Reduced-Form Vector Autoregressions

Frank Schorfheide

University of Pennsylvania

Econ 722 – Part 1

January 29, 2019

Model Specification

- A vector autoregression is a generalization of the AR(p) model to the multivariate case:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t$$

- y_t is a $n \times 1$ random vector that takes values in $\mathbb{R}^n$.
- $u_t \sim iid(0, \Sigma)$ is a vector of reduced-form innovations.

Likelihood Function

- Define the $(np+1) \times 1$ vector x_t as

$$x_t = [y'_{t-1}, \dots, y'_{t-p}, 1]'$$

- Moreover, define the matrixes

$$Y = \left[\begin{array}{c} y'_1 \\ \vdots \\ y'_T \end{array} \right], \quad X = \left[\begin{array}{c} x'_1 \\ \vdots \\ x'_T \end{array} \right], \quad \Phi = [\Phi_1, \dots, \Phi_p, \Phi_c]'$$

- The conditional density of y_t:

$$\begin{aligned} & p(y_t | Y_{1:t-1}, Y_{1-p:0}, \Phi, \Sigma) \\ & \quad \propto \quad |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (y'_t - x'_t \Phi) \Sigma^{-1} (y'_t - x'_t \Phi)' \right\}, \end{aligned}$$

where $Y_{t_0:t_1} = [y_{t_0}, \dots y_{t_1}]$.

Some Matrix Algebra

- **Definition:** Let A and B be 2×2 matrices with the elements

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

  The *vec* operator is defined as the operator that stacks the columns of a matrix, that is,

$$vec(A) = [a_{11}, a_{21}, a_{12}, a_{22}]'$$

  and the Kronecker product is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix} \quad \square$$

- **Lemma:** Let A, B, C be matrices whose dimension are such that the product ABC exists. Then

$$vec(ABC) = (C' \otimes A)vec(B) \quad \square$$

Some Matrix Algebra

- **Lemma:** Let A and B be two $n \times n$ matrices, then

$$tr[A + B] = tr[A] + tr[B] \quad \square$$

- **Lemma:** Let a be a $n \times 1$ vector, B be a symmetric positive definite $n \times n$ matrix, and tr the trace operator that sums the diagonal elements of a matrix. Then

$$a'Ba = tr[Baa'] \quad \square$$

Likelihood Function

- Write

$$\exp\left\{-\frac{1}{2}(y_t' - x_t'\Phi)\Sigma^{-1}(y_t' - x_t'\Phi)'\right\}$$

- as

$$\exp\left\{-\frac{1}{2}tr[\Sigma^{-1}(y_t' - x_t'\Phi)'(y_t' - x_t'\Phi)]\right\}.$$

Likelihood Function

- Take the product of the conditional densities of $y_1, \ldots, y_T$ to obtain the joint density.
- Let $Y_{1-p:0}$ be a vector with initial observations

$$
\begin{aligned}
& p(Y_{1:T}|Y_{1-p:0}, \Phi, \Sigma) \\
= & \prod_{t=1}^{T} p(y_t|Y_{1:t-1}, Y_{1-p:0}, \Phi, \Sigma) \\
\propto & |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} tr[\Sigma^{-1}(y_t' - x_t'\Phi)'(y_t' - x_t'\Phi)] \right\} \\
\propto & |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr \left[\Sigma^{-1} \sum_{t=1}^{T} (y_t' - x_t'\Phi)'(y_t' - x_t'\Phi) \right] \right\} \\
\propto & |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(Y - X\Phi)'(Y - X\Phi)] \right\}
\end{aligned}
$$

Likelihood Function

- Define the "OLS" estimator

$$\hat{\Phi} = (X'X)^{-1}X'Y.$$

- Define the sum of squared OLS residual matrix

$$\hat{S} = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}) = Y'Y - Y'X(X'X)^{-1}X'Y.$$

- It can be verified that

$$(Y - X\Phi)'(Y - X\Phi) = (\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi}) + \hat{S}$$

Likelihood Function

- This leads to the following representation of the likelihood function

$$p(Y|\Phi,\Sigma) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}\hat{S}]\right\} \times \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi-\hat{\Phi})'X'X(\Phi-\hat{\Phi})]\right\}.$$

- Let $\beta = vec(\Phi)$ and $\hat{\beta} = vec(\hat{\Phi})$. It can be verified that

$$tr[\Sigma^{-1}(\Phi-\hat{\Phi})'X'X(\Phi-\hat{\Phi})] = (\beta-\hat{\beta})'[\Sigma \otimes (X'X)^{-1}]^{-1}(\beta-\hat{\beta}).$$

- The likelihood function has the alternative representation

$$p(Y|\Phi,\Sigma) \propto |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}\hat{S}]\right\} \times \exp\left\{-\frac{1}{2}(\beta-\hat{\beta})'[\Sigma \otimes (X'X)^{-1}]^{-1}(\beta-\hat{\beta})\right\}.$$

"Inverting" the Likelihood Function

- Suppose that

$$p(\Phi, \Sigma) \propto c.$$

- Then

$$p(\Phi, \Sigma | Y) \propto p(Y | \Phi, \Sigma).$$

Background: Matricvariate Normal Distribution

- Suppose that the random matrix Φ has density

$$p(\Phi|\Sigma, X'X) \propto |\Sigma \otimes (X'X)^{-1}|^{-1/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})]\right\}$$

  then $\Phi|(\Sigma, X'X)$ is matricvariate normal

$$MN(\hat{\Phi}, \Sigma \otimes (X'X)^{-1}).$$

- Let $\beta = vec(\Phi)$ and $\hat{\beta} = vec(\hat{\Phi})$. Then

$$\beta|\Sigma, X'X \sim N\left(\hat{\beta}, \Sigma \otimes (X'X)^{-1}\right).$$

- To generate a draw Z from a multivariate $N(\mu, \Sigma)$, decompose $\Sigma = CC'$. E.g., C could be the lower triangular Cholesky factor. Then let $Z = \mu + C \cdot N(0, I)$.

Background: Inverted Wishart Distribution

- Let Σ be a $n \times n$ positive definite random matrix. Σ has the Inverted Wishart $IW(S, \nu)$ distribution if its density is of the form

$$p(\Sigma|S,\nu) \propto |S|^{\nu/2}|\Sigma|^{-(\nu+n+1)/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}S]\right\}$$

- To sample a Σ from an Inverted Wishart $IW(S, \nu)$ distribution, draw $n \times 1$ vectors $Z_1, \ldots, Z_\nu$ from a multivariate normal $N(0, S^{-1})$ and let

$$\Sigma = \left[\sum_{i=1}^{\nu} Z_i Z_i'\right]^{-1}$$

Likelihood Function Interpreted as PDF for (Φ, Σ) – Step 1: $p(\Phi|\Sigma, Y)$

- Interpret the likelihood as density for (Φ, Σ):

$$
\begin{aligned}
& p(\Phi, \Sigma | Y) \\
& \quad \propto \quad |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} \hat{S}] \right\} \\
& \qquad \quad \times \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} (\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi})] \right\} \\
& \quad \propto \quad |\Sigma|^{-T/2} |\Sigma \otimes (X'X)^{-1}|^{1/2} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} \hat{S}] \right\} \\
& \qquad \quad \times (2\pi)^{-nk/2} |\Sigma \otimes (X'X)^{-1}|^{-1/2} \\
& \qquad \quad \times \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} (\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi})] \right\}.
\end{aligned}
$$

- Thus,

$$
\Phi | (\Sigma, Y) \sim MN(\hat{\phi}, \Sigma \otimes (X'X)^{-1}).
$$

Likelihood Function Interpreted as PDF for (Φ, Σ) – Step 2: $p(\Sigma|Y)$

- Compute $p(\Sigma|Y) \propto \int p(Y|\Phi,\Sigma) d\Phi$...
- Note that:

$$|\Sigma \otimes (X'X)^{-1}|^{1/2} = |\Sigma|^{k/2}|X'X|^{-n/2}.$$

- Therefore,

$$p(\Sigma|Y) \propto |\Sigma|^{-(T-k)/2}|X'X|^{-n/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}\hat{S}]\right\}.$$

- Deduce

$$\Sigma|Y \sim IW(\hat{S}, T-k-n-1), \quad \Phi|(\Sigma, Y) \sim MN(\hat{\phi}, \Sigma \otimes (X'X)^{-1}).$$

- Write as

$$(\Phi, \Sigma)|Y \sim MNIW\left(\hat{\Phi}, (X'X)^{-1}, \hat{S}, T-k-n-1\right).$$

Bayesian Analysis with Improper Prior

- Replace (improper) prior $p(\Phi, \Sigma) \propto c$ by (improper) prior:

$$p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2}.$$

- Then, the posterior is obtained from

$$p(\Phi, \Sigma | Y) \propto p(Y | \Phi, \Sigma) p(\Phi, \Sigma).$$

- Our previous analysis implies that

$$(\Phi, \Sigma) | Y \sim MNIW\left(\hat{\Phi}, (X'X)^{-1}, \hat{S}, T - k\right).$$

Algorithm: Direct Sampling of VAR Parameters

For $s = 1, \ldots, n_{sim}$:

1. Draw $\Sigma^{(s)}$ from an $IW(\hat{S}, T - k)$ distribution.
2. Draw $\Phi^{(s)}$ from the conditional distribution $MN(\hat{\Phi}, \Sigma^{(s)} \otimes (X'X)^{-1})$. $\square$

The Use of Priors for VARs

- Priors are used to "regularize" the VAR likelihood and cope with the dimensionality problem: the number of free parameters is often large relative to the number of observations.
- Priors add information to the estimation problem.

Prior with Kronecker Structure

- Consider the prior:

$$\Sigma \sim IW(\underline{\nu}, \underline{S}), \quad \Phi|\Sigma \sim MN\big(\underline{\mu}_{\Phi}, \Sigma \otimes \underline{P}_{\Phi}^{-1}\big), \quad .$$

- Prior density:

$$\begin{aligned} p(\Phi, \Sigma) &= (2\pi)^{-nk/2}|\Sigma|^{-k/2}|\underline{P}_{\Phi}|^{n/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\big[\Sigma^{-1}(\Phi - \underline{\mu}_{\Phi})'\underline{P}_{\Phi}(\Phi - \underline{\mu}_{\Phi})\right\} \\ &\quad \times \underline{C}_{IW}|\Sigma|^{-(\underline{\nu}+n+1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\big[\Sigma^{-1}\underline{S}\big]\right\}. \end{aligned}$$

- $\underline{C}_{IW}$ is the normalization constant of the IW prior.

Posterior

$$
\begin{aligned}
& p(Y|\Phi,\Sigma)p(\phi,\Sigma) \\
\propto \quad & (2\pi)^{-nk/2}|\Sigma|^{-k/2}|\underline{P}_\Phi|^{n/2}\underline{C}_{IW}|\Sigma|^{-(\underline{\nu}+n+1)/2}(2\pi)^{-nT/2}|\Sigma|^{-T/2} \\
& \times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(\Phi-\underline{\mu}_\Phi)'\underline{P}_\Phi(\Phi-\underline{\mu}_\Phi)\right]\right\}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(\Phi-\hat{\Phi})X'X(\Phi-\hat{\Phi})\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\underline{S}\right]\right\}\exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\hat{S}\right]\right\}
\end{aligned}
$$

Posterior

- Define

$$\bar{P}_\Phi = \underline{P}_\Phi + X'X, \quad \bar{\mu}_\Phi = \bar{P}_\Phi^{-1}\big[\underline{P}_\Phi \underline{\mu}_\Phi + X'X\hat{\Phi}\big].$$

- Write

$$\begin{aligned} & p(Y|\Phi,\Sigma)p(\phi,\Sigma) \\ = \; & (2\pi)^{-nk/2}|\Sigma|^{-k/2}|\underline{P}_\Phi|^{n/2}\underline{C}_{IW}|\Sigma|^{-(\underline{\nu}+n+1)/2}(2\pi)^{-nT/2}|\Sigma|^{-T/2}|\bar{P}_\phi|^{n/2}|\bar{P}_\phi|^{-n/2} \\ & \times \exp\left\{-\frac{1}{2}\text{tr}\big[\Sigma^{-1}(\Phi-\bar{\mu}_\Phi)\bar{P}_\phi(\Phi-\bar{\mu}_\Phi)\big]\right\} \\ & \times \exp\left\{-\frac{1}{2}\text{tr}\big[\Sigma^{-1}\big(\underline{S}+\underline{\mu}_\Phi'\underline{P}_\Phi\underline{\mu}_\Phi + Y'Y - \bar{\mu}_\Phi'\bar{P}_\Phi\bar{\mu}_\Phi\big)\big]\right\} \end{aligned}$$

Posterior

- Define

$$\bar{S} = \underline{S} + \underline{\mu}_\Phi' \underline{P}_\Phi \underline{\mu}_\Phi + Y'Y - \bar{\mu}_\Phi' \bar{P}_\Phi \bar{\mu}_\Phi, \quad \bar{\nu} = \underline{\nu} + T$$

- Deduce that

$$\Sigma | Y \sim IW(\bar{S}, \bar{\nu}), \quad \Phi | (\Sigma, Y) \sim MN(\bar{\mu}_\Phi, \Sigma \otimes \bar{P}^{-1}).$$

- Draws from the posterior can be generated by direct sampling.

Marginal Data Density

- Let $\bar{C}_{IW}$ be the normalization constant of the posterior IW distribution. Then,

$$
\begin{aligned}
p(Y) &= \int\int p(Y|\Phi,\Sigma)p(\phi,\Sigma)d\Phi d\Sigma \\
&= (2\pi)^{-nT/2}\int |\underline{P}_\Phi|^{k/2}|\bar{P}_\phi|^{-k/2}\underline{C}_{IW}|\Sigma|^{-(\bar{\nu}+n+1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}\bar{S}]\right\}d\Sigma \\
&= (2\pi)^{-nT/2}\frac{|\underline{P}_\Phi|^{n/2}}{|\bar{P}_\Phi|^{n/2}}\frac{\underline{C}_{IW}}{\bar{C}_{IW}},
\end{aligned}
$$

- where

$$
\frac{\underline{C}_{IW}}{\bar{C}_{IW}} = \frac{|\underline{S}|^{\underline{\nu}/2}}{|\bar{S}|^{\bar{\nu}/2}}\frac{2^{n_y\bar{\nu}/2}\prod_{i=1}^{n_y}\Gamma((\bar{\nu}+1-i)/2)}{2^{n_y\underline{\nu}/2}\prod_{i=1}^{n_y}\Gamma((\underline{\nu}+1-i)/2)}.
$$

Constructing a Prior from "Dummy Observations"

- Suppose we have T^* dummy observations (Y^*, X^*), plug the dummy observations into the likelihood function and multiply the likelihood function by the (initial) improper prior:

$$p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2} p(Y^*|\Phi, \Sigma).$$

- Define

$$\Phi^* = (X^{*'} X^*)^{-1} X^{*'} Y^*, \quad S^* = (Y^* - X^*\Phi^*)'(Y^* - X^*\Phi^*).$$

- This leads to a prior

$$\Sigma \sim IW(\underline{\nu}, \underline{S}), \quad \Phi|\Sigma \sim MN\big(\underline{\mu}_\Phi, \Sigma \otimes \underline{P}_\Phi^{-1}\big), \quad .$$

with

$$\underline{\nu} = T^* - k, \quad \underline{S} = S^*, \quad \underline{\mu}_\Phi = \Phi^*, \quad \underline{P}_\Phi = X^{*'} X^*.$$

Posterior with Dummy Observation Prior

- Posterior is proportional to

$$p(\Phi, \Sigma, Y) \propto p(Y|\Phi, \Sigma) p(Y^*|\Phi, \Sigma) |\Sigma|^{-(n+1)/2}$$

- Define $\bar{T} = T^* + T$ and

$$\begin{aligned}
\bar{\Phi} &= (X^{*'} X^* + X'X)^{-1}(X^{*'} Y^* + X'Y) \\
\bar{S} &= \Big[Y^{*'} Y^* + Y'Y) - (X^{*'} Y^* + X'Y)'(X^{*'} X^* + X'X)^{-1}(X^{*'} Y^* + X'Y) \Big].
\end{aligned}$$

- Then, let $\bar{X} = [X^{*'}, X']'$ and deduce:
- Deduce that

$$\Sigma | Y \sim IW(\bar{S}, \bar{T} - k), \quad \Phi | (\Sigma, Y) \sim MN(\bar{\Phi}, \Sigma \otimes (\bar{X}'\bar{X})^{-1}).$$

Example: Minnesota Prior

- Reference: Doan, Litterman, and Sims (1984), Sims and Zha (1998).
- Consider the following Gaussian bivariate VAR(2).

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

- Define $y_t = [y_{1,t}, y_{2,t}]'$, $x_t = [y'_{t-1}, y'_{t-2}, 1]'$, and $u_t = [u_{1,t}, u_{2,t}]'$ and

$$\Phi = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \gamma_{11} & \gamma_{21} \\ \gamma_{12} & \gamma_{22} \\ \alpha_1 & \alpha_2 \end{bmatrix}.$$

Example: Minnesota Prior

- Dummies for the β coefficients:

$$\begin{aligned} Y^* &= X^*\Phi + U \\ \left[\begin{array}{cc} \lambda_1 \underline{s}_1 & 0 \\ 0 & \lambda_1 \underline{s}_2 \end{array}\right] &= \left[\begin{array}{ccccc} \lambda_1 \underline{s}_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 \underline{s}_2 & 0 & 0 & 0 \end{array}\right] \Phi + \left[\begin{array}{cc} u_{11} & u_{12} \\ u_{21} & u_{22} \end{array}\right] \end{aligned}$$

The first observation implies, for instance, that

$$\begin{aligned} \lambda_1 \underline{s}_1 = \lambda_1 \underline{s}_1 \beta_{11} + u_{11} \quad &\Longrightarrow \quad \beta_{11} = 1 - \frac{u_{11}}{\lambda_1 \underline{s}_1} \\ &\Longrightarrow \quad \beta_{11} \sim \mathcal{N}\left(1, \frac{\Sigma_{11}}{\lambda_1^2 \underline{s}_1^2}\right) \\ 0 = \lambda_1 \underline{s}_1 \beta_{21} + u_{12} \quad &\Longrightarrow \quad \beta_{21} = -\frac{u_{12}}{\lambda_1 \underline{s}_1} \\ &\Longrightarrow \quad \beta_{21} \sim \mathcal{N}\left(0, \frac{\Sigma_{22}}{\lambda_1^2 \underline{s}_1^2}\right) \end{aligned}$$

Example: Minnesota Prior

- Dummies for the γ coefficients:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \lambda_1 \underline{s}_1 2^{\lambda_2} & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 \underline{s}_2 2^{\lambda_2} & 0 \end{bmatrix} \Phi + U$$

- For lags of order p the entry above would be $\lambda_1 \underline{s}_i p_2^{\lambda}$.
- The prior for the covariance matrix is implemented by λ_3 replications of

$$\begin{bmatrix} \underline{s}_1 & 0 \\ 0 & \underline{s}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + U$$

