Tutorial 5: Bayesian VARs

Review the Concepts and Proofs

1. Why is shrinkage particularly important for VAR models?
2. What are the marginal distributions of a Normal-inverse Wishart distribution?
3. Explain the specification of the Minnesota prior.
4. Explain the fictitious sample interpretation of the natural conjugate prior for normally distributed VAR models.
5. In what respect in the natural conjugate prior more restrictive than the independent Normal-Wishart prior? Give an example.
6. How can you find a symmetric 90 percent H step ahead forecast interval for a VAR model that is estimated (a) with a natural conjugate prior and (b) with an independent Normal-Wishart prior?

Paper-pen exercises

1. Consider the M-dimensional VAR(p) model

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$.

(a) Show that (1) can be cast in *matrix representation*

$$Y = XA + E$$

and in *vector representation*

$$y = \mathbf{X}\alpha + \varepsilon.$$

Find the dimensions of the vectors and matrices.

(b) Find the distribution of ε.

(c) Show that the joint pdf of y (conditional on p pre-sample observations $y_0, \dots, y_{1-p}$) can be written as

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\}.$$

To this end, use an appropriate recursive factorization of the joint distribution in conditionals and marginals.

(d) (*) Find the distribution of y conditional on p pre-sample observations $y_0, \dots, y_{1-p}$. To this end, apply the transformation technique

$$f(y) = \left|\frac{\partial \varepsilon}{\partial y'}\right| f_\varepsilon(\varepsilon)$$

and use without proof $\left|\frac{\partial \varepsilon}{\partial y'}\right| = 1$.

(e) (*) Show that the likelihood of the VAR(p) model has a Normal-Wishart structure.

2. Consider the VAR(p) model (1). Find the posterior distribution of α and Σ if you use the improper prior $f(\alpha, \Sigma) = |\Sigma|^{-\frac{M+\phi}{2}}$.

3. Consider the VAR(p) model (1). Find the posterior distribution of α if you use the Minnesota prior $\Sigma = \hat{\Sigma}$ and $\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{V}_M)$.

4. Consider the VAR(p) model (1). Find the posterior distribution of α and Σ if you use the natural conjugate prior $\Sigma \sim iW(\underline{S}, \underline{\nu})$ and $\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$.

5. Consider the VAR(p) model (1). Show that the following two estimation procedures yield the same posterior distributions of α and Σ: Applying the natural conjugate prior $\Sigma \sim iW(\underline{S}, \underline{\nu})$ and $\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$ to the data Y and X, or applying the diffuse prior $f(\alpha, \Sigma) \propto |\Sigma|^{-\frac{M+1}{2}}$ to the augmented data

$$Y_* = \begin{pmatrix} Y \\ Y_0 \end{pmatrix}, \quad X_* = \begin{pmatrix} X \\ X_0 \end{pmatrix}, \quad E_* = \begin{pmatrix} E \\ E_0 \end{pmatrix}$$

where Y_0 and X_0 contain T_0 fictitious observations chosen to satisfy the conditions (C.1) $\underline{V} = (X_0'X_0)^{-1}$, (C.2) $\underline{\alpha} = \text{vec}(\underline{A}) = \text{vec}(\hat{A}_0) \equiv \text{vec}[(X_0'X_0)^{-1}X_0'Y_0]$, and (C.3) $\underline{S} = (Y_0 - X_0\hat{A}_0)'(Y_0 - X_0\hat{A}_0)$.

6. Consider the VAR(p) model (1) together with the independent Normal-Wishart prior $\Sigma \sim iW(\underline{S}, \underline{\nu})$ and $\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{W})$.

   (a) Find the posterior $f(\alpha, \Sigma|y)$.

   (b) Show that the posterior can be factorized into $f(\alpha|y, \Sigma)f(\Sigma|y)$, where $f(\alpha|y, \Sigma)$ is a normal distribution with posterior mean $\bar{\alpha}$ and posterior variance matrix $\bar{W}$ while $f(\Sigma|y)$ is an unknown distribution.

   (c) Show that the posterior can be factorized into $f(\Sigma|y, \alpha)f(\alpha|y)$, where $f(\Sigma|y, \alpha)$ is an inverse Wishart distribution with parameters $\bar{S}$ and $\bar{\nu}$ while $f(\alpha|y)$ is an unknown distribution.

   (d) How would you simulate the posterior?

Appendix 1: The Wishart distribution

A good reference is Steven W. Nydick (2012), The Wishart and Inverse Wishart Distributions, downloadable here.

Let H be an $M \times M$ random matrix that follows a Wishart distribution with parameters S and ν, $H \sim W(S, \nu)$. Then it has pdf

$$f_W(H|S,\nu) = c_W^{-1}|S|^{-\frac{\nu}{2}}|H|^{\frac{\nu-M-1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(S^{-1}H)\right],$$

where $c_W = 2^{\frac{\nu M}{2}}\pi^{\frac{M(M-1)}{4}}\prod_{i=1}^{M}\Gamma(\frac{\nu+1-i}{2})$ is an integration constant, $\nu > M-1$ is a scalar parameter, and S is an $M \times M$ symmetric and positive definite scale matrix. The expectation is

$$\mathrm{E}(H) = \nu S.$$

The Wishart distribution is often used as a prior for the precision matrix (=inverse of the variance matrix).

Let Σ be an $M \times M$ random matrix that follows an inverse Wishart distribution with parameters Ψ and δ, $\Sigma \sim iW(\Psi, \delta)$. Then it has pdf

$$f_{iW}(\Sigma|\Psi,\delta) = c_{iW}^{-1}|\Psi|^{\frac{\delta}{2}}|\Sigma|^{-\frac{\delta+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\Psi\Sigma^{-1})\right],$$

where $c_{iW} = 2^{\frac{\delta M}{2}}\pi^{\frac{M(M-1)}{4}}\prod_{i=1}^{M}\Gamma(\frac{\delta+1-i}{2})$ is an integration constant, $\delta > M-1$ is a scalar parameter, and Ψ is an $M \times M$ symmetric and positive definite scale matrix. The expectation is

$$\mathrm{E}(\Sigma) = \Psi/(\delta - M - 1).$$

The inverse Wishart distribution is often used as a prior for the variance matrix.

Relationship between Wishart and inverse Wishart distribution: Let $H \sim W(S,\nu)$. Then $\Sigma \equiv H^{-1} \sim iW(S^{-1},\nu)$.

Appendix 2: The matrix variate t distribution

The $p \times q$ random matrix X has matrix variate t distribution, $X \sim MT(\mu, V, S, \nu)$ if it has pdf

$$f(X) = c_{MT}^{-1}|S|^{\frac{\nu}{2}}|V|^{-\frac{q}{2}}|S + (X-\mu)'V^{-1}(X-\mu)|^{-\frac{\nu+p}{2}}$$

where μ is a $p \times q$ symmetric and positive definite matrix, S is a symmetric and positive definite $q \times q$ matrix, V is a $p \times p$ matrix, $\nu > q-1$ is a scalar and

$$c_{MT} = \pi^{\frac{pq}{2}} \prod_{i=1}^{q} \frac{\Gamma(\frac{\nu+1-i}{2})}{\Gamma(\frac{\nu+p+1-i}{2})}.$$

The matrix variate t distribution has mean

$$\mathrm{E}(X) = \mu, \qquad \nu > 1,$$

and variance

$$\mathrm{Var}[\mathrm{vec}(X)] = \frac{1}{\nu - q - 1} S \otimes V, \qquad \nu > q+1,$$

see Dreze and Richard (1983), Bayesian analysis of simultaneous equation systems, in: Griliches and Intriligator (eds.), Handbook of Econometrics, Vol., Chapter 9, p. 517-598.

A single element X_{ij} has mean

$$\mathrm{E}(X_{ij}) = \mu_{ij}, \qquad \nu > 1,$$

and variance

$$\mathrm{Var}(X_{ij}) = \frac{1}{\nu - q - 1} V_{ii} S_{jj}, \qquad \nu > q+1.$$

Its marginal distribution is the non-standardized t distribution, $X_{ij} \sim t(\mu_{ij}, \sigma_{ij}^2, \nu - q + 1)$, where $\sigma_{ij}^2 = V_{ii}S_{jj}/(\nu - q + 1)$.

The standardized element

$$T_{ij} = \frac{X_{ij} - \mu_{ij}}{\sigma_{ij}} = \frac{X_{ij} - \mu_{ij}}{\sqrt{V_{ii}S_{jj}/(\nu - q + 1)}}$$

has student t distribution, i.e., $T_{ij} \sim t(\nu - q + 1)$.

In a Bayesian context, the matrix variate t distribution appears as a marginal distribution of a normal-inverse Wishart distribution. Suppose

$$\operatorname{vec}(X)|\Sigma \sim \mathcal{N}(\operatorname{vec}(\mu), \Sigma \otimes V), \quad \Sigma \sim iW(S, \nu).$$

Then

$$X \sim MT(\mu, V, S, \nu).$$

Appendix 3: Rules for matrices

Suppose all matrices P, Q, R, and S are of appropriate dimensions. Then

1. $\operatorname{vec}(QPR) = (R' \otimes Q) \operatorname{vec}(P)$
2. $\operatorname{tr}(PQR) = \operatorname{tr}(RPQ)$
3. $\operatorname{tr}(PQR) = \operatorname{vec}(P')'(C' \otimes I) \operatorname{vec}(Q)$
4. $\operatorname{tr}(PQRS) = \operatorname{vec}(P')'(S' \otimes Q) \operatorname{vec}(R)$
5. $|P \otimes Q| = |P|^n |Q|^m$ for an $m \times m$ matrix P and an $n \times n$ matrix Q