The Econometrics of DSGE Models

Giuseppe Ragusa
Luiss University

EIEF
Lecture 6: (Bayesian) VAR

March 15, 2017

Vector Autoregressions

- $VAR(p)$

$$y_t = \underset{n\times 1}{C} + \underset{n\times n}{B_1} y_{t-1} + \ldots + B_p y_{t-p} + \underset{n\times 1}{u_t}$$

- Flexible multivariate model
- Bridge between reduced-form and structural models

Vector Autoregressions

Notation (I)

Rewrite the VAR as

$$\underset{(1\times n)}{y_t} = \underset{(1\times q)}{z_t}\underset{(q\times n)}{\Gamma} + u_t$$

where $q = np + 1$ and

$$z_t = (1, y'_{t-1}, \ldots, y'_{t-p})$$
$$\Gamma = (C'\, B'_1 \ldots B'_p)$$

Stacking along the time dimension we can write

$$\underset{(T\times n)}{Y} = \underset{(T\times q)}{Z}\Gamma + \underset{(T\times n)}{U}$$

where

$$Y \equiv (y_1\, y_2 \ldots y_T) \quad U \equiv (u_1\, u_2 \ldots u_T)$$

Vector Autoregressions

Notation (II)

We can also write the VAR as

$$
\begin{aligned}
vec(Y) &= vec(Z\Gamma + U) \\
&= vec(Z\Gamma) + vec(U) \\
&= (I_n \otimes Z)vec(\Gamma) + vec(U)
\end{aligned}
$$

$$\boxed{\underset{(nT\times 1)}{y} = \underset{(nT\times qn)}{(I_n \otimes Z)} \underset{(qn\times 1)}{\beta} + u}$$

wh

$$y = vec(Y), \beta \equiv \mathrm{vec}(\Gamma), u = vec(U).$$

Vector Autoregressions

Notation IV

Let

$$\xi_t = \begin{bmatrix} y_t - C \\ y_{t-1} - C \\ \vdots \\ \vdots \\ y_{t-p+1} - C \end{bmatrix}, F = \begin{bmatrix} B_1 & B_2 & B_3 & \cdots & B_{p-1} & B_p \\ I_n & 0 & 0 & \cdots & \cdots & 0 \\ 0 & I_n & 0 & \cdots & 0 & 0 \\ & & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_n & 0 \end{bmatrix}, v_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.$$

Then, the VAR(p) can be written as VAR(1) in ξ_t

$$\xi_t = F\xi_{t-1} + v_t$$

Assumptions

- Variance

$$\begin{aligned} E[u_t u_t'] &= \Sigma \\ E[uu'] &= \Sigma \otimes I_T \\ E[uu'|z] &= \Sigma \otimes I_T \end{aligned}$$

- Distribution

$$u \sim N(0, \Sigma \otimes I_T)$$

  or, using the matrix notation

$$U \sim MN(0, I_T, \Sigma)$$

  where $MN(M, R, C)$ denote the **matric-variate** normal distribution with mean M, row-wise variance R and column wise variance C.

Covariance Stationarity - MA(∞)

From $\xi_t = F\xi_{t-1} + u_t$, we have

$$\xi_{t+s} = v_{t+s} + Fv_{t+s-1} + F^2 v_{t+s-2} + \ldots + F^{s-1} v_{t+1} + F^s \xi_t.$$

Definition

A VAR(p) is covariance stationary if the eigenvalues of the matrix F satisfies

$$|I_n \lambda^p - B_1 \lambda^{p-1} - B_2 \lambda^{p-2} - \cdots - B_p| = 0$$

In this case, $F^s \to 0$, as $s \to \infty$. Thus,

$$y_t = C + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \cdots = C + \Psi(L) u_t$$