Example: Minnesota Prior

- Sums-of-coefficients dummy observations, reflecting the belief that when y_i has been stable at its initial level, it will tend to persist at that level, regardless of the value of other variables:

$$\begin{bmatrix} \lambda_4 \underline{y}_1 & 0 \\ 0 & \lambda_4 \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda_4 \underline{y}_1 & 0 & \lambda_4 \underline{y}_1 & 0 & 0 \\ 0 & \lambda_4 \underline{y}_2 & 0 & \lambda_4 \underline{y}_2 & 0 \end{bmatrix} \Phi + U$$

- Co-persistence prior dummy observations, reflecting the belief that when data on all y's are stable at their initial levels, thy will tend to persist at that level:

$$\begin{bmatrix} \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 & \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 & \lambda_5 \end{bmatrix} \Phi + U$$

Example: DSGE Model Prior

- See Del Negro and Schorfheide (2004)
- **Idea:** simulate dummy observations from a DSGE model to shrink toward DSGE model restrictions.

Hierarchical Models and Hyperparameter Selection

- "Performance" of Bayesian VAR is sensitive to prior variance.
- Choose prior variance in a data-driven way.
- Hierarchical model

$$p(Y|\Phi, \Sigma)p(\Phi, \Sigma|\lambda)p(\lambda),$$

where λ controls features of the prior.

Example: Minnesota Prior

- λ_1 is the overall tightness of the prior. Large values imply a small prior covariance matrix.
- λ_2: the variance for the coefficients of lag h is scaled down by the factor $(l^{-\lambda_2})^2$.
- λ_3: determines the weight for the prior on Σ. Suppose that $Z_i = \mathcal{N}(0, \sigma^2)$. Then an estimator for σ^2 is $\hat{\sigma}^2 = \frac{1}{\lambda_3} \sum_{i=1}^{\lambda_3} Z_i^2$. The larger λ_3, the more informative the estimator, and in the context of the VAR, the tighter the prior.
- λ_4 and λ_5: tuning parameters for sums-of-coefficients and co-persistence dummies.
- In addition: $\underline{s} = \text{std}(Y_{-\tau,0})$ and $\underline{y} = \text{mean}(Y_{-\tau,0})$, where $Y_{-\tau,0}$ is a short presample.

Selection vs. Averaging

- **Selection:**
  - Compute

$$p(Y|\lambda) = \int p(Y|\Phi,\Sigma)p(\Phi,\Sigma|\lambda)d(\Phi,\Sigma).$$

  - Define: $\hat{\lambda} = \text{argmax } p(Y|\lambda)$.
  - Work with $p(\Phi,\Sigma|Y,\hat{\lambda})$.
- **Averaging:**
  - Use prior $p(\lambda)$
  - Factorize posterior as

$$p(\Phi,\Sigma,\lambda|Y) = p(\Phi,\Sigma|Y,\lambda)p(\lambda|Y),$$

where $p(\lambda|Y) \propto p(Y|\lambda)p(\lambda)$.

Example: Minnesota Prior w/ Dummy Observations

- The marginal likelihood can be calculated from the normalization constants of the MNIW distribution (see Zellner (1971, Appendix)):

$$p(Y|\lambda) = (2\pi)^{-nT/2} \frac{|\bar{X}'\bar{X}|^{-\frac{n}{2}} |\bar{S}|^{-\frac{\bar{T}-k}{2}}}{|X^{*\prime} X^{*}|^{-\frac{n}{2}} |S^{*}|^{-\frac{T^{*}-k}{2}}} \frac{2^{\frac{n(\bar{T}-k)}{2}} \prod_{i=1}^{n} \Gamma[(\bar{T}-k+1-i)/2]}{2^{\frac{n(T^{*}-k)}{2}} \prod_{i=1}^{n} \Gamma[(T^{*}-k+1-i)/2]}.$$

- The hyperparameters $(\bar{y}, \bar{s}, \lambda)$ enter through the dummy observations X^{*} and Y^{*}.

