SCOTTISH GRADUATE PROGRAMME IN ECONOMICS

ECNM11060

Bayesian Econometrics

A Bayesian vector autoregression (BVAR)

Time series modelling for empirical macroeconomics

$\Rightarrow$ Vector Autoregressive (VAR) models are a popular way of summarizing inter-relationships between macroeconomic variables (see Stock & Watson article on VARs)

$\Rightarrow$ Stock & Watson say the job of a macroeconometrician consists of the following tasks:

- $\rightarrow$ Describe and summarise macroeconomic time series
- $\rightarrow$ **Predictive inference:** Produce forecasts based on historical data
- $\rightarrow$ **Structural inference:** Recover the structure of the macroeconomy from the data and study dynamic causal effects between variables
- $\rightarrow$ Advise policymakers (e.g., at central banks or government institutions)

$\Rightarrow$ A **VAR** is the tool used to perform these tasks (the workhorse in empirical macroeconomics)

A VAR model

⇒ A VAR(P) model can be written as follows:

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{p=1}^{P} \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

⇒ $\boldsymbol{y}_t$ is an $M \times 1$ vector of endogenous variables

⇒ $\boldsymbol{c}$ is an $M \times 1$ vector of intercepts

⇒ $\boldsymbol{A}_p$, for $p = 1, \ldots, P$, are $M \times M$ matrices of coefficients

⇒ ε_t is an $M \times 1$ vector of Gaussian errors (zero mean and variance $\boldsymbol{\Sigma}$)

⇒ Exogenous variables or additional deterministic terms can be added, but are omitted here to keep the notation simple

Matrix-form of the VAR

⇒ Let $\boldsymbol{Y}$ and $\boldsymbol{E}$ be $T \times M$ matrices stacking the T observations on each variable in columns next to each other

⇒ The VAR can then be written as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A} + \boldsymbol{E}$$

⇒ $\boldsymbol{X}$ collects $\boldsymbol{x}_t'$ on the tth row with $\boldsymbol{x}_t = (\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-P}', 1)'$

⇒ $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_P, \boldsymbol{c})'$ is $K \times M$ and $K = MP + 1$ is the number of explanatory variables per equation

⇒ Let α be the $KM \times 1$ vector of VAR coefficients

⇒ $\boldsymbol{\alpha} = \mathrm{vec}(\boldsymbol{A})$

Likelihood function of a VAR

- Similar to the standard regression model, the likelihood function breaks into two parts: one for α given Σ and one for Σ:

- $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \boldsymbol{y}$ is multivariate Normal:

$$\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \boldsymbol{y} \sim \mathcal{N}\Big(\widehat{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\Big)$$

- $\boldsymbol{\Sigma}^{-1}$ has Wishart form:

$$\boldsymbol{\Sigma}^{-1} \mid \boldsymbol{y} \sim \mathcal{W}\Big(T - K - M - 1,\ \boldsymbol{S}^{-1}\Big)$$

- OLS quantities $\widehat{\boldsymbol{A}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\widehat{\boldsymbol{\alpha}} = \mathrm{vec}(\widehat{\boldsymbol{A}})$ and $\boldsymbol{S} = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{A}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{A}})$

In R these OLS quantities are computed straightforwardly

```r
M   <- ncol(Y)    # No. of endogenous variables
N   <- nrow(Y)    # No. of observations
K   <- ncol(X)    # No. of coefficients per equation

tXX  <- crossprod(X)                     # Compute X'X
tXy  <- crossprod(X,y)                   # Compute X'y

# OLS quantities
A.ols      <- solve(tXX)%*%tXy           # Reduced-form OLS coefficients
eps.ols    <- Y - X %*%A.ols             # Reduced-form shocks
SSR.ols    <- crossprod(eps.ols)         # SSR based on OLS
SIG.ols    <- SSR.ols/(N - K - M - 1)    # OLS variance-covariance matrix
```