where

$$\Psi_j = F_{11}^{(j)}$$

and

$$\Psi(L) = I_n - \Psi_1 L - \Psi_2 L^2 - \ldots$$

Impulse-Response Function

For

$$y_t = C + u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \cdots$$

the matrix Ψ_s has the interpretation

$$\frac{\partial y_{t+s}}{\partial u_t'} = \Psi_s.$$

Thus, the row i, column j element of Ψ_s identifies the consequences of a one unit increase in the jth variable's innovation at date t (u_{jt}) for the value of the ith variable at time $t+s$ ($y_{i,t+s}$)

$$\frac{\partial y_{i,t+s}}{\partial u_{jt}} = \Psi_{ij}.$$

Impulse-response Function

A plot of

$$\frac{\partial y_{i,t+s}}{\partial u_{jt}} = \Psi_{ij},$$

as a function of s is called the *impulse-response function.* It describes the response of $y_{j,t+s}$ to a one-time impulse in y_{jt} with all the other variables dated t or earlier held constant.

- Is there a sense in which this multiplier can be viewed as measuring the casual effect of y_j on y_i? (Not really!)

Classical inference

Since

$$F_{11}^{(j)} = f(C, B_1, \ldots, B_p),$$

we need to estimate $\Gamma = (C', B_1', \ldots, B_p')$

$$Y = Z\Gamma + U$$

- We can estimate $\beta = vec(B)$ by OLS

$$\begin{aligned}
\hat{\Gamma} &= (Z'Z)^{-1}Z'Y, \\
\hat{\beta} &= \mathrm{vec}(\hat{\Gamma}) \\
\sqrt{T}(\hat{\beta} - \beta) &\xrightarrow{d} N(0, \Omega), \\
\Omega &= \Sigma \otimes E(z_t' z_t)^{-1}
\end{aligned}$$

- Since each equations has the same regressors

$$OLS = SUR = MLE$$

Bayesian inference

Bayesian inference will be based on the posterior distribution

$$p(\beta, \Sigma | y) \propto \underbrace{p(y|\beta, \Sigma)}_{\text{likelihood}} \underbrace{p(\beta|\Sigma) p(\Sigma)}_{\text{prior}}$$

- What is the likelihood?
- What are appropriate priors?

Bayesian inference

Likelihood function

Condition on the initial p observations:

$$p(Y|\beta,\Sigma) = \prod_{t=1}^{T} p(y_t|y^{t-1},\beta,\Sigma)$$

$$y_t|y^{t-1},\beta,\Sigma \sim N(z_t\Gamma,\Sigma)$$

Thus,

$$\begin{aligned} p(Y|\beta,\Sigma) &\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - z_t\Gamma)'\Sigma^{-1}(y_t - z_t\Gamma)\right\} \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y-(I\otimes Z)\beta)'\{\Sigma\otimes I_T\}^{-1}(y-(I\otimes Z)\beta)\right\} \end{aligned}$$

Bayesian inference

Posterior with flat prior

Flat prior on β and Σ:

- The posterior distribution is Normal-Inverse Wishart

$$\beta|\Sigma, Y \sim N\left(\hat{\beta}, \Sigma \otimes (Z'Z)^{-1}\right)$$

$$\Sigma|Y \sim IW(\hat{S}, T-n-k-1)$$

where

$$\hat{S} = (y-(I_T \otimes Z)\hat{\beta})'(y-(I_T \otimes Z)\beta)$$

Inverse Wishart distribution

- Inverse Wishart distribution is a probability distribution defined on real-valued positive-definite matrices.
- If $\underset{n\times n}{\Sigma} \sim IW(S, \nu)$, then density of Σ is given by

$$p(\Sigma) = \frac{|S|^{\frac{\nu}{2}}}{2^{\frac{\nu n}{2}}\Gamma_n\left(\frac{\nu}{2}\right)}|\Sigma|^{-\frac{\nu+n+1}{2}}\exp\left\{-\frac{1}{2}\mathrm{Tr}(S\Sigma^{-1})\right\},$$

  where $\Gamma_n(\cdot)$ is the n-variate gamma function
- The moments are

$$\begin{aligned} E[\Sigma] &= \frac{S}{\nu-n-1}, \quad \nu > n+1 \\ \mathrm{mode}(\Sigma) &= \frac{S}{\nu+n+1}. \end{aligned}$$

