Macroeconometrics

Lecture 14 SVARs: Bayesian estimation II

Tomasz Woźniak

Department of Economics
University of Melbourne

Maximum likelihood estimation

Structural VARs with exclusion restrictions

Useful distribution: normal-generalized-normal

Prior and posterior distributions

Gibbs sampler

Normalization

Compulsory reading:

Woźniak (2021) Bayesian Structural VARs: Algorithms and Inference, Lecture notes

Useful readings:

Waggoner & Zha (2003) A Gibbs sampler for structural vector autoregressions, Journal of Economic Dynamics & Control

Waggoner & Zha (2003) Likelihood preserving normalization in multiple equation models, Journal of Econometrics

Objectives.

- To present the most useful frequentists and Bayesian methods
- To introduce an efficient Bayesian estimation algorithm
- To provide tools to work reliably with SVARs

Learning outcomes.

- Implementing concentrated likelihood
- Generating random draws from a normal-generalized-normal distribution
- Applying normalisation to estimation output

Maximum likelihood estimation

Maximum likelihood estimation: model setup

$$
\begin{aligned}
y_t &= \mu_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B u_t \\
&= A x_t + B u_t \\
u_t | y_{t-1}, \ldots, y_{t-p} &\sim iid\mathcal{N}\left(\mathbf{0}_N, I_N\right)
\end{aligned}
$$

Matrix notation.

$$
\begin{aligned}
Y &= AX + BU \\
U|X &\sim \mathcal{MN}_{N\times T}\left(\mathbf{0}_{N\times T}, I_T, I_N\right)
\end{aligned}
$$

$$
\underset{N\times T}{Y} = \begin{bmatrix} y_1 & \ldots & y_T \end{bmatrix} \qquad \underset{N\times K}{A} = \begin{bmatrix} \mu_0 & A_1 & \ldots & A_p \end{bmatrix}
$$

$$
\underset{K\times T}{X} = \begin{bmatrix} x_1 & \ldots & x_T \end{bmatrix} \qquad \underset{K\times 1}{x_t} = \begin{bmatrix} 1 & y_{t-1}' & \ldots & y_{t-p}' \end{bmatrix}'
$$

$$
\underset{N\times T}{U} = \begin{bmatrix} u_1 & \ldots & u_T \end{bmatrix}
$$

MLE for SVARs with exclusion restrictions

Step 1: Obtain the MLE of the RF model $\left(\widehat{A}, \widehat{\Sigma}\right)$

Step 2a: If lower-triangular identification is required compute the SF model MLE by $\widehat{B} = chol(\widehat{\Sigma})$ and return $\left(\widehat{A}, \widehat{B}\right)$

Step 2b: If other identifying exclusion restrictions are required that are available by the application of **Algorithm 1**, then apply this algorithm to find matrix Q, compute $\widehat{B}^* = \widehat{B}Q'$, return: $\left(\widehat{A}, \widehat{B}^*\right)$

Step 2c: If non-linear restrictions or restrictions that are not feasible through the application of **Algorithm 1** are required, then use the **concentrated maximum likelihood estimation** method presented below.

Step 3: Apply the invariance property of the MLE to compute

$$\left(\widehat{B}_+, \widehat{B}_0\right) = \left(\widehat{B}^{-1}\widehat{A}, \widehat{B}^{-1}\right)$$

and other functions of the parameters such as the IRFs and FEVD

Step 4: Use bootstrap procedure to compute the MLE's standard errors MacKinnon (2006, ER) and Herwartz, Lange (2020, OREEF)

Maximum likelihood estimation

Log-likelihood function.

$$
\begin{aligned}
l(A,B|Y,X) &\propto -T\log\det(B) - \frac{1}{2}\mathrm{tr}\left[(BB')^{-1}(Y-AX)(Y-AX)'\right] \\
&= -T\log\det(B) - \frac{1}{2}\mathrm{tr}\left[(BB')^{-1}(YY' - 2YX'A' + AXX'A'\right]
\end{aligned}
$$

MLE of A.

$$
\begin{aligned}
\frac{\partial l(A,B|Y,X)}{\partial A} &= -\frac{1}{2}\left[-2(BB')^{-1}YX' + 2(BB')^{-1}AXX'\right] \\
&= (BB')^{-1}YX' - (BB')^{-1}AXX'
\end{aligned}
$$

$$
\begin{aligned}
(\widehat{B}\widehat{B}')^{-1}YX' - (\widehat{B}\widehat{B}')^{-1}\widehat{A}XX' &= \mathbf{0}_{N\times(1+Np)} \\
YX' - \widehat{A}XX' &= \mathbf{0}_{N\times(1+Np)} \\
\widehat{A}XX' &= YX' \\
\widehat{A} &= YX'(XX')^{-1}
\end{aligned}
$$

Maximum likelihood estimation

MLE of B using concentrated log-likelihood function.

Analytical solution for $\widehat{A}$ is plugged in the log-likelihood function
Consistency of $\widehat{A}$ assures the consistency of $\widehat{B}$
Numerical optimization is used to find $\widehat{B}$

$$\begin{aligned} l(B|\widehat{A}, \widehat{\Sigma}, Y, X) &\propto -T \log\det(B) - \frac{1}{2}\operatorname{tr}\left[(BB')^{-1}(Y - \widehat{A}X)(Y - \widehat{A}X)'\right] \\ &= -T \log\det(B) - \frac{T}{2}\operatorname{tr}\left[(BB')^{-1}\frac{\widehat{U}\widehat{U}'}{T}\right] \\ &= -T \log\det(B) - \frac{T}{2}\operatorname{tr}\left[(BB')^{-1}\widehat{\Sigma}\right] \end{aligned}$$

$$\widehat{B} = \underset{B \in \mathbb{B}}{\arg\max}\; l(B|\widehat{A}, \widehat{\Sigma}, Y, X)$$

Use invariance property of the MLE to estimate

$$\widehat{B}_0 = \widehat{B}^{-1} \qquad \text{and} \qquad \widehat{B}_+ = \widehat{B}^{-1}\widehat{A}$$

Maximum likelihood estimation

A small fiscal policy model.

$$B_0 y_t = b_0 + B_1 y_{t-1} + \cdots + B_p y_{t-p} + B u_t$$
$$u_t | y_{t-1}, \ldots, y_{t-p} \sim iid\mathcal{N}\left(\mathbf{0}_N, \text{diag}\left(\sigma^2\right)\right)$$

Identification.

$$B_0 y_t = \cdots + B u_t$$

$$\begin{bmatrix} 1 & 0 & -\theta_{gdp} \\ 0 & 1 & -\gamma_{gdp} \\ -\rho_{ttr} & -\rho_{gs} & 1 \end{bmatrix} \begin{bmatrix} ttr_t \\ gs_t \\ rgdp_t \end{bmatrix} = \cdots + \begin{bmatrix} 1 & \theta_{gs} & 0 \\ \gamma_{ttr} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_t^{(ttr)} \\ u_t^{(gs)} \\ u_t^{(gdp)} \end{bmatrix}$$

ttr_t – total tax revenue, gs_t – government spendings, gdp_t – gross domestic product. All variables are expressed in log, real, per capita terms.

The system is not identified. See the identification of this system as discussed in Mertens, Ravn (2014, JME) *A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers* and the follow up literature.

Maximum likelihood estimation

Premultiply the model equation by B_0^{-1} to obtain:

$$y_t = \mu_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0^{-1} B u_t$$
$$u_t | y_{t-1}, \ldots, y_{t-p} \sim iid\mathcal{N}\left(\mathbf{0}_N, \operatorname{diag}\left(\sigma^2\right)\right)$$

RF and SF parameters relationships.

$$A = B_0^{-1} B_+$$
$$\Sigma = B_0^{-1} B \operatorname{diag}\left(\sigma^2\right) B' B_0^{-1\prime}$$

Use the MLE.

$$\widehat{A} = YX'(XX')^{-1}$$
$$\widehat{U} = Y - \widehat{A}X$$
$$\widehat{\Sigma} = T^{-1}\widehat{U}\widehat{U}'$$

Maximum likelihood estimation

Concentrated likelihood function.

$$
\begin{aligned}
l(B|\widehat{A},\widehat{\Sigma},Y,X) \propto & -T\log\det(B) + T\log\det(B_0) - \frac{T}{2}\sum_{n=1}^{N}\log\sigma_n^2 \\
& - \frac{T}{2}\operatorname{tr}\left[\left(B_0^{-1}B\operatorname{diag}\left(\sigma^2\right)B'B_0^{-1\prime}\right)^{-1}\widehat{\Sigma}\right]
\end{aligned}
$$

SF parameters MLE.

$$
\begin{aligned}
\left(\widehat{B}_0,\widehat{B},\widehat{\sigma}^2\right) &= \underset{B_0,B,\sigma^2\in\Theta}{\arg\max}\; l(B_0,B,\sigma^2|\widehat{A},\widehat{\Sigma},Y,X) \\
\widehat{B}_+ &= \widehat{B}_0\widehat{A}
\end{aligned}
$$

Useful formulae.

$$
\det(AB) = \det(A)\det(B), \quad \det(A') = \det(A)
$$

$$
\det\left(A^{-1}\right) = \left(\det(A)\right)^{-1}, \quad \det(\operatorname{diag}(a)) = \prod_{n=1}^{N} a_n
$$

Structural VARs with exclusion restrictions

Structural VARs with exclusion restrictions

Algorithm 1 by Rubio-Ramírez, Waggoner, & Zha (2010) was based on some initial values $(\tilde{B}_+, \tilde{B}_0)$

In Lecture 13 we transformed the RF parameters (A, Σ)

Gibbs sampler by Waggoner & Zha (2003) presented today allows reasonable flexibility of imposing zero restrictions for

- recursive and non-recursive identification patterns
- over-identifying restrictions (more than $N(N-1)/2$) restrictions

Structural VARs with exclusion restrictions

Gibbs sampler by Waggoner & Zha (2003) features:

A great flexibility in the model specification includes:

- conditional heteroskedasticity
- time-variation in the parameters

Facilitates Bayesian model comparison and selection

Can be also used to sample initial values of parameters $(\tilde{B}_+, \tilde{B}_0)$ if the model specification does not allow for sampling (A, Σ) easily.

Further extensions include SVARs:

- identified through heteroskedasticity and non-normal residuals
- identified by zero and sign restrictions
- identified using instrumental variables (Proxy SVARs)

Structural VARs with exclusion restrictions

$$\begin{aligned} B_0 y_t &= b_0 + B_1 y_{t-1} + \cdots + B_p y_{t-p} + u_t \\ &= B_+ x_t + u_t \end{aligned}$$

$$\begin{aligned} B_+ &= \begin{bmatrix} b_0 & B_1 & \ldots & B_p \end{bmatrix} \\ x_t &= \begin{bmatrix} 1 & y_{t-1}' & \ldots & y_{t-p}' \end{bmatrix}' \end{aligned}$$

Exclusion restrictions on the rows of B_0.

$$\underset{(1\times N)}{B_{0[n\cdot]}} = \underset{(1\times r_n)}{b_n}\underset{(r_n\times N)}{V_n} \qquad \text{such that} \qquad B_0 = \begin{bmatrix} b_1 V_1 \\ \vdots \\ b_N V_N \end{bmatrix}$$

b_n – a $1 \times r_n$ vector of unrestricted elements of n row of B_0
V_n – an $r_n \times N$ fixed matrix of ones and zeros

E.g.: $b_n = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \quad V_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \rightarrow \quad B_{0[n\cdot]} = \begin{bmatrix} b_1 & 0 & b_2 \end{bmatrix}$

Structural VARs with exclusion restrictions

The nth equation.

$$
\begin{aligned}
b_n V_n y_t &= B_n x_t + u_{n.t} \\
u_{n.t} &\sim \mathcal{N}(0, 1)
\end{aligned}
$$

$$\downarrow$$

$$
\begin{aligned}
b_n V_n Y &= B_n X + U_n \\
U_n &\sim \mathcal{N}\left(\mathbf{0}_T, I_T\right)
\end{aligned}
$$

$$
\begin{aligned}
\underset{(N\times T)}{Y} &= \begin{bmatrix} y_1 & \dots & y_T \end{bmatrix} \\
\underset{(K\times T)}{X} &= \begin{bmatrix} x_1 & \dots & x_T \end{bmatrix} \\
\underset{(1\times T)}{U_n} &= \begin{bmatrix} u_{n.1} & \dots & u_{n.T} \end{bmatrix} \\
\underset{(1\times K)}{B_n} &= B_{+[n\cdot]}
\end{aligned}
$$

Structural VARs with exclusion restrictions

Likelihood function.

$$
\begin{aligned}
L(B_+, B_0|Y, X) &\propto \det\left(B_0^{-1}B_0^{-1\prime}\right)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_n V_n Y - B_n X)(b_n V_n Y - B_n X)'\right\} \\
&= |\det\left(B_0\right)|^{T} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_n V_n Y - B_n X)(b_n V_n Y - B_n X)'\right\} \\
&= \left|\det\left(\begin{bmatrix} b_1 V_1 \\ \vdots \\ b_N V_N \end{bmatrix}\right)\right|^{T} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_n V_n Y - B_n X)(b_n V_n Y - B_n X)'\right\}
\end{aligned}
$$

Useful distribution: Normal-generalized-normal

Let $1 \times K$ vectors B_n and $1 \times r_n$ vectors b_n for $n = 1, \ldots, N$ follow:

$$p(B_+, B_0) \sim \mathcal{NGN}\left(\tilde{B}, \Omega, S, \nu\right)$$

$$= \left(\prod_{n=1}^{N} p(B_n | b_n)\right) p(b_1, \ldots, b_N)$$

$$p(B_n | b_n) = \mathcal{N}_K\left(b_n V_n \tilde{B}, \Omega\right) \quad \text{for } n = 1, \ldots, N$$

$$p(b_1, \ldots, b_n) \propto |\det(B_0)|^{\nu - N} \exp\left\{-\frac{1}{2} \sum_{n=1}^{N} b_n V_n S^{-1} V_n' b_n'\right\}$$

$\tilde{B}$ – an $N \times K$ matrix; $\underset{K \times K}{\Omega}$ and $\underset{N \times N}{S}$ – positive definite matrices; $\nu \geq N$

Kernel.

$$\underbrace{|\det(B_0)|^{\nu - N} \exp\left\{-\frac{1}{2} \sum_{n=1}^{N} b_n V_n S^{-1} V_n' b_n'\right\}}_{\text{generalized-normal}} \underbrace{\exp\left\{-\frac{1}{2} \sum_{n=1}^{N} (B_n - b_n V_n \tilde{B}) \Omega^{-1} (B_n - b_n V_n \tilde{B})'\right\}}_{\text{conditional normal}}$$

Structural VARs with exclusion restrictions

Likelihood function as $\mathcal{NGN}$ distribution.

$$L(B_+, B_0|Y,X) \propto |\det(B_0)|^T \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_nV_nY - B_nX)(b_nV_nY - B_nX)'\right\}$$
$$= |\det(B_0)|^T \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_nV_nYY'V_n'b_n' - 2b_nV_nYX'B_n' + B_nXX'B_n')\right\}$$

Expand the expression from the parentheses:

$$b_nV_nYY'V_n'b_n' - 2b_nV_nYX'B_n' + B_nXX'B_n' =$$
$$= b_nV_nYY'V_n'b_n' + B_nXX'B_n' - 2b_nV_nYX'(XX')^{-1}(XX')B_n' \pm b_nV_nYX'(XX')^{-1}(XX')(XX')^{-1}XY'V_n'b_n'$$

