

Improving the Performance of the Gibbs Sampler: The Use of Different Centerings in the LVR-HM3

In implementing the Gibbs sampler in HM settings, we need to be alert to the possible occurrence of poor mixing, i.e., situations in which successive values in the chains generated for one or more parameters in a given model are highly autocorrelated. When mixing is poor, it can be difficult to assess convergence of the Gibbs sampler. Even if one is reasonably confident that the sampler has converged, it is difficult to know whether all regions of the joint posterior have been adequately traversed in a given set of M iterations.

One strategy to reduce high posterior correlations and improve mixing is to center the covariates in one's model (see Spiegelhalter et al., 2003; Gilks et al., Chpt 6., 1996). In this section we focus on the importance of centering latent variable predictors. Thus in the case of our models, one option would be to center student initial status around its school mean (i.e., $\pi_{0ij} - \beta_{00j}$) and to center school mean initial status around the grand mean (i.e., $\beta_{00j} - \gamma_{000}$). We explored the effects of using various centerings on the performance of the Gibbs sampler in estimating the following model. Note that this model is very similar to Model 2 presented in our article; it contains all of the predictors specified in Model 2, and, in addition, includes a student-level measure of home resources as a predictor in the level-2 equations.

$$Y_{ij} = \pi_{0ij} + \pi_{1ij} (\text{GRADE}_{ii} - 7) + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (1)$$

$$\begin{aligned} \pi_{0ij} = & \beta_{00j} + \beta_{01j} (\text{HOMERES}_{ij} - \overline{\text{HOMERES}_{..}}) + \beta_{02j} (\text{ED_EXPEC}_{ij} - \overline{\text{ED_EXPEC}_{..}}) \\ & + \beta_{03j} (\text{BEHAV_PBLMS}_{ij} - \overline{\text{BEHAV_PBLMS}_{..}}) + r_{0ij} \quad r_{0ij} \sim N(0, \tau_{\pi 0j}) \end{aligned}$$

$$\begin{aligned} \pi_{1ij} = & \beta_{10j} + \text{Bw}_j (\pi_{0ij} - \beta_{00j}) + \beta_{11j} (\text{HOMERES}_{ij} - \overline{\text{HOMERES}_{..}}) + \\ & \beta_{12j} (\text{ED_EXPEC}_{ij} - \overline{\text{ED_EXPEC}_{..}}) + \beta_{13j} (\text{BEHAV_PBLMS}_{ij} - \overline{\text{BEHAV_PBLMS}_{..}}) + r_{1ij} \end{aligned}$$

$$r_{1ij} \sim N(0, \tau_{\pi 1j}) \quad \text{Cov}(r_{0ij}, r_{1ij}) = 0 \quad (2)$$

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + u_{00j} \quad u_{00j} \sim N(0, \tau_{\beta00})$$

$$\beta_{01j} = \gamma_{010}$$

$$\beta_{02j} = \gamma_{020}$$

$$\beta_{03j} = \gamma_{030}$$

$$\beta_{10j} = \gamma_{100} + Bb^*(\beta_{00j} - \gamma_{000}) + \gamma_{101}(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + \gamma_{102}(\overline{TCHEFF}_{.j} - \overline{TCHEFF}_{..}) + u_{10j} \quad u_{10j} \sim N(0, \tau_{\beta10})$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

$$\beta_{13j} = \gamma_{130}$$

$$Bw_j = Bw_0 + Bw_1(\beta_{00j} - \gamma_{000}) + Bw_2(\overline{HOMERES}_{.j} - \overline{HOMERES}_{..}) + Bw_3(\overline{TCHEFF}_{.j} - \overline{TCHEFF}_{..}) + u_{Bwj} \quad u_{Bwj} \sim N(0, \tau_{Bw}) \quad (3)$$

$$Cov(u_{00j}, u_{10j}) = 0, Cov(u_{00j}, u_{Bwj}) = 0, Cov(u_{10j}, u_{Bwj}) = \tau_{\beta10, Bw}$$

The following are the four different centering conditions we explored:

1. No centering of latent variable predictors at levels 2 and 3
2. Level-2 centering : $\pi_{0ij} - \beta_{00j}$
3. Level-3 centering: $\beta_{00j} - \gamma_{000}$
4. Centering at levels 2 and 3: $\pi_{0ij} - \beta_{00j}$ & $\beta_{00j} - \gamma_{000}$.

Note that the performance of the Gibbs sampler for each of the above conditions was monitored by means of computing autocorrelations among the deviates in a chain. Plots of a series of autocorrelations (i.e., autocorrelation functions (ACF) