RANDOM EFFECTS AS A WEIGHTED AVERAGE

The model:

\[ y_{ig} = X_{ig} \beta + v_g + \epsilon_{ig}, \quad i = 1, \ldots, n, \quad g = 1, \ldots, M \]  

\[ v_g | X \sim N(0, \tau^2), \quad \epsilon_{ig} \sim N(0, \sigma^2). \]

In addition we assume that conditional on \( X \) (the full \( nm \times k \) matrix of \( X_{ig} \) values) the \( \epsilon \) and \( v \) vectors are jointly normal with diagonal covariance matrix.

As is explained in earlier notes, in this model the GLS estimate of \( \beta \) with known \( \sigma^2 \) and \( \tau^2 \) takes the form of a weighted average of the between and within regression estimates, where the within estimate is the fixed effects estimator (or equivalently the estimate based on the deviations from group means of \( y \) and \( X \)) and the between estimate is the estimate based on group-mean data (i.e. the \( M \) data points generated by taking means across \( i \) for each \( g \)). The formula derived in the earlier notes was

\[ \hat{\beta}_{GLS} = \left( X'X \right)^{-1} X'y^* = \left( \sigma^{-2} \hat{X}' \hat{X} + \delta^2 \hat{X}' \hat{X} \right)^{-1} \left( \sigma^{-2} \hat{X}' \hat{X} \hat{\beta}_w + \delta^2 \hat{X}' \hat{X} \hat{\beta}_b \right). \]

The earlier notes did not give an explicit formula for \( \delta \) as a function of \( \sigma^2 \) and \( \tau^2 \) or for the likelihood function as a function of the between and within residual sums of squares. These notes fill in these gaps.

Since the residual covariance matrix has the form

\[ \Omega = I_M \otimes \hat{\Omega} = I_M \otimes \left( \sigma^2 I_n + \tau^2 1 \right), \]

the GLS estimator can be described as OLS on transformed data, where \( y \) and \( X \) are pre-multiplied by \( W \)

\[ W = I_m \otimes \tilde{W} = I_m \otimes \left( \sigma^{-1} (I - \frac{1}{n} 1) + \frac{\delta^2}{n} 1 \right) \]

\[ \tilde{W}^2 \hat{\Omega}^{-1} = \sigma^{-2} (I - \frac{1}{n} 1) + \frac{\delta^2}{n} 1 \]

\[ \therefore \tilde{W}^2 \hat{\Omega} = (\sigma^{-2} (I - \frac{1}{n} 1) + \frac{\delta^2}{n} 1) (\sigma^2 I + \tau^2 1) = I \]

\[ \therefore - \frac{1}{n} + \frac{\delta^2 \sigma^2}{n} + \tau^2 \delta^2 = 0. \]

From this we can conclude that \( \delta^2 = 1/ \left( \sigma^2 + n \tau^2 \right) \).

The \( \hat{\Omega} \) matrix has \( n - 1 \) eigenvalues of \( \sigma^2 \) (corresponding to eigenvectors that sum to one) and 1 eigenvalue of \( \tau^2 n + \sigma^2 \). The full \( \Omega \) matrix therefore has \( M(n-1) \) eigenvalues of 1 and \( M \) of \( \tau^2 n + \sigma^2 \). The log likelihood function can be written as

\[ - \frac{Mn}{2} \log(2\pi) - \frac{M}{2} \log(\tau^2 n + \sigma^2) - \frac{M(n-1)}{2} \log(\sigma^2) - \frac{\bar{u}' \bar{u}}{2\sigma^2} - \frac{\bar{u}' \bar{u}}{2(\tau^2 + \sigma^2/n)}, \]

where \( \bar{u} \) are the residuals from the within regression and \( \bar{u} \) are the residuals from the between regression.
The conditional posterior distribution on $\beta$ given $\sigma^2$ and $\tau^2$ is $N(\hat{\beta}_{GLS}, (X'\Omega^{-1}X)^{-1})$, where
\begin{equation}
\hat{\beta}_{GLS} = (\sigma^{-2}\tilde{X}'\tilde{X} + (\sigma^2 + \tau^2n)^{-1}X'X)^{-1}(\sigma^{-2}\tilde{X}'\tilde{X}\hat{\beta}_w + (\sigma^2 + \tau^2n)^{-1}X'X\hat{\beta}_b) \tag{10}
\end{equation}
\begin{equation}
X'\Omega^{-1}X = \sigma^{-2}\tilde{X}'\tilde{X} + (\sigma^2 + \tau^2n)^{-1}\bar{X}'\bar{X} \tag{11}
\end{equation}

Conditional on $\beta$,
\begin{equation}
\sigma^{-2} \sim \text{Gamma}(\frac{1}{2}M(n - 1) - 1, \frac{1}{2}\bar{u}'\bar{u}) \tag{12}
\end{equation}
\begin{equation}
(\tau^2 + \sigma^2)^{-1} \sim \text{Gamma}(\frac{1}{2}M - 1, \frac{1}{2}\bar{u}'\bar{u}) , \tag{13}
\end{equation}
with these two random variables independent. Of course this means that $\sigma^2$ and $\tau^2$ themselves are dependent.

These results suggest a particularly simple way to sample from the posterior on $\sigma^2$, $\tau^2$ and $\beta$. Assuming we have some initial estimates — for example by applying OLS to estimate $\beta$ and estimating $\tau^2 + \sigma^2/n$ as $\bar{u}'\bar{u}/M$, $\sigma^2$ as $\bar{u}'\bar{u}/(Mn)$:

1. Draw $\beta$ from its normal conditional posterior above, and use the draw to construct $\bar{u}$ and $\tilde{u}$.
2. Draw $\sigma^{-2}$ and $\tau^2 + \sigma^2/n$ from their conditional posteriors above.
3. Return to 1.

With this scheme, $\bar{X}'\bar{X}$, $\tilde{X}'\tilde{X}$, $\beta_w$, and $\beta_b$ can be computed once, before the iterations start. The MCMC sampled values are constructed by reweighting these objects.

Important note about singularity: There may be variables (like a time trend) that show no variation in group means, as well as variables that are constant within states. Thus either $\bar{X}'\bar{X}$ or $\tilde{X}'\tilde{X}$ or both could be singular, making $\hat{\beta}_w$ and/or $\hat{\beta}_b$ undefined. But the expression (11) can also be written in a form that is insensitive to these singularities:
\begin{equation}
\hat{\beta}_{GLS} = (\sigma^{-2}\tilde{X}'\tilde{X} + (\sigma^2 + \tau^2n)^{-1}\tilde{X}'\tilde{y})^{-1}(\sigma^{-2}\tilde{X}'\tilde{y} + (\sigma^2 + \tau^2n)^{-1}\tilde{X}'\tilde{y}) \tag{14}
\end{equation}

It is this form that should be used for actual computation.

Note also that in forming $\tilde{u}$ and $\bar{u}$, one is using the same $\beta$ value for both, and this $\beta$ is always well defined. The fact that, e.g., a trend variable shows no variation in state means does not create a problem when forming the between residuals $\bar{u}$. The coefficient on trend just becomes a contribution to the constant term.