Bayesian Econometrics for Everybody

User Guide: @ll BEsmarter*

BEsmarter–Team

Universidad EAFIT

October 24, 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 ($\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})'$) 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 $\mu_i \stackrel{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ 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

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

$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \tag{2}$$

1.1 Natural Conjugate Family Priors: Normal-Inverse Gamma Model

The conjugate priors for the parameters are:

$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 B_0) \tag{3}$$

$$\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2) \tag{4}$$

*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

$$B = (B_0^{-1} + X'X)^{-1} \tag{5}$$

$$\beta^* = B(B_0^{-1}\beta_0 + X'X\hat{\beta}) = B(B_0^{-1}\beta_0 + X'y) \tag{6}$$

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

$$\beta|\sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\beta^*, \sigma^2 B) \tag{7}$$

$$\sigma^2|\mathbf{y}, \mathbf{X} \sim \mathcal{IG}(\alpha^*, \delta^*) \tag{8}$$

where

$$\alpha^* = \alpha_0 + n \tag{9}$$

$$\delta^* = \delta_0 + y'y + \beta_0' B_0^{-1}\beta_0 - \beta^{*'} B^{-1}\beta^* \tag{10}$$

We can see that the posterior distributions, equations 66 and 67, are from the same family as the prior distributions, equations 63 and 4.

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

$$\begin{aligned}
\beta^* &= (I - (B_0^{-1} + X'X)^{-1}X'X)\beta_0 + (B_0^{-1} + X'X)^{-1}X'X\hat{\beta} \\
&= (I - W)\beta_0 + W\hat{\beta} \qquad (11)
\end{aligned}$$

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

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

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

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):

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

$$(D + E)^{-1} = D^{-1}(E^{-1} + D^{-1})E^{-1} \tag{13}$$

Using 72 and 6.2

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

Then we express the updated parameters for the variance as

$$
\begin{aligned}
\delta^* &= \delta_0 + (y - X\hat{\beta})'(y - X\hat{\beta}) + \hat{\beta}'[(X'X)^{-1} + B_0]^{-1}\hat{\beta} \\
&\quad - 2\hat{\beta}[(X'X)^{-1} + B_0]^{-1}\beta_0 + \beta_0'[(X'X)^{-1} + B_0]^{-1}\beta_0 \\
&= \delta_0 + (n - k)\hat{\sigma}^2_{LSE} + (\hat{\beta} - \beta_0)'[(X'X)^{-1} + B_0]^{-1}(\hat{\beta} - \beta_0)
\end{aligned} \tag{14}
$$

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^*.

1.2 Marginal Distribution

First consider the following result:

$$\begin{aligned}
&(y - X\beta)^{'}(y - X\beta) + (\beta - \beta_o)B_o^{-1}(\beta - \beta_o) \\
&= \beta^{'}(X^{'}X + B_o^{-1})\beta - 2\beta^{'}(X^{'}y + B_o^{-1}\beta_o) + y^{'}y + \beta_o^{'}V^{-1}\beta_o
\end{aligned} \tag{15}$$

let

$$\begin{aligned}
B^{*-1} =&(X^{'}X + B_o^{-1}) \\
B^* =&(X^{'}X + B_o^{-1})^{-1} \\
\beta^* =&B(X^{'}y + B_o^{-1}\beta_o) \\
=&(X^{'}X + B_o^{-1})^{-1}(X^{'}y + B_o^{-1}\beta_o)
\end{aligned}$$

Replacing in 15

$$\begin{aligned}
&\beta^{'}(X^{'}X + B_o^{-1})\beta - 2\beta^{'}(X^{'}y + B_o^{-1}\beta_o) + y^{'}y + \beta_o^{'}B_o^{-1}\beta_o \\
&= \beta^{'}B^{-1}\beta - 2\beta^{'}B^{-1}\beta^* + \beta^{*\prime}B^{-1}\beta^* - \beta^{**\prime}B^{-1}\beta^* + y^{'}y + \beta_o^{'}B_o^{-1}\beta_o \\
&= (\beta - \beta^*)^{'}B^{-1}(\beta - \beta^*) - \beta^{'}B_o^{*-1}\beta + y^{'}y + \beta_o^{'}B_o^{-1}\beta_o \\
&= (\beta - \beta^*)^{'}B^{-1}(\beta - \beta^*) + M
\end{aligned} \tag{16}$$

Taking into account (16), (1), (63) and (4) we calculate the marginal as:

$$\int_0^{\infty} \int_{\beta} \pi(\beta \mid \sigma^2,\ B_o,\ \beta_o)\pi(\sigma^2 \mid \alpha_0/2, \delta_0/2)\mathscr{L}(\mathbf{y}|\beta, \sigma^2, \mathbf{X})d\sigma^2 d\beta =$$

$$= \int_0^\infty \pi(\sigma^2 \mid \alpha_0/2, \delta_0/2) \frac{1}{(2\pi\sigma^2)^{(p+n)/2} \mid B_o \mid^{1/2}} \int_\beta exp\left\{-\frac{1}{2\sigma^2}((\beta-\beta_o)^{'} B_o^{-1}(\beta-\beta_o)+M))\right\} d\sigma^2 d\beta$$
$$= \int_0^\infty \pi(\sigma^2 \mid \alpha_0/2, \delta_0/2) \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left\{-\frac{1}{2\sigma^2}M\right\} \frac{1}{(2\pi\sigma^2)^{p/2} \mid B_o \mid^{1/2}} \int_\beta exp\left\{-\frac{1}{2\sigma^2}((\beta-\beta^*)^{'} B^{-1}(\beta-\beta^*))\right\} d\sigma^2 d\beta$$
$$= \int_0^\infty \pi(\sigma^2 \mid \alpha_0/2, \delta_0/2) \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left\{-\frac{1}{2\sigma^2}M\right\} \frac{\mid B \mid^{1/2}}{\mid B_o \mid^{1/2}} d\sigma^2$$

The last step was performed completing the multivariate pdf of a p-dimensional normal distribution with mean β^* and covariance matrix $\sigma^2 B$. Replacing 4 we obtain:

$$\begin{aligned}
&= \int_0^\infty \frac{(\alpha_0/2)^{\delta_0/2}}{\Gamma(\delta_0/2)(\sigma^2)^{\delta_0/2+1}} exp\left\{(\alpha_0/2\sigma^2\right\} \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left\{-\frac{1}{2\sigma^2}M\right\} \frac{\mid B \mid^{1/2}}{\mid B_o \mid^{1/2}} \\
&= \frac{\mid B \mid^{1/2} (\alpha_0/2)^{\delta_0/2}}{(2\pi)^{n/2} \mid B_o \mid^{1/2} \Gamma(\delta_0/2)} \int_0^\infty (\sigma^2)^{-(n+\delta_0)/2-1} exp\left\{-\frac{1}{2\sigma^2}(M+\alpha_0)\right\} \\
&\quad making \;\; u = \frac{\sigma^2(M+\alpha_0)}{2} \\
&= \frac{\mid B \mid^{1/2} (\alpha_0/2)^{\delta_0/2} 2^{n/2+(\delta_0/2)+1-1}}{(2\pi)^{n/2} \mid B_o \mid^{1/2} \Gamma(\delta_0/2)(M+\alpha_0)^{n/2+(\delta_0/2)+1-1}} \int_0^\infty u^{(n+\delta_0)/2-1} e^{-u} du \\
&= \frac{\mid B \mid^{1/2} (\alpha_0/2)^{\delta_0/2} 2^{\delta_0/2}}{(\pi)^{n/2} \mid B_o \mid^{1/2} \Gamma(\delta_0/2)(M+\alpha_0)^{n/2+\delta_0/2}} \Gamma(\frac{n+\delta_0}{2}) \\
&\quad making \;\; M+\alpha_0 = \alpha_0 \left(1 + \frac{(\delta_0/2)M}{\delta_0(\alpha_0/2)}\right) \\
&= \frac{\mid B \mid^{1/2} \Gamma(\frac{n+\delta_0}{2})(\delta_0/2)^{n/2}}{\pi^{n/2} \mid B_o \mid^{1/2} (\alpha_0/2)^{n/2} \delta_0^{n/2} \Gamma(\delta_0/2) \left(1 + \frac{(\delta_0/2)M}{\delta_0(\alpha_0/2)}\right)^{(n+\delta_0)/2}}
\end{aligned} \tag{17}$$

Applying some extra algebra over M we obtain

$$\begin{aligned}
M =& y^{'}y - \beta^{*'} B^{-1}\beta^* + \beta_0^{'} B_0^{-1}\beta_0 \\
=& y^{'}y + \beta_0^{'} B_o^{-1}\beta_0 - (X^{'}y + B_o^{-1}\beta_0)^{'} B^*(X^{'}y + B_o^{-1}\beta_0) \\
=& y^{'}y + \beta_0^{'} B_o^{-1}\beta_0 - y^{'}XBX^{'}y - 2y^{'}XBB_0^{-1}\beta_0 + \beta_0^{'}B_o^{-1}BB_o^{-1}\beta_0 \\
=& y^{'}(I - XBX)y - 2y^{'}XBB_0^{-1}\beta_0 + \beta_0^{'}(B_o^{-1} + B_o^{-1}BB_o^{-1})\beta_0
\end{aligned} \tag{18}$$

Define $\Sigma_e^{-1} = I - XBX^{'}$ and consider the following lema:

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}$$

Taking into account the latter and the definition of B

$$B_o^{-1} + B_o^{-1}BB_o^{-1} = B_o^{-1} + B_o^{-1}(B_o^{-1} + X^{'}X)^{-1}B_o^{-1}$$
$$Making \;\; A = B_o \;\; B = C = I \;\; and \;\; D^{-1} = X^{'}X$$

$$
\begin{aligned}
&= (B_o + (X^{'}X)^{-1}) \\
&Making \;\; A^{-1} = X^{'}X \;\; B = C = I \;\; and \;\; D = B_o \\
&= (X^{'}X) - (X^{'}X)(B_o^{-1} + (X^{'}X))^{-1}X^{'}X \\
&= X^{'}(I - X(B_o^{-1} + (X^{'}X))^{-1}X^{'})X \\
&= X^{'}(I - XBX^{'})X \\
&= X^{'}\Sigma_e^{-1}X
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
&BB^{-1} = I = B(X^{'}X + B_o^{-1}) \\
&BB_o^{-1} = I - BX^{'}X \qquad\qquad Pre-multipliying \;\; X \\
&XBB_o^{-1} = X - XBX^{'}X = \Sigma_e^{-1}X \\
&XBB_o^{-1} = \Sigma_e^{-1}X
\end{aligned} \tag{20}
$$

Replacing (19) and (20) in (18) we obtaing:

$$
\begin{aligned}
M =& y^{'}\Sigma_e^{-1}y - 2y^{'}\Sigma_e^{-1}X\beta_0 + \beta_0^{'}X^{'}\Sigma_e^{-1}X\beta_0 \\
=& (y - X\beta_0)^{'}\Sigma_e^{-1}(y - X\beta_0)
\end{aligned} \tag{21}
$$

From the inversion lemma we also obtain

$$
\begin{aligned}
\Sigma_e^{-1} =& I - XBX^{'} \\
=& I - X(B_0^{-1} + X^{'}X)^{-1}X^{'} \\
& Making \;\; A = I \;\; B = X = C^{'} \;\; and \;\; D = B_0 \\
\Sigma_e =& I + XB_0X^{'} = (I - XBX^{'})^{-1}
\end{aligned} \tag{22}
$$

Consider the following lemma:

$$
\mid A + BDC \mid = \mid A \mid\mid D \mid\mid D^{-1} + CA^{-1}B \mid \tag{23}
$$

Let $A = I$, $B = X$, $D = B_o$, $C = X^{'}$, then we have

$$
\begin{aligned}
\mid \Sigma_e \mid =& = \mid I + XB_oX^{'} \mid \\
&= \mid B_o \mid\mid B_o^{-1} + X^{'}X \mid \\
&= \mid B_o \mid\mid B^{-1} \mid \\
&= \frac{\mid B_o \mid}{\mid B \mid}
\end{aligned}
$$

$$\frac{1}{\mid \Sigma_e \mid^{1/2}} = \left(\frac{\mid B \mid}{\mid B_o \mid} \right)^{1/2} \tag{24}$$

Finally consider $\Sigma_M = \frac{\alpha_0 \Sigma_e}{\delta_0}$, (21) and (24) and replace in (17)

$$\begin{aligned}
&\int_0^\infty \int_\beta \pi(\beta \mid \sigma^2,\ B_o,\ \beta_o)\pi(\sigma^2 \mid \alpha_0/2, \delta_0/2)\mathscr{L}(\mathbf{y}|\beta, \sigma^2, \mathbf{X})d\sigma^2 d\beta = \\
&= \frac{\Gamma(\frac{n+\delta_0}{2})}{\pi^{n/2}\Gamma(\delta_0/2)\delta_0^{n/2} \mid \Sigma_M \mid^{1/2} \left[1 + \frac{1}{\delta_0}(y - X\beta_0)'\Sigma_M^{-1}(y - X\beta_0)\right]^{(n+\delta_0)/2}} \\
&= T\left[y \mid X\beta_0, \frac{\alpha_0(I + XB_0X^{'})}{\delta_0}, \delta_0\right]
\end{aligned} \tag{25}$$

Where $T[t \mid m, S, f]$ is the pdf of a multivariate t distribution with location parameter m, scale matrix S and f degrees of freedom evaluated at t.