Illustration: Marginal Likelihood of λ

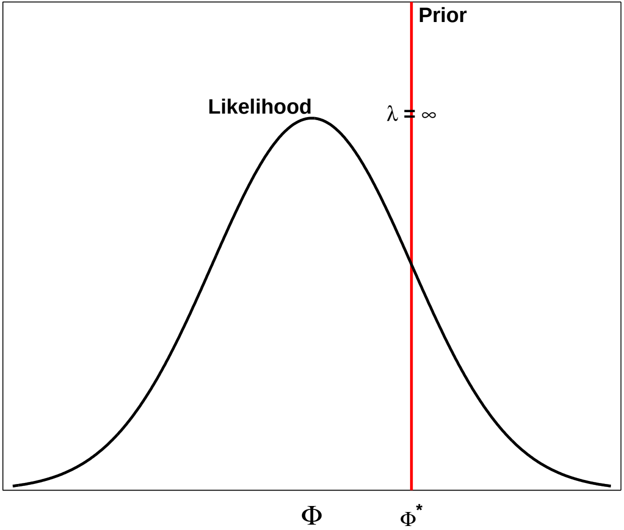
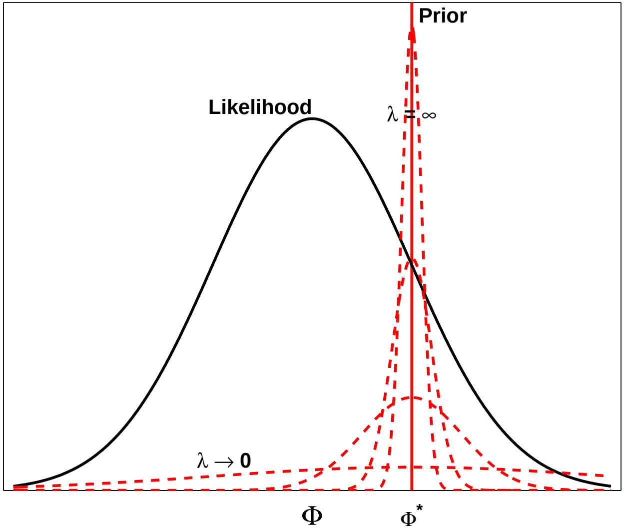


Illustration: Marginal Likelihood of λ



Example: Marginal Likelihood of λ

- Suppose the VAR takes the special form of an AR(1) model:

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim iidN(0,1)$$

- Suppose the prior takes the form

$$\phi \sim N\left(\phi^*, \frac{1}{\lambda T \gamma_0}\right).$$

  where $\gamma_0 = 1/(1-\phi_*^2)$
- Define $\gamma_1 = \phi_* \gamma_0$ and denote the sample autocovariances by

$$\hat{\gamma}_0 = \frac{1}{T}\sum y_t^2, \quad \hat{\gamma}_1 = \frac{1}{T}\sum y_t y_{t-1}.$$

- For convenience, we standardized the prior variance by T.

Example: Marginal Likelihood of λ

- After some algebra it can be shown that marginal likelihood of λ takes the following form

$$\ln p(Y|\lambda, \phi^*) = -T/2 \ln(2\pi) - \frac{T}{2}\tilde{\sigma}^2(\lambda, \phi^*) - \frac{1}{2}c(\lambda, \phi^*).$$

- The term $\tilde{\sigma}^2(\lambda, \phi^*)$ measures the in-sample one-step-ahead forecast error:

$$\begin{aligned}
\lim_{\lambda \longrightarrow 0} \tilde{\sigma}^2(\lambda, \phi^*) &= \frac{1}{T}\sum (y_t - \hat{\phi} y_{t-1})^2 \\
\lim_{\lambda \longrightarrow \infty} \tilde{\sigma}^2(\lambda, \phi^*) &= \frac{1}{T}\sum (y_t - \phi^* y_{t-1})^2.
\end{aligned}$$

- The third term above can be interpreted as a penalty for model complexity and is of the form

$$c(\lambda, \phi^*) = \ln\left(1 + \frac{\hat{\gamma}_0}{\lambda \gamma_0}\right).$$

- As λ approaches zero, the marginal log likelihood function tends to minus infinity.

