Bayesian Econometrics for Everybody

User Guide: @ll BEsmarter*

BEsmarter–Team

Universidad EAFIT

May 16, 2016

1 Normal Model

The Gaussian linear model is the workhorse in econometrics, its is specification is given by a dependent variable y_i which is related to a set of exogenous variables ($\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iK})'$) in a linear way, that is, $y_i = \beta_1 x_{i1} + \beta_1 x_{i2} + \ldots + \beta_1 x_{iK} + \mu_i = \mathbf{x}_i'\beta + \mu_i$ where $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$ and $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ is an stochastic error such that $\mathbf{x}_i \perp \mu_i$.

Writing this model in matrix form we get $\mathbf{y} = \mathbf{X}\beta + \mu$ such that $\mu \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ which implies that $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \boldsymbol{\sigma^2}\mathbf{I})$. So the likelihood function is

$$\begin{aligned}\mathscr{L}(\mathbf{y}|\beta, \sigma^2, \mathbf{X}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \end{aligned} \tag{1}$$

1.1 Natural Conjugate Family Priors: Normal-Inverse Gamma Model

The conjugate priors for the parameters are:

$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 B_0) \tag{2}$$

$$\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2) \tag{3}$$

*Bayesian Econometrics: Simulations, Models and Applications to Research, Teaching, Encoding with Responsibility.

The joint posterior distribution for the parameters is then

$$
\begin{aligned}
\pi(\beta,\sigma^2|\mathbf{y},\mathbf{X}) &\propto \mathscr{L}(\mathbf{y}|\beta,\sigma^2,\mathbf{X})\pi(\beta|\sigma^2)\pi(\sigma^2)\\
&= (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right\}\\
&\quad\times(\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta_0)'B_0^{-1}(\beta-\beta_0)\right\}\\
&\quad\times\frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)}\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\\
&= (\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}[y'y-y'X\beta-\beta'X'y+\beta'X'X\beta]\right\}\\
&\quad\times(\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}[\beta'B_0^{-1}\beta-\beta'B_0^{-1}\beta_0-\beta_0'B_0^{-1}\beta+\beta_0'B_0^{-1}\beta_0]\right\}\\
&\quad\times\frac{(\delta_0/2)^{(\alpha_0/2)}}{\Gamma(\alpha_0/2)}\frac{1}{(\sigma^2)^{(\alpha_0/2+1)}}\exp\left\{-\frac{\delta_0}{2\sigma^2}\right\}\\
&\propto (\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}[\beta'(B_0^{-1}+X'X)\beta-2\beta'(B_0^{-1}\beta_0+X'X\hat{\beta})]\right\}\\
&\quad\times\frac{1}{(\sigma^2)^{(\alpha_0+n)/2+1}}\exp\left\{-\frac{\delta_0+y'y+\beta_0'B_0^{-1}\beta_0}{2\sigma^2}\right\}
\end{aligned}
$$

where $\hat{\beta}=(X'X)^{-1}X'y$.