From (25) we can also conclude that when $\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2)$, $\beta \sim \mathcal{N}(\beta_0, \sigma^2 B_0)$ and $y \sim \mathcal{N}(X\beta, \sigma^2 I)$

$$\int_0^\infty \int_\beta \mathcal{IG}(\alpha_0/2, \delta_0/2)\mathcal{N}(\beta_0, \sigma^2 B_0)\mathcal{N}(X\beta, \sigma^2 I)d\sigma^2 d\beta \sim T\left(X\beta_0, \frac{\alpha_0(I + XB_0X^{'})}{\delta_0}, \delta_0\right) \tag{26}$$

1.3 Predictive Distribution

Consider the vector of future q observations $\hat{y}^{'} = (\hat{y}_1, ..., \hat{y}_q) \sim \mathcal{N}(\hat{X}\beta, \sigma^2 I)$ the joint pdf for σ^2, β and $\hat{y}$ would be

$$P(\hat{y},\ \beta,\ \sigma^2\ \mid\ y,\ X,\ \hat{X}) = P(\hat{y} \mid\ \beta,\ \sigma^2,\ \hat{X})P(\beta,\ \sigma^2\ \mid\ y,\ \ X) \tag{27}$$

The predictive distribution of $\hat{y}$ could be obtained integrating over σ^2 and β in (27) if we also consider the results in (66) and (67) we can calculate the predictive distribution as:

$$P(\hat{y} \mid\ y,\ X,\ \hat{X}) = \int_0^\infty \int_\beta \mathcal{IG}(\alpha^*, \delta^*)\mathcal{N}(\beta^*, \sigma^2 B)\mathcal{N}(\hat{X}\beta, \sigma^2 I)d\sigma^2 d\beta$$

Considering (26) we finally obtain

$$P(\hat{y} \mid\ y,\ X,\ \hat{X}) \sim T\left(\hat{X}\beta^*, \frac{\alpha^*(I + \hat{X}B\hat{X}^{'})}{\delta^*}, 2\delta^*\right) \tag{28}$$

1.4 Independent Priors: Normal-Inverse Gamma Model

1.5 Diffuse Priors: Jeffreys prior

2 Probit Model

The binary probit model is designed to deal with situtions in which the dependent variabel can take one of two values, coded as 0 and 1. Assume we have a non-observable endogenous variable y_i^* related to other K exogenous variables $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})'$ in a linear model given by $y_i^* = \mathbf{x}_i'\beta + \mu_i$, where $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. To make tractable this situation, it is used the perspective of Greenberg (2012) in which some latent variable

$$y_i = \begin{cases} 0 & \text{if } y_i^* < 0 \\ 1 & \text{otherwise} \end{cases}$$

Or, being more concise, $y_i = 1[y_i^* > 0]$.

To model these situations, one should limit the outcomes to be bounded between zero and one. More specifically, the probit model uses a normal standard distribution $\Phi(Z)$.

It is important to note three implications of this context. First, the variance of the stochastic perturbation is nonnegative. It is because one should assure the identifiability of all K parameters $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$ when standardizing to estimate. The simplest way to identify them is setting the variance of the perturbation to $\sigma_\mu^2 = 1$. Second, we are assuming homoskedasticity. Nonetheless, as long as there is exchangeability between the arguments of the posteriors, it is easy to generalize this model to take into account for heteroskedasticity. Third, following Albert and Chib (1993), the contribution of the *ith* observation to the likelihood function is

$$p(y_i|y_i^*) = 1[y_i = 0]1[y_i^* < 0] + 1[y_i = 1]1[y_i^* {\geq} 0]$$

In other words, the likelihood function would be:

$$\mathscr{L}(\mathbf{y}|\beta, \mathbf{X}) = \prod_{i=1}^{n} \{1[y_i = 0]1[y_i^* < 0] + 1[y_i = 1]1[y_i^* \geq 0]\} \tag{29}$$

It would mean that, as long as the latent variable had been correctly specified, this expression would be one. Of course, it would be zero otherwise. So, if the direction of the variable y_i^* is correct in the *ith* observation, then the priors specified for that observation would be taken into account for the posterior calculates.

2.1 Prior distributions

As long as we are concerned only for the latent variable $y^* = (y_1^*, y_2^*, \dots, y_n^*)$ and the parameters $\beta = (\beta_1, \beta_2, \dots, \beta_K)$, we assume

$$\pi(\beta, y^*) = \pi(\beta)\pi(y^*|\beta)$$

β is assumed to be normal multivariate of dimension K. Because of the specification $y_i^* = \mathbf{x}_i'\beta + \mu_i$ where $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. We know that y^* is normal multivariate of dimension n. Thus,

$$\pi(\beta, y^*) = \mathcal{N}_k(\beta|b_0, B_0)\mathcal{N}_n(y^*|X\beta, I_n) \tag{30}$$

where b_0 and B_0 are hyperparameters, and X is the design matrix of the exogenous variables.

If there had been heteroskedasticity, one would take into account another parameter λ_i. This would yield a prior such as $\pi(\beta, y^*, \lambda_i^{-1}) = \pi(\beta)\pi(y^*|\beta, \lambda_i)\pi(\lambda_i)$. Note that the parameter is nothing but the variance of each observation. It distributes gamma $\lambda_i \sim \mathcal{G}(v/2, v/2)$.

2.2 Posterior distributions

The posterior, assuming homoskedasticity, is

$$\pi(\beta, y^*|y) \propto \pi(\beta)\pi(y^*|\beta)\prod_{i=1}^{n}\{1[y_i = 0]1[y_i^* < 0] + 1[y_i = 1]1[y_i^* \geq 0]\}$$

$$\pi(\beta, y^*|y) \propto \mathcal{N}_k(\beta|b_0, B_0)\mathcal{N}_n(y^*|X\beta, I_n)\prod_{i=1}^{n}\{1[y_i = 0]1[y_i^* < 0] + 1[y_i = 1]1[y_i^* \geq 0]\} \tag{31}$$

2.2.1 Conditional posterior distribution of β

The conditional posterior distribution of β is gotten by considering only the terms in the posterior which contains β. Then we have

$$\pi(\beta|y^*, y) \propto \pi(\beta)\pi(y^*|\beta)$$

$$= \frac{1}{(2\pi)^{k/2}|B_0|^{1/2}} \exp\left\{-\frac{1}{2}(\beta - b_0)'B_0^{-1}(\beta - b_0)\right\} \frac{1}{(2\pi)^{n/2}|I_n|^{1/2}} \exp\left\{-\frac{1}{2}(y^* - X\beta)'I_n^{-1}(y^* - X\beta)\right\}$$

Absorbing in the proportional constant the terms $\frac{1}{(2\pi)^{k/2}|B_0|^{1/2}}$ and $\frac{1}{(2\pi)^{n/2}|I_n|^{1/2}}$ we get

$$\propto \exp\left\{-\frac{1}{2}[(\beta - b_0)'B_0^{-1}(\beta - b_0) + (y^* - X\beta)'I_n^{-1}(y^* - X\beta)]\right\}$$

$$= \exp\left\{-\frac{1}{2}[\beta' B_0^{-1}\beta - b_0' B_0^{-1}\beta - \beta' B_0^{-1} b_0 + b_0' B_0^{-1} b_0 + y^{*\prime} I_n^{-1} y^* - y^{*\prime} I_n^{-1} X\beta - \beta' X' I_n^{-1} y^* + \beta' X' I_n^{-1} X\beta]\right\}$$

Note that $b_0' B_0^{-1}\beta$ is the transpose of $\beta' B_0^{-1} b_0$. Also note that $y^{*\prime} I_n^{-1} X\beta$ is the transpose of $\beta' X' I_n^{-1} y^*$. Given that all four expressions are matrices of size 1×1 it is true that $A = A'$. Then we could add them up. Thus, absorbing in the proportional constant the terms $b_0' B_0^{-1} b_0$ and $y^{*\prime} I_n^{-1} y^*$, we get

$$\propto \exp\left\{-\frac{1}{2}[\beta' B_0^{-1}\beta - 2b_0' B_0^{-1}\beta - 2y^{*\prime} I_n^{-1} X\beta + \beta' X' I_n^{-1} X\beta]\right\}$$

$$= \exp\left\{-\frac{1}{2}[\beta'(B_0^{-1} + X'X)\beta - 2(b_0' B_0^{-1} + y^{*\prime} I_n^{-1} X)\beta]\right\}$$

Then we transpose $[(b_0' B_0^{-1} + y^{*\prime} X)\beta]' = \beta'(X'y^* + B_0^{-1} b_0)$ to get

$$= \exp\left\{-\frac{1}{2}[\beta'(B_0^{-1} + X'X)\beta - 2\beta'(X'y^* + B_0^{-1} b_0)]\right\}$$

Defining $B_1 = (B_0^{-1} + X'X)^{-1}$ and $b = B_1(X'y^* + B_0^{-1} b_0)$, and adding and subtracting $b' B_1^{-1} b$; we would have

$$= \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1} B_1(X'y^* + B_0^{-1} b_0) + b' B_1^{-1} b - b' B_1^{-1} b]\right\}$$

$$= \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1} b + b' B_1^{-1} b]\right\} \exp\left\{\frac{b' B_1^{-1} b}{2}\right\}$$

Absorbing in the proportional constant $\exp\left\{\frac{b' B_1^{-1} b}{2}\right\}$ we have

$$\propto \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1} b + b' B_1^{-1} b]\right\}$$

Finally, factoring the square,

$$= \exp\left\{-\frac{1}{2}[(\beta - b)' B_1^{-1}(\beta - b)]\right\}$$

Which is the kernel of a Normal distribution with mean b and variance B_1. Given that this mean is a vector of size $K \times 1$ and the variance is a matrix of size $K \times K$, we are talking about a normal multivariate of dimension K. Thus, the posterior would be

$$\beta | y^*, y \sim \mathcal{N}_k(b, B_1)$$

2.2.2 Conditional posterior distribution of y^*

On the other hand, the conditional posterior distribution of y^* is gotten by only considering the terms that have y^*

$$\pi(y^*|\beta, y) \propto \mathcal{N}_n(y^*|X\beta, I_n) \prod_{i=1}^{n} \{1[y_i = 0]1[y_i^* < 0] + 1[y_i = 1]1[y_i^* \geq 0]\}$$