- To simulate from $IW(S, \nu)$
  - simulate n times from $N(0, S^{-1}) \Longrightarrow \underset{n\times\nu}{\upsilon} \Longrightarrow (\upsilon'\upsilon)^{-1} \sim IW(S, \nu)$

Bayesian inference

Posterior with flat prior

Flat prior on β and Σ:

- The posterior distribution is Normal-Inverse Wishart

$$\beta|\Sigma, Y \sim N\left(\hat{\beta}, \Sigma \otimes (Z'Z)^{-1}\right)$$
$$\Sigma|Y \sim IW(\hat{S}, T-n-k-1)$$

The posterior mode of the parameters are given by

$$\mathrm{mode}(\beta) = \hat{\beta}$$
$$\mathrm{mode}(\Sigma) = \frac{\hat{S}}{T-k}$$

Bayesian inference

Posterior with conjugate priors

Assume a N-IW prior:

$$\beta|\Sigma \sim N(\gamma_0, \Sigma \otimes \Omega_0), \quad \gamma_0 = \text{vec}(\Gamma_0)$$
$$\Sigma \sim IW(\Psi_0, n)$$

N-IW prior on β and Σ:

- The posterior distribution is Normal-Inverse Wishart

$$\beta|\Sigma, Y \sim N\left(\tilde{\beta}, \Sigma \otimes (Z'Z + \Omega_0^{-1})^{-1}\right)$$
$$\Sigma|Y \sim IW(\hat{S}, T + n)$$

where

$$\tilde{\Gamma} = \left(Z'Z + \Omega_0^{-1}\right)^{-1}(Z'y + \Omega_0^{-1}\gamma_0), \quad \tilde{\beta} = \text{vec}(\tilde{\Gamma})$$
$$\hat{S} = \Psi_0 + \hat{u}'\hat{u} + (\tilde{\Gamma} - \Gamma_0)'\Omega_0^{-1}(\tilde{\Gamma} - \Gamma_0)$$

Summary

Flat priors

$$\beta|\Sigma, Y \sim N\left(\hat{\beta}, \Sigma \otimes (Z'Z)^{-1}\right)$$

$$\Sigma|Y \sim IW(\hat{S}, T-n-k-1)$$

Conjugate priors

$$\beta|\Sigma, Y \sim N\left(\tilde{\beta}, \Sigma \otimes (Z'Z + \Omega_0^{-1})^{-1}\right)$$

$$\Sigma|Y \sim IW(\hat{S}, T+n)$$

Informative priors

The case for informative priors

- VAR models are densely parameterized
- Classical methods of Bayesian inference with flat priors imply
  - High estimation uncertainty
  - Overfitting
  - Poor out-of-sample forecasting performance

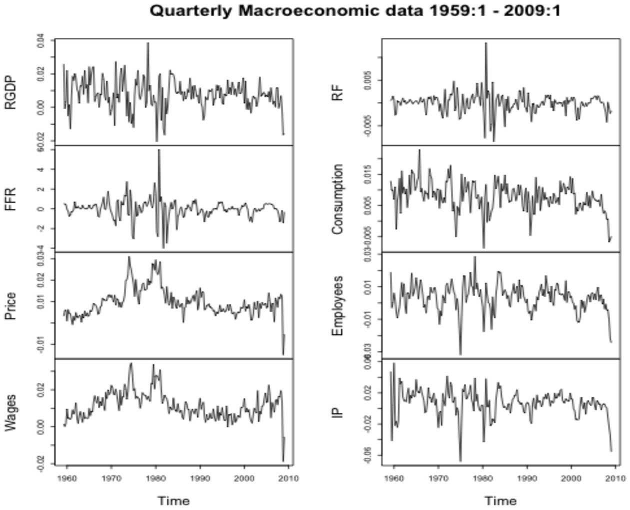
Flat prior VARs

An example

- Quarterly macroeconomic data fro the US (1959:1 to 2009:1)
  - Real GDP
  - Consumption
  - Employees
  - Wages
  - Federal Funds Rate
  - US-Tbill rate 1 month
  - Industrial production
- $p = 5$
- Total number of parameters: 180

