Structural Macroeconometrics

Chapter I
Vector Autoregressions (VARs)

Johannes Pfeifer

Kobe University

Summer 2018

Outline

Vector Autoregressive (VAR) Model

- Development of VARs as a modelling tool in the early 1980s

  $\rightarrow$ concerns about validity of some of the assumptions used in traditional **systems of equations** (SOE) macroeconometric models

- Sims (1980): Identifying restrictions in traditional models "incredible"

  $\rightarrow$ often based on partial-equilibrium analyses that do not hold in a general-equilibrium framework

  $\rightarrow$ models likely to be under-identified

- **VAR**: dynamic systems of equations in which all variables are endogenous

  $\rightarrow$ current level of each variable depends on past movements in variable and in all other variables

  $\rightarrow$ few assumptions about the underlying structure of the economy

VAR model

- Time series vector with K observables:

$$y_t = \begin{bmatrix} y_{1t}, & y_{2t}, & \dots, & y_{Kt} \end{bmatrix}'$$

- Vector autoregressive (VAR) model of order p:

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t$$

- Assumptions:

$$E(u_t) = 0,\ E(u_t u_s) = 0 \quad \forall s \neq t$$

$$E(u_t u_t') = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & & & \vdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{bmatrix}$$

where σ_{ij} denotes the (co)variance between the error terms in equation i and j

Companion Form

- Write VAR(p) as VAR(1) in **companion form**:

$$Y_t = \boldsymbol{\nu} + \mathbf{A} Y_{t-1} + U_t$$

- where

$$Y_t \equiv (y_t', y_{t-1}' \ldots, y_{t-p+1}')' \qquad (Kp \times 1)$$

$$\boldsymbol{\nu} \equiv (\nu', 0, \ldots, 0)' \qquad (Kp \times 1)$$

$$\mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix} \qquad (Kp \times Kp)$$

$$U \equiv (u_t', 0, \ldots, 0)' \qquad (Kp \times 1)$$

Stability

- A VAR(1) is **stable** if all eigenvalues of A_1 have modulus less than 1
  $\rightarrow$ equivalent to $\det(I_K - A_1 z) \neq 0$ for $|z| \leq 1$
- Just showed that any VAR(p) can be written as VAR(1)
  $\rightarrow$ VAR(p) stable if $\det(I_K - \mathbf{A}z) \neq 0$ for $|z| \leq 1$
- Can show that

$$\det(I_K - \mathbf{A}z) = \det(I_K - A_1 z - \cdots - A_p z^p)$$

- Stability condition for VAR(p)

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } |z| \leq 1 \tag{1}$$

Moving Average Representation

- VAR model using lag operator L with $L^i y_t = y_{t-i}$:

$$\begin{aligned} y_t &= \nu + (A_1 L + \cdots + A_p L^p) y_t + u_t \\ A(L) y_t &= \nu + u_t \end{aligned} \tag{2}$$

  where the lag polynomial $A(L) \equiv I_K - A_1 L - \cdots - A_p L^p$
- Let

$$\Phi(L) \equiv \sum_{i=0}^{\infty} \Phi_i L^i \qquad \text{such that} \qquad \Phi(L) A(L) = I_K$$

- Pre-multiplying (2)

$$\begin{aligned} y_t &= \Phi(L)\nu + \Phi(L) u_t \\ &= \underbrace{\left(\sum_{i=0}^{\infty} \Phi_i\right)\nu}_{\equiv \mu} + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \end{aligned}$$

Moving Average Representation

- $\Phi(L)$ often denoted as $A(L)^{-1} \rightarrow$ *inverse* of $A(L)$
- $A(L)$ invertible if $|A(z)| \neq 0$ for $|z| \leq 1$, which is the stability condition (1)
- Wold Decomposition Theorem: stationary process has an infinite order moving average representation
- Φ_i can be computed recursively

$$\Phi_0 = I_K$$

$$\Phi_i = \sum_{j=0}^{i} \Phi_{i-j} A_j \quad i = 1, 2, \ldots$$

where $A_j = 0$ for $j > p$ (see Lütkepohl 2005, Eq. 2.1.22)

- For stable processes: $\Phi_i \rightarrow 0$ as $i \rightarrow \infty$

Impulse Response Function (IRF)

- **Question**: How does the system react to a shock?
- Shifting y_t by h periods into future

$$y_{t+h} = \mu + u_{t+h} + \Phi_1 u_{t+h-1} + \ldots + \Phi_h u_t + \Phi_{h+1} u_{t-1} + \ldots$$

- Hence

$$\frac{\partial y_{t+h}}{\partial u_t'} = \Phi_h$$

- Reaction of the j-th element of y_{t+h} to a unit change in k-th element of u_t

$$\frac{\partial y_{j,t+h}}{\partial u_{kt}} = \phi_{jk,h}$$

  where ϕ is respective element of Φ
- Accumulated effect/**Long-run effect**: $\Psi_\infty \equiv \sum_{i=0}^{\infty} \Phi_i$, calculated as

$$\Psi_\infty = \Phi(1) = (I_K - A_1 - \cdots - A_p)^{-1}$$

- Long-run effect mostly used in the context of variables in first differences

Impulse Response Function (IRF)

- Forecast errors (“innovations” or **reduced form shocks**)

$$u_t = y_t - E(y_t | y_{t-1}, y_{t-2}, \ldots)$$

- **Problem:** contemporaneous correlation between reduced form shocks across equations

  $\rightarrow$ unrealistic to assume change in one element of u_t without a change in the other elements

- **Way to go:** need orthogonal **structural shocks**

Structural (identified) VAR models - "A-Model"

- **Idea:** Model contemporaneous relations between variables

$$y_{1t} = \nu_1 - a_{10} y_{2t} + a_{11} y_{1,t-1} + a_{12} y_{2,t-1} + \varepsilon_{1t}$$
$$y_{2t} = \nu_2 - a_{20} y_{1t} + a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + \varepsilon_{2t}$$

- Error terms (**structural shocks**) ε_{1t} and ε_{2t} are uncorrelated
- **Problem:** Variables y_{1t} and y_{2t} are endogenous
- In matrix form

$$\begin{bmatrix} 1 & a_{10} \\ a_{20} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- More compactly:

$$A y_t = \nu^* + A_1^* y_{t-1} + \varepsilon_t \tag{3}$$

Structural (identified) VAR models - "A-Model"

- Premultiplication of (3) by A^{-1} allows us to obtain reduced form VAR(1):

$$\underbrace{A^{-1}A}_{=I_2} y_t = \underbrace{A^{-1}\nu^*}_{\equiv\nu} + \underbrace{A^{-1}A_1^*}_{\equiv A_1} y_{t-1} + \underbrace{A^{-1}\varepsilon_t}_{\equiv u_t}$$

$$y_t = \nu + A_1 y_{t-1} + u_t$$

- Reduced form shocks are a linear combination of structural shocks, determined by A^{-1}
- If contemporaneous relation between endogenous variables is known, linear combination is also known $\Rightarrow$ VAR identified

Structural (identified) VAR models - "B-Model"

- **Idea:** forecast errors u_t as linear combinations of structural shocks ε_t:

$$u_t = B\varepsilon_t$$

- Assuming $\varepsilon_t \sim (0, I_K)$ gives $\Sigma_u = BI_KB'$
- How to identify the K^2 elements of B?
- **Problem:** Symmetry of Σ_u so that $\Sigma_u = BB'$ only specifies $K(K+1)/2$ different elements
- To identify K^2 elements of B, need $K(K-1)/2$ further relations

Structural (identified) VAR models - "B-Model"

- One way is choosing B to be lower-triangular

  $\rightarrow$ **recursive scheme**

$$
\begin{aligned}
u_{1t} &= b_{11}\varepsilon_{1t} \\
u_{2t} &= b_{21}\varepsilon_{1t} + b_{22}\varepsilon_{2t} \\
&\vdots \\
u_{Kt} &= b_{K1}\varepsilon_{1t} + b_{K2}\varepsilon_{2t} + \ldots + b_{KK}\varepsilon_{Kt}
\end{aligned}
$$