Note we know the posterior distribution in function of the value given by y_i. Hence, we could conclude it follows a truncated normal specified as:

$$y_i^* \sim \begin{cases} \mathcal{TN}_{(-\infty,0)}(\mathbf{x}_i\beta, 1) & \text{if } y_i = 0 \\ \mathcal{TN}_{(0,\infty)}(\mathbf{x}_i\beta, 1) & \text{otherwise} \end{cases}$$

3 Ordered Probit Model

The *Ordered Probit Model* is an extension of the probit model. The ordered probit model allows for many alternatives. However, these alternatives must be ordered. For instance, in markting surveys, consumers are often asked for their impresion of a product and must choose from alternatives Very Bad, Bad, Indifferent, Good and Very Good. In this case, the five choices have a logical ordering from Very Bad through Very Good (Koop et al., 2007)

Assume we have a non-observable endogenous variable y_i^* related to other K exogenous variables $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})'$ in a linear model given by $y_i^* = \mathbf{x}_i'\beta + \mu_i$, where $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$. Which impply $y^* \sim \mathcal{N}_n(X\beta, I_n)$ where $y^* = (y_1^*, y_2^*, \ldots, y_n^*)$.
The model specification is as follow:

$$y_i = j \Leftrightarrow \alpha_{j-1} < y_i^* \leqslant \alpha_j, j = 1, 2, \ldots, m \tag{32}$$

where $\alpha_0 = -\infty$, $\alpha_1 = 0$ and $\alpha_m = \infty$. We define $\alpha = \begin{bmatrix} \alpha_2 & \alpha_3 & \ldots & \alpha_{m-1} \end{bmatrix}'$.

So the likelihood function is:

$$\mathscr{L}(\mathbf{y}|\beta, \mathbf{X}) = \prod_{i=1}^{n} p(y_i^*|y_i, \beta, \alpha) \tag{33}$$

where $y_i^*|y_i, \beta, \alpha \sim \mathcal{TN}_{(\alpha_{y_i-1}, \alpha_{y_i})}(\mathbf{x}_i'\beta, 1)$

3.1 Priors Distributions

We employ independent prios for the parameters:

$$\pi(\beta, y^*, \alpha) = \pi(\beta)\pi(\alpha) \tag{34}$$

We assume a Gaussian prior for the location parameters, i.e., $\beta \sim \mathcal{N}_K(b_0, B_0)$ where $b_0 = 0_K$ and $B_0 = 1000I_K$ and The prior of α is assumed to be improper.

3.2 Posteriors Distributions

The posterios is:

$$\begin{aligned}
\pi(\beta, y^*, \alpha|y) \propto& \pi(\beta)\pi(\alpha)\prod_{i=1}^{n} p(y_i^*|y_i, \beta, \alpha) \\
\pi(\beta, y^*\alpha|y) \propto& \mathcal{N}_k(\beta|b_0, B_0)\mathcal{N}_n(y^*|X\beta, I_n)\prod_{i=1}^{n}(1[y_i = j]1[\alpha_{y_i-1} < y_i^* \leq \alpha_{y_i}])
\end{aligned} \tag{35}$$

3.2.1 Conditional Posterior Distributions of β

For the posterios of β we have to take in account only the terms that include β in the equations (35). So we get:

$$\begin{aligned}
\pi(\beta|y^*, y) \propto& \pi(\beta)\pi(y^*|\beta) \\
=& \frac{1}{(2\pi)^{k/2}|B_0|^{1/2}} \exp\left\{-\frac{1}{2}(\beta - b_0)' B_0^{-1}(\beta - b_0)\right\} \frac{1}{(2\pi)^{n/2}|I_n|^{1/2}} \exp\left\{-\frac{1}{2}(y^* - X\beta)' I_n^{-1}(y^* - X\beta)\right\} \\
=& \frac{1}{(2\pi)^{k/2}|B_0|^{1/2}(2\pi)^{n/2}|I_n|^{1/2}} \exp\left\{-\frac{1}{2}[(\beta - b_0)' B_0^{-1}(\beta - b_0) + (y^* - X\beta)' I_n^{-1}(y^* - X\beta)]\right\} \\
\propto& \exp\left\{-\frac{1}{2}[(\beta - b_0)' B_0^{-1}(\beta - b_0) + (y^* - X\beta)' I_n^{-1}(y^* - X\beta)]\right\} \\
\propto& \exp\left\{-\frac{1}{2}[\beta' B_0^{-1}\beta - b_0' B_0^{-1}\beta - \beta' B_0^{-1} b_0 + b_0' B_0^{-1} b_0 + y^{*\prime} I_n^{-1} y^* - y^{*\prime} I_n^{-1} X\beta - \beta' X' I_n^{-1} y^* + \beta' X' I_n^{-1} X\beta]\right\}
\end{aligned}$$

the terms that does not have β can be absorbed by the proportional contrant and grouping terms we get

$$\begin{aligned}
\propto& \exp\left\{-\frac{1}{2}[\beta' B_0^{-1}\beta - 2b_0' B_0^{-1}\beta - 2y^{*\prime} I_n^{-1} X\beta + \beta' X' I_n^{-1} X\beta]\right\} \\
=& \exp\left\{-\frac{1}{2}[\beta'(B_0^{-1} + X'X)\beta - 2(b_0' B_0^{-1} + y^{*\prime} X)\beta]\right\} \\
=& \exp\left\{-\frac{1}{2}[\beta'(B_0^{-1} + X'X)\beta - 2\beta'(X'y^* + B_0^{-1} b_0)]\right\}
\end{aligned}$$

Defining $B_1 = (B_0^{-1} + X'X)^{-1}$ and $b = B_1(X'y^* + B_0^{-1}b_0)$, and adding and subtracting $b'B_1^{-1}b$; we would have

$$
\begin{aligned}
&= \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1}B_1(X'y^* + B_0^{-1}b_0) + b'B_1^{-1}b - b'B_1^{-1}b]\right\} \\
&= \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1}b + b'B_1^{-1}b]\right\} \times \exp\left\{\frac{b'B_1^{-1}b}{2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}[\beta' B_1^{-1}\beta - 2\beta' B_1^{-1}b + b'B_1^{-1}b]\right\} \\
&= \exp\left\{-\frac{1}{2}[(\beta - b)' B_1^{-1}(\beta - b)]\right\}
\end{aligned}
$$

Which is the kernel of a Normal distribution with mean b and variance B_1. The mean vector b has a dimension of $K \times 1$ and the variance is a matrix of dimension $K \times K$. Thus, the posterior distribution of β is:

$$\beta | y^*, \alpha, y \sim \mathcal{N}_k(b, B_1) \tag{36}$$

3.2.2 Conditional Posterior Distributions of α

The oredered probit model gives for the probability that $y = j$

$$
\begin{aligned}
P(y = j \mid \beta, y^*, \alpha) =& P(\alpha_{j-1} < y_i^* \leq \alpha_j \mid \beta, y^*, \alpha) \\
=& P(\alpha_{j-1} < \mathbf{x}_i'\beta + \mu_i \leq \alpha_j \mid \beta, y^*, \alpha) \\
=& P(\alpha_{j-1} - \mathbf{x}_i'\beta < \mu_i \leq \alpha_j - \mathbf{x}_i'\beta \mid \beta, y^*, \alpha) \\
=& \Phi(\alpha_j - \mathbf{x}_i'\beta) - \Phi(\alpha_{j-1} - \mathbf{x}_i'\beta)
\end{aligned}
$$

where Φ is the normal standar cumulative distribution.

Given that the priof of α is improper, the posterior distribution for α must only satify the ordered restriction $\alpha_2 < \alpha_3, \ldots, \alpha_{m-1}$. Then the posterior distribution for α might be:

$$\alpha \mid \beta, y* \sim \prod_i \Phi(\alpha_{y_i} - \mathbf{x}_i'\beta) - \Phi(\alpha_{y_{i-1}} - \mathbf{x}_i'\beta)$$

Finally, we can draw y^* from a normal trucated distribution as follow:

$$y_i^* \mid \beta, \alpha \sim \mathcal{TN}_{(\alpha_{y_{i-1}}, \alpha_{y_i})}(\mathbf{x}_i'\beta, 1) \tag{37}$$

4 Negative Binomial Model

Assume that y is a random variable with parameter α such that $y = f(y|\alpha)$. Likewise, assume that α is a random variable with parameters (λ, γ); where λ is deterministic and γ is randomly distributed with parameter α. [2]. Under this assumptions, the marginal distribution of y can be expressed as

$$f(y|\lambda,\alpha) = \int (y|\lambda,\alpha)\, f(\gamma|\alpha)\, d\gamma$$

Now, if we assume that

$$y \sim Poisson\ (y|\delta) = \frac{e^{-\delta}\,\delta^{y}}{y!}$$

and that $\delta = \lambda\,\gamma$, where

$$\gamma \sim Gamma\ (\alpha,\alpha) = \frac{\alpha^{\alpha}}{\Gamma(\alpha)}\,\gamma^{(\alpha-1)}\,e^{-\alpha\gamma}$$

We would have that,

$$\begin{aligned}
f(y|\lambda,\alpha) &= \int f(y|\lambda,\alpha)\, f(\gamma|\alpha)\, d\gamma \\
&= \int_0^\infty \frac{e^{-\delta}\delta^{y}}{y!}\frac{\alpha^{\alpha}}{\Gamma(\alpha)}\gamma^{\alpha-1}e^{-\alpha\gamma}d\gamma \\
&= \int_0^\infty \frac{e^{-\lambda\gamma}(\lambda\gamma)^{y}}{y!}\frac{\alpha^{\alpha}}{\Gamma(\alpha)}\gamma^{\alpha-1}e^{-\alpha\gamma}d\gamma \\
&= \frac{\alpha^{\alpha}}{\Gamma(\alpha)\ y!}\int_0^\infty e^{-\lambda\gamma}\lambda^{y}\gamma^{y}\gamma^{\alpha-1}e^{-\alpha\gamma}d\gamma \\
&= \frac{\alpha^{\alpha}\lambda^{y}}{\Gamma(\alpha)\ y!}\int_0^\infty \gamma^{(y+\alpha-1)}\,e^{-(\lambda+\alpha)\gamma}d\gamma \\
&= \frac{\alpha^{\alpha}\lambda^{y}}{\Gamma(\alpha)\ y!}\int_0^\infty \gamma^{\phi-1}e^{-\psi\gamma}d\gamma
\end{aligned}$$

Where $\phi = y + \alpha$ and $\psi = \lambda + \alpha$. Additionally, recall that the Gamma function $\Gamma(\phi,\psi)$ is defined as $\int_0^\infty x^{\phi-1}e^{-\psi x}dx = \frac{\Gamma(\phi)}{\psi^{\phi}}$. Hence, the above expression can be rewrote as,

[2] The section for the Negative Binomial model was developed based primarily, but not solely, on Casella and Berger (2002), Greenberg (2008), Bolker (2007), Cameron and Trivedi (2005), and Rossi et al. (2005) and his R package *bayesm*, see Rossi (2015).

$$
\begin{aligned}
&= \frac{\lambda^y \alpha^\alpha}{\Gamma(\alpha)\ y!} \frac{\Gamma(y+\alpha)}{(\lambda+\alpha)^{(y+\alpha)}} \\
&= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)\, y!} \frac{\alpha^\alpha\, \lambda^y}{(\lambda+\alpha)^\alpha\ (\lambda+\alpha)^y} \\
&= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)\, y!} \left[\frac{\alpha}{\lambda+\alpha}\right]^\alpha \left[\frac{\lambda}{\lambda+\alpha}\right]^y
\end{aligned}
$$

This last expression, which is the likelihood function in our model, can be seen as a Negative Binomial distribution with parameters $\lambda = exp(x'\beta)$ where $\beta \sim N\left(\bar{\beta}, A^{-1}\right)$[3] and α where $\alpha \sim \Gamma(a,b)$[4].