Let $\widehat{A} = YX'(XX')^{-1}$

$$= b_nV_n\left[YY' - YX'(XX')^{-1}XY'\right]V_n'b_n' + \left(B_n - b_nV_n\widehat{A}\right)(XX')\left(B_n - b_nV_n\widehat{A}\right)'$$

Likelihood function can be presented as:

$$L(B_+, B_0|Y,X) \sim \mathcal{NGN}\left(\widehat{A},\ (XX')^{-1},\ (YY' - YX'(XX')^{-1}XY')^{-1},\ T+N\right)$$

Prior and posterior distributions

Natural-conjugate prior distribution

The $\mathcal{NGN}$ distribution can be used as a natural-conjugate prior:

$$p(B_+, B_0) \sim \mathcal{NGN}\left(\underline{B}, \underline{\Omega}, \underline{S}, \underline{\nu}\right)$$

$$p(B_+, B_0) = \left(\prod_{n=1}^{N} p(B_n|b_n)\right) p(b_1, \ldots, b_N)$$

$$p(B_n|b_n) \sim \mathcal{N}_K\left(b_n V_n \underline{B}, \underline{\Omega}\right)$$

$$p(b_1, \ldots, b_N) \propto |\det(B_0)|^{\underline{\nu}-N} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} b_n V_n \underline{S}^{-1} V_n' b_n'\right\}$$

Kernel.

$$|\det(B_0)|^{\underline{\nu}-N} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} b_n V_n \underline{S}^{-1} V_n' b_n'\right\} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} (B_n - b_n V_n \underline{B}) \underline{\Omega}^{-1} (B_n - b_n V_n \underline{B})'\right\}$$

Posterior distribution

$$p(B_+, B_0|Y, X) \propto L(B_+, B_0|Y, X)p(B_+, B_0)$$

$$= |\det(B_0)|^T \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(b_n V_n Y - B_n X)(b_n V_n Y - B_n X)'\right\}$$

$$\times |\det(B_0)|^{\underline{\nu}-N} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} b_n V_n \underline{S}^{-1} V_n' b_n'\right\}$$

$$\times \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}(B_n - b_n V_n \underline{B})\underline{\Omega}^{-1}(B_n - b_n V_n \underline{B})'\right\}$$

$$= |\det(B_0)|^{T+\underline{\nu}-N}$$

$$\times \exp\left\{-\frac{1}{2}\sum_{n=1}^{N}\left(b_n V_n Y Y' V_n' b_n' - 2 b_n V_n Y X' B_n' + B_n X X' B_n'\right.\right.$$

$$\left.\left.+ b_n V_n \underline{S}^{-1} V_n' b_n' + B_n \underline{\Omega}^{-1} B_n' - 2 b_n V_n \underline{B}\underline{\Omega}^{-1} B_n' + b_n V_n \underline{B}\underline{\Omega}^{-1}\underline{B}' V_n' b_n'\right)\right\}$$

Posterior distribution

Expression from the parentheses:

$$
\begin{aligned}
& b_n V_n Y Y' V_n' b_n' - 2 b_n V_n Y X' B_n' + B_n X X' B_n' + b_n V_n \underline{S}^{-1} V_n b_n' \\
& \qquad\qquad + B_n \underline{\Omega}^{-1} B_n' - 2 b_n V_n \underline{B}\underline{\Omega}^{-1} B_n' + b_n V_n \underline{B}\underline{\Omega}^{-1} \underline{B}' V_n' b_n' \\
& = B_n \left[X X' + \underline{\Omega}^{-1} \right] B_n' - 2 b_n V_n \left[Y X' + \underline{B}\underline{\Omega}^{-1} \right] B_n' + b_n V_n \left[Y Y' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1} \underline{B}' \right] V_n' b_n' \\
& = B_n \overline{\Omega}^{-1} B_n' - 2 b_n V_n \left[Y X' + \underline{B}\underline{\Omega}^{-1} \right] \overline{\Omega}\overline{\Omega}^{-1} B_n' + b_n V_n \left[Y Y' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1} \underline{B}' \right] V_n' b_n' \\
& = B_n \overline{\Omega}^{-1} B_n' - 2 b_n V_n \left[Y X' + \underline{B}\underline{\Omega}^{-1} \right] \overline{\Omega}\overline{\Omega}^{-1} B_n' + b_n V_n \left[Y Y' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1} \underline{B}' \right] V_n' b_n' \\
& = B_n \overline{\Omega}^{-1} B_n' - 2 b_n V_n \overline{B}\overline{\Omega}^{-1} B_n' \pm b_n V_n \overline{B}\overline{\Omega}^{-1} \overline{B}' V_n' b_n' + b_n V_n \left[Y Y' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1} \underline{B}' \right] V_n' b_n' \\
& = \left(B_n - b_n V_n \overline{B} \right) \overline{\Omega}^{-1} \left(B_n - b_n V_n \overline{B} \right)' + b_n V_n \left[Y Y' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1} \underline{B}' - \overline{B}\overline{\Omega}^{-1} \overline{B}' \right] V_n' b_n' \\
& = \left(B_n - b_n V_n \overline{B} \right) \overline{\Omega}^{-1} \left(B_n - b_n V_n \overline{B} \right)' + b_n V_n \overline{S}^{-1} V_n' b_n'
\end{aligned}
$$

Posterior distribution

$$p(B_+, B_0|Y, X) \sim \mathcal{NGN}\left(\overline{B}, \overline{\Omega}, \overline{S}, \overline{\nu}\right)$$

$$\begin{aligned}
\overline{\Omega} &= \left[XX' + \underline{\Omega}^{-1}\right]^{-1} \\
\overline{B} &= \left[YX' + \underline{B}\underline{\Omega}^{-1}\right]\overline{\Omega} \\
\overline{S} &= \left[YY' + \underline{S}^{-1} + \underline{B}\underline{\Omega}^{-1}\underline{B}' - \overline{B}\overline{\Omega}^{-1}\overline{B}'\right]^{-1} \\
\overline{\nu} &= T + \underline{\nu}
\end{aligned}$$

$$V_1, \dots, V_N$$

Gibbs sampler for normal-generalized-normal distribution

Gibbs sampler

Posterior distribution.

$$p(B_+, B_0|Y, X) \sim \mathcal{NGN}\left(\overline{B}, \overline{\Omega}, \overline{S}, \overline{\nu}\right)$$

$$= p(B_0|Y, X) \prod_{n=1}^{N} p(B_n|Y, X, b_n)$$

$$p(B_0|Y, X) \propto |\det(B_0)|^{\overline{\nu}-N} \exp\left\{-\frac{1}{2}\sum_{n=1}^{N} b_n V_n \overline{S}^{-1} V_n' b_n'\right\}$$

$$p(B_n|Y, X, b_n) \sim \mathcal{N}_K\left(b_n V_n \overline{B}, \overline{\Omega}\right)$$

Sampler for $p(B_0|Y, X)$ proceeds row by row and draws from full conditional posterior distributions:

$$p\left(b_n|Y, X, b_1, \dots, b_{n-1}, b_{n+1}, \dots, b_N\right)$$

Sampling from a multivariate normal $p(B_n|Y, X, b_n)$ is simple

Gibbs sampler

Sampling from the $\mathcal{NGN}$ distribution is performed in two steps:

Gibbs sampler for B_0

Sample S draws from
$p\left(b_n^{(s)} | Y, X, b_1^{(s)}, \dots, b_{n-1}^{(s)}, b_{n+1}^{(s-1)}, \dots, b_N^{(s-1)}\right)$ obtaining
posterior sample $\left\{b_1^{(s)}, \dots, b_N^{(s)}\right\}_{s=1}^{S}$

Normalize the draws $\left\{b_1^{(s)}, \dots, b_N^{(s)}\right\}_{s=1}^{S}$

Direct sampling for B_+

For each draw of $b_n^{(s)}$ draw B_n from $p(B_n | Y, X, b_n)$
for $n = 1, \dots, N$ and $s = 1, \dots, S$