Adding and subtracting $\beta^{*'}B^{-1}\beta^*$ where

$$B=(B_0^{-1}+X'X)^{-1} \tag{4}$$

$$\beta^*=B(B_0^{-1}\beta_0+X'X\hat{\beta})=B(B_0^{-1}\beta_0+X'y) \tag{5}$$

And completing the square

$$
\begin{aligned}
\pi(\beta,\sigma^2|\mathbf{y},\mathbf{X}) &\propto (\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}[\beta'(B_0^{-1}+X'X)\beta-2\beta'B^{-1}B(B_0^{-1}\beta_0+X'X\hat{\beta})+\beta^{*'}B^{-1}\beta^*-\beta^{*'}B^{-1}\beta^*]\right\}\\
&\quad\times\frac{1}{(\sigma^2)^{(\alpha_0+n)/2+1}}\exp\left\{-\frac{\delta_0+y'y+\beta_0'A^{-1}\beta_0}{2\sigma^2}\right\}\\
&= (\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}[\beta'B^{-1}\beta-2\beta'B^{-1}\beta^*+\beta^{*'}B^{-1}\beta^*]\right\}\\
&\quad\times\frac{1}{(\sigma^2)^{(\alpha_0+n)/2+1}}\exp\left\{-\frac{\delta_0+y'y+\beta_0'B_0^{-1}\beta_0-\beta^{*'}B^{-1}\beta^*}{2\sigma^2}\right\}\\
&= (\sigma^2)^{-\frac{k}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta^*)'B^{-1}(\beta-\beta^*)\right\}\\
&\quad\times(\sigma^2)^{-\left(\frac{\alpha^*}{2}+1\right)}\exp\left\{-\frac{\delta^*}{2\sigma^2}\right\}
\end{aligned}
$$

This means posterior distributions for β and σ^2 of the form

$$\beta|\sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\beta^*, \sigma^2 B) \tag{6}$$

$$\sigma^2|\mathbf{y}, \mathbf{X} \sim \mathcal{IG}(\alpha^*, \delta^*) \tag{7}$$

where

$$\alpha^* = \alpha_0 + n \tag{8}$$

$$\delta^* = \delta_0 + y'y + \beta_0' B_0^{-1} \beta_0 - \beta^{*'} B^{-1} \beta^* \tag{9}$$

We can see that the posterior distributions, equations 6 and 7, are from the same family as the prior distributions, equations 2 and 3.

We can also express β^* as a weighted average of the prior information β_0 and the Maximum Likelihood estimate $\hat{\beta} = (X'X)^{-1}X'y$, the Maximum Likelihood estimate, as

$$\begin{aligned}\beta^* &= (B_0^{-1} + X'X)^{-1}(B_0^{-1}\beta_0 + X'X\hat{\beta}) \\ &= (B_0^{-1} + X'X)^{-1}B_0^{-1}\beta_0 + (B_0^{-1} + X'X)^{-1}X'X\end{aligned}$$

Given the following property of inverse matrices $(D+E)^{-1}E = I - (D+E)^{-1}D$ (Smith, 1973), taking $D = X'X$, $E = B_0^{-1}$, then

$$(X'X + B_0^{-1})^{-1}B_0^{-1} = I - (X'X + B_0^{-1})^{-1}X'X$$

And, setting $W = (B_0^{-1} + X'X)^{-1}X'X$,

$$\begin{aligned}\beta^* &= (I - (B_0^{-1} + X'X)^{-1}X'X)\beta_0 + (B_0^{-1} + X'X)^{-1}X'X\hat{\beta} \\ &= (I - W)\beta_0 + W\hat{\beta} \end{aligned} \tag{10}$$

Observe that when the prior covariance matrix is highly non–informative, such that $B_0^{-1} \to \mathbf{0}$, we obtain $W \to I$, such that $\beta^* \to \hat{\beta}$, that is, the posterior mean location parameter converges to the Maximum Likelihood estimate.

Finally, there is another alternative representation that can be specially useful for the linear regression model. First, take

$$\begin{aligned}\delta^* &= \delta_0 + y'y + \beta_0' B_0^{-1}\beta_0 - \beta^{*'} B^{-1}\beta^* \\ &= \delta_0 + y'y + \beta_0' B_0^{-1}\beta_0 - (B_0^{-1}\beta_0 + X'X\hat{\beta})'B(B_0^{-1}\beta_0 + X'X\hat{\beta}) \\ &= \delta_0 + y'y + \beta_0' B_0^{-1}\beta_0 - \hat{\beta}'X'XBX'X\hat{\beta} - 2\hat{\beta}'X'XBB_0^{-1}\beta_0 - \beta_0' B_0^{-1}BB_0^{-1}\beta_0\end{aligned}$$