If we define $\theta = \frac{\alpha}{\lambda+\alpha}$ and consequently, $(1-\theta) = 1 - \frac{\alpha}{\lambda+\alpha} = \frac{\lambda+\alpha-\alpha}{\lambda+\alpha} = \frac{\lambda}{\lambda+\alpha}$; we obtain the usual functional form of the negative binomial model as presented in equation 38.

$$
\mathcal{L}(y|\lambda,\alpha) = \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)\, y!} \theta^\alpha (1-\theta)^y \tag{38}
$$

Note that $\mathbb{E}[y] = \lambda$ and that $Var(y) = \lambda + \frac{1}{\alpha}\lambda^2$.

$$
\begin{aligned}
\mathbb{E}(y) &= \frac{\alpha(1-\theta)}{\theta} \\
&= \frac{\frac{\alpha\lambda}{\lambda+\alpha}}{\frac{\alpha}{\lambda+\alpha}} \\
&= \lambda
\end{aligned}
\qquad\qquad
\begin{aligned}
Var(y) &= \frac{\alpha(1-\theta)}{\theta^2} \\
&= \frac{\frac{\alpha\lambda}{\lambda+\alpha}}{\left(\frac{\alpha}{\lambda+\alpha}\right)^2} \\
&= \frac{\frac{\lambda}{1}}{\frac{\alpha}{\lambda+\alpha}} \\
&= \frac{(\lambda+\alpha)\lambda}{\alpha} \\
&= \frac{\lambda\alpha+\lambda^2}{\alpha} \\
&= \lambda + \frac{1}{\alpha}\lambda^2
\end{aligned}
$$

4.1 Priors Distributions

We employ independent prios for the parameters:

$$
\pi(\beta,\alpha) = \pi(\beta)\pi(\alpha) \tag{39}
$$

[3] This is the prior distibutions for β.

[4] This is the prior distribution for α.

We assume Gaussian distribution for β, i.e $\beta \sim \mathcal{N}(b_0,\, \Sigma^{-1})$ and the parameter α has a Gamma prior distribution $\alpha \sim \Gamma\,(a, b)$.

4.2 Posterior Distribution

The posterior for the parameter β is defined as

$$\begin{aligned} \pi\,(\beta|y) &\propto \mathcal{L}\,(y|\lambda, \beta)\,\pi\,(\beta) \\ &\propto \theta^{\alpha}\;(1-\theta)^{y}\;exp\left[-\frac{1}{2}\,(\beta - b_0)'\,\Sigma\,(\beta - b_0)\right] \end{aligned} \tag{40}$$

We might use the Metropolis-Hasting algorithm for sampling from the last distribution as follow. Let $u \sim U(0, 1)$

$$\begin{aligned} u \le acc &= min\left\{\frac{\pi\,(\beta_{can})}{\pi\,(\beta_s)}\,, 1\right\} \\ &= min\left\{\frac{\theta^{\alpha}\;(1-\theta)^{y}\;exp\left[-\frac{1}{2}\,(\beta_{can} - b_0)'\,\Sigma\,(\beta_{can} - b_0)\right]}{\theta^{\alpha}\;(1-\theta)^{y}\;exp\left[-\frac{1}{2}\,(\beta_s - b_0)'\,\Sigma\,(\beta_s - b_0)\right]}\,, 1\right\} \\ &= min\left\{exp\left\{-\frac{1}{2}\left[\left(\beta_{can} - \hat{\beta}\right)'\,\Sigma\,(\beta_{can} - b_0) - (\beta_s - b_0)'\,\Sigma\,(\beta_s - b_0)\right]\right\}\,, 1\right\} \end{aligned}$$

where β_s is a value avaiable from the β distribution and β_{can} is a candidate.
Finally, the posterior for the parameter α is,

$$\begin{aligned} \pi\,(\alpha|y) &\propto \mathcal{L}\,(y|\lambda, \alpha)\,\pi\,(\alpha) \\ &\propto \left[\frac{\alpha}{\alpha + \lambda}\right]^{\alpha}\;\left[\frac{\lambda}{\alpha + \lambda}\right]^{y}\;\alpha^{a-1}\,e^{-b\alpha} \end{aligned} \tag{41}$$

Using the Metropolis-Hasting algorithm for sampling we have,

$$\begin{aligned} u \le acc &= min\left\{\frac{\pi\,(\alpha_{can})}{\pi\,(\alpha_s)}\,, 1\right\} \\ &= min\left\{\frac{\left[\frac{\alpha_{can}}{\alpha_{can}+\lambda}\right]^{\alpha_{can}}\;\left[\frac{\lambda}{\alpha_{can}+\lambda}\right]^{y}\;\alpha_{can}^{a-1}\,e^{-b\alpha_{can}}}{\left[\frac{\alpha_s}{\alpha_s+\lambda}\right]^{\alpha_s}\;\left[\frac{\lambda}{\alpha_s+\lambda}\right]^{x}\;\alpha_s^{a-1}\,e^{-b\alpha_s}}\,, 1\right\} \end{aligned}$$

5 Multivariate Regression Model

We assume that our observations $Y = (\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_m})$ an $n \times m$ matrix. It is generated by the following model:

$$Y = XB + U \tag{42}$$

where X is an $n \times k$ matrix, $B = (\beta_1, \beta_2, \ldots, \beta m)$ is a $k \times m$ matrix of regression parameters, and $U = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_m})$ matrix of unobserved random disturbance terms with each row of U has a distribution $\mathcal{N}(\mathbf{0}_m, \Sigma)$. Under these assumptions the pdf for Y, given X, B and Σ is

$$p(Y \mid X, B, \Sigma) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} tr (Y - XB)'(Y - XB)\Sigma^{-1}\right] \tag{43}$$

"tr" denotes the trace operation. Noting that,

$$\begin{aligned}
(Y - XB)'(Y - XB) =& Y'Y - 2Y'XB + B'X'XB \\
=& Y'Y - 2Y'XB + B'X'XB - 2Y'X[(X'X)^{-1}X'Y] + 2Y'X[(X'X)^{-1}X'Y] \\
=& Y'Y - 2Y'X(X'X)^{-1}X'XB + B'X'XB - 2Y'X[(X'X)^{-1}X'Y] + 2Y'X[(X'X)^{-1}X'Y] \\
=& Y'Y - 2Y'X[(X'X)^{-1}X'Y] + Y'X[(X'X)^{-1}X'Y] \\
& + B'X'XB - 2Y'X(X'X)^{-1}X'XB + Y'X[(X'X)^{-1}X'Y] \\
=& (Y - X\hat{B})'(Y - X\hat{B}) + (B - \hat{B})'X'X(B - \hat{B}) \\
=& S + (B - \hat{B})'X'X(B - \hat{B})
\end{aligned} \tag{44}$$

where

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

and

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

We can write the likelihood function of B and Σ as follow:

$$\mathcal{L}(B, \Sigma \mid Y, X) \propto |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} tr\, S\Sigma^{-1} - \frac{1}{2} tr\, (B - \hat{B})'X'X(B - \hat{B})\Sigma^{-1}\right] \tag{45}$$

5.1 Priors Distributions

We assume diffuse prior pdf and that the elements of B and Σ are independently distributed; that is,

$$p(B, \Sigma) = p(B)p(\Sigma)$$

we take

$$p(B) = const$$

and

$$p(\Sigma) = |\Sigma|^{-(m+1)/2}$$

which implies the following prior pdf on Σ^{-1}

$$p(\Sigma^{-1}) = \left|\Sigma^{-1}\right|^{-(m+1)/2}$$

5.2 Posterior Distribution

Combining the likelihood function and the priors above we obtain the following joint posterior pdf for the parameters:

$$p(B, \Sigma \mid Y, X) \propto |\Sigma|^{-(m+1)/2} \times \exp\left[-\frac{1}{2} tr[S + (B - \hat{B})' X' X (B - \hat{B})] \Sigma^{-1}\right] \tag{46}$$

and

$$p(B, \Sigma^{-1} \mid Y, X) \propto \left|\Sigma^{-1}\right|^{-(m+1)/2} \times \exp\left[-\frac{1}{2} tr[S - (B - \hat{B})' X' X (B - \hat{B})] \Sigma^{-1}\right] \tag{47}$$

We can write

$$p(B, \Sigma \mid Y, X) = p(B \mid \Sigma, Y, X) p(\Sigma \mid Y, X)$$

then,

$$p(B \mid \Sigma, Y, X) \propto |\Sigma|^{-k/2} \times \exp\left[-\frac{1}{2} tr[(B - \hat{B})' X' X (B - \hat{B})] \Sigma^{-1}\right] \tag{48}$$

and

$$p(\Sigma \mid Y, X) \propto |\Sigma|^{-\nu/2} \times \exp\left[-\frac{1}{2} tr\, S \Sigma^{-1}\right] \tag{49}$$

with $\nu = n - k + m + 1$

5.3 Predictive pdf

We wish to derive the predictive pdf on the dependent variable for next p times periods W, a $p \times m$ matrix, assumed to be generate by the same model generating Y; that is,

$$W = ZB + V \tag{50}$$

where Z is a $p \times k$ matrix. Then the predictive pdf for W is given by

$$p(W \mid Y, X, Z) \propto \int \int p(B, \Sigma^{-1} \mid Y, X) p(W \mid Z, B, \Sigma^{-1}) dB d\Sigma^{-1} \tag{51}$$

where

$$p(W \mid Z, B, \Sigma^{-1}) \propto \left|\Sigma^{-1}\right|^{p/2} \exp\left\{-\frac{1}{2} tr[(W - ZB)' (W - ZB) \Sigma^{-1}]\right\} \tag{52}$$

combining (47) and (52) we get,

$$\left|\Sigma^{-1}\right|^{\frac{n+p-m-1}{2}} exp\left\{-\frac{1}{2} tr A \Sigma^{-1}\right\}$$

with $A = (Y - XB)^{'}(Y - XB) + (W - ZB)^{'}(W - ZB)$. We can use the inverse wishart properties in order to integrate (51) with respet to Σ^{-1}, which yields

$$|A|^{-\frac{n+p}{2}} = \left|(Y - XB)^{'}(Y - XB) + (W - ZB)^{'}(W - ZB)\right|^{-\frac{n+p}{2}} \tag{53}$$

To integate wiht respect to B we complete the square on B,

$$\begin{aligned} A =& Y'Y + W'W - 2Y'XB - 2W'ZB + B'X'XB + B'Z'ZB \\ =& Y'Y + W'W + B'(X'X + Z'Z)B - 2(Y'X + W'Z)B \\ =& Y'Y + W'W + B'MB - 2M\tilde{B}B + \tilde{B}'M\tilde{B} - \tilde{B}'M\tilde{B} \\ =& Y'Y + W'W - \tilde{B}'M\tilde{B} + (B - \tilde{B})'M(B - \tilde{B}) \end{aligned} \tag{54}$$

where $M = X'X + Z'Z$ and $\tilde{B} = M^{-1}(X'Y + Z'W)$. So we have that,

$$|A|^{-\frac{n+p}{2}} = \left|Y'Y + W'W - \tilde{B}'M\tilde{B} + (B - \tilde{B})'M(B - \tilde{B})\right|^{-\frac{n+p}{2}}$$

Using properties of the generalized multivariate Student t pdf, we can perform the integration with respect to the parameter B, which yield,

$$p(W \mid Y, X, Z) \propto \left|Y'Y + W'W - \tilde{B}'M\tilde{B}\right|^{-\frac{n+p-k}{2}} \tag{55}$$

To simplify this expression we complete square on W as follows,

$$\begin{aligned} Y'Y + W'W - \tilde{B}'M\tilde{B} =& Y'Y + W'W - (X'Y + Z'W)'M^{-1}(X'Y + Z'W) \\ =& Y'(I - XM^{-1}X')Y + W'(I - ZM^{-1}Z')W \\ & - Y'XM^{-1}Z'W - W'ZM^{-1}X'Y \\ =& Y'(I - XM^{-1}X' - XM^{-1}Z'C^{-1}ZM^{-1}X')Y \\ & + (W - C^{-1}ZM^{-1}X'Y)'C(W - C^{-1}ZM^{-1}X'Y) \end{aligned} \tag{56}$$

where $C = I - ZM^{-1}Z'$ and $C^{-1} = I + Z(X'X)^{-1}Z'$. Also

$$\begin{aligned} C^{-1}ZM^{-1} =& [I - ZM^{-1}Z']ZM^{-1} \\ =& Z[I + (X'X)^{-1}Z'Z]M^{-1} \\ =& Z(X'X)^{-1}(X'X + Z'Z)M^{-1} \\ =& Z(X'X)^{-1} \end{aligned} \tag{57}$$

then,

$$XM^{-1}X' - XM^{-1}Z'C^{-1}ZM^{-1}X' = X[M^{-1} + M^{-1}Z'Z(X'X)^{-1}]X'$$

$$
\begin{aligned}
&= XM^{-1}(X'X + Z'Z)(X'X)^{-1}X' \\
&= X(X'X)^{-1}X' \qquad (58)
\end{aligned}
$$

finally, substituting (57) and (58) in (56) we get,

$$Y'Y + W'W - \tilde{B}'M\tilde{B} = Y'[I - X(X'X)^{-1}X]Y + (W - Z\tilde{B})'C(W - Z\tilde{B})$$

Noting that

$$Y'[I - X(X'X)^{-1}X]Y = (Y - X\hat{B})^{'}(Y - X\hat{B}) = S$$

we can write the preditive pdf (55) as

$$p(W \mid Y, X, Z) \propto \left| S + (W - Z\tilde{B})'(I - ZM^{-1}Z')(W - Z\tilde{B} \right|^{-\frac{n+p-k}{2}} \qquad (59)$$

The matrix of future observations has a pdf in the generalized multivariate Student t.

6 Endogenous Covariate Model

6.1 Instrumental Variable

$$y_i = \mathbf{x}_i'\beta_1 + \beta_s \mathbf{x}_{is} + u_i \qquad (60)$$

$$\mathbf{x}_{is} = \mathbf{x}_i'\gamma_1 + \mathbf{z}_i'\gamma_2 + v_i \qquad (61)$$

We have k_1 regressors in $\mathbf{x}_i'$, and k_2 regressors in $\mathbf{z}_i'$ which are the instruments, that need to be orthogonal to u_i and v_i. In order to devise an MCMC algorithm to sample the parameters, it is necessary to specify prior distributions for the parameters. Let $\beta = (\beta_1', \beta_s')'$,$\gamma = (\gamma_1', \gamma_2')'$, and we assume that $(u_i, v_i) \sim \mathcal{N}(\mathbf{0}_2, \Sigma)$.

The variance-covariance matrix is the same as the one assume for $c_i = (x_{is}, u_i)'$. $c_i \sim \mathcal{N}(\xi, \Sigma)$ because as we have said before, conditioning on the regressors and considering constant the parameters, the response variable, in this case $\mathbf{x}_{is}$ has the same variance than the stochastic disturbance, however not the same expected value, as you can notice. Now, we can assume:

$$\gamma \sim \mathcal{N}_{k_1+k_2}(\gamma_0, G_0)$$

$$\beta \sim \mathcal{N}_{k_1+1}(\beta_0, B_0)$$

$$\Sigma^{-1} \sim \mathcal{W}(\Sigma_0, v_0)$$

$\gamma \sim \mathcal{N}_{k_1+k_2}(\gamma_0, G_0)$ corresponds to a normal multivariate distribution of dimension $k_1 + k_2$ because we want to sample k_1 parameters corresponding to the k_1 covariates ($\mathbf{x}_i'$) present in equations (60) and (61), and k_2 parameters corresponding to the k_2 covariates ($\mathbf{z}_i'$) present only in equation (61).
$\beta \sim \mathcal{N}_{k_1+1}(b_0, B_0)$ corresponds to a normal multivariate distribution of dimension $k_1 + 1$ because we have $k_1 + 1$ covariates in equation 60 , k_1 in $\mathbf{x}_i'$ and the other one corresponds to $\mathbf{x}_{is}$.
$\Sigma^{-1} \sim \mathcal{W}(v_0, \Sigma_0)$ corresponds to a Wishart distribution of dimension 2, because it corresponds to the variance-covariance matrix of the normally bivariate distributed function of u_i and v_i with parameters v_0 (degrees of freedom) and Σ_0 (scale matrix).

6.1.1 Prior Distribution

The priors are as follow:

$$\pi(\gamma) = \frac{exp\big[-\frac{1}{2}(\gamma-\gamma_0)G_0^{-1}(\gamma-\gamma_0)'\big]}{(2\pi)^{(k_1+k_2)}|G_0|^{\frac{1}{2}}}$$
$$\pi(\beta) = \frac{exp\big[-\frac{1}{2}(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)'\big]}{(2\pi)^{(K_1+1)}|B_0|^{\frac{1}{2}}}$$
$$\pi(\Sigma) = \frac{|\Sigma|^{-\frac{v_0-3}{2}}|\Sigma_0^{-1}|^{\frac{v_0}{2}}exp\big[-\frac{1}{2}tr(\Sigma_0^{-1}\Sigma^{-1})\big]}{4\Gamma_2(\frac{v_0}{2})}$$

Notice that we assumed a prior for Σ^{-1} but we are sampling Σ, nonetheless we know that if $\Sigma^{-1} \sim \mathcal{W}(\Sigma_0, v_0)$, then $\Sigma \sim \mathcal{IW}_2(\Sigma_0^{-1}, v_0)$. In the Inverse-Wishart PDF $\frac{v_0-3}{2}$ was initially $\frac{v_0-p-1}{2}$ but p (the dimension of the multivariate distribution) is 2. And 4 was initially $2^{\frac{np}{2}}$ but n (degrees of freedom) is v_0 and $p = 2$.

6.1.2 Sampling Σ

In order to sample Σ we need to find the posterior distribution, which is: $\pi(\Sigma|\beta, \gamma, y) = f(y|\Sigma)\pi(\Sigma)$. We therefore need the likelihood function $f(y|\Sigma)$. We are assuming a sample size of n, and we assume that the u_i are independent and identically normally distributed. Assume $u_i \sim \mathcal{N}(0, \sigma^2)$.
If the each vector of y_i is independent, then the likelihood of the sample size is the product of the likelihood of each y_i from $i = 1$ to n. Then, the likelihood can be written as:

$$f(y_1, y_2, ..., y_n|\sigma^2) = f(y_1|\sigma^2)f(y_2|\sigma^2)...f(y_n|\sigma^2)$$

$$
\begin{aligned}
&= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} exp\bigg[-\frac{1}{2\sigma^2}(y_1 - \mathbf{x}_1'\beta)^2\bigg]\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} exp\bigg[-\frac{1}{2\sigma^2}(y_2 - \mathbf{x}_2'\beta)^2\bigg]...\\
&\times \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} exp\bigg[-\frac{1}{2\sigma^2}(y_n - \mathbf{x}_n'\beta)^2\bigg]\\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_1'\beta)^2\bigg]\\
f(y_1, y_2, ..., y_n|\sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} exp\bigg[-\frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)\bigg]
\end{aligned}
$$

However this likelihood function corresponds to a normal univariate distribution, and we are conditioning on y, which means we are conditioning on u_i and v_i. Let $X_i' = (\mathbf{x}_i', \mathbf{x}_{is})$ and $Z_i' = (\mathbf{x}_i', \mathbf{z}_i')$. Recall that $c_i = (\mathbf{x}_{is}, u_i)'$. $c_i \sim \mathcal{N}_2(\xi, \Sigma)$

We know that a normal bivariate distribution has the following likelihood function:

$$
f(x|\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} exp\bigg[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\bigg]
$$

We can express $u_i = y_i - X_i'\beta$ and $v_i = \mathbf{x}_{is} - Z_i'\gamma)$. Then we can write the likelihood function for u_i and v_i as:

$$
f(y|\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{n}{2}}} exp\bigg[-\frac{1}{2}\sum_{i}^{n}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\bigg]
$$

We would like to find the posterior distribution of Σ, that is:

$$
\pi(\Sigma|\beta, \gamma, y) = \pi(\Sigma)f(y|\Sigma)
$$

Because β and γ are present in $\pi(y|\Sigma)$ as we saw earlier. Then, the posterior can be written as:

$$
\begin{aligned}
\pi(\Sigma|\beta, \gamma, y) &= \frac{|\Sigma|^{-\frac{v_0-3}{2}}|\Sigma_0^{-1}|^{\frac{v_0}{2}} exp\big[-\frac{1}{2}tr(\Sigma_0^{-1}\Sigma^{-1})\big]}{4\Gamma_2(\frac{v_0}{2})}\\
&\times \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{n}{2}}} exp\bigg[-\frac{1}{2}\sum_{i}^{n}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\bigg]\\
\pi(\Sigma|\beta, \gamma, y) &\propto |\Sigma|^{-\frac{v_0-3}{2}}|\Sigma|^{-\frac{n}{2}} exp\bigg[-\frac{1}{2}tr(\Sigma_0^{-1}\Sigma^{-1})\bigg]\\
&\times exp\bigg[-\frac{1}{2}\sum_{i}^{n}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\bigg]
\end{aligned}
$$

The constant absorbed by the constant was:

$$
\frac{|\Sigma_0^{-1}|^{\frac{v_0}{2}}}{4\Gamma_2(\frac{v_0}{2})}(2\pi)^{-\frac{n}{2}}
$$

In order to find the posterior, we would like to include in the trace operator the term corresponding to the normal distribution in the exponential. We can say that:

$$exp\Big[-\frac{1}{2}\sum_i^n (y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\Big]$$

$$= exp\Big[-\frac{1}{2}\sum_i^n tr\Big((y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\Big)\Big]$$

$$= exp\Big[-\frac{1}{2}tr\Big(\sum_i^n (y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}\Big)\Big]$$

$$= exp\Big[-\frac{1}{2}tr\Big(\sum_i^n \begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\Big)\Big]$$

Now, we can write the posterior as:

$$\pi(\Sigma|\beta,\gamma,y) \propto |\Sigma|^{-\frac{v_0-3}{2}}|\Sigma|^{-\frac{n}{2}} exp\Big[-\frac{1}{2}tr(\Sigma_0^{-1}\Sigma^{-1})\Big]$$

$$\times\, exp\Big[-\frac{1}{2}tr\Big(\sum_i^n \begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Sigma^{-1}\Big)\Big]$$

$$\pi(\Sigma|\beta,\gamma,y) \propto |\Sigma|^{-\frac{v_0+n-3}{2}} exp\Big[-\frac{1}{2}tr\Big[\Big(\Sigma_0^{-1} + \sum_i^n \begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Big)\Sigma^{-1}\Big]\Big]$$

Which has the form of a $\mathcal{W} \sim (v_1, R_1)$ (A Wishart Bivariate distribution) where $v_1 = v_0 + n$ and $R_1 = \Big[\Sigma_0^{-1} + \sum_i^n \begin{pmatrix} y - X_i'\beta \\ \mathbf{x}_{is} - z_i'\gamma \end{pmatrix}(y_i - X_i'\beta, \mathbf{x}_{is} - Z_i'\gamma)'\Big]^{-1}$