Return the sample drawn from the $\mathcal{NGN}$ posterior distribution

$$\left\{B_+^{(s)}, B_0^{(s)}\right\}_{s=1}^{S}$$

Gibbs sampler

Gibbs sampler for $b_n^{(s)} \sim p\left(b_n | Y, X, b_1^{(s)}, \dots, b_{n-1}^{(s)}, b_{n+1}^{(s-1)}, \dots, b_N^{(s-1)}\right)$

Compute the following values:

$$U_n = chol\left(\overline{\nu}\left(V_n \overline{S}^{-1} V_n'\right)^{-1}\right) - \text{ an } r_n \times r_n \text{ upper-triangular matrix}$$

$$w = \left[B_{0[-n.\cdot]}^{(s)}\right]_{\perp} - \text{where } w \text{ is a } 1 \times N \text{ matrix}$$

$$w_1 = wV_n'U_n' \cdot \left(wV_n'U_n'U_nV_nw'\right)^{-\frac{1}{2}} - \text{where } w_1 \text{ is a } 1 \times r_n \text{ vector}$$

$$W_n = \begin{bmatrix} w_1' & w_{1\perp}' \end{bmatrix}' - \text{where } W_n \text{ is a } r_n \times r_n \text{ matrix}$$

Draw the elements of a $1 \times r_n$ vector α_n:

Draw the first element of α_n by:

drawing $u \sim \mathcal{N}\left(\mathbf{0}_{\nu+1}, \overline{\nu}^{-1} I_{\nu+1}\right)$, and

setting $\alpha_{n[\cdot.1]} = \begin{cases} \sqrt{u'u} & \text{with probability } 0.5 \\ -\sqrt{u'u} & \text{with probability } 0.5 \end{cases}$

Draw the remaining $r_n - 1$ elements of α_n from $\mathcal{N}\left(\mathbf{0}_{r_n-1}, \overline{\nu}^{-1} I_{r_n-1}\right)$.

Compute the draw from the full conditional posterior distribution of b_n by:

$$b_n^{(s)} = \alpha_n W_n U_n$$

Gibbs sampler

In the description of the Gibbs sampler in the previous slide:

$X_\perp$ – is an orthogonal complement matrix of X

$B^{(s)}_{0[-n.\cdot]}$ – denotes matrix $B^{(s)}_0$ without its nth row:

$$B^{(s)}_{0[-n.\cdot]} = \begin{bmatrix} b^{(s)}_1 V_1 \\ \vdots \\ b^{(s)}_{n-1} V_{n-1} \\ b^{(s-1)}_{n+1} V_{n+1} \\ \vdots \\ b^{(s-1)}_N V_N \end{bmatrix}$$

Normalization

Normalization

The output of the Gibbs sampler perfectly reproduces the symmetry of the posterior distribution of matrices B_0 and B_+ around zero being the effect of local identification.

The likelihood function and the posterior distribution have 2^N equal modes as a consequence of local identification

Such an output represents the global shape of the posterior distribution (that is inherited from the likelihood function) and provides the full statistical characterization of these parameters.

It cannot be used to compute rotation-dependent characteristics of the values of interest such as the estimates of the parameters of the model or the impulse response functions.

The means of these values computed from the unnormalized output will all be equal to zero.

Normalization

Waggoner & Zha (2003, JOE) provide the optimal normalizing rule that preserves the shape of the likelihood function.

Let $\widehat{B}_0$ be one of the 2^N modes of the posterior distribution with normalized signs of the shocks, that is, with the chosen sign of the rows

Normalization of the posterior sampler output is performed with respect to $\widehat{B}_0$ so that its only mode is equal to $\widehat{B}_0$

Normalization

Define a set of $N \times N$ diagonal normalizing matrices D_i for $i = 1, \ldots, 2^N$

The diagonal elements of D_i are equal to 1 or -1 in all of the possible combinations across i. D_i are orthogonal matrices.