Digression: From regression to VAR

- $\Rightarrow$ In the regression model the parameters are β and σ^2
- $\Rightarrow$ It has proved convenient to work with $h = 1/\sigma^2$ (the precision)
- $\Rightarrow$ In a VAR it proves convenient to work with $\mathbf{\Sigma}^{-1}$ instead
- $\Rightarrow$ In regression, h typically has a Gamma distribution
- $\Rightarrow$ With a VAR, $\mathbf{\Sigma}^{-1}$ will typically have a Wishart distribution

The Wishart distribution

$\Rightarrow$ The Wishart is the matrix generalization of the Gamma distribution

$\Rightarrow$ Details: See the appendix to the textbook

$\Rightarrow$ If $\mathbf{\Sigma}^{-1} \sim \mathcal{W}(c, \boldsymbol{C})$, with c degrees of freedom and $\boldsymbol{C}$ scaling

$\Rightarrow$ Note: It is easy to take random draws from a Wishart distribution

Prior issue 1: Over-parameterisation

- $\Rightarrow$ VARs are not parsimonious: α contains *KM* parameters
- $\Rightarrow$ A VAR(4) with 5 dependent variables has 105 parameters
- $\Rightarrow$ Large VARs with 100$+$ variables: Thousands (or tens of thousands) of parameters
- $\Rightarrow$ Macro data sets (e.g., FRED-QD): The number of observations on each variable might be only a few hundred

Prior issue 1: Over-parameterisation (cont.)

- $\Rightarrow$ Without prior information it is hard to obtain precise estimates (degrees of freedom coming from likelihood are $T - K - M - 1$)
- $\Rightarrow$ Features such as impulse responses and forecasts will tend to be imprecisely estimated
- $\Rightarrow$ It is desirable to “shrink” estimates
- $\Rightarrow$ Prior information offers a natural (and sensible) way of doing this
- $\Rightarrow$ Different priors do shrinkage in different ways

Prior issue 2:
Analytical (conjugate) vs MCMC (non-conjugate)

$\Rightarrow$ Some priors lead to analytical results for the posterior and predictive densities

$\Rightarrow$ Other priors require MCMC methods, which increase computational burden

$\Rightarrow$ A recursive forecasting exercise, e.g., requires repeated calculation of posterior and predictive distributions; MCMC can be very demanding in that setting

$\Rightarrow$ One may want to go with "not-so-good" prior which leads to analytical results, if "ideal" prior leads to slow computation

Prior issue 3: Extensions of the VAR

- $\Rightarrow$ Priors differ in how easily they can accommodate extensions of a basic VAR
- $\Rightarrow$ Restricted VARs: Different equations have different explanatory variables
- $\Rightarrow$ TVP-VARs: VAR coefficients vary over time
- $\Rightarrow$ Heteroskedasticity
- $\Rightarrow$ Such extensions typically require MCMC, so there is no need to restrict attention to priors with analytical posteriors in the basic VAR

The Minnesota prior

- $\Rightarrow$ The classic shrinkage prior developed by Litterman, Sims, and others at the University of Minnesota and the Federal Reserve Bank of Minneapolis
- $\Rightarrow$ Key simplification: Replace Σ with an estimate $\widehat{\Sigma}$
- $\Rightarrow$ The original Minnesota prior simplifies further by assuming Σ is diagonal, with $\widehat{\sigma}_{ii} = s_i^2$ (the OLS error variance in the *i*th equation)
- $\Rightarrow$ If Σ is not assumed diagonal, one can use, e.g., $\widehat{\Sigma} = \boldsymbol{S}/T$

Minnesota prior: Specifying the prior mean

⇒ **Prior mean:** "Towards what should we shrink?"

⇒ The Minnesota prior assumes:

$$\boldsymbol{\alpha} \sim \mathcal{N}(\underline{\boldsymbol{\alpha}}_{Min}, \underline{\boldsymbol{V}}_{Min})$$

⇒ The explanatory variables in any equation fall into three groups: own lags, lags of other dependent variables, and exogenous or deterministic terms

⇒ $\underline{\boldsymbol{\alpha}}_{Min} = \mathbf{0}$ implies shrinkage towards zero (avoids overfitting); typically used for differenced/stationary time series, e.g., GDP growth or stock market returns

⇒ For levels data (e.g., GDP), the element of $\underline{\boldsymbol{\alpha}}_{Min}$ for the first own lag is typically set to 1 (centred over a random walk)

Minnesota prior: Specifying the prior variance

⇒ **Prior variance:** "By how much should we shrink?"

⇒ $\underline{\boldsymbol{V}}_{Min}$ is diagonal

⇒ Let $\underline{\boldsymbol{V}}_i$ be the block for equation i, with diagonal elements $\underline{V}_{i,jj}$

⇒ A common implementation (for lag $\ell = 1, \ldots, P$):

$$\underline{V}_{i,jj} = \begin{cases} \dfrac{\underline{a}_1}{\ell^2} & \text{coefficients on own lags} \\ \dfrac{\underline{a}_2\,\sigma_{ii}}{\ell^2\,\sigma_{jj}} & \text{coefficients on lags of variable } j \neq i \\ \underline{a}_3\,\sigma_{ii} & \text{coefficients on exogenous variables} \end{cases}$$

Minnesota prior: Properties

$\Rightarrow$ Choosing $\frac{KM(KM+1)}{2}$ elements of $\underline{\boldsymbol{V}}_{Min}$ is reduced to choosing just $\underline{a}_1, \underline{a}_2, \underline{a}_3$

$\Rightarrow$ As lag length ℓ increases, coefficients are increasingly shrunk towards zero

$\Rightarrow$ Setting $\underline{a}_1 > \underline{a}_2$ means own lags are shrunk less than lags of other variables

$\Rightarrow$ The factor σ_{ii}/σ_{jj} adjusts for differences in the scale of the variables

$\Rightarrow$ Work by Giannone, Lenza and Primiceri (*ReStat*, 2015) develops methods for estimating the hyperparameters from the data

Posterior Inference with the Minnesota prior

$\Rightarrow$ Simple analytical results involving only the Normal distribution (as $\boldsymbol{\Sigma}$ is treated as known):

$$\boldsymbol{\alpha} \mid \boldsymbol{y} \sim \mathcal{N}\left(\overline{\boldsymbol{\alpha}}_{Min}, \overline{\boldsymbol{V}}_{Min}\right)$$

$\Rightarrow$ Formulae for $\overline{\boldsymbol{\alpha}}_{Min}$ and $\overline{\boldsymbol{V}}_{Min}$ are available in standard sources (e.g., Bayesian Econometric Methods)

$\Rightarrow$ The Minnesota prior works well in practice and remains widely used

Natural conjugate prior

⇒ A drawback of the Minnesota prior is its treatment of $\boldsymbol{\Sigma}$: Ideally we want to treat $\boldsymbol{\Sigma}$ as an unknown parameter

⇒ The natural conjugate prior allows us to do this while retaining analytical results:

$$\boldsymbol{\alpha} \mid \boldsymbol{\Sigma} \sim \mathcal{N}(\underline{\boldsymbol{\alpha}}, \boldsymbol{\Sigma} \otimes \underline{\boldsymbol{V}}) \qquad \boldsymbol{\Sigma}^{-1} \sim \mathcal{W}\left(\underline{s}, \underline{\boldsymbol{S}}^{-1}\right)$$

⇒ $\underline{\boldsymbol{\alpha}}$, $\underline{\boldsymbol{V}}$, $\underline{s}$, and $\underline{\boldsymbol{S}}$ are prior hyperparameters chosen by the researcher

⇒ Noninformative version: $\underline{s} = 0$, $\underline{\boldsymbol{S}} = \underline{\boldsymbol{V}}^{-1} = \tau \boldsymbol{I}$, $\tau \rightarrow 0$

Posterior under the natural conjugate prior

⇒ The posterior has analytical form:

$$\boldsymbol{\alpha} \mid \boldsymbol{\Sigma}, \boldsymbol{y} \sim \mathcal{N}(\overline{\boldsymbol{\alpha}}, \ \boldsymbol{\Sigma} \otimes \overline{\boldsymbol{V}}) \qquad \boldsymbol{\Sigma}^{-1} \mid \boldsymbol{y} \sim \mathcal{W}\left(\overline{s}, \overline{\boldsymbol{S}}^{-1}\right)$$

⇒ The marginal posterior for $\boldsymbol{\alpha}$ is a multivariate t-distribution with mean $\overline{\boldsymbol{\alpha}}$, degrees of freedom $\overline{s}$, and covariance:

$$\mathbb{V}(\boldsymbol{\alpha} \mid \boldsymbol{y}) = \frac{1}{\overline{s} - M - 1} \overline{\boldsymbol{S}} \otimes \overline{\boldsymbol{V}}$$

⇒ Predictive inference can also be done analytically (for one-step ahead forecasts)

Implementation of the natural conjugate VAR

```
tXX <- crossprod(X)                         # Compute X'X # K x K
tXY <- crossprod(X,Y)                       # Compute X'Y # K x M
tYY <- crossprod(Y)                         # Compute Y'Y # M x M

# Posterior moments
V_po.inv    <- (tXX + V_pr.inv)       # Posterior precision
V_po        <- solve(V_po.inv)        # Posterior variance-covariance
A_po        <- V_po%*%(tXY + V_pr.inv%*%A_pr) # Posterior mean
a_po        <- as.vector(A_po)        # Vectorise coefficients

s_po        <- s_pr + N/2                       # Posterior DoF
S_po        <- S_pr + (tYY + t(A_pr)%*%V_pr.inv%*%A_pr
                 - t(A_po)%*%V_po.inv%*%A_po)/2  # Posterior scaling
```

Implementation of the natural conjugate VAR (cont.)

```
# Monte carlo integration
nsave <- 10000 # No. of draws
A_store   <- array(0,c(nsave,K,M))
SIG_store <- array(0,c(nsave,M,M))

for(irep in 1:nsave){

#Step 1: Draw SIG|Y marginally from an inverse Wishart
SIGinv_draw <- matrix(rWishart(1,s_po,solve(S_po)),M,M)
SIG_draw    <- solve(SIGinv_draw)

#Step 2: Draw coefficients from conditional posterior vec(A)|SIG,Y
bigV_po  <- kronecker(SIG_draw,V_po)
cholV_po <- t(chol(bigV_po)) # Cholesky of bigV_po
a_draw   <- a_po + cholV_po%*%rnorm(K*M)
A_draw   <- matrix(a_draw,K,M)  # Reshape in matrix form

A_store[irep,,]   <- A_draw
SIG_store[irep,,] <- SIG_draw

}
```

Problems with the natural conjugate prior

⇒ The natural conjugate prior has the great advantage of analytical results, but has restrictive properties

⇒ To make problems concrete: For example, a VAR includes output growth and money supply growth, and the researcher wants to impose the neutrality of money

⇒ Neutrality implies coefficients on lagged money growth are zero in the output growth equation, but not necessarily in other equations

Problems with the natural conjugate prior (cont.)

- $\Rightarrow$ **Problem 1:** Cannot simply impose neutrality; if we allow different equations to have different explanatory variables, analytical results are not available

- $\Rightarrow$ **Problem 2:** Cannot "almost impose" neutrality through the prior; the prior covariance $\Sigma \otimes \underline{\boldsymbol{V}}$ forces the prior covariance of coefficients in any two equations to be proportional (one cannot tighten the prior on a single equation)

- $\Rightarrow$ Note that the Minnesota prior form $\underline{\boldsymbol{V}}_{Min}$ is not consistent with the natural conjugate prior

- $\Rightarrow$ "Asymmetric conjugate priors for large Bayesian VARs" in *Quantitative Economics* (Joshua Chan) addresses many of the problems mentioned above by treating the VAR as a set of M conjugate linear regression problems

Other Bayesian VAR priors

⇒ Many other Bayesian VAR priors have been proposed

⇒ Independent Normal-Wishart prior, steady-state VAR, priors based on macro theory (e.g., "Priors from General Equilibrium Models for VARs")

⇒ Many machine learning VAR priors (e.g., Bayesian Lasso VAR, Normal Gamma, Triple Gamma, Dirichlet-Laplace, Horseshoe, etc.), see in "Adaptive Shrinkage in Bayesian VARs"

⇒ The BEAR Toolbox provides details of several of them

Stochastic Search Variable Selection (SSVS) in VARs

- ⇒ Many approaches seek parsimony in VARs through global-local shrinkage priors; SSVS is one example

- ⇒ SSVS places a hierarchical prior on each VAR coefficient α_j; a mixture of two Normals:

$$\alpha_j \mid \gamma_j \sim (1-\gamma_j)\,\mathcal{N}\left(0, \kappa_{0j}^2\right) + \gamma_j \mathcal{N}\left(0, \kappa_{1j}^2\right)$$

- ⇒ $\gamma_j \in \{0, 1\}$ is an unknown binary indicator estimated from the data

- ⇒ "Bayesian stochastic search for VAR model restrictions" in *Journal of Econometrics* (George, Sun and Ni)

SSVS: Choosing the mixture variances

⇒ If $\gamma_j = 1$: $\alpha_j \sim \mathcal{N}(0, \kappa_{1j}^2)$ with **large** variance; i.e., relatively noninformative

⇒ If $\gamma_j = 0$: $\alpha_j \sim \mathcal{N}(0, \kappa_{0j}^2)$ with **small** variance; i.e., coefficient shrunk to virtually zero

⇒ "Default semi-automatic approach":

$$\kappa_{0j} = c_0 \sqrt{\widehat{\mathrm{var}}(\alpha_j)}, \qquad \kappa_{1j} = c_1 \sqrt{\widehat{\mathrm{var}}(\alpha_j)}$$

where $\widehat{\mathrm{var}}(\alpha_j)$ is estimated from an unrestricted VAR (e.g., with OLS)

⇒ Constants $c_0 \ll c_1$ (e.g. $c_0 = 0.1$, $c_1 = 10$)

SSVS: Matrix form and prior for γ

- $\Rightarrow$ The SSVS prior can be written as: $\boldsymbol{\alpha} \mid \gamma \sim \mathcal{N}(\mathbf{0}, \boldsymbol{DD})$, where $\boldsymbol{D}$ is diagonal with (j, j) element:

$$d_j = \begin{cases} \kappa_{0j} & \text{if } \gamma_j = 0 \\ \kappa_{1j} & \text{if } \gamma_j = 1 \end{cases}$$

- $\Rightarrow$ Prior for γ: $\Pr(\gamma_j = 1) = \underline{q}_j$; setting $\underline{q}_j = \frac{1}{2}$ makes each coefficient equally likely *a priori* to be included or excluded

- $\Rightarrow$ Use the same Wishart prior for $\boldsymbol{\Sigma}^{-1}$

- $\Rightarrow$ George, Sun and Ni also show how to apply SSVS to the off-diagonal elements of $\boldsymbol{\Sigma}$

Gibbs Sampler for the SSVS VAR

⇒ The Gibbs sampler sequentially draws from:

$$p(\boldsymbol{\alpha} \mid \boldsymbol{y}, \gamma, \boldsymbol{\Sigma}), \quad p(\gamma \mid \boldsymbol{y}, \alpha, \boldsymbol{\Sigma}), \quad p(\boldsymbol{\Sigma}^{-1} \mid \boldsymbol{y}, \gamma, \boldsymbol{\alpha})$$

⇒ $\boldsymbol{\alpha} \mid \boldsymbol{y}, \gamma, \boldsymbol{\Sigma} \sim \mathcal{N}(\overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{V}})$ and $\boldsymbol{\Sigma}^{-1} \mid \boldsymbol{y}, \gamma, \boldsymbol{\alpha} \sim \mathcal{W}(\overline{s}, \overline{\boldsymbol{S}}^{-1})$

⇒ Posterior inclusion probabilities $\overline{q}_j = \Pr(\gamma_j = 1 \mid \boldsymbol{y}, \alpha, \boldsymbol{\Sigma})$ have simple closed forms

⇒ For efficient Gibbs Sampling in non-conjugate VARs, see "Large Bayesian VARs with SV and non-conjugate priors" in *Journal of Econometrics* (Carriero, Chan, Clark and Marcellino)

Illustration: Bayesian VAR for small macro data set

⇒ Data: Standard quarterly US data set from 1953:Q1 to 2006:Q3

⇒ Variables: inflation rate $\Delta\pi_t$, unemployment rate u_t, interest rate r_t (a New Keynesian VAR)

⇒ Model: Unrestricted VAR with intercept and 4 lags (common with quarterly data)

⇒ Consider different VAR priors: Noninformative versus independent SSVS-Wishart

Results: Posterior Means and Inclusion Probabilities

$\Rightarrow$ Point estimates of VAR coefficients are often not of primary interest, but Table 1 presents them for two priors

$\Rightarrow$ With SSVS, $\Pr(\gamma_j = 1 \mid \boldsymbol{y})$ is the **posterior inclusion probability** (see Table 2)

$\Rightarrow$ Model selection using $\Pr(\gamma_j = 1 \mid \boldsymbol{y}) > \frac{1}{2}$ restricts 25 of 39 coefficients to zero

Table 1. Posterior mean of VAR coefficients

Posterior mean of VAR Coefficients for Two Priors						
	Noninformative			SSVS-VAR		
	$\Delta\pi_t$	u_t	r_t	$\Delta\pi_t$	u_t	r_t
Intercept	0.292	0.322	−0.014	0.205	0.317	0.014
$\Delta\pi_{t-1}$	1.509	0.004	0.549	1.504	0.004	0.395
u_{t-1}	−0.266	1.273	−0.719	−0.142	1.256	−0.565
r_{t-1}	−0.057	−0.021	0.775	−0.001	−0.009	0.786
$\Delta\pi_{t-2}$	−0.468	0.101	−0.775	−0.505	0.006	−0.226
u_{t-2}	0.197	−0.310	0.788	0.074	−0.325	0.537
r_{t-2}	0.063	−0.023	−0.029	0.002	−0.008	−0.000
$\Delta\pi_{t-3}$	−0.077	−0.188	0.817	−0.007	0.005	0.002
u_{t-3}	−0.014	−0.129	−0.355	0.023	−0.044	−0.008
r_{t-3}	−0.007	0.097	0.100	−0.000	0.056	0.112
$\Delta\pi_{t-4}$	0.037	0.115	−0.485	−0.001	0.003	−0.058
u_{t-4}	0.037	0.067	0.311	0.016	0.014	0.056
r_{t-4}	−0.001	−0.025	0.059	−0.001	−0.003	0.001

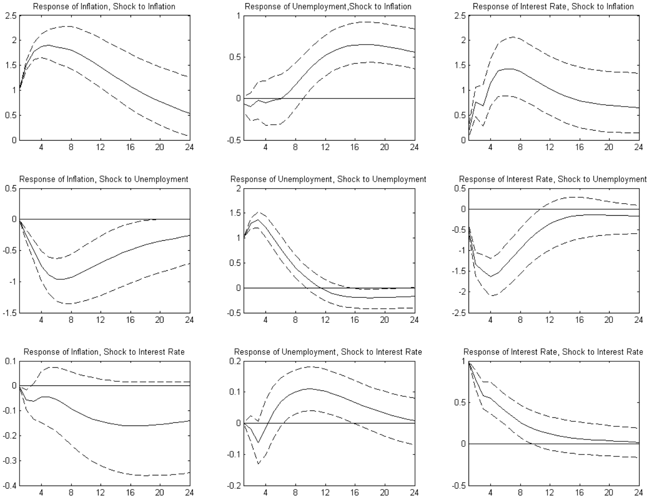
Table 2. Posterior inclusion probabilities