$$
\begin{aligned}
&= \delta_0 + y'y - \hat{\beta}'X'XBX'X\hat{\beta} - 2\hat{\beta}'X'XBB_0^{-1}\beta_0 + \beta_0'(B_0^{-1} - B_0^{-1}BB_0^{-1})\beta_0 \\
&\quad - \hat{\beta}'X'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
&= \delta_0 + y'y - \hat{\beta}'X'X\hat{\beta} + \hat{\beta}'(X'X - X'XBX'X)\hat{\beta} - 2\hat{\beta}'X'XBB_0^{-1}\beta_0 \\
&\quad + \beta_0'(B_0^{-1} - B_0^{-1}BB_0^{-1})\beta_0
\end{aligned}
$$

We can rewrite some of these terms as follows:

$$
(y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2\hat{\beta}'X'(X\hat{\beta} + \hat{\mu}) + \hat{\beta}'X'X\hat{\beta} = y'y - \hat{\beta}'X'X\hat{\beta}
$$

where $y = X\hat{\beta} + \hat{\mu}$ and taking into account that $X'\hat{\mu} = 0$.

The following matrix identities will also prove useful (Smith, 1973):

$$
(D + E)^{-1} = D^{-1} - D^{-1}(D^{-1} + E^{-1})^{-1}D^{-1} \tag{11}
$$

$$
(D + E)^{-1} = D^{-1}(E^{-1} + D^{-1})E^{-1} \tag{12}
$$

Using 11 and 12

$$
\begin{aligned}
[(X'X)^{-1} + B_0]^{-1} &= X'X - X'X(X'X + B_0^{-1})^{-1}X'X \\
&= B_0^{-1} - B_0^{-1}(X'X + B_0^{-1})^{-1}B_0^{-1} \\
&= X'X(X'X + B_0^{-1})^{-1}B_0^{-1}
\end{aligned}
$$

Then we express the updated parameters for the variance as

$$
\begin{aligned}
\delta^* &= \delta_0 + (y - X\hat{\beta})'(y - X\hat{\beta}) + \hat{\beta}'[(X'X)^{-1} + B_0]^{-1}\hat{\beta} \\
&\quad - 2\hat{\beta}[(X'X)^{-1} + B_0]^{-1}\beta_0 + \beta_0'[(X'X)^{-1} + B_0]^{-1}\beta_0 \\
&= \delta_0 + (n - k)\hat{\sigma}^2_{LSE} + (\hat{\beta} - \beta_0)'[(X'X)^{-1} + B_0]^{-1}(\hat{\beta} - \beta_0)
\end{aligned} \tag{13}
$$

where $\hat{\sigma}^2_{LSE} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - k}$ is the Least Squares estimate of the variance.

This representation shows how the variance is affected by the Least Squared estimates and prior information for both the coefficients and the variance.[1]

[1] Remember that the least squares estimates and the Maximum Likelihood estimates of the location parameters are the same in the Gaussian model.

We can find the marginal posterior distributions for β and σ^2 by integrating out the other parameter. For σ^2, we can then take

$$\pi(\sigma^2|y) = \int \pi(\beta, \sigma^2|y)d\beta$$

Since β is present only in the normal distribution, this is immediately equivalent to

$$\sigma^2|\mathbf{y}, \mathbf{X} \sim \mathcal{IG}(\alpha^*, \delta^*)$$

The marginal posterior distribution for β can be obtained similarly by integrating out σ^2,

$$\begin{aligned}
\pi(\beta|\mathbf{y}, \mathbf{X}) &= \int \pi(\beta, \sigma^2|y)d\sigma^2 \\
&= \int \left(\frac{1}{\sigma^2}\right)^{\frac{\alpha^*+k}{2}+1} \exp\left\{-\frac{R}{2\sigma^2}\right\} d\sigma^2
\end{aligned}$$