- **Choleski-decomposition** of Σ_u solves identification problem
- But: doesn't say anything about economic sense of identification scheme!
- Any other scheme that yields $K(K-1)/2$ restrictions is fine

Structural (identified) VAR models - Other schemes

- **Long-run restrictions**: restrictions on the long-run effects of economic shock (Blanchard and Quah 1989), i.e. restrictions on

$$\Psi_\infty = \Phi(1) = (I_K - A_1 - \cdots - A_p)^{-1} B$$

- **Sign restrictions**: impose sign restrictions on the responses of a subset of the endogenous variables to a particular structural shock (see e.g. Mountford and Uhlig 2009)

Structural IRFs

- Consider again MA representation

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

- Decompose Σ_u as $\Sigma_u = PP'$ using identification scheme (e.g. Cholesky in which case P is lower triangular)

- Then

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i P P^{-1} u_{t-i} = \mu + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i}$$

$$\begin{aligned} \Theta_i &= \Phi_i P \\ \varepsilon_t &= P^{-1} u_t \\ \Sigma_\varepsilon &= P^{-1} \Sigma_u \left(P^{-1}\right)' = I_k \end{aligned}$$

Structural IRFs

- Shifting y_t by h periods into future

$$\begin{aligned} y_{t+h} &= \mu + u_{t+h} + \Phi_1 u_{t+h-1} + \ldots + \Phi_h u_t + \Phi_{h+1} u_{t-1} + \ldots \\ &= \mu + P\varepsilon_{t+h} + \Phi_1 P\varepsilon_{t+h-1} + \ldots + \Phi_h P\varepsilon_t + \Phi_{h+1} P\varepsilon_{t-1} + \ldots \\ &= \mu + \Theta_0\varepsilon_{t+h} + \Theta_1\varepsilon_{t+h-1} + \ldots + \Theta_h\varepsilon_t + \Theta_{h+1}\varepsilon_{t-1} + \ldots \end{aligned}$$

- Hence

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t'} = \Theta_h$$

- Reaction of the j-th element of y_{t+h} to a unit change in k-th element of ε_t

$$\frac{\partial y_{j,t+h}}{\partial \varepsilon_{kt}} = \theta_{jk,h}$$

where θ denotes respective element of Θ

Estimating the VAR model

- Each equation of the VAR is, by itself, a classical regression

  $\rightarrow$ OLS is a consistent (and efficient?) estimator

- Equations linked only by their disturbances

  $\rightarrow$ **Seemingly Unrelated Regression** (SUR) model

- Efficiency gain by using Generalized Least Squares (GLS)

- Zellner (1962): If equations have **identical explanatory variables**

  $\rightarrow$ OLS identical to GLS (see also Greene 2011, chap. 10)

- Unrestricted VAR is SUR model with identical regressors

Convenient Notation

$$
\begin{aligned}
Y &\equiv (y_1, \ldots, y_T) && (K \times T) \\
B &\equiv (\nu, A_1, \ldots, A_p) && (K \times (Kp+1)) \\
Z_t &\equiv \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && ((Kp+1) \times 1) \\
Z &\equiv (Z_0, \ldots, Z_{T-1}) && ((Kp+1) \times T) \\
U &\equiv (u_1, \ldots, u_T) && (K \times T)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y} &\equiv vec(Y) && (KT \times 1) \\
\boldsymbol{\beta} &\equiv vec(B) && ((K^2p + K) \times 1)
\end{aligned}
$$

Multivariate LS Estimation

- Multivariate LS estimator given by

$$\widehat{\boldsymbol{\beta}} = [(ZZ')^{-1}Z \otimes I_K]\mathbf{y}$$

(see Lütkepohl 2005, Eq. 3.2.7)