Then, we can say that: $\Sigma \sim \mathcal{W}(v_1, R_1)$

6.1.3 Sampling β

We assumed a Normal distribution $\mathcal{N}(b_0, B_0)$ for beta. The likelihood function corresponding to y is also normal, therefore the posterior will be also Normal. The posterior of β can be written as:

$$\pi(\beta|\Sigma,\gamma,y) = \pi(\beta)f(y_i, \mathbf{x}_{is}|\beta,\Sigma,\gamma)$$

For the sake of aesthetics lets call henceforth $\theta = (\beta, \gamma, \Sigma)$. In order to find this posterior, we use the fact that:

$$f(y_i, \mathbf{x}_{is}|\theta) = f(\mathbf{x}_{is}|\theta)f(y_i|\mathbf{x}_{is},\theta)$$

To write the likelihood function we need to know the conditioned expected value of y_i:

$$E(y_i|\mathbf{x}_{is},\theta) = E(\mathbf{x}_i'\beta_1 + \beta_s\mathbf{x}_{is} + u_i|\mathbf{x}_{is},\theta) = x_i'\beta_1 + \beta_s\mathbf{x}_{is} + E(u_i|\mathbf{x}_{is},\theta)$$

But the conditioned expected value of u_i is not zero, because we assumed $(u_i, v_i) \sim \mathcal{N}_2(0, \Sigma)$ .Hence,

$$E(u_i|\mathbf{x}_{is}, \theta) = E(u_i|v_i)$$

The last expected value could be calculated as the expected value of X given Y, if X and Y come from a normal bivariate distribution. Then

$$E(u_i|v_i) = \mu_u + \frac{(v_i - \mu_v)\rho\sigma_{22}^{\frac{1}{2}}}{\sigma_{11}^{\frac{1}{2}}} = \frac{v_i\sigma_{12}}{\sigma_{11}}$$

Because by construction, $\mu_u = \mu_v = 0$ and by definition $\rho = \frac{\sigma_{12}}{\sigma_{11}^{\frac{1}{2}}\sigma_{22}^{\frac{1}{2}}}$ However, we can write v_i as: $\mathbf{x}_{is} - x_i'\gamma_1 - \mathbf{z}_i'\gamma_2$, or following the notation above: $v_i = \mathbf{x}_{is} - Z_i'\gamma$. Then, we can write the expected value as:

$$E(u_i|v_i) = \frac{(\mathbf{x}_{is} - Z_i'\gamma)\sigma_{12}}{\sigma_{11}}$$

Then:

$$E(y_i|\mathbf{x}_{is}, \theta) = \mathbf{x}_i'\beta_1 + \beta_s\mathbf{x}_{is} + \frac{(\mathbf{x}_{is} - Z_i'\gamma)\sigma_{12}}{\sigma_{11}}$$

Following the notation above:

$$E(y_i|\mathbf{x}_{is}, \theta) = X_i'\beta + \frac{(\mathbf{x}_{is} - Z_i'\gamma)\sigma_{12}}{\sigma_{11}}$$

We know that y_i has a normal distribution because it inherits the distribution from u_i. To write the likelihood, we need the variance. That corresponds to a conditional variance of a normal bivariate distribution, which is:

$$\begin{aligned}
Var[y_i|\mathbf{x}_{is}] &= \sigma_y^2(1 - \rho^2) \\
&= \sigma_{11}\left(1 - \frac{\sigma_{12}^2}{\sigma_{22}\sigma_{11}}\right) \\
&= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} \\
\omega = Var[y_i|\mathbf{x}_{is}] &= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}
\end{aligned}$$

We can say thar:

$$y_i|\mathbf{x}_{is} \sim \mathcal{N}(X_i'\beta - \frac{(\mathbf{x}_{is} - Z_i'\gamma)\sigma_{12}}{\sigma_{11}}, \omega)$$

We are assuming a sample size of n, and we assume that the u_i are independent and identically normally distributed. We then can find the likelihood for all y_i as we did before, writing the probability density for the observed sample:

$$f(y_1, y_2, ..., y_n|\mathbf{x}_{is}, \theta) = \left(\frac{1}{2\pi\omega}\right)^{\frac{n}{2}} exp\left[-\frac{1}{2\omega}\sum_{i=1}^{n}\left(y_i - X_i'\beta - \frac{(\mathbf{x}_{is} - Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\right)^2\right]$$

We also know that the likelihood function for $\mathbf{x}_{is}$ is:

$$f(\mathbf{x}_{1s}, \mathbf{x}_{2s}, ...\mathbf{x}_{ns}|\theta) = \left(\frac{1}{2\pi\sigma_{22}}\right)^{\frac{n}{2}} exp\Big[-\frac{1}{2\sigma_{22}}\sum_{i=1}^{n}(\mathbf{x}_{is} - Z_i'\gamma)^2\Big]$$

The variance is not conditional, because we are not conditioning on y_i.

As we said before, the posterior distribution is:

$$\pi(\beta|\Sigma, \gamma, y) = \pi(\beta) f(y_i, \mathbf{x}_{is}|\beta, \Sigma, \gamma)$$

$$\pi(\beta|\Sigma,\gamma,y) = \frac{exp\big[-\frac{1}{2}(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)'\big]}{(2\pi)^{(k_1+1)}|B_0|^{\frac{1}{2}}}\left(\frac{1}{2\pi\sigma_{22}}\right)^{\frac{n}{2}} exp\Big[-\frac{1}{2\sigma_{22}}\sum_{i=1}^{n}(\mathbf{x}_{is} - Z_i'\gamma)^2\Big]$$
$$\times\left(\frac{1}{2\pi\omega}\right)^{\frac{n}{2}} exp\Big[-\frac{1}{2\omega}\sum_{i=1}^{n}\Big(y_i - X_i'\beta - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2\Big]$$
$$\pi(\beta|\Sigma,\gamma,y) \propto exp\big[-\frac{1}{2}(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)'\big] exp\Big[-\frac{1}{2\omega}\sum_{i=1}^{n}\Big(y_i - X_i'\beta - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2\Big]$$

We can rewrite $exp\Big[-\frac{1}{2\omega}\sum_{i=1}^{n}\Big(y_i - X_i'\beta - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2\Big]$ as

$$exp\Big[-\frac{1}{2\omega}\sum_{i=1}^{n}\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}} - X_i'\beta\Big)^2\Big]$$

And expanding we yield:

$$exp\Big[-\frac{1}{2\omega}\sum_{i=1}^{n}\Big(\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2 + \beta' X_i X_i'\beta - 2\beta' X_i\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)\Big)\Big]$$

Now, because the summatory is a linear operator, we can rewrite the posterior as:

$$\pi(\beta|\Sigma,\gamma,y) \propto exp\Big[-\frac{1}{2}\Big[(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)' + \omega^{-1}\sum_{i=1}^{n}\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2$$
$$+\omega^{-1}\beta'\sum_{i=1}^{n} X_i X_i'\beta - 2\omega^{-1}\beta'\sum_{i=1}^{n}\Big[X_i\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)\Big]\Big]$$

We can expand the term $(\beta-\beta_0)B_0^{-1}(\beta-\beta_0)'$ as:

$$\beta' B_0^{-1}\beta - 2\beta' B_0^{-1} b_0 + b_0' B_0^{-1} b_0$$

Focusing in the big term on the brackets we have:

$$\beta' B_0^{-1}\beta - 2\beta' B_0^{-1} b_0 + b_0' B_0^{-1} b_0 + \omega^{-1}\sum_{i=1}^{n}\Big(y_i - \frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Big)^2$$

$$+\omega^{-1}\beta'\sum_{i=1}^{n}X_iX_i'\beta-2\omega^{-1}\beta'\sum_{i=1}^{n}\Bigg[X_i\Bigg(y_i-\frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Bigg)\Bigg]$$

Factorizing, we yield:

$$\beta'\Bigg(B_0^{-1}+\omega^{-1}\sum_{i=1}^{n}X_iX_i'\Bigg)\beta-2\beta'\Bigg(B_0^{-1}b_0+\omega^{-1}\sum_{i=1}^{n}\Bigg[X_i\Bigg(y_i-\frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Bigg)\Bigg]\Bigg)$$

We do not write the terms $b_0'B_0^{-1}b_0$ and $\omega^{-1}\sum_{i=1}^{n}\Bigg(y_i-\frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Bigg)^2$ because they are absorbed by the constant.

Lets call $B_1=[B_0^{-1}+\omega^{-1}\sum_{i=1}^{n}X_iX_i']^{-1}$ and

$$\beta^1=B_1\Bigg(B_0^{-1}b_0+\omega^{-1}\sum_{i=1}^{n}\Bigg[X_i\Bigg(y_i-\frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Bigg)\Bigg]\Bigg)$$

We add and subtract to the expression of interest

$$\beta^{1\prime}B_1^{-1}\beta^1$$

Because we known that

$$(x-\mu)'\Sigma^{-1}(x-\mu)=x'\Sigma^{-1}x-2x'\Sigma^{-1}\mu+\mu'\Sigma^{-1}\mu$$

Which is what we have if we add $\beta^{1\prime}B_1^{-1}\beta^1$ and forget the same expression with the negative sign, because it does not deppend on β

We then can factorize

$$\beta'\Bigg(B_0^{-1}+\omega^{-1}\sum_{i=1}^{n}X_iX_i'\Bigg)\beta-2\beta'\Bigg(B_0^{-1}b_0+\omega^{-1}\sum_{i=1}^{n}\Bigg[X_i\Bigg(y_i-\frac{(\mathbf{x}_{is}-Z_i'\gamma)\sigma_{12}}{\sigma_{11}}\Bigg)\Bigg]\Bigg)+\beta^{1\prime}B_1^{-1}\beta^1$$

As

$$(\beta-\beta^1)'B_1^{-1}(\beta-\beta^1)$$

We then have:

$$\pi(\beta|\Sigma,\gamma,y)\propto exp\Bigg[-\frac{1}{2}(\beta-\beta^1)'B_1^{-1}(\beta-\beta^1)\Bigg]$$

Therefore, we can say that $\beta\sim\mathcal{N}(\beta^1,B_1)$

6.1.4 Sampling γ

Basically, we have to change y_i for $\mathbf{x}_{is}$ and β for γ.

$$E(v_i|\beta,\gamma,y)=\frac{u_i\sigma_{12}}{\sigma_{11}}$$

$$E(\mathbf{x}_{is}|\beta,\gamma,y) = Z_i'\gamma + \frac{(y_i - X_i'\beta)\sigma_{12}}{\sigma_{11}}$$
$$Var(\mathbf{x}_{is}|\beta,\gamma,y) = \omega_1 = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$$

We can say that:

$$\mathbf{x}_{is}|y_i \sim \mathcal{N}(Z_i'\gamma + \frac{(y_i - X_i'\beta)\sigma_{12}}{\sigma_{11}}, \omega_1)$$