Flat prior VARs

Macroeconomic data



Flat prior VARs

Fitted values

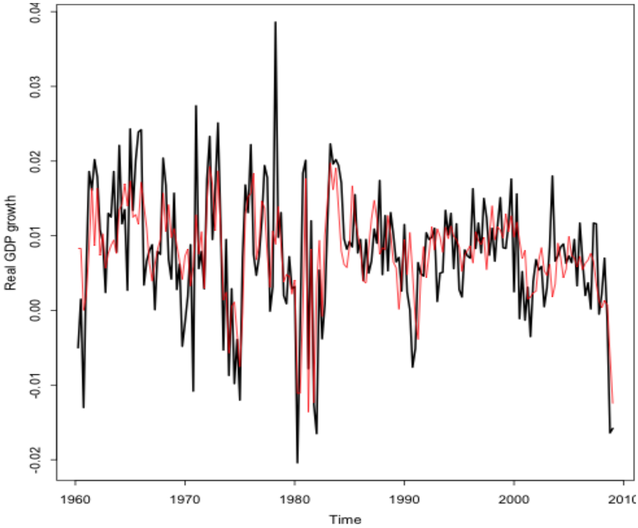
Calculate:

$$y_t = z_t \hat{\Gamma}$$

for $t = 1, \ldots, T$

Flat prior VARs

Fitted values



Flat prior VARs

Out-of-sample

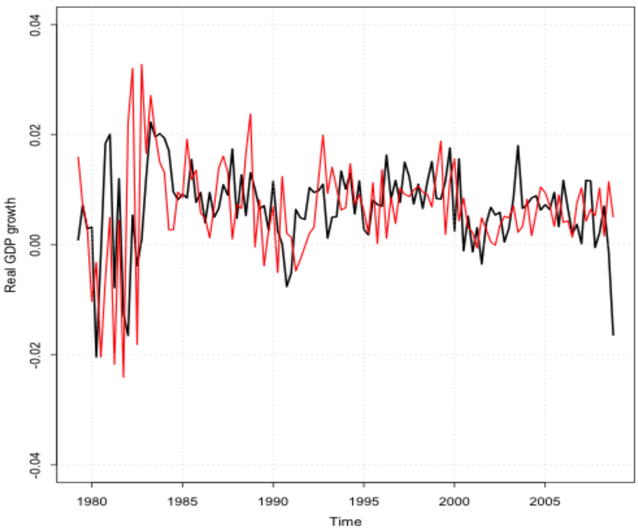
Estimate

$$\begin{aligned}
\hat{y}_{T+1} &= z_T \hat{\Gamma} \\
\hat{y}_{T+2} &= (1, \hat{y}'_{T+1}, y'_T, \ldots, y_{T-p-1})' \hat{\Gamma} \\
\vdots &= \vdots \\
\hat{y}_{T+h} &= (1, \hat{y}'_{T+1}, \hat{y}'_T, \ldots, \hat{y}_{T+h-1}, \ldots, y_{T-p-h-1})' \hat{\Gamma}
\end{aligned}$$

- To evaluate the out-of-sample forecast we use either a rolling window or a recursive scheme