These 2^N matrices D_i exhaustively represent all of the possible combinations of ones and minus ones on the N elements of the main diagonal

Normalization

The algorithm for normalization utilizes a distance measure:

$$d\left[X|\widehat{\Sigma}\right] = \sum_{n=1}^{N} X_{[n.\cdot]}\widehat{\Sigma}^{-1}X'_{[n.\cdot]}$$

X – an $N \times N$ matrix

$X_{[n.\cdot]}$ – the nth row of X

$\widehat{\Sigma}$ – an $N \times N$ positive definite matrix

Normalization

To normalize the sth draw from the posterior distribution, $B_0^{(s)}$:

Compute the distance between $D_i B_0^{(s)}$ and $\widehat{B}_0$:

$$d\left[\left(D_i B_0^{(s)}\right)^{-1\prime} - \widehat{B}_0^{-1\prime} \middle| \left(\widehat{B}_0' \widehat{B}_0\right)^{-1}\right]$$

for $i = 1, \ldots, 2^N$

Return the normalized draw given by

$$D_{i*} B_0^{(s)}$$

where $i*$ is the value of i that minimizes the distance above

Normalize $B_0^{(s)}$ before sampling $B_+^{(s)}$ dependent on $B_0^{(s)}$

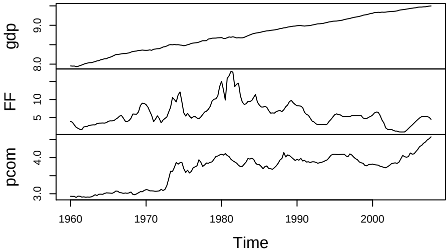
Normalization

An illustration

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} gdp_t \\ FF_t \\ pcom_t \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} gdp_{t-1} \\ FF_{t-1} \\ pcom_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1.t} \\ u_{3.t} \\ u_{3.t} \end{bmatrix}$$

Restrictions are imposed using

$$V_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \qquad V_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad V_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



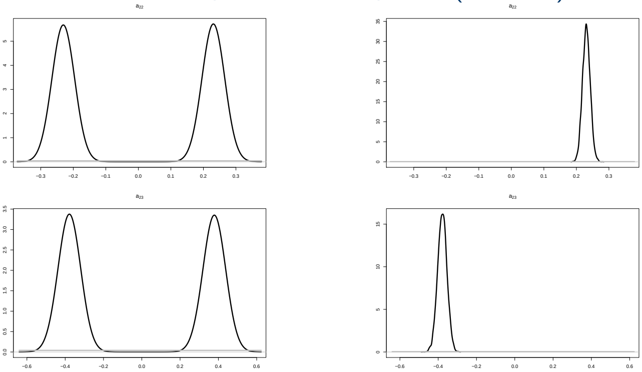
Normalization

non-normalized	normalized

posterior means

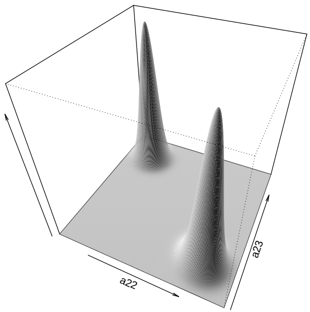
$$\begin{bmatrix} -0.001 & 0 & 0 \\ 0 & 0.001 & -0.001 \\ 0.023 & 0.002 & -0.058 \end{bmatrix} \quad \begin{bmatrix} 0.081 & 0 & 0 \\ 0 & 0.231 & -0.379 \\ -0.994 & -0.059 & 2.462 \end{bmatrix}$$

densities: posterior vs. prior: (a_{22}, a_{23})



Normalization

non-normalized	normalized



bivariate densities: (a_{22}, a_{23})

Structural VARs: Gibbs sampling

The Gibbs sampler by Waggoner & Zha (2003)

Thanks to simple derivations an efficient sampler is feasible

The sampler draws the SF parameters directly

Many extensions in the specification of the model are possible

Normalization provides a solution to identification of the shocks up to their signs (a rotation matrix D)