The posterior distribution of γ is:

$$\pi(\gamma|\beta,\Sigma,y) = \pi(\gamma)f(y_i,\mathbf{x}_i s|\theta)$$

We write $f(y_i,\mathbf{x}_{is}|\theta)$ as $f(\mathbf{x}_{is}|y_i,\theta)f(y|\theta)$. Then:

$$\pi(\gamma|\beta,\Sigma,y) = \pi(\gamma)f(\mathbf{x}_{is}|y_i,\theta)f(y|\theta)$$

$$\pi(\gamma|\beta,\Sigma,y) = \frac{exp\big[-\frac{1}{2}(\gamma-\gamma_0)G_0^{-1}(\gamma-\gamma_0)'\big]}{(2\pi)^{(k_1+k_2)}|G_0|^{\frac{1}{2}}}\left(\frac{1}{2\pi\omega_1}\right)^{\frac{n}{2}} exp\left[-\frac{1}{2\omega_1}\sum_{i=1}^{n}\left(\mathbf{x}_{is} - Z_i'\gamma - \frac{(y_i - X_i'\beta)\sigma_{12}}{\sigma_{11}}\right)^2\right]$$
$$\times\left(\frac{1}{2\pi\sigma_{11}}\right)^{\frac{n}{2}} exp\left[-\frac{1}{2\sigma_{11}}\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i'\beta\right)^2\right]$$
$$\pi(\gamma|\beta,\Sigma,y) \propto exp\big[-\frac{1}{2}(\gamma-\gamma_0)G_0^{-1}(\gamma-\gamma_0)'\big]\left(\frac{1}{2\pi\sigma_{11}}\right)^{\frac{n}{2}} exp\left[-\frac{1}{2\sigma_{11}}\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i'\beta\right)^2\right]$$

We expand the term in brackets and omit the terms that do not depend on γ to yield:

$$\gamma' G_0^{-1}\gamma - 2\gamma' G_0^{-1}\gamma_0 + \omega_1^{-1}\gamma'\sum_{i=1}^{n} Z_i Z_i'\gamma - 2\omega_1^{-1}\gamma'\sum_{i=1}^{n}\left[Z_i\left(\mathbf{x}_{is} - \frac{(y_i - X_i'\beta)\sigma_{12}}{\sigma_{22}}\right)\right]$$

Defining $G_1 = [G_0^{-1} + \omega_1^{-1}\sum_{i=1}^{n} Z_i Z_i']^{-1}$ and

$$\gamma^1 = G_1\left(G_0^{-1}\gamma_0 + \omega_1^{-1}\sum_{i=1}^{n}\left[Z_i\left(\mathbf{x}_{is} - \frac{(y_i - X_i'\beta)\sigma_{12}}{\sigma_{22}}\right)\right]\right)$$

If we add and substract $\gamma^{1\prime}G_1^{-1}\gamma^1$ (completing the square) and do the same as above we yield:

$$\gamma \sim \mathcal{N}(\gamma^1, G_1)$$

6.2 Endogenous Switching Model

A latent variable y_{1i}^* determines whether the outcome observed is y_{2i} or y_{3i}. Specifically, we observe whether y_{1i}^* is positive or negative,

$$y_{1i} = \begin{cases} 1, & y_{1i}^* > 0 \\ 0, & y_{1i}^* \leqslant 0 \end{cases} \tag{62}$$

and observe exactly one of y_{2i} or y_{3i}, according to

$$y_i = \begin{cases} y_{2i}, & y^*_{1i} > 0 \\ y_{3i}, & y^*_{1i} \leqslant 0 \end{cases} \tag{63}$$

where

$y_{2i} = j \Leftrightarrow \alpha_{2,j-1} < y^*_{2i} \leqslant \alpha_{2,j}, j = 1, 2, \ldots, m$; $\alpha_{2,0} = -\infty$, $\alpha_{2,1} = 0$ and $\alpha_{2,m} = \infty$.

$y_{3i} = k \Leftrightarrow \alpha_{3,k-1} < y^*_{3i} \leqslant \alpha_{3,k}, k = 1, 2, \ldots, m$; $\alpha_{3,0} = -\infty$, $\alpha_{3,1} = 0$ and $\alpha_{3,m} = \infty$, y^*_{2i} and y^*_{3i} are latent variables, m is the number of categories, and $\alpha_{2,1} = \alpha_{3,1} = 0$ for identification purposes.

We assume that the latent variables in Definitions 62 and 63 are linear in the regressors with additive errors, i.e., $y^*_{1i} = x'_{1i}\beta_1 + \varepsilon_{1i}$, $y^*_{2i} = x'_{2i}\beta_2 + \varepsilon_{2i}$ and $y^*_{3i} = x'_{3i}\beta_3 + \varepsilon_{3i}$. In addition, the correlated errors are joint normal, $\varepsilon \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$,

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & 1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & 1 \end{bmatrix} \right) \tag{64}$$

The Likelihood

Setting $\alpha_2 = \begin{bmatrix} \alpha_{2,2} & \alpha_{2,3} & \ldots & \alpha_{2,m-1} \end{bmatrix}'$ and $\alpha_3 = \begin{bmatrix} \alpha_{3,2} & \alpha_{3,3} & \ldots & \alpha_{3,m-1} \end{bmatrix}'$ the cut-point vectors of the states, and $\theta = (\beta_1, \beta_2, \beta_3, \alpha_2, \alpha_3, \Sigma)$, the likelihood of this model is

$$f(y_1, y|\theta : x_{1i}, x_{2i}, x_{3i}) = \prod_{i:y_{1i}=1} P(y_{2i} = y_i, y_{1i} = 1|\theta : x_{1i}, x_{2i}) \prod_{i:y_{1i}=0} P(y_{3i} = y_i, y_{1i} = 0|\theta : x_{1i}, x_{3i}) \tag{65}$$

where the joint probabilities are from bivariate Normal cumulative distribution functions (the assumption in 64).

Given that our econometric framework is Bayesian, it is a good idea to reparametrize the model to improve the mixing properties of the algorithm In particular, we first separated out the largest cut-point from both states, $\alpha_{2,m-1}$ and $\alpha_{3,m-1}$, and define the transformation:

$$\lambda_2 = \frac{1}{[\alpha_{2,m-1}]^2} \quad and \quad \lambda_3 = \frac{1}{[\alpha_{3,m-1}]^2} \tag{66}$$

let

$$\tilde{\beta}_2 = \sqrt{\lambda_2}\beta_2, \quad \tilde{y}_{2i} = \sqrt{\lambda_2}y^*_{2i}, \quad \tilde{\varepsilon}_{2i} = \sqrt{\lambda_2}\varepsilon_{2i}$$

$$\tilde{\beta}_3 = \sqrt{\lambda_3}\beta_3, \quad \tilde{y}_{3i} = \sqrt{\lambda_3}y^*_{3i}, \quad \tilde{\varepsilon}_{3i} = \sqrt{\lambda_3}\varepsilon_{3i} \tag{67}$$

The error variance for the transformed disturbances now takes the following form:

$$\begin{bmatrix} \varepsilon_{1i} \\ \tilde{\varepsilon}_{2i} \\ \tilde{\varepsilon}_{3i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{12} & \lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{13} & \tilde{\sigma}_{23} & \lambda_3 \end{bmatrix} \right) \equiv \mathcal{N}(0, \tilde{\Sigma}) \tag{68}$$

where $\tilde{\sigma}_{12} = \sqrt{\lambda_2}\sigma_{12}$, $\tilde{\sigma}_{13} = \sqrt{\lambda_3}\sigma_{13}$ and $\tilde{\sigma}_{23} = \sqrt{\lambda_2\lambda_3}\sigma_{23}$ (Definitions 66 and 67). The transformed cut-point vectors are defined by $\tilde{\alpha}_2 = \begin{bmatrix} \tilde{\alpha}_{2,2} & \tilde{\alpha}_{2,3} & \dots & \tilde{\alpha}_{2,m-2} \end{bmatrix}'$ and $\tilde{\alpha}_3 = \begin{bmatrix} \tilde{\alpha}_{3,2} & \tilde{\alpha}_{3,3} & \dots & \tilde{\alpha}_{3,m-2} \end{bmatrix}'$.

Prior Distributions

We employ independent priors for the parameters of $\tilde{\theta} = (\beta_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\alpha_2}, \tilde{\alpha_3}, \tilde{\Sigma})$:

$$\pi(\tilde{\theta}) = \pi(\tilde{\beta})\pi(\tilde{\alpha}_2)\pi(\tilde{\alpha}_3)\pi(\tilde{\Sigma}). \tag{69}$$

where $\tilde{\beta} = (\beta_1', \tilde{\beta}_2', \tilde{\beta}_3')'$.

We assume a Gaussian prior for the location parameters, i.e., $\tilde{\beta} \sim \mathcal{N}_K(\tilde{b}_0, \tilde{B}_0)$ where $\tilde{b}_0 = 0_K$ and $\tilde{B}_0 = 1000I_K$, that is, a priori there is no effect of the regressors on the outcome, and the prior information is vague. The priors of $\tilde{\alpha}_2$ and $\tilde{\alpha}_3$ are assumed to be improper, and finally we assume an inverse Wishart prior $\tilde{\Sigma} \sim \mathcal{IW}(\rho, R)$ with the restriction that the element $(1,1)$ of $\tilde{\Sigma}$ is equal to 1, and $\rho = 6$ and $R = I_3$

Posterior Distributions

Setting,

$$\tilde{s} = \begin{bmatrix} y^*_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \end{bmatrix}, X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}, \tilde{s}_i = \begin{bmatrix} y^*_{1i} \\ \tilde{y}_{2i} \\ \tilde{y}_{3i} \end{bmatrix}, X_i = \begin{bmatrix} x'_{1i} & 0 & 0 \\ 0 & x'_{2i} & 0 \\ 0 & 0 & x'_{3i} \end{bmatrix} \tag{70}$$

we assume the same set of controls for the outcome variable, i.e., $x_{2i} = x_{3i}$, with $\tilde{\theta}_{-\tau}$ denoting all parameters other than τ. Using the likelihood (65), the priors (69), and the previous definitions (70), we can get the posterior distributions.

Posterior Distribution of $\widetilde{\beta}$

$$
\begin{aligned}
&P(\tilde{\beta} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\beta}}, y_1, y) \propto N_k(\tilde{b}_0, \tilde{B}_0) \times \prod_i N_3(X_i\tilde{\beta}, \widetilde{\Sigma}) \\
&\propto \exp\left[-\frac{1}{2}(\tilde{\beta}-\tilde{b}_0)'\tilde{B}_0(\tilde{\beta}-\tilde{b}_0)\right] \times \frac{1}{|\widetilde{\Sigma}|^{N/2}} \exp\left[-\frac{1}{2}\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta})\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta}) + (\tilde{\beta}-\tilde{b}_0)'\tilde{B}_0^{-1}(\tilde{\beta}-\tilde{b}_0)\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\sum_i(\tilde{s}_i'\widetilde{\Sigma}^{-1}\tilde{s}_i - 2\tilde{\beta}'X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{\beta}'X_i'\widetilde{\Sigma}^{-1}X_i\tilde{\beta}) + \tilde{\beta}'\tilde{B}_0^{-1}\tilde{\beta} - 2\tilde{\beta}'\tilde{B}_0^{-1}\tilde{b}_0 + \tilde{b}_0'\tilde{B}_0^{-1}\tilde{b}_0\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})\tilde{\beta} - 2\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0) + \sum_i \tilde{s}_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{b}_0'\tilde{B}_0^{-1}\tilde{b}_0\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})\tilde{\beta} - 2\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)\right]
\end{aligned}
$$

Completing the square we get,

$$
\begin{aligned}
&\propto \exp\left[-\frac{1}{2}\left(\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})\tilde{\beta} - 2\tilde{\beta}'(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)\right] \\
&\quad \times \exp\left[-\frac{1}{2}\left((\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)'(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})^{-1}(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})\right.\right. \\
&\quad \left.\left.(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)\right] \\
&\quad \times \exp\left[\frac{1}{2}\left((\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)'(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})^{-1}(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})\right.\right. \qquad (71) \\
&\quad \left.\left.(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)\right] \\
&\propto \exp\left[-\frac{1}{2}\left(\tilde{\beta} - (\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})^{-1}(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)'\right. \\
&\quad \left.\left(\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1}\right)\left(\tilde{\beta} - (\sum_i X_i'\widetilde{\Sigma}^{-1}X_i + \tilde{B}_0^{-1})^{-1}(\sum_i X_i'\widetilde{\Sigma}^{-1}\tilde{s}_i + \tilde{B}_0^{-1}\tilde{b}_0)\right)\right]
\end{aligned}
$$

That is the kernel of a multivariate Normal distribution.

