Macroeconometrics

Lecture 8 Bayesian VARs

Tomasz Woźniak

Department of Economics
University of Melbourne

Compulsory reading:

Woźniak (2016) Bayesian Vector Autoregressions, Australian Economic Review

Objectives.

- To start working with Bayesian Vector Autoregression
- To introduce the Minnesota prior
- To derive the joint posterior distribution of VAR parameters

Learning outcomes.

- Specifying the prior distribution
- Working with matrix-variate normal–Wishart distribution
- Completing the squares to derive the posterior distribution

Vector autoregressions

VAR(p) model.

$$
\begin{aligned}
y_t &= \mu_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t \\
\epsilon_t | Y_{t-1} &\sim iid\mathcal{N}_N\left(\mathbf{0}_N, \Sigma\right)
\end{aligned}
$$

Matrix notation.

$$
\begin{aligned}
Y &= X{\color{red}A} + E \\
E|X &\sim \mathcal{MN}_{T\times N}\left(\mathbf{0}_{T\times N}, {\color{red}\Sigma}, I_T\right)
\end{aligned}
$$

$$
\underset{(K\times N)}{{\color{red}A}} = \begin{bmatrix} \mu_0' \\ A_1' \\ \vdots \\ A_p' \end{bmatrix} \quad \underset{(T\times N)}{Y} = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_T' \end{bmatrix} \quad \underset{(K\times 1)}{x_t} = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad \underset{(T\times K)}{X} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix} \quad \underset{(T\times N)}{E} = \begin{bmatrix} \epsilon_1' \\ \epsilon_2' \\ \vdots \\ \epsilon_T' \end{bmatrix}
$$

where $K = 1 + pN$

Useful distributions

Normal inverse Wishart family of distributions

Matrix-variate normal distribution

A $K \times N$ matrix A is said to follow a matrix-variate normal distribution:

$$A \sim \mathcal{MN}_{K \times N}(M, Q, P),$$

where M is a $K \times N$ matrix and

Q $N \times N$ row-specific covariance matrix

P $K \times K$ column-specific covariance matrix

if $\text{vec}(A)$ is multivariate normal:

$$\text{vec}(A) \sim \mathcal{N}_{KN}(\text{vec}(M), Q \otimes P)$$

Density function.

$$\mathcal{MN}_{K \times N}(M, Q, P) = c_{mn}^{-1} \exp\left\{-\frac{1}{2}\text{tr}\left[Q^{-1}(A-M)'P^{-1}(A-M)\right]\right\}$$

$$c_{mn} = (2\pi)^{\frac{KN}{2}} \det(Q)^{\frac{K}{2}} \det(P)^{\frac{N}{2}}$$

Inverse Wishart distribution

An $N \times N$ square symmetric and positive definite matrix Σ follows an inverse Wishart distribution:

$$\Sigma \sim \mathcal{IW}_N(S, \nu)$$

where S is $N \times N$ positive definite symmetric matrix called the scale matrix and $\nu \geq N$ denotes degrees of freedom, if its density is given by:

$$\mathcal{IW}_N(S, \nu) = c_{iw}^{-1} \det(\Sigma)^{-\frac{\nu+N+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1} S\right]\right\}$$

$$c_{iw} = 2^{\frac{\nu N}{2}} \pi^{\frac{N(N-1)}{4}} \prod_{n=1}^{N} \Gamma\left(\frac{\nu+1+n}{2}\right) \det(S)^{-\frac{\nu}{2}}$$

Moments.

$$\mathbb{E}[\Sigma] = \frac{1}{\nu - N - 1} S \qquad \text{for } \nu > N + 1$$

Normal-Inverse Wishart distribution

$$A|\Sigma \sim \mathcal{MN}_{K\times N}\left(M, \Sigma, P\right)$$
$$\Sigma \sim \mathcal{IW}_{N}\left(S, \nu\right)$$

then the joint distribution of (A, Σ) is normal-inverse Wishart

$$p(A, \Sigma) = \mathcal{NIW}_{K\times N}\left(M, P, S, \nu\right)$$

with the density given by:

$$\begin{aligned} p\left(A, \Sigma\right) = c_{nw}^{-1} &\det(\Sigma)^{-(\nu+N+K+1)/2} \\ &\times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}(A-M)'P^{-1}(A-M)\right]\right\} \\ &\times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\Sigma^{-1}S\right]\right\} \end{aligned}$$

$$c_{nw} = 2^{\frac{N(K+\nu)}{2}}\pi^{\frac{N(N+2K-1)}{4}}\left[\prod_{n=1}^{N}\Gamma\left(\frac{\nu+1-n}{2}\right)\right]\det(P)^{\frac{N}{2}}\det(S)^{-\frac{\nu}{2}}$$

Likelihood function

Likelihood Function

VAR model.

$$Y = XA + E$$
$$Y|X, A, \Sigma \sim \mathcal{MN}_{T\times N}(XA, \Sigma, I_T)$$

Likelihood function.

$$\begin{aligned} L(A, \Sigma|Y, X) &\propto \det(\Sigma)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(Y - XA)'(Y - XA)\right]\right\} \\ &= \det(\Sigma)^{-\frac{T}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(A - \widehat{A})'X'X(A - \widehat{A})\right]\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(Y - X\widehat{A})'(Y - X\widehat{A})\right]\right\} \end{aligned}$$

where $\widehat{A} = (X'X)^{-1}X'Y$

Likelihood Function

Likelihood function.

The likelihood function can be presented as a normal-inverse Wishart distribution for (A, Σ)

$$L\left(A, \Sigma | Y, X\right) = \mathcal{NIW}_{K\times N}\left(\widehat{A}, (X'X)^{-1}, (Y - X\widehat{A})'(Y - X\widehat{A}), T - N - K - 1\right)$$

Prior distribution

Natural-conjugate prior distribution

Leads to joint posterior distribution for (A, Σ) of the same form

$$p(A, \Sigma) = p(A|\Sigma)\, p(\Sigma)$$
$$A|\Sigma \sim \mathcal{MN}_{K\times N}\left(\underline{A}, \Sigma, \underline{V}\right)$$
$$\Sigma \sim \mathcal{IW}_N\left(\underline{S}, \underline{\nu}\right)$$

Kernel.

$$p(A, \Sigma) \propto \det(\Sigma)^{-\frac{N+K+\underline{\nu}+1}{2}}$$
$$\times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(A-\underline{A})'\underline{V}^{-1}(A-\underline{A})\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\underline{S}\right]\right\}$$

Minnesota prior

Minnesota prior

Proposed originally by

Doan, Litterman, Sims (1984) Forecasting and Conditional Projection Using Realistic Prior Distributions, Econometric Reviews

Presented in a version by

Karlsson (2013) Forecasting with Bayesian vector autoregression, in: Handbook of Economic Forecasting

Objective.

Use a natural-conjugate prior for VARs for feasible posterior computations and set its parameters to express stylised facts about macroeconomic time series

Minnesota prior

Stylised fact #1
Macroeconomic variables are unit-root nonstationary and are well-characterised by a multivariate random walk process

$$y_t = y_{t-1} + \epsilon_t$$

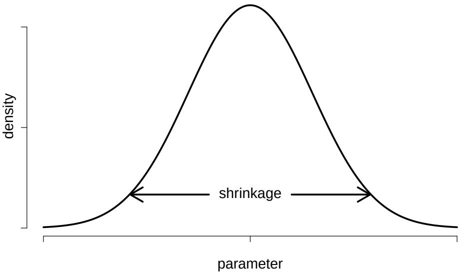
Set the prior mean of A to

$$\underline{A} = \begin{bmatrix} \mathbf{0}_{N\times 1} & I_N & \mathbf{0}_{N\times(p-1)N} \end{bmatrix}'$$

Minnesota prior: shrinkage

Prior shrinkage

is the dispersion of the prior distribution around the prior mean, $\underline{A}$.
It is determined e.g. by the diagonal elements of $\underline{V}$.



Prior covariance matrix.

$$\mathbb{V}\text{ar}[\text{vec}(A)] = \Sigma \otimes \underline{V}$$

Minnesota prior

Set the column-specific prior covariance of A to

$$\underline{V} = \text{diag}\left(\begin{bmatrix}\kappa_2 & \kappa_1\left(\mathbf{p}^{-2} \otimes \imath_N'\right)\end{bmatrix}\right)$$

$$\mathbf{p} = \begin{bmatrix}1 & 2 & \dots & p\end{bmatrix}$$

$\imath_N = \texttt{rep(1,N)}$ – an $N \times 1$-vector of ones

κ_1 – overall shrinkage level for autoregressive slopes

κ_2 – overall shrinkage for the constant term

Prior covariance matrix.

$$\mathbb{V}\text{ar}[\text{vec}(A)] = \mathbf{\Sigma} \otimes \underline{V}$$

Minnesota prior

$$\underline{V} = \text{diag}\left(\begin{bmatrix}\kappa_2 & \kappa_1\left(\mathbf{p}^{-2} \otimes \imath_N'\right)\end{bmatrix}\right)$$

Stylised fact #2.
Strongly favour the unit-root hypothesis
Set prior variances for autoregressive slopes to a small number
$\kappa_1 = 0.02^2$

Stylised fact #3.
The effect of more distant lags of y_t is smaller and smaller
Shrink parameters of autoregressive slopes more towards $\underline{A}$ with incleasing lag l by dividing the corresponding prior variances by l^2 for l corresponding to subsequent elements of $\mathbf{p}$

Stylised fact #4.
The data are little informative about the values of μ_0
Shrink μ_0 much less than autoregressive parameters $\kappa_2 \gg \kappa_1$
For instance $\kappa_2 = 100$

Posterior distribution

Posterior distribution

$$
\begin{aligned}
p\left(A, \Sigma | Y, X\right) &\propto L(A, \Sigma | Y, X) p\left(A, \Sigma\right) \\
&= L(A, \Sigma | Y, X) p\left(A | \Sigma\right) p\left(\Sigma\right)
\end{aligned}
$$

Kernel.

$$
\begin{aligned}
p\left(A, \Sigma | Y, X\right) \propto \det(\Sigma)^{-\frac{T}{2}} \\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(A - \widehat{A})'X'X(A - \widehat{A})\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(Y - X\widehat{A})'(Y - X\widehat{A})\right]\right\} \\
&\times \det(\Sigma)^{-\frac{N+K+\underline{\nu}+1}{2}} \\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(A - \underline{A})'\underline{V}^{-1}(A - \underline{A})\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\underline{S}\right]\right\},
\end{aligned}
$$

Posterior distribution

Kernel.

$$
\begin{aligned}
p\left(A, \Sigma | Y, X\right) &\propto \det(\Sigma)^{-\frac{T+N+K+\underline{\nu}+1}{2}} \\
&\times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\left[(A-\widehat{A})'X'X(A-\widehat{A}) + (A-\underline{A})'\underline{V}^{-1}(A-\underline{A})\right.\right.\right. \\
&\qquad \left.\left.\left. +(Y-X\widehat{A})'(Y-X\widehat{A}) + \underline{S}\right]\right]\right\}
\end{aligned}
$$

Apply the transformations and complete the squares to show that

$$
\begin{aligned}
&(A-\widehat{A})'X'X(A-\widehat{A}) + (A-\underline{A})'\underline{V}^{-1}(A-\underline{A}) + (Y-X\widehat{A})'(Y-X\widehat{A}) + \underline{S} \\
&\qquad = (A-\overline{A})'\overline{V}^{-1}(A-\overline{A}) + \underline{S} + Y'Y + \underline{A}'\underline{V}^{-1}\underline{A} - \overline{A}'\overline{V}^{-1}\overline{A}
\end{aligned}
$$

Now, present the kernel as the normal-inverse Wishart distribution

Joint posterior distribution

$$p\left(A, \Sigma | Y, X\right) = p(A|Y, X, \Sigma) p\left(\Sigma | Y, X\right)$$

$$\begin{aligned} p(A|Y, X, \Sigma) &= \mathcal{MN}_{K\times N}\left(\overline{A}, \Sigma, \overline{V}\right) \\ p(\Sigma|Y, X) &= \mathcal{IW}_N\left(\overline{S}, \overline{\nu}\right) \end{aligned}$$

$$\begin{aligned} \overline{V} &= \left(X'X + \underline{V}^{-1}\right)^{-1} \\ \overline{A} &= \overline{V}\left(X'Y + \underline{V}^{-1}\underline{A}\right) \\ \overline{\nu} &= T + \underline{\nu} \\ \overline{S} &= \underline{S} + Y'Y + \underline{A}'\underline{V}^{-1}\underline{A} - \overline{A}'\overline{V}^{-1}\overline{A} \end{aligned}$$

Posterior mean of A

Posterior mean of matrix A is:

$$
\begin{aligned}
\overline{A} &= \overline{V}\left(X'Y + \underline{V}^{-1}\underline{A}\right) \\
&= \overline{V}\left(X'X\widehat{A} + \underline{V}^{-1}\underline{A}\right) \\
&= \overline{V}X'X\widehat{A} + \overline{V}\underline{V}^{-1}\underline{A}
\end{aligned}
$$

a linear combination of the MLE $\widehat{A}$ and the prior mean $\underline{A}$

Bayesian VARs

Bayesian VARs are benchmark models for macroeconomic forecasting

Closed-form solutions to the estimation problem allow fast computations

Minnesota prior reflects stylized facts about macroeconomic time series

Large Bayesian VARs use the Kronecker structure of covariance matrix and the application of shrinkage for precise economic forecasting