Example: Marginal Likelihood of λ

- Recall that marginal likelihood of λ takes the following form

$$\ln p(Y|\lambda, \phi^*) = -T/2 \ln(2\pi) - \frac{T}{2}\tilde{\sigma}^2(\lambda, \phi^*) - \frac{1}{2}c(\lambda, \phi^*).$$

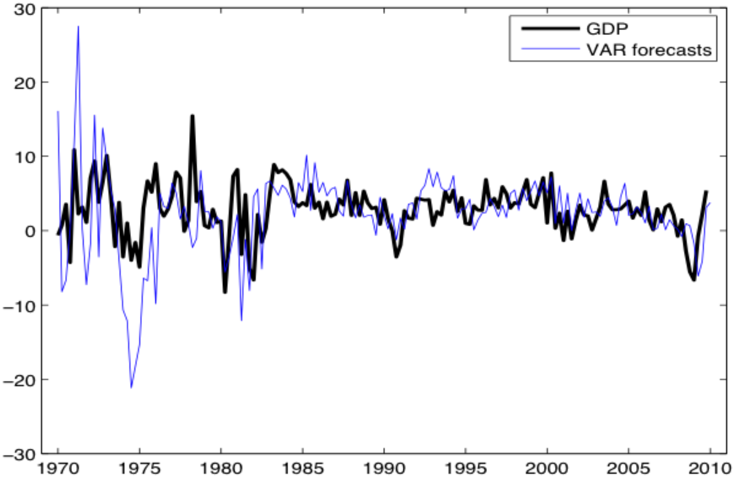
- If an interior maximum of marginal likelihood exists, it is given by

$$\hat{\lambda} = \frac{\gamma_0 \hat{\gamma}_0^2}{T(\hat{\gamma}_0\gamma_1 - \gamma_0\hat{\gamma}_1)^2 - (\gamma_0)^2\hat{\gamma}_0}.$$

The following pages are taken from Giannone, Lenza, and Primiceri (2012).

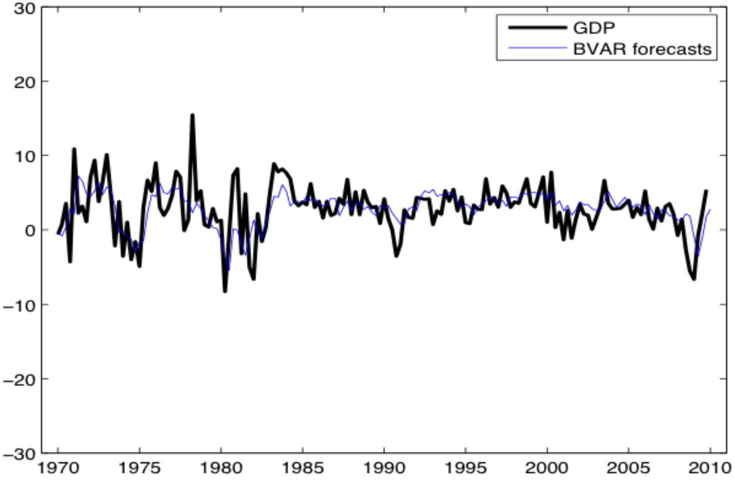
US GDP growth and VAR forecast (1-step ahead)

Flat-prior VAR



US GDP growth and BVAR forecast (1-step ahead)

BVAR (MN+SOC+DIO priors + hyperparameter selection)



BVARs: Forecasting performance

Mean Squared Forecast Errors

		7-variable VAR		
		Flat-prior	BVAR with MN prior (λ=0.2)	BVAR with MN+SOC+DIO
1 Quarter Ahead	Real GDP	19.18	9.61	7.97
	GDP Deflator	2.27	1.53	1.35
	Federal Funds Rate	1.83	1.08	1.03
1 Year Ahead	Real GDP	11.90	5.48	3.42
	GDP Deflator	2.22	1.85	1.58
	Federal Funds Rate	0.56	0.40	0.31