

II. Vector autoregressions

A. Introduction

1. VARs as forecasting models

Forecasting:

Let y_{1t} be the first element of a vector \mathbf{y}_t

$$\hat{y}_{1|t-1} = f_t(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1)$$

Convenient assumptions:

- 1) time invariant
- 2) linear
- 3) finite parameter

$$\hat{y}_{1|t-1} = c + \phi_1' \mathbf{y}_{t-1} + \phi_2' \mathbf{y}_{t-2} + \dots + \phi_p' \mathbf{y}_{t-p}$$

Useful result:

Define the “true values”

of parameters $(c, \phi_1, \phi_2, \dots, \phi_p)$

to be the values that minimize

$$\text{plim } T^{-1} \sum_{t=1}^T (y_{1t} - \hat{y}_{1|t-1})^2$$

Then under fairly general conditions,
the true values are estimated consistently
by OLS regression of y_{1t} on a constant and
 $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{t-p}$

Estimate separate OLS regression
for each element of $\mathbf{y}_t = \text{VAR}$
One-step-ahead forecast:

$$\hat{y}_{it-1} = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p}$$

II. Vector autoregressions

- A. Introduction
 - 1. VARs as forecasting models
 - 2. Gaussian VARs as data-generating process

Consider the following process whereby the vector \mathbf{y}_t might have been generated:

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t \sim \text{i.i.d. } N(\mathbf{0}, \Omega)$$

log likelihood:

$$\mathcal{L} = \log f(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}, \boldsymbol{\theta})$$

$$= -(Tn/2) \log(2\pi) - (T/2) \log |\Omega|$$

$$- (1/2) \sum_{t=1}^T \boldsymbol{\varepsilon}_t' \Omega^{-1} \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{c} - \Phi_1 \mathbf{y}_{t-1} - \Phi_2 \mathbf{y}_{t-2} - \dots - \Phi_p \mathbf{y}_{t-p}$$

$\boldsymbol{\theta}$ = vector containing elements of

$$\mathbf{c}, \Phi_1, \Phi_2, \dots, \Phi_p, \Omega$$

Proposition:

(1) The values of $\hat{\mathbf{c}}, \hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p$ that maximize the likelihood function are identical to results from OLS estimation, equation by equation

(2) The MLE $\hat{\Omega}$ is given by

$$\hat{\Omega} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$$
$$\hat{\varepsilon}_t = \mathbf{y}_t - \hat{\mathbf{c}} - \hat{\Phi}_1 \mathbf{y}_{t-1} - \hat{\Phi}_2 \mathbf{y}_{t-2} - \dots - \hat{\Phi}_p \mathbf{y}_{t-p}$$

Gaussian VAR is nice example of where quasi-maximum likelihood estimation can be justified

II. Vector autoregressions

A. Introduction

1. VARs as forecasting models
2. Gaussian VARs as data-generating process
3. VARs as ad hoc dynamic structural models

Example:

f_t = federal funds rate

y_t = output growth

π_t = inflation

m_t = money growth rate

Represent Fed behavior by

$$f_t = \alpha_0 + \alpha_1 y_t + \alpha_2 \pi_t + \alpha_3 f_{t-1} + \alpha_4 y_{t-1} + \alpha_5 \pi_{t-1} + \alpha_6 m_{t-1} + v_t$$

Fed responds to current output and inflation but not current money growth

$$\mathbf{y}_t = (f_t, y_t, \pi_t, m_t)'$$

$$\mathbf{B}_0 \mathbf{y}_t = \mathbf{k} + \mathbf{B}_1 \mathbf{y}_{t-1} + \mathbf{v}_t$$

$$\mathbf{B}_0 = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_2 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

$$\mathbf{B}_1 = \begin{bmatrix} \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

If $\mathbf{v}_t \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{D})$, log likelihood is

$$\begin{aligned} \mathcal{L} &= \log f(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T | \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1}, \boldsymbol{\theta}) \\ &= -(Tn/2) \log(2\pi) - (T/2) \log |\mathbf{D}| \\ &\quad + T \log |\mathbf{B}_0| - (1/2) \sum_{t=1}^T \mathbf{v}_t' \mathbf{D}^{-1} \mathbf{v}_t \end{aligned}$$

$$\mathbf{D} = E(\mathbf{v}_t \mathbf{v}_t')$$

$$\mathbf{v}_t = \mathbf{B}_0 \mathbf{y}_t - \mathbf{k} - \mathbf{B}_1 \mathbf{y}_{t-1} - \mathbf{B}_2 \mathbf{y}_{t-2} - \dots - \mathbf{B}_p \mathbf{y}_{t-p}$$

$\boldsymbol{\theta}$ = vector containing elements of

$$\mathbf{k}, \mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_p, \mathbf{D}$$

If model is just-identified, the MLE's

$$\hat{\mathbf{k}}, \hat{\mathbf{B}}_0, \hat{\mathbf{B}}_1, \dots, \hat{\mathbf{B}}_p, \hat{\mathbf{D}}$$

are transformations of the VAR MLE's

$$\hat{\mathbf{c}}, \hat{\boldsymbol{\Phi}}_1, \hat{\boldsymbol{\Phi}}_2, \dots, \hat{\boldsymbol{\Phi}}_p, \hat{\boldsymbol{\Omega}}$$

II. Vector autoregressions

- A. Introduction
- B. Normal-Wishart priors for VARs

$$\mathbf{y}_t = \mathbf{\Gamma}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

$n \times 1$ $n \times k$ $k \times 1$ $n \times 1$

$$k = np + 1$$

$$\mathbf{x}_t' = (1, \mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \dots, \mathbf{y}_{t-p}')'$$

$$\boldsymbol{\varepsilon}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1 \sim N(\mathbf{0}, \boldsymbol{\Omega})$$

Likelihood function:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_T' \end{bmatrix}$$

$T \times n$

$$p(\mathbf{Y} | \boldsymbol{\Gamma}, \boldsymbol{\Omega}) = (2\pi)^{-Tn/2} |\boldsymbol{\Omega}|^{-T/2}$$

$$\exp \left[-\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\Gamma}' \mathbf{x}_t)' \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\Gamma}' \mathbf{x}_t) \right]$$

Prior for $\boldsymbol{\Omega}$:

univariate regression

$$\sigma^2 = E(\boldsymbol{\varepsilon}_t^2)$$

$$Z_i \sim N(0, \lambda^{-1})$$

$$W = (Z_1^2 + Z_2^2 + \dots + Z_N^2)$$

W has gamma distribution

with parameters N, λ

$$\sigma^{-2} \sim \Gamma(N, \lambda)$$

Vector autoregression:

$$\mathbf{\Omega} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$$

$$\mathbf{z}_t \sim N(\mathbf{0}, \mathbf{\Lambda}^{-1})$$

$n \times 1$

$$\mathbf{W} = (\mathbf{z}_1 \mathbf{z}_1' + \mathbf{z}_2 \mathbf{z}_2' + \dots + \mathbf{z}_N \mathbf{z}_N')$$

\mathbf{W} has Wishart distribution
with parameters $N, \mathbf{\Lambda}$

$$\mathbf{\Omega}^{-1} \sim W(N, \mathbf{\Lambda})$$

$$\mathbf{W} \sim W(N, \mathbf{\Lambda}) \Rightarrow$$

$$p(\mathbf{w}) = c |\mathbf{\Lambda}|^{N/2} |\mathbf{w}|^{(N-n-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\mathbf{w}\mathbf{\Lambda})\right]$$

$$c = \left[2^{Nm/2} \pi^{n(n-1)/4} \prod_{j=1}^n \Gamma\left(\frac{N+1-j}{2}\right) \right]^{-1}$$

so prior takes the form

$$p(\mathbf{\Omega}^{-1}) \propto |\mathbf{\Omega}|^{-(N-n-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\mathbf{\Omega}^{-1} \mathbf{\Lambda})\right]$$

$$\mathbf{y}_t = \mathbf{\Gamma}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

$n \times 1$ $n \times k$ $k \times 1$ $n \times 1$

$$\boldsymbol{\gamma} = \text{vec}(\mathbf{\Gamma})$$

$nk \times 1$

first k components of $\boldsymbol{\gamma}$ =
coefficients to explain y_{1t}

Prior for γ :

univariate regression

$$\beta | \sigma^{-2} \sim N(\mathbf{m}, \sigma^2 \mathbf{M})$$

vector autoregression

$$\gamma | \Omega^{-1} \sim N \left(\mathbf{m}, \begin{matrix} \Omega & & \\ \text{nk} \times 1 & \text{n} \times \text{n} & \\ & & \mathbf{M} \\ & & \text{k} \times \text{k} \end{matrix} \right)$$

$$p(\gamma | \Omega^{-1}) \propto (2\pi)^{-nk/2} |\Omega|^{-k/2}$$

$$\exp \left[-\frac{1}{2} (\gamma - \mathbf{m})' (\Omega \otimes \mathbf{M})^{-1} (\gamma - \mathbf{m}) \right]$$

$$p(\gamma, \Omega^{-1} | \mathbf{Y}) \propto p(\gamma, \Omega^{-1}, \mathbf{Y})$$

$$= p(\mathbf{Y} | \gamma, \Omega^{-1}) p(\gamma | \Omega^{-1}) p(\Omega^{-1})$$

$$\propto |\Omega|^{-T/2} \exp$$

$$\left[-\frac{1}{2} \sum (\mathbf{y}_t - \Gamma' \mathbf{x}_t)' \Omega^{-1} (\mathbf{y}_t - \Gamma' \mathbf{x}_t) \right]$$

$$|\Omega|^{-k/2} \exp \left[-\frac{1}{2} (\gamma - \mathbf{m})' (\Omega \otimes \mathbf{M})^{-1} (\gamma - \mathbf{m}) \right]$$

$$|\Omega|^{-(N-n-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\Omega^{-1} \Lambda) \right]$$

After a lot of algebra, this can be

rewritten as

$$p(\gamma, \Omega^{-1} | \mathbf{Y}) \propto$$

$$|\Omega|^{-k/2} \exp \left[-\frac{1}{2} (\gamma - \mathbf{m}^*)' (\Omega \otimes \mathbf{M}^*)^{-1} (\gamma - \mathbf{m}^*) \right]$$

$$|\Omega|^{-(T+N-n-1)/2} \exp \left[-\frac{1}{2} \text{tr}(\Omega^{-1} \Lambda^*) \right]$$

$$\text{i.e., } \gamma | \Omega^{-1}, \mathbf{Y} \sim N(\mathbf{m}^*, \Omega \otimes \mathbf{M}^*)$$

$$\Omega^{-1} | \mathbf{Y} \sim W(T+N, \Lambda^*)$$

$$\boldsymbol{\gamma}|\boldsymbol{\Omega}^{-1}, \mathbf{Y} \sim N(\mathbf{m}^*, \boldsymbol{\Omega} \otimes \mathbf{M}^*)$$

$$\mathbf{M}^* = \left(\mathbf{M}^{-1} + \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1}$$

$$\mathbf{m}^* = (\mathbf{I}_n \otimes \mathbf{M}^* \mathbf{M}^{-1}) \mathbf{m} + (\mathbf{I}_n \otimes \mathbf{M}^* \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t') \hat{\boldsymbol{\gamma}}$$

$$\hat{\boldsymbol{\gamma}} = \text{vec}(\hat{\boldsymbol{\Gamma}})$$

$$\hat{\boldsymbol{\Gamma}} = \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{y}_t' \right)$$

Diffuse prior: $\mathbf{M}^{-1} = \mathbf{0}$

$$\boldsymbol{\gamma}_i|\boldsymbol{\Omega}^{-1}, \mathbf{Y} \sim N\left(\hat{\boldsymbol{\gamma}}_i, \boldsymbol{\Omega} \otimes \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right]^{-1}\right)$$

Estimate i th equation of VAR by OLS

$$\hat{\boldsymbol{\gamma}}_i = \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^T \mathbf{x}_t y_{it} \right)$$

(corresponds to elements $k(i-1) + 1$ through ki of $\hat{\boldsymbol{\gamma}}$)

$\hat{\boldsymbol{\gamma}}_i$ is posterior mean of $\boldsymbol{\gamma}_i$ and

$\sigma_{ii} \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1}$ is posterior variance (conditional on $\boldsymbol{\Omega}$)

$$\boldsymbol{\Omega}^{-1}|\mathbf{Y} \sim W(T+N, \boldsymbol{\Lambda}^*)$$

$$\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda} + \hat{\mathbf{S}} + \mathbf{Q}$$

$$\hat{\mathbf{S}} = \sum_{t=1}^T \left(\mathbf{y}_t - \hat{\boldsymbol{\Gamma}} \mathbf{x}_t \right) \left(\mathbf{y}_t - \hat{\boldsymbol{\Gamma}} \mathbf{x}_t \right)'$$

$$\mathbf{Q} = \mathbf{V}' \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right) \mathbf{M}^* \mathbf{M}^{-1} \mathbf{V}$$

$$\text{vec} \begin{pmatrix} \mathbf{V} \\ k \times n \end{pmatrix} = \mathbf{m} - \hat{\boldsymbol{\gamma}}$$

Diffuse prior: $N = 0, \mathbf{M}^{-1} = \mathbf{0}, \boldsymbol{\Lambda} = \mathbf{0}$

Diffuse prior:

$$\Omega^{-1} | \mathbf{Y} \sim W(T, \hat{\mathbf{S}})$$

$$\boldsymbol{\gamma} | \Omega^{-1}, \mathbf{Y} \sim N(\hat{\boldsymbol{\gamma}}, \Omega \otimes [\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t']^{-1})$$

To generate a draw from the posterior distribution of $(\boldsymbol{\gamma}, \Omega^{-1})$ with a diffuse prior:

(1) Estimate i th equation of VAR by OLS for $i = 1, 2, \dots, n$

$$\hat{\boldsymbol{\gamma}}_i = \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^T \mathbf{x}_t y_{it} \right)$$

$$\mathbf{M}^* = \left(\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right)^{-1}$$

(2) Calculate residual of i th equation and sum of outer products of residuals:

$$\hat{\boldsymbol{\varepsilon}}_{it} = y_{it} - \hat{\boldsymbol{\gamma}}_i' \mathbf{x}_t$$

$$\hat{\mathbf{S}}_{n \times n} = \sum_{t=1}^T \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$$

(3) Generate an artificial sample of size T of $(n \times 1)$ vector \mathbf{z}_i where $\mathbf{z}_i \sim N(\mathbf{0}, \hat{\mathbf{S}}^{-1})$ and calculate the $(n \times n)$ matrix

$$\mathbf{W} = \sum_{i=1}^T \mathbf{z}_i \mathbf{z}_i'$$

(4) Set $\mathbf{\Omega} = \mathbf{W}^{-1}$.
 (5) Generate a draw for $\boldsymbol{\gamma}$ from a $N(\hat{\boldsymbol{\gamma}}, \mathbf{\Omega} \otimes \mathbf{M}^*)$ distribution.
 The values of $\mathbf{\Omega}$ from step (4) and $\boldsymbol{\gamma}$ from step (5) represent a single draw from the posterior distribution.
 To generate a Monte Carlo sample of D draws, repeat steps (3)-(5) D times.

To get standard errors on impulse-response functions, for each draw calculate the impulse-response function implied by that value of $\boldsymbol{\gamma}$.

Usual procedure: Ω is fixed at $T^{-1}\hat{\mathbf{S}}$ for all draws = asymptotic classical distribution or approximation to Bayesian distribution with diffuse prior.

Example of using a non-diffuse prior (Del Negro and Schorfheide, 2002).

They use a log-linearization of a real business cycle model to get theoretical values for VAR parameters:

$$\mathbf{y}_t = \Gamma_0' \mathbf{x}_t + \varepsilon_t$$
$$E(\varepsilon_t \varepsilon_t') = \Omega_0$$

Suppose we had an artificial “sample” of observations $\{\mathbf{y}_t\}_{t=p+1}^{\tilde{T}}$ and base \mathbf{m} , \mathbf{M} , N , and Λ on a diffuse inference from this sample:

$$\mathbf{M} = \left(\sum_{i=1}^{\tilde{T}} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' \right)^{-1}$$

$$\tilde{\mathbf{\Gamma}} = \mathbf{M} \sum_{i=1}^{\tilde{T}} \tilde{\mathbf{x}}_i \tilde{\mathbf{y}}_i'$$

$$\mathbf{m} = \text{vec}(\tilde{\mathbf{\Gamma}})$$

$$N = \tilde{T}$$

$$\mathbf{\Lambda} = \sum_{i=1}^{\tilde{T}} \left(\mathbf{y}_i - \tilde{\mathbf{\Gamma}}' \mathbf{x}_i \right) \left(\mathbf{y}_i - \tilde{\mathbf{\Gamma}}' \mathbf{x}_i \right)'$$

We could then use these values of \mathbf{m} , \mathbf{M} , N , and $\mathbf{\Lambda}$ together with the actual observed data $\{\mathbf{y}_t\}_{t=p+1}^T$ to form a posterior inference using the earlier formulas, where choice of \tilde{T} would reflect how much weight to put on the “real business cycle prior.”

Actually, we can use the RBC's implied values of $\mathbf{\Gamma}_0$ and $\mathbf{\Omega}_0$ not just to simulate a few observations, but we can use them to calculate analytically the moments:

$$E(\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t') = \mathbf{A}_0$$

$$E(\tilde{\mathbf{x}}_t \tilde{\mathbf{y}}_t') = \mathbf{B}_0$$

$$E(\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t') = \mathbf{C}_0$$

where \mathbf{A}_0 , \mathbf{B}_0 and \mathbf{C}_0 are functions of $\mathbf{\Gamma}_0$ and $\mathbf{\Omega}_0$.

$$\mathbf{M} = \left(\sum_{i=1}^{\tilde{T}} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i' \right)^{-1} \leftrightarrow (\tilde{T} \mathbf{A}_0)^{-1}$$

$$\tilde{\Gamma} = \mathbf{M} \sum_{i=1}^{\tilde{T}} \tilde{\mathbf{x}}_i \tilde{\mathbf{y}}_i' \leftrightarrow \mathbf{A}_0^{-1} \mathbf{B}_0$$

$$\mathbf{m} = \text{vec}(\tilde{\Gamma})$$

$$N = \tilde{T}$$

$$\Lambda = \sum_{i=1}^{\tilde{T}} \left(\mathbf{y}_i - \tilde{\Gamma}' \mathbf{x}_i \right) \left(\mathbf{y}_i - \tilde{\Gamma}' \mathbf{x}_i \right)' \\ \leftrightarrow \tilde{T} (\mathbf{C}_0 - \mathbf{B}_0' \mathbf{A}_0^{-1} \mathbf{B}_0)$$

Del Negro and Schorfheide find that putting equal weights on prior and data ($\tilde{T} = T$) results in substantially better forecasts than unrestricted VAR, and better than Minnesota prior for horizons greater than 1 quarter.

RBC = good simplification (shrinkage).

Problem with Normal-Wishart prior

$$\gamma | \Omega^{-1} \sim N \left(\mathbf{m}, \Omega \otimes \mathbf{M} \right)$$

$\begin{matrix} nk \times 1 & n \times n & k \times k \end{matrix}$

$y_{1,t-1}$ = first element of \mathbf{x}_t

$y_{2,t-1}$ = second element of \mathbf{x}_t

$$\gamma|\Omega^{-1} \sim N\left(\mathbf{m}, \Omega \otimes \mathbf{M}\right)$$

confidence in γ_1 = coefficient relating

y_{1t} to $y_{1,t-1}$ is $\sigma_{11}m_{11}$

confidence in γ_2 = coefficient relating

y_{1t} to $y_{2,t-1}$ is $\sigma_{11}m_{22}$

$$\gamma\Omega^{-1} \sim N\left(\mathbf{m}, \Omega \otimes \mathbf{M}\right)$$

confidence in γ_{k+1} = coefficient relating

y_{2t} to $y_{1,t-1}$ is $\sigma_{22}m_{11}$

confidence in γ_{k+2} = coefficient relating

y_{2t} to $y_{2,t-1}$ is $\sigma_{22}m_{22}$

Problem: If $\sigma_{11}m_{11} > \sigma_{11}m_{22}$ (pretty confident variable 2 doesn't matter for variable 1), then must have

$\sigma_{22}m_{11} > \sigma_{22}m_{22}$ (think variable 2 doesn't matter for variable 2 either)

Ways to get around this problem:

(1) Assume that Ω is diagonal. Then single-equation methods are equivalent to full-system inference, use different M_i for each equation.

(2) Drop natural conjugates, turn to numerical Bayesian methods.

II. Vector autoregressions

- A. Introduction
- B. Normal-Wishart priors for VARs
- C. Bayesian analysis of structural VARs

Consider structural VAR:

$$\begin{matrix} \mathbf{A}' & \mathbf{y}_t & = & \mathbf{B}' & \mathbf{x}_t & + & \mathbf{v}_t \\ n \times n & n \times 1 & & n \times k & k \times 1 & & n \times 1 \end{matrix}$$

$$\mathbf{x}'_t = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \dots, \mathbf{y}'_{t-p}, 1)$$

$$k = np + 1$$

$$\mathbf{v}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1} \sim N(\mathbf{0}, \mathbf{I}_n)$$

$$p(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1})$$

$$= p(\mathbf{v}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{-p+1}) \left| \frac{\partial \mathbf{y}_t}{\partial \mathbf{v}_t} \right|^{-1}$$

$$\frac{\partial \mathbf{y}_t}{\partial \mathbf{v}_t} = (\mathbf{A}')^{-1}$$

$$p(\mathbf{y}_1, \dots, \mathbf{y}_T | \mathbf{A}, \mathbf{B}; \mathbf{y}_0, \mathbf{y}_{-1}, \dots, \mathbf{y}_{-p+1})$$

$$= (2\pi)^{-Tn/2} |\mathbf{A}|^T$$

$$\exp \left[-\frac{1}{2} \sum_{t=1}^T (\mathbf{A}' \mathbf{y}_t - \mathbf{B}' \mathbf{x}_t)' (\mathbf{A}' \mathbf{y}_t - \mathbf{B}' \mathbf{x}_t) \right]$$

Take transpose

$$\mathbf{y}'_t \mathbf{A} = \mathbf{x}'_t \mathbf{B} + \mathbf{v}'_t$$

and stack rows on top of each other

for $t = 1, 2, \dots, T$:

$$\begin{matrix} \mathbf{Y} & \mathbf{A} & = & \mathbf{X} & \mathbf{B} & + & \mathbf{V} \\ T \times n & n \times n & & T \times k & k \times n & & T \times n \end{matrix}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_n \end{bmatrix}$$

$\mathbf{b}_i = (k \times 1)$ vector of coefficients

on the lags in the i th structural equation:

$$\mathbf{y}'_t \mathbf{a}_i = \mathbf{x}'_t \mathbf{b}_i + v_{it}$$

The vec operator stacks the columns of a $(k \times n)$ matrix on top of each other, from left to right, to form a $(kn \times 1)$ vector:

$$\text{vec}(\mathbf{B}) = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \equiv \mathbf{b}$$

So first k elements of the $(nk \times 1)$ vector \mathbf{b} correspond to the coefficients on lags for the first equation in the VAR.

$$\mathbf{Y} \mathbf{A} = \mathbf{X} \mathbf{B} + \mathbf{V}$$

$T \times n \quad n \times n \quad T \times k \quad k \times n \quad T \times n$

$$\text{vec}(\mathbf{XB}) = (\mathbf{I}_n \otimes \mathbf{X}) \text{vec}(\mathbf{B})$$

$Tn \times 1 \quad n \times n \quad T \times k \quad kn \times 1$

$$= \tilde{\mathbf{X}} \mathbf{b}$$

$Tn \times nk \quad nk \times 1$

for $\tilde{\mathbf{X}} \equiv (\mathbf{I}_n \otimes \mathbf{X})$

Likewise define

$$\begin{aligned}\tilde{\mathbf{y}} &\equiv \text{vec}(\mathbf{Y}\mathbf{A}) \\ &_{Tn \times 1} \\ &= (\mathbf{I}_n \otimes \mathbf{Y})\mathbf{a}\end{aligned}$$

for $\mathbf{a} \equiv \text{vec}(\mathbf{A})$. First n elements of \mathbf{a}
 $_{n^2 \times 1}$

correspond to coefficients on contemporaneous variables in first equation of VAR.

$$\begin{matrix} \mathbf{Y} & \mathbf{A} & = & \mathbf{X} & \mathbf{B} & + & \mathbf{V} \\ T \times n & n \times n & & T \times k & k \times n & & T \times n \end{matrix}$$

Taking vec of full system,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\mathbf{b} + \tilde{\mathbf{v}}$$

where $\tilde{\mathbf{v}} \sim N(\mathbf{0}, \mathbf{I}_{Tn})$. Note that conditional on \mathbf{a} , this is a classical regression model with unit variance for the residual.

$$\begin{aligned}p(\mathbf{Y}|\mathbf{a}, \mathbf{b}) &= (2\pi)^{-Tn/2} |\mathbf{A}|^T \\ &\exp\left[-\frac{1}{2}(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b})'(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\mathbf{b})\right]\end{aligned}$$

$$\mathbf{b}|\mathbf{a} \sim N(\mathbf{m}(\mathbf{a}), \mathbf{M}(\mathbf{a}))$$

$$p(\mathbf{b}|\mathbf{a}) = (2\pi)^{-nk/2} |\mathbf{M}(\mathbf{a})|^{-1/2}$$

$$\exp\left\{-\frac{1}{2}[\mathbf{b} - \mathbf{m}(\mathbf{a})]'[\mathbf{M}(\mathbf{a})]^{-1}[\mathbf{b} - \mathbf{m}(\mathbf{a})]\right\}$$

$p(\mathbf{a})$ arbitrary

$$p(\mathbf{b}, \mathbf{a}|\mathbf{Y}) = p(\mathbf{b}|\mathbf{a}, \mathbf{Y})p(\mathbf{a}|\mathbf{Y})$$

$$\mathbf{b}|\mathbf{a}, \mathbf{Y} \sim N(\mathbf{m}^*(\mathbf{a}), \mathbf{M}^*(\mathbf{a}))$$

$$\mathbf{M}^*(\mathbf{a}) = \{[\mathbf{M}(\mathbf{a})]^{-1} + \tilde{\mathbf{X}}'\tilde{\mathbf{X}}\}^{-1}$$

$$\tilde{\mathbf{X}} \equiv (\mathbf{I}_n \otimes \mathbf{X})$$

$$\mathbf{M}^*(\mathbf{a}) = \{[\mathbf{M}(\mathbf{a})]^{-1} + [\mathbf{I}_n \otimes \mathbf{X}'\mathbf{X}]\}^{-1}$$

$$\mathbf{b}|\mathbf{a}, \mathbf{Y} \sim N(\mathbf{m}^*(\mathbf{a}), \mathbf{M}^*(\mathbf{a}))$$

$$\mathbf{m}^*(\mathbf{a}) = \mathbf{M}^*(\mathbf{a})\{[\mathbf{M}(\mathbf{a})]^{-1}\mathbf{m}(\mathbf{a}) + \tilde{\mathbf{X}}'\tilde{\mathbf{y}}\}$$

$$\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = (\mathbf{I}_n \otimes \mathbf{X}')(\mathbf{I}_n \otimes \mathbf{Y})\mathbf{a}$$

$$= (\mathbf{I}_n \otimes \mathbf{X}'\mathbf{Y})\mathbf{a}$$

Still, we need to “invert” ($nk \times nk$) matrix

$$\{[\mathbf{M}(\mathbf{a})]^{-1} + [\mathbf{I}_n \otimes \mathbf{X}'\mathbf{X}]\}^{-1}$$

Suppose our prior for equation i takes the form

$$\mathbf{b}_i | \mathbf{a} \sim N(\mathbf{m}_i(\mathbf{a}), \mathbf{M}_i(\mathbf{a}))$$

for $\mathbf{M}_i(\mathbf{a})$ a $(k \times k)$ matrix, with priors independent across equations.

$$\{[\mathbf{M}(\mathbf{a})]^{-1} + [\mathbf{I}_n \otimes \mathbf{X}'\mathbf{X}]\}^{-1} = \left\{ \begin{bmatrix} \mathbf{M}_1(\mathbf{a})^{-1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{M}_n(\mathbf{a})^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{X}'\mathbf{X} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{X}'\mathbf{X} \end{bmatrix} \right\}^{-1}$$

$$\{[\mathbf{M}(\mathbf{a})]^{-1} + [\mathbf{I}_n \otimes \mathbf{X}'\mathbf{X}]\}^{-1} = \begin{bmatrix} \mathbf{M}_1^*(\mathbf{a}) & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{M}_n^*(\mathbf{a}) \end{bmatrix}$$

for $\mathbf{M}_i^*(\mathbf{a}) = [\mathbf{M}_i(\mathbf{a})^{-1} + \mathbf{X}'\mathbf{X}]^{-1}$

$$\mathbf{m}^*(\mathbf{a}) = \mathbf{M}^*(\mathbf{a}) \{ [\mathbf{M}(\mathbf{a})]^{-1} \mathbf{m}(\mathbf{a}) + (\mathbf{I}_n \otimes \mathbf{X}'\mathbf{Y})\mathbf{a} \}$$

$$\begin{bmatrix} \mathbf{m}_1^*(\mathbf{a}) \\ \vdots \\ \mathbf{m}_n^*(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1^*(\mathbf{a})\mathbf{q}_1(\mathbf{a}) \\ \vdots \\ \mathbf{M}_n^*(\mathbf{a})\mathbf{q}_n(\mathbf{a}) \end{bmatrix}$$

$$\mathbf{q}_i(\mathbf{a}) = [\mathbf{M}_i(\mathbf{a})]^{-1} \mathbf{m}_i(\mathbf{a}) + \mathbf{X}'\mathbf{Y}\mathbf{a}_i$$

Specification of $p(\mathbf{b}|\mathbf{a})$:
 expect each series to behave like a random walk.

Structural equations:

$$\underset{T \times n}{\mathbf{Y}} \underset{n \times n}{\mathbf{A}} = \underset{T \times k}{\mathbf{X}} \underset{k \times n}{\mathbf{B}} + \underset{T \times n}{\mathbf{V}}$$

Reduced form:

$$\underset{T \times n}{\mathbf{Y}} = \underset{T \times k}{\mathbf{X}} \underset{k \times n}{\mathbf{\Pi}} + \underset{T \times n}{\mathbf{E}}$$

$$\mathbf{\Pi} = \mathbf{B}\mathbf{A}^{-1}$$

$$\mathbf{E} = \mathbf{V}\mathbf{A}^{-1}$$

Random walk:

$$\Pi_{k \times n} = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} = \mathbf{B}_{k \times n} \mathbf{A}^{-1}_{n \times n}$$
$$E(\mathbf{B}|\mathbf{A}) = \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

$$\mathbf{b}_i | \mathbf{a} \sim N(\mathbf{m}_i(\mathbf{a}), \mathbf{M}_i(\mathbf{a}))$$

$$\mathbf{m}_i(\mathbf{a}) = \begin{bmatrix} \mathbf{a}_i \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

Let $m_i^{(r)}$ denote the r th element of this vector (prior expectation of r th element of \mathbf{b}_i) and $M_i^{(r)}$ its variance:

$$b_i^{(r)} | \mathbf{a} \sim N(m_i^{(r)}, M_i^{(r)})$$

with priors across coefficients taken to be independent.

Multiplying the likelihood by the prior density

$$p(b_i^{(r)}|\mathbf{a}) = (2\pi)^{-1/2} [M_i^{(r)}]^{-1/2} \exp\left[-\frac{[b_i^{(r)} - m_i^{(r)}]^2}{2M_i^{(r)}}\right]$$

is numerically identical to acting as if I had observed a variable $q_i^{(r)}$ from the system

$$q_i^{(r)} = b_i^{(r)} / \sqrt{M_i^{(r)}} + v_i^{(r)}$$

where the observed value of $q_i^{(r)}$ is

$$m_i^{(r)} / \sqrt{M_i^{(r)}} \text{ and } v_i^{(r)} \sim N(0, 1).$$

Or, since $m_i^{(r)}$ is the r th element of \mathbf{a}_i if $r \leq n$ and is zero otherwise, this is numerically identical to having observed the $(n \times 1)$ vector $\mathbf{y}_i^{(r)}$ and $(k \times 1)$ vector $\mathbf{x}_i^{(r)}$ from the following system:

$$[\mathbf{y}_i^{(r)}]' \mathbf{a}_i = [\mathbf{x}_i^{(r)}]' \mathbf{b}_i + v_i^{(r)}$$

$$\mathbf{y}_i^{(r)} = \begin{cases} \mathbf{e}_r(n) / \sqrt{M_i^{(r)}} & \text{if } r \leq n \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$\mathbf{x}_i^{(r)} = \mathbf{e}_r(k) / \sqrt{M_i^{(r)}}$$

for $\mathbf{e}_r(n)$ the r th column of \mathbf{I}_n

$\mathbf{e}_r(k)$ the r th column of \mathbf{I}_k

Stack these “dummy observations”

for $r = 1, 2, \dots, k$ in matrices

$$\mathbf{Y}_{id} = \begin{bmatrix} [\mathbf{y}_i^{(1)}]' \\ \vdots \\ [\mathbf{y}_i^{(k)}]' \end{bmatrix} \quad \mathbf{X}_{id} = \begin{bmatrix} [\mathbf{x}_i^{(1)}]' \\ \vdots \\ [\mathbf{x}_i^{(k)}]' \end{bmatrix}$$

The posterior $p(\mathbf{b}_i | \mathbf{a}, \mathbf{Y})$ is then:

$$\mathbf{b}_i | \mathbf{a}, \mathbf{Y} \sim N(\mathbf{m}_i^*(\mathbf{a}), \mathbf{M}_i^*(\mathbf{a}))$$

$$\mathbf{M}_i^*(\mathbf{a}) = [\mathbf{M}_i(\mathbf{a})^{-1} + \mathbf{X}'\mathbf{X}]^{-1}$$
$$= (\mathbf{X}'_{id}\mathbf{X}_{id} + \mathbf{X}'\mathbf{X})^{-1}$$

$$\mathbf{m}_i^*(\mathbf{a}) = \mathbf{M}_i^*(\mathbf{a}) \{ [\mathbf{M}_i(\mathbf{a})^{-1}\mathbf{m}_i(\mathbf{a}) + \mathbf{X}'\mathbf{Y}\mathbf{a}_i] \}$$
$$= (\mathbf{X}'_{id}\mathbf{X}_{id} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'_{id}\mathbf{Y}_{id} + \mathbf{X}'\mathbf{Y})\mathbf{a}_i$$

$$b_i^{(r)} | \mathbf{a} \sim N(m_i^{(r)}, M_i^{(r)})$$

Remaining question: how choose $M_i^{(r)}$.

Let $j(r)$ denote which variable the

r th coefficient refers to ($j = 1, 2, \dots, n$)

and $\ell(r)$ its lag ($\ell = 1, 2, \dots, p$)

i.e., $b_i^{(r)}$ is the coefficient relating

$\mathbf{y}'_i \mathbf{a}_i$ to $y_{j,t-\ell}$

(1) If units in which y_{jt} is measured are doubled, value of $b_i^{(r)}$ is cut in half
 \Rightarrow make $M_i^{(r)}$ inversely proportional to σ_j , the standard deviation of univariate autoregression for y_{jt}

(2) have more confidence in zero priors for bigger ℓ
 \Rightarrow make $M_i^{(r)}$ inversely proportional to ℓ^{λ_3} ($\lambda_3 > 0$)

$$b_i^{(r)} | \mathbf{a} \sim N(m_i^{(r)}, M_i^{(r)})$$

$$M_i^{(r)} = \begin{cases} \frac{\lambda_0 \lambda_1}{\sigma_{j(r)} [\ell^{(r)}]^{\lambda_3}} & \text{for } r = 1, 2, \dots, k-1 \\ \lambda_0 \lambda_4 & \text{for } r = k \end{cases}$$

where λ_0 controls tightness of prior for \mathbf{a} and λ_1 controls tightness of random walk prior