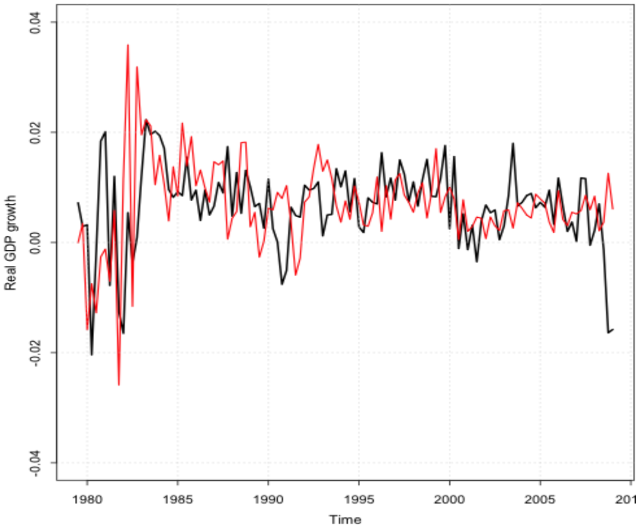
Flat prior VARs

Out-of-sample - 1-step ahead



Flat prior VARs

Out-of-sample - 2-step-ahead



"Minnesota" prior

- $VAR(p)$

$$y_t = \underset{n\times 1}{C} + \underset{n\times n}{B_1} y_{t-1} + \ldots + B_p y_{t-p} + \underset{n\times 1}{u_t}$$

- "Shrink" VAR coefficients towards naive benchmark models
  - stationary variables

$$y_t = C + u_t$$

  - non-stationary variables

$$y_t = C + y_{t-1} + u_t$$

"Minnesota" prior

VAR(p)

$$
\begin{aligned}
y_t &= C + B_1 y_{t-1} + \ldots + B_p y_{t-p} + u_t \\
u_t &\sim N(0, \Sigma)
\end{aligned}
$$

- Shrink VAR coefficients toward naive model

$$y_t = C + y_{t-1} + u_t$$

- Do so using a conjugate prior

$$\beta | \Sigma \sim N(b, \Sigma \otimes \Omega_0)$$

- Specifically

$$[B_k]_{ij} | \Sigma \sim N\left(b_{k,ij}, v_{k,ij}^2\right),$$

$$b_{k,ij} = \begin{cases} 1 & \text{if } k = 1 \text{ and j=i} \\ 0 & \text{otherwise.} \end{cases}, \quad v_{k,ij} = \begin{cases} \frac{\lambda}{k^2} & i = j \\ \frac{\lambda^2}{k^2} \frac{\sigma_i}{\hat{\sigma}_j} & i \neq j \end{cases}$$

where

- λ : controls overall tightness
- $B_1, \ldots, B_p$ are a priori independent and normally distributed
- the prior on the intercept is diffuse $C \sim N(0, \varsigma^2)$, $\varsigma \to \infty$

“Minnesota” priors

- Key hyperparameter: λ, which controls the informativeness of this prior
- Minnesota priors substantially improves forecasting performance of the model
  - Litterman (1980) and Doan, Litterman, and Sims (1984)
- Even large scale VARs do well
  - Banbura, Giannone, and Reichlin (2010)
- N-IW version: Minnesota prior can be coupled with

$$\Sigma \sim IW(\Psi_0, n)$$

- Other priors can be super-imposed; e.g., inexact differencing

$$\Pi = I_n - B_1 - B_2 - \ldots - B_p$$

```
 [1] "Real GDP, quantity index (2000 = 100)"
 [2] "Interest rate: federal funds (pct per annum)"
 [3] "Consumption Price Index Non Durable and Service"
 [4] "Interest rate: US T-bills, sec mkt, 3-month"
 [5] "Real Consumption Non Durable and Service"
 [6] "S&P 500 Composite Stock Index"
 [7] "Consumer Price Index All Items"
 [8] "Employees, nonfarm: total private"
 [9] "Real spot market price index: all commodities"
[10] "Depository inst reserves: nonborrowed (mil USD)"
[11] "Depository inst reserves: total (mil USD)"
[12] "Money stock: M2 (bil USD)"
[13] "Industrial production index: total"
[14] "Capacity utilization: manufacturing (SIC)"
[15] "Unemp. rate: All workers, 16 and over (%)"
[16] "Housing starts: Total (thousands)"
[17] "Real avg hrly earnings, non-farm prod. workers"
[18] "Money stock: M1 (bil USD)"
[19] "Interest rate: US treasury const. mat., 5-yr"
[20] "Interest rate: US treasury const. mat., 10-yr"
[21] "US effective exchange rate: index number"
```