- Covariance matrix estimated from residuals:

$$\begin{aligned}\widetilde{\Sigma}_u &= \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t' \\ &= \frac{1}{T}\widehat{U}\widehat{U}' \\ &= \frac{1}{T}(Y-\widehat{B}Z)(Y-\widehat{B}Z)',\end{aligned}$$

where $vec(\widehat{B}) = \widehat{\boldsymbol{\beta}}$

- Small-sample correction

$$\widehat{\Sigma}_u = \frac{T}{T-Kp-1}\widetilde{\Sigma}_u$$

Principle of maximum likelihood - step-by-step

- Start with (assumed) distribution of the data (e.g. y or y given x)
- Determine likelihood of observing given sample as function of unknown parameters
- Choose as ML estimates values of parameters that give highest likelihood

Example: normal linear regression model

- Simple regression model with assumptions A1-A4 that assure estimator is BLUE

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{4}$$

- Need distributional assumption

  $\rightarrow$ ε_i normally and independently distributed: $\varepsilon_i \sim NID(0, \sigma^2)$

- y_i has continuous distribution

  $\rightarrow$ probability of observing particular outcome is zero

- Density function at observed point y_i

$$f(y_i|x_i; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \frac{(\overbrace{y_i - \beta_1 - \beta_2 x_i}^{y_i - E(y_i|x_i)})^2}{\sigma^2} \right\} \tag{5}$$

$\rightarrow$ contribution of observation i to likelihood function

Example: normal linear regression model

- Due to independence assumption, joint density of $y_1, \ldots, y_N$ conditional on $X = (x_1, \ldots, x_N)'$ product of the individual densities

$$f(y_1, \ldots, y_N | X; \beta, \sigma^2) = \prod_{i=1}^{N} f(y_i | x_i; \beta, \sigma^2)$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \prod_{i=1}^{N} \exp\left\{ -\frac{1}{2} \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2} \right\} \tag{6}$$

- Loglikelihood as function of unknown parameters

$$\log L(\beta, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2} \tag{7}$$

$\rightarrow$ first term does not depend on β

$\rightarrow$ max. w.r.t. β's corresponds to minimizing residual sum of squares

$\rightarrow$ β_1 and β_2 identical to OLS estimator

Example: normal linear regression model

- To obtain estimator for σ^2, plug $\hat{\beta}$'s in loglikelihood

$$\log L(\hat{\beta}, \sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(\overbrace{y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i}^{e_i})^2}{\sigma^2} \tag{8}$$

- Maximizing w.r.t. σ^2 yields first-order condition

$$-\frac{N}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{e_i^2}{\sigma^4} = 0 \tag{9}$$

and, hence, ML estimator for σ^2

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \tag{10}$$

$\rightarrow$ consistent estimator for σ^2

$\rightarrow$ *does not* correspond to unbiased OLS estimator

Maximum Likelihood Estimation of VAR I

- Assumption:

$$\mathbf{u} = vec(U) \sim \mathcal{N}\left(0, I_T \otimes \Sigma_u\right)$$

- Probability density of $\mathbf{u}$ hence

$$f_u(\mathbf{u}) = \frac{1}{(2\pi)^{KT/2}} \left|I_T \otimes \Sigma_u\right|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{u}'\left(I_T \otimes \Sigma_u^{-1}\right)\mathbf{u}\right]$$

- Requires model to be not **stochastically singular**, i.e. forecast error variance matrix Σ_u must have full rank
- Log-likelihood

$$\begin{aligned}\ln \mathcal{L}(\boldsymbol{\beta}, \Sigma_u) = &-\frac{KT}{2}\ln 2\pi - \frac{T}{2}\log\left(|\Sigma_u|\right) \\ &-\frac{1}{2}\left[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}\right]'\left(I_T \otimes \Sigma_u^{-1}\right)\left[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}\right]\end{aligned}$$