Where $R = \delta^* + (\beta - \beta^*)'B^{-1}(\beta - \beta^*)$. Then we can write

$$\begin{aligned}
\pi(\beta|y) &= \int \left(\frac{1}{\sigma^2}\right)^{\frac{\alpha^*+k}{2}+1} \exp\left\{-\frac{R}{2\sigma^2}\right\} d\sigma^2 \\
&= \frac{\Gamma([\alpha^* + k]/2)}{(R/2)^{(\alpha^*+k)/2}} \int \frac{(R/2)^{(\alpha^*+k)/2}}{\Gamma([\alpha^* + k]/2)} (\sigma^2)^{-(\alpha^*+k)/2-1} \exp\left\{-\frac{R}{2\sigma^2}\right\} d\sigma^2
\end{aligned}$$

The right term is the integral of the probability density function of an inverse gamma distribution with parameters $\nu = (\alpha^* + k)/2$ and $\tau = R/2$. Since we are integrating over the whole support of σ^2, the integral is equal to 1 and therefore

$$\begin{aligned}
\pi(\beta|y) &= \frac{\Gamma([\alpha^* + k]/2)}{(R/2)^{(\alpha^*+k)/2}} \\
&\propto R^{-(\alpha^*+k)/2} \\
&= [\delta^* + (\beta - \beta^*)'B^{-1}(\beta - \beta^*)]^{-(\alpha^*+k)/2} \\
&= \left[1 + \frac{(\beta - \beta^*)'(\delta^* B)^{-1}(\beta - \beta^*)}{\alpha^*}\right]^{-(\alpha^*+k)/2} (\delta^*\alpha^*)^{-(\alpha^*+k)/2} \\
&\propto \left[1 + \frac{(\beta - \beta^*)'(H)^{-1}(\beta - \beta^*)}{\alpha^*}\right]^{-(\alpha^*+k)/2}
\end{aligned}$$

with $H = \delta^* B$. This last expression signifies a multivariate t-distribution for β as

$$\beta|y \sim t_k(\alpha^*, \beta^*, H)$$

Intuitively, since we have incorporated the uncertainty of the variance, the posterior for β changes from a normal to a t-distribution, which has heavier tails.

Finally, if we want to obtain the posterior distribution for a single β parameter, say β_1 we can proceed in two ways as in Zellner (1996). We can either integrate out $\beta_2, \ldots, \beta_k$ and then σ^2, or integrate

$\beta_2, \ldots, \beta_k$ directly from the previous expression. In this context, we can write the marginal posterior for the vector β as

$$\pi(\beta|\mathbf{y}, \mathbf{X}) \propto \left[1 + \frac{\delta' H \delta}{\alpha^*}\right]^{-(\alpha^*+k)/2}$$

with $\delta' = (\beta - \beta^*)$, $\delta_1 = \beta_1 - \beta_1^*$, a scalar term, and $\delta_2' = (\beta_2 - \beta_2^*, \ldots, \beta_k - \beta_k^*)$. The procedure is then the same as the derivation of equation (4.12) and we can write the marginal posterior for β_1 as

$$\frac{\beta_1 - \beta_1^*}{(h^{11})^{1/2}}|\mathbf{y}, \mathbf{X} \sim t_{\alpha^*}$$

Or in general for β_j as

$$\frac{\beta_j - \beta_j^*}{(h^{jj})^{1/2}}|\mathbf{y}, \mathbf{X} \sim t_{\alpha^*}$$

where h^{jj} is the jth diagonal element of H^{-1} and β_j^* is the jth element of β^*.

MARGINAL LIKELIHOOD

PREDICTIVE DISTRIBUTION

1.2 Independent Priors: Normal-Inverse Gamma Model

1.3 Diffuse Priors: Jeffreys prior

2 Probit Model

3 Ordered Probit Model

4 Negative Binomial Model

5 Endogenous Covariate Model

References

Smith, A. F. M. (1973). A General Bayesian Linear Model. *Journal of the Royal Statistical Society*, 35(1):67–75.

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics.* Wiley Classics Library. John Wiley & Sons.