Then the conditional posterior distribution:

$$\tilde{\beta} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\beta}}, X \sim \mathcal{N}(\tilde{b}, \tilde{B}_1)$$

where

$$\tilde{B}_1 = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1} \quad and \quad \tilde{b} = \left[X'(I_n \otimes \tilde{\Sigma}^{-1})X + \tilde{B}_0^{-1}\right]^{-1}\left[X'(I_n \otimes \tilde{\Sigma}^{-1})\tilde{s} + \tilde{B}_0^{-1}\tilde{b}_0\right]$$

Posterior Distribution of $\widetilde{\Sigma}$

$$\begin{aligned}
&P(\widetilde{\Sigma} \mid \tilde{s}, \tilde{\theta}_{-\widetilde{\Sigma}}, y_1, y) \propto I(\widetilde{\Sigma}_{11} = 1) \times IW(\rho, R) \times \prod_i N_3(X_i\tilde{\beta}, \widetilde{\Sigma}) \\
\propto&\frac{1}{|\tilde{\Sigma}|^{N/2}} \exp\left[-\frac{1}{2}\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta})\right] \\
&\times \frac{|R|^{\rho/2}}{|\widetilde{\Sigma}^{\frac{\rho+p+1}{2}} 2^{\frac{p\rho}{2}}\Gamma_p(\frac{m}{2})} \exp\left[-\frac{1}{2}tr(R\widetilde{\Sigma}^{-1})\right] \\
\propto&|\tilde{\Sigma}|^{-\frac{N+\rho+p+1}{2}} \exp\left[-\frac{1}{2}\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta})\right] \times \exp\left[-\frac{1}{2}tr(R\widetilde{\Sigma}^{-1})\right]
\end{aligned}$$

To finish we need to apply some trace properties. First, an scalar is equal to the trace of itself, second, the trace operator is invariant under cyclic permutation. Finally, the sum of traces is equal to the trace of a sum. Thus we get

$$\begin{aligned}
\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta}) =&tr\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta})\right) \\
=&tr\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}\right)
\end{aligned}$$

then,

$$\begin{aligned}
\propto&|\tilde{\Sigma}|^{-\frac{N+\rho+p+1}{2}} \exp\left[-\frac{1}{2}\sum_i(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}(\tilde{s}_i - X_i\tilde{\beta})\right] \times \exp\left[-\frac{1}{2}tr(R\widetilde{\Sigma}^{-1})\right] \\
\propto&|\tilde{\Sigma}|^{-\frac{N+\rho+p+1}{2}} \exp\left[-\frac{1}{2}tr\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1}\right)\right] \times \exp\left[-\frac{1}{2}tr(R\widetilde{\Sigma}^{-1})\right] \\
\propto&|\tilde{\Sigma}|^{-\frac{N+\rho+p+1}{2}} \exp\left[-\frac{1}{2}tr\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})'\widetilde{\Sigma}^{-1} + R\widetilde{\Sigma}^{-1}\right)\right] \\
\propto&|\tilde{\Sigma}|^{-\frac{N+\rho+p+1}{2}} \exp\left[-\frac{1}{2}tr\left(\left(\sum_i(\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})' + R\right)\widetilde{\Sigma}^{-1}\right)\right]
\end{aligned}$$

That is the kernel of an Iverse-Wishat distribution.

Then the conditional posterior distribution:

$$\tilde{\Sigma} \mid \tilde{s}, \tilde{\theta}_{-\tilde{\Sigma}}, X \sim \mathcal{IW}\left(N+\rho, \left[\sum_i (\tilde{s}_i - X_i\tilde{\beta})(\tilde{s}_i - X_i\tilde{\beta})' + R\right]\right) I(\tilde{\Sigma}_{11} = 1).$$

Posterior Distriburion of $\tilde{\alpha_2}$

$$\tilde{\alpha}_2 \mid \tilde{\theta}_{-\tilde{\alpha}_2}, \tilde{y}_2, y_1^*, X \sim \prod_{i:y_1=1} \Phi\left(\frac{\tilde{\alpha}_{2,y_i} - \tilde{\mu}_{2i}}{\sqrt{\tilde{\sigma}_2^2}}\right) - \Phi\left(\frac{\tilde{\alpha}_{2,y_{i-1}} - \tilde{\mu}_{2i}}{\sqrt{\tilde{\sigma}_2^2}}\right)$$

where $\Phi(.)$ is the cumulative normal distribution,

$$\tilde{\mu}_2 = X_2\tilde{\beta}_2 + \left(I_N \otimes \begin{bmatrix}\tilde{\sigma}_{12} & \tilde{\sigma}_{23}\end{bmatrix}\right)\left(I_N \otimes \begin{bmatrix}1 & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{13} & \lambda_3\end{bmatrix}^{-1}\right) H_2$$

$\tilde{\mu}_{i2}$ being the component i of the vector $\tilde{\mu}_2$, $H_2 = [y_{11}^* - x_{11}'\beta_1, \tilde{y}_{31} - x_{31}'\tilde{\beta}_3, y_{12}^* - x_{12}'\beta_1, \tilde{y}_{32} - x_{32}'\tilde{\beta}_3, \cdots, y_{1N}^* - x_{1N}'\beta_1, \tilde{y}_{3N} - x_{3N}'\tilde{\beta}_3]'$ of dimension $2N \times 1$ and

$$\tilde{\sigma}_2^2 = \lambda_2 - \begin{bmatrix}\tilde{\sigma}_{12} & \tilde{\sigma}_{23}\end{bmatrix}\begin{bmatrix}1 & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{13} & \lambda_3\end{bmatrix}^{-1}\begin{bmatrix}\tilde{\sigma}_{12} \\ \tilde{\sigma}_{23}\end{bmatrix}$$

and we ample $\tilde{y}_{2i}$ independently from:

$$\tilde{y}_{2i} \mid \tilde{\theta}, \tilde{y}_{3i}, y_{1i}^*, X \sim \begin{cases}\mathcal{TN}_{(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})}(\tilde{\mu}_{i2}, \tilde{\sigma}_2^2) & \text{if } y_{1i} = 1 \\ \mathcal{N}(\tilde{\mu}_2, \tilde{\sigma}_2^2) & \text{if } y_{1i} = 0\end{cases}, i = 1, 2, ..., n$$

where $\mathcal{TN}_{(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})}(\tilde{\mu}_{i2}, \tilde{\sigma}_2^2)$ is a truncated normal distribution in the interval $(\tilde{\alpha}_{2,y_{i-1}}, \tilde{\alpha}_{2,y_i})$ with mean $\tilde{\mu}_{i2}$ and variance $\tilde{\sigma}_2^2$.

Posterior Distriburion of $\tilde{\alpha_3}$

$$\tilde{\alpha}_3 \mid \tilde{\theta}_{-\tilde{\alpha}_3}, \tilde{y}_3, y_1^*, X \sim \prod_{i:y_1=0} \Phi\left(\frac{\tilde{\alpha}_{3,y_i} - \tilde{\mu}_{i3}}{\sqrt{\tilde{\sigma}_3^2}}\right) - \Phi\left(\frac{\tilde{\alpha}_{3,y_{i-1}} - \tilde{\mu}_{i3}}{\sqrt{\tilde{\sigma}_3^2}}\right)$$

where

$$\tilde{\mu}_3 = X_3\tilde{\beta}_3 + \left(I_N \otimes \begin{bmatrix}\tilde{\sigma}_{13} & \tilde{\sigma}_{23}\end{bmatrix}\right)\left(I_N \otimes \begin{bmatrix}1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \lambda_2\end{bmatrix}^{-1}\right) H_3$$

$\tilde{\mu}_{i3}$ is the component i of the vector $\tilde{\mu}_3$, $H_3 = [y^*_{11} - x'_{11}\beta_1, \tilde{y}_{21} - x'_{21}\tilde{\beta}_2, y^*_{12} - x'_{12}\beta_1, \tilde{y}_{22} - x'_{22}\tilde{\beta}_2, \cdots, y^*_{1,N} - x'_{1N}\beta_1, \tilde{y}_{2N} - x'_{2N}\tilde{\beta}_2]'$ of dimension $2N \times 1$ and

$$\tilde{\sigma}_3^2 = \lambda_2 - \begin{bmatrix} \tilde{\sigma}_{13} & \tilde{\sigma}_{23} \end{bmatrix} \begin{bmatrix} 1 & \tilde{\sigma}_{12} \\ \tilde{\sigma}_{12} & \lambda_2 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{13} \\ \tilde{\sigma}_{23} \end{bmatrix}$$

We can sample $\tilde{y}_{3i}$ independently from:

$$\tilde{y}_{3i} \mid \tilde{\theta}, \tilde{y}_{2i}, y^*_{1i}, X \sim \begin{cases} \mathcal{TN}_{(\tilde{\alpha}_{3,y_{i-1}}, \tilde{\alpha}_{3,y_i})}(\tilde{\mu}_{i3}, \tilde{\sigma}_3^2) & \text{if } y_{1i} = 0 \\ \mathcal{N}(\tilde{\mu}_3, \tilde{\sigma}_3^2) & \text{if } y_{1i} = 1 \end{cases}, i = 1, 2, ..., n$$

And sample y^*_{1i} independently from:

$$y^*_{1i} \mid \tilde{\theta}, \tilde{y}_{2i}, \tilde{y}_{3i}, X \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(\tilde{\mu}_{i1}, \tilde{\sigma}_1^2) & \text{if } y_{1i} = 1 \\ \mathcal{TN}_{(-\infty,0)}(\tilde{\mu}_{i1}, \tilde{\sigma}_1^2) & \text{if } y_{1i} = 0 \end{cases}, i = 1, 2, ..., n$$

where

$$\tilde{\mu}_1 = X_1\beta_1 + \left(I_N \otimes \begin{bmatrix} \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \end{bmatrix}\right) \left(I_N \otimes \begin{bmatrix} \lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{23} & \lambda_3 \end{bmatrix}^{-1}\right) H_1$$

$\tilde{\mu}_{i1}$ is the component i of the vector $\tilde{\mu}_1$, $H_1 = [\tilde{y}_{21} - x'_{21}\tilde{\beta}_2, \tilde{y}_{31} - x'_{31}\tilde{\beta}_3, \tilde{y}_{22} - x_{22'}\tilde{\beta}_2, \tilde{y}_{32} - x'_{22}\tilde{\beta}_3, \cdots, \tilde{y}_{2N} - x'_{2N}\tilde{\beta}_2, \tilde{y}_{3N} - x'_{3N}\tilde{\beta}_3]'$ and

$$\tilde{\sigma}_1^2 = 1 - \begin{bmatrix} \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \end{bmatrix} \begin{bmatrix} \lambda_2 & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{32} & \lambda_3 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\sigma}_{12} \\ \tilde{\sigma}_{13} \end{bmatrix}$$

Finally, we invert the mappings described by the reparametrization to recover the structural coefficients of interest.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

Bolker, B. (2007). *Ecological Models and Data in R*. Princeton University Press.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics*. Cambridge University Press.

Casella, G. and Berger, R. L. (2002). *Statistical inference*. Duxbury Press, second edition.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*. Cambridge University Press.

Greenberg, E. (2012). *Introduction to Bayesian econometrics*. Cambridge University Press.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

Rossi, P. E. (2015). bayesm: Bayesian Inference for Marketing/Micro-Econometrics.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian statistics and marketing*. John Wiley and Sons.

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.