Maximum Likelihood Estimation of VAR II

- Log-likelihood

$$\begin{aligned}\ln \mathcal{L}(\boldsymbol{\beta}, \Sigma_u) = &- \frac{KT}{2} \ln 2\pi - \frac{T}{2} \log\left(|\Sigma_u|\right) \\ &- \frac{1}{2}\left[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}\right]' \left(I_T \otimes \Sigma_u^{-1}\right)\left[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}\right]\end{aligned}$$

- We are searching for estimates of $\boldsymbol{\beta}, \Sigma_u$ that maximize this likelihood function, i.e. look for point where partial derivatives are 0
- Can be done in the computer using Newton-type optimizers by optimizing over unique elements of $\boldsymbol{\beta}, \Sigma_u$
- Note: Σ_u only has $N(N+1)/2$ independent entries
- We are looking for global maximum, but many optimizers can get stuck at local maxima

Application: Dynamic Effects of Gov. Spending Shocks

- Blanchard and Perotti (2002), "*An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output*" (QJE)
- Baseline: Trivariate VAR

$$Y_t = A(L)Y_{t-1} + U_t$$

- Vector of endogenous variables $Y_t \equiv [T_t, G_t, X_t]$: log of real, per-capita taxes, spending, and GDP (later add consumption /investment)
- Vector of corresponding reduced-form residuals $U_t \equiv [t_t, g_t, x_t]$

Identification

- Residual structure (without loss of generality)

$$t_t = a_1 x_t + a_2 e_t^g + e_t^t$$
$$g_t = b_1 x_t + b_2 e_t^t + e_t^g$$
$$x_t = c_1 t_t + c_2 g_t + e_t^x$$

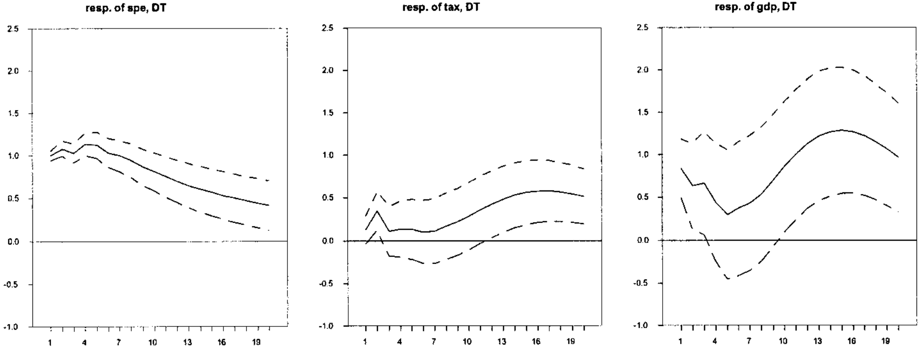
  where e's are structural shocks to recover

- a_1 and b_1 elasticities that capture automatic responses and can be constructed from institutional information
- For U.S. data:

$$b_1 = 0 \tag{11}$$

- Crucial assumption: within quarter, government does not discretionarily react to changes in conditions (as opposed to automatic response captured by b_1)
- a_2 and b_2: do taxes or spending come first?
  $\rightarrow$ test robustness by setting both in turn to zero (doesn't matter empirically)

Let's do the dirty work ourselves. . .



- This is the goal: Figure 5 of Blanchard and Perotti (2002)

Bibliography I

Blanchard, Olivier J. and Roberto Perotti (2002). "An empirical characterization of the dynamic effects of changes in government spending and taxes on output". *Quarterly Journal of Economics 117* (4), 1329–1368.

Blanchard, Olivier J. and Danny Quah (1989). "The dynamic effects of aggregate demand and supply disturbances". *American Economic Review 79* (4), 655–73.

Greene, William H. (2011). *Econometric analysis*. 7th ed. Pearson.

Lütkepohl, Helmut (2005). *New introduction to multiple time series analysis*. Berlin: Springer.

Mountford, Andrew and Harald Uhlig (2009). "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics 24* (6), 960–992.

Sims, Christopher A. (1980). "Macroeconomics and reality". *Econometrica 48* (1), 1–48.

Bibliography II

Zellner, Arnold (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias". *Journal of the American Statistical Association 57* (298), 348–368.