Distribution of \mathbf{a}

$$p(\mathbf{a}, \mathbf{b}|\mathbf{Y}) = p(\mathbf{b}|\mathbf{a}, \mathbf{Y})p(\mathbf{a}|\mathbf{Y})$$

$$\mathbf{b}|\mathbf{a}, \mathbf{Y} \sim N(\mathbf{m}^*(\mathbf{a}), \mathbf{M}^*(\mathbf{a}))$$

$$p(\mathbf{a}|\mathbf{Y}) \propto p(\mathbf{a})|\mathbf{A}|^T$$

$$|\mathbf{I}_{Tn} + (\mathbf{I}_n \otimes \mathbf{X})\mathbf{M}(\mathbf{a})(\mathbf{I}_n \otimes \mathbf{X}')^{-1/2}$$

$$\exp\left\{-\frac{1}{2}[\mathbf{a}'(\mathbf{I}_n \otimes \mathbf{Y}'\mathbf{Y})\mathbf{a} +$$

$$+ \mathbf{m}(\mathbf{a})'\mathbf{M}(\mathbf{a})^{-1}\mathbf{m}(\mathbf{a})$$

$$- \mathbf{m}^*(\mathbf{a})'(\mathbf{M}^*(\mathbf{a}))^{-1}\mathbf{m}^*(\mathbf{a})]\right\}$$

What is this distribution $p(\mathbf{a}|\mathbf{Y})$?

Use numerical methods.

Example: Sims-Zha, distribution of $p(\mathbf{a}|\mathbf{Y})$:

$$p(\mathbf{a}|\mathbf{Y}) \propto p(\mathbf{a})|\mathbf{A}|^T$$

$$|\mathbf{I}_{Tn} + (\mathbf{I}_n \otimes \mathbf{X})\mathbf{M}(\mathbf{a})(\mathbf{I}_n \otimes \mathbf{X}')^{-1/2}$$

$$\exp\left\{-\frac{1}{2}[\mathbf{a}'(\mathbf{I}_n \otimes \mathbf{Y}'\mathbf{Y})\mathbf{a} +$$

$$+ \mathbf{m}(\mathbf{a})'\mathbf{M}(\mathbf{a})^{-1}\mathbf{m}(\mathbf{a})$$

$$- \mathbf{m}^*(\mathbf{a})'(\mathbf{M}^*(\mathbf{a}))^{-1}\mathbf{m}^*(\mathbf{a})]\right\}$$

$$\equiv q(\mathbf{a})$$

Importance density $g(\mathbf{a})$ should be similar to $q(\mathbf{a})$ but with fatter tails.

(1) Find

$$\mathbf{a}_0 \equiv \arg \max_{\mathbf{a}} q(\mathbf{a})$$

$$\mathbf{H}_0 \equiv - \left. \frac{\partial^2 \log q(\mathbf{a})}{\partial \mathbf{a} \partial \mathbf{a}'} \right|_{\mathbf{a}=\mathbf{a}_0}$$

(2) Let $g(\mathbf{a})$ be n^2 -dimensional Student tdistribution centered at \mathbf{a}_0 with scale matrix \mathbf{H}_0^{-1} and 9 degrees of freedom.

(3) Generate j th draw $\mathbf{a}^{(j)}$ from $g(\mathbf{a})$.
E.g., generate

$$\mathbf{u}^{(j)} \sim N(\mathbf{0}, \mathbf{I}_{n^2})$$

$$\mathbf{v}^{(j)} = \mathbf{P}_0^{-1} \mathbf{u}^{(j)} + \mathbf{a}_0$$

\mathbf{P}_0 is Cholesky factor of \mathbf{H}_0

$$\mathbf{e}^{(j)} \sim N(\mathbf{0}, \mathbf{I}_9)$$

$$v^{(j)} = (1/9)[\mathbf{e}^{(j)}]'[\mathbf{e}^{(j)}]$$

$$\mathbf{a}^{(j)} = \mathbf{v}^{(j)}/v^{(j)}$$

(4) Calculate

$$g(\mathbf{a}^{(j)}) =$$

$$c \left\{ 1 + [\mathbf{a}^{(j)} - \mathbf{a}_0]' \mathbf{H}_0^{-1} [\mathbf{a}^{(j)} - \mathbf{a}_0] \right\}^{-(n^2+9)/2}$$

choosing c so that $\max_{j=1, \dots, D} g(\mathbf{a}^{(j)})$ is same

scale as $\max_{j=1, \dots, D} q(\mathbf{a}^{(j)})$

(5) Calculate the weight

$$\omega^{(j)} = \frac{g(\mathbf{a}^{(j)})}{g(\mathbf{a}^{(j)})}$$

(6) Repeat steps (3)-(5) for

$$j = 1, 2, \dots, D.$$

(7) The sample $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(D)}$, when weighted by $K^{-1}\omega^{(1)}, \dots, K^{-1}\omega^{(D)}$ for $K = \omega^{(1)} + \dots + \omega^{(D)}$, is a sample from $p(\mathbf{a}|\mathbf{Y})$, e.g.,

$$\hat{E}(\mathbf{a}|\mathbf{Y}) = \frac{\sum_{j=1}^D \omega^{(j)} \mathbf{a}^{(j)}}{\sum_{j=1}^D \omega^{(j)}}$$

(8) For each j , generate $\mathbf{b}^{(j)}$ from the $N(\mathbf{m}^*(\mathbf{a}^{(j)}), \mathbf{M}^*(\mathbf{a}^{(j)}))$ distribution and weight with $K^{-1}\omega^{(j)}$ for a draw from $p(\mathbf{b}|\mathbf{Y})$, e.g.

$$\hat{E}(\mathbf{b}|\mathbf{Y}) = \frac{\sum_{j=1}^D \omega^{(j)} \mathbf{b}^{(j)}}{\sum_{j=1}^D \omega^{(j)}}$$