BVAR Application

The objective is to forecast

- Real GDP (RGDP)
- Federal Funds Rate (FFR)
- Consumer Price Index All Items (CPI)
- We use a VAR(5)

$$y_t = C + B_1 y_{t-1} + B_2 y_{t-2} + \ldots + B_p y_{t-p} + u_t$$

  with $p = 5$
- There are a total of $(21 \times 5 + 1) \times 21 = \mathbf{2226}$
- OLS is infeasible

BVAR Application

j	Estimate	Evaluate
1	$1959:Q1-1978:Q4$	$1979:Q1-1980:Q4$
2	$1959:Q2-1979:Q1$	$1979:Q2-1981:Q1$
3	$1959:Q3-1979:Q2$	$1979:Q3-1981:Q2$
$\vdots$	$\vdots$	
114	$1987:Q2-2007:Q1$	$2007:Q2-2009:Q1$

- For each variable report, the out-of-sample forecast

$$
\begin{aligned}
\bar{y}^j_{T+1} &= \bar{C}^j + \bar{B}^j_1 y_T + \bar{B}^j_2 y_{T-1} + \ldots + \bar{B}^j_5 y_{T-5} \\
\bar{y}^j_{T+2} &= \bar{C}^j + \bar{B}^j_1 \bar{y}_{T+1} + \bar{B}^j_2 y_T + \ldots + \bar{B}^j_5 y_{T-4} \\
\bar{y}^j_{T+3} &= \bar{C}^j + \bar{B}^j_1 \bar{y}_{T+2} + \bar{B}^j_2 \bar{y}_{T+1} + \ldots + \bar{B}^j_5 y_{T-3} \\
\bar{y}^j_{T+4} &= \bar{C}^j + \bar{B}^j_1 \bar{y}_{T+3} + \bar{B}^j_2 \bar{y}_{T+2} + \ldots + \bar{B}^j_5 y_{T-2}
\end{aligned}
$$

where $\bar{C}^j, \bar{B}^j_1, \ldots, \bar{B}^j_5$ are the posterior mean of the matrix of parameters that use the sample j.

Results

Table: Mean of out-of-sample errors. One step ahead. Random walk with drift vs. BVAR.

RGDP	FFR	CPI
Random Walk with Drift		
0.0015	0.0763	0.0017
0.0029	0.1584	0.0036
0.0043	0.2465	0.0056
0.0057	0.3530	0.0078
Bayesian BVAR		
0.0008	0.0438	0.0017
0.0015	0.1171	0.0040
0.0020	0.2119	0.0069
0.0024	0.3328	0.0103

Results

Table: Relative Mean square forecasting errors. Random walk with drift vs. BVAR.

RGDP	FFR	CPI
1.30802	1.03348	2.22025
1.43418	0.97702	1.98485
1.41020	1.00995	1.68671
1.39083	1.00949	1.39658

Structural VAR

A structural autoregressive model is

$$A_0 y_t = C_0 + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t$$

where

$$E[\varepsilon_t] = 0, \quad E[\varepsilon_t \varepsilon_t'] = I_n$$

Equivalently the model can be written compactly as

$$A(L) y_t = C_0 + \varepsilon_t$$

where

$$A(L) = A_0 - A_1 L - A_2 L^2 - \ldots$$

is the autoregressive lag order polynomial

Structural VAR

Reduced-form representation

Pre-multiply both sides of the structural VAR

$$A_0 y_t = C_0 + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t$$

by A_0^{-1}, we arrive at

$$A_0^{-1} A_0 y_t = A_0^{-1} C_0 + A_0^{-1} A_1 y_{t-1} + \ldots + A_0^{-1} A_p y_{t-p} + A_0^{-1} \varepsilon_t$$

which can be represented as

$$y_t = \underbrace{C}_{A_0^{-1} C_0} + \underbrace{B_1}_{A_0^{-1} A_1} y_{t-1} + \ldots + \underbrace{B_p}_{A_0^{-1} B_p} y_{t-p} + \underbrace{u_t}_{A_0^{-1} \varepsilon_t}$$

Structural VAR

Identification

The question is:

- Under what conditions given

$$\Gamma = (C, B_1, B_2, \ldots, B_p), \text{ and } \Sigma_u$$

and Σ_u can be consistently estimated can we recover

$$A_0, A_1, A_2, \ldots, A_p,$$

the "structural" VAR parameters?

Structural VAR

Identification

All we need is to recover the elements of A_0, since

- knowledge of A_0 would allow us to recover

$$\varepsilon_t = A_0 u_t$$

- knowledge of A_0 would allow us to recover

$$A_i = A_0 B_i, \quad i = 1, \ldots, n$$

Structural VAR

Identification

The variance of u_t is

$$E[u_t u_t'] = A_0^{-1} E[\varepsilon_t \varepsilon_t'] A_0^{-1'}$$
$$\Sigma_u = A_0^{-1} A_0^{-1'}$$

where we make use of the fact that $E[\varepsilon_t \varepsilon_t'] = I_n$.

- We can think of $\Sigma_u = A_0^{-1} A_0^{-1'}$ as a system of nonlinear equation in the unknown parameter A_0 (Σ_u can be consistently estimated and thus can be treated as known)
- This system of nonlinear equations can be solved using numerical method, provided that
    - the number of equations is equal to the number of unknown (order condition)

Structural VAR

Identification

$$\Sigma_u = A_0^{-1} A_0^{-1'}$$

- There are

  $$n(n+1)/2$$

  non-redundant equations (Σ_u is symmetric)
- There are

  $$n^2$$

  parameter to estimate
- We have to restrict at least

  $$n^2 - n(n+1)/2 = n(n-1)/2$$

  parameters

Structural VAR

Types of identification schemes

Different approaches to identify a structural VAR:

- Recursively identified models
- Long run restrictions
- Sign restrictions
- DSGE source of identifications

Structural VAR

Recursively identified models

- Popular way of identifying SVAR
- Consider the Cholesky decomposition of Σ_u

$$\Sigma_u = LL'$$

  where the matrix L is lower triangular (is has $n(n+1)/2$ non-zero elements)
- Setting

$$A_0^{-1} = L$$

  solves the identification problem
- Given lower triangularity of L there is no need to use numerical solution methods in this case
- Appropriate if "orthogonalization" can be justified on economic grounds

Structural VAR

Recursively identified models (Example)

Consider the following model

$$A_0 y_t = C_0 + A_1 y_{t-1} + \varepsilon_t$$

where

$$y_t = \begin{pmatrix} p_t \\ gdp_t \\ m_t \\ i_t \end{pmatrix}$$

- p_t is the log price level,
- gdp_t is log real GDP,
- m_t the log of a monetary aggregate such as $M1$,
- i_t the federal funds rate.

Structural VAR

Recursively identified models (Example)

Given the reduced-form

$$y_t = C + B_1 y_{t-1} + u_t$$

The proposed identification is

$$\begin{pmatrix} u_t^p \\ u_t^{gdp} \\ u_t^m \\ u_t^i \end{pmatrix} = \underbrace{\begin{bmatrix} a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \\ g & h & i & j \end{bmatrix}}_{A_0^{-1}} \begin{pmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \\ \varepsilon_t^3 \\ \varepsilon_t^4 \end{pmatrix}$$

What does it mean economically?

Structural VAR

Recursively identified models (Example)

$$\begin{bmatrix} a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \\ g & h & i & j \end{bmatrix} \begin{pmatrix} p_t \\ gdp_t \\ m_t \\ i_t \end{pmatrix} = C_0 + A_1 \begin{pmatrix} p_t \\ gdp_t \\ m_t \\ i_t \end{pmatrix} + \varepsilon_t$$

- Prices do not respond to *gdp*, *m*, *i* (horizontal AS)
- AD is downward sloping
- Money demand do not respond to interest rates

Structural VAR

Long run restrictions