum of squares

Lag Specification: Practitioner's Advice

Rules of Thumb:

- $p = 4$ when working with quarterly data
- $p = 12$ with monthly data
- The effective constraint is $np < T/3$

Example:

$T = 100$

$p = 4$

$n \leq 7$

Session 8. Forecasting using VARs

Forecasting Using the Estimated VAR

- Let $\mathbf{Y_{t\text{-}1}}$ be a matrix containing all information available up to time *t* (before realizations of $\mathbf{e_t}$ are known):

$$\mathbf{Y_{t\text{-}1}} = (\mathbf{y_{t\text{-}1}}, \mathbf{y_{t\text{-}2}}, \ldots, \mathbf{y_{t\text{-}T}})$$

- Then:

$$E[\mathbf{y_t} \mid \mathbf{Y_{t\text{-}1}}] = \hat{\mathbf{G}}_0 + \hat{\mathbf{G}}_1 \mathbf{y_{t\text{-}1}} + \hat{\mathbf{G}}_2 \mathbf{y_{t\text{-}2}} + \ldots + \hat{\mathbf{G}}_p \mathbf{y_{t\text{-}p}}$$

Forecasting Using the Estimated VAR

- The forecast error can be decomposed into the sum of $\mathbf{e_t}$, the unexpected innovation of $\mathbf{y_t}$, and the coefficient estimation error:

$$\mathbf{y}_{\mathbf{t}} - E[\mathbf{y}_{\mathbf{t}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] = \mathbf{e}_{\mathbf{t}} + \nu(\mathbf{Y}_{\mathbf{t\text{-}1}})$$

- If the estimator for the coefficients is consistent and estimates are based on many data observations, the coefficient estimation error tends to be small, and:

$$\mathbf{y}_{\mathbf{t}} - E[\mathbf{y}_{\mathbf{t}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] \cong \mathbf{e}_{\mathbf{t}}$$

Iterated Forecasts

- Iterating one period forward:

$$E[\mathbf{y}_{\mathbf{t+1}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] = \mathbf{G}_0 + \mathbf{G}_1 \mathbf{E}[\mathbf{y}_{\mathbf{t}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] + \mathbf{G}_2 \mathbf{y}_{\mathbf{t\text{-}1}} + \ldots + \mathbf{G}_p \mathbf{y}_{\mathbf{t\text{-}p+1}}$$

- Iterating *j* periods forward:

$$E[\mathbf{y}_{\mathbf{t+j}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] = \mathbf{G}_0 + \mathbf{G}_1 \mathbf{E}[\mathbf{y}_{\mathbf{t+j\text{-}1}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] + \mathbf{G}_2 \mathbf{E}[\mathbf{y}_{\mathbf{t+j\text{-}2}} \mid \mathbf{Y}_{\mathbf{t\text{-}1}}] + \ldots + \mathbf{G}_p \mathbf{y}_{\mathbf{t\text{-}p+j}}$$