	$\Delta\pi_t$	u_t	r_t
Intercept	0.726	0.967	0.103
$\Delta\pi_{t-1}$	1.000	0.065	0.953
u_{t-1}	0.793	1.000	0.875
r_{t-1}	0.061	0.239	1.000
$\Delta\pi_{t-2}$	0.994	0.034	0.513
u_{t-2}	0.429	0.905	0.781
r_{t-2}	0.058	0.206	0.104
$\Delta\pi_{t-3}$	0.081	0.030	0.128
u_{t-3}	0.223	0.216	0.102
r_{t-3}	0.042	0.859	0.662
$\Delta\pi_{t-4}$	0.065	0.051	0.278
u_{t-4}	0.213	0.141	0.237
r_{t-4}	0.056	0.172	0.110

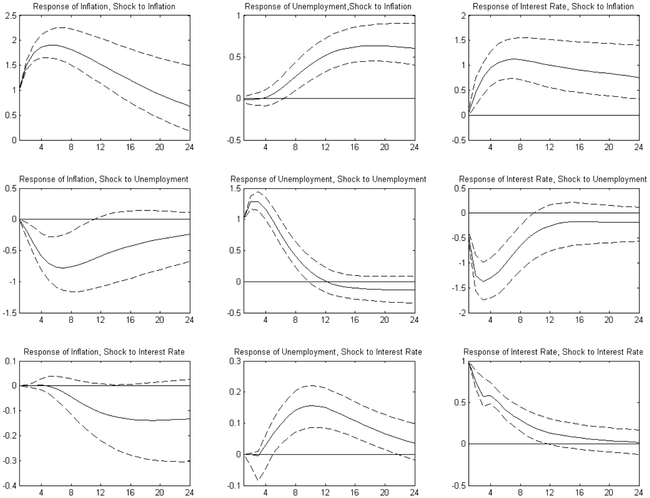
Impulse response analysis

- ⇒ Impulse response analysis is commonly performed with VARs
- ⇒ A standard identifying assumption is used, allowing the interest rate shock to be interpreted as a monetary policy shock
- ⇒ Figures 2 and 3 present impulse responses of all variables to shocks using a noninformative and a SSVS prior
- ⇒ The posterior median is the solid line; dotted lines are the 10th and 90th percentiles (80% credible bands)
- ⇒ The two priors give broadly similar results
- ⇒ A careful examination reveals that SSVS leads to slightly more precise inferences (narrower credible bands) due to shrinkage

Impulse responses for noninformative prior



Impulse responses for SSVS prior



Useful R Packages for conjugate/non-conjugate BVARs

⇒ **Conjugate priors**

→ `BVAR`: Hierarchical Bayesian VARs with conjugate priors (Minnesota-type)

⇒ **Non-conjugate shrinkage priors**

→ `bayesianVARs`: Efficient MCMC with a range of global-local shrinkage priors

→ `bsvars`: Bayesian structural VARs; see `bsvars.org` and `bsvarSIGNs` for sign-identified models

⇒ **Large-scale and multi-country models**

→ `BGVAR` and `bpvars`: Global and panel VARs for large-scale, multi-country macro-financial data

Summary

$\Rightarrow$ This lecture covered the basic methods and issues in Bayesian VAR modelling:

- $\rightarrow$ Why is shrinkage necessary?
- $\rightarrow$ How should shrinkage be done?

$\Rightarrow$ With the recent explosion of interest in large (flexible) VARs, the need for answers to such questions has greatly increased

$\Rightarrow$ Many researchers are now developing models and methods to address them

$\Rightarrow$ Algorithms developed in Carriero et al. (*JoE*, 2019) and Chan (*QE*, 2022) allow one to treat the VAR as a set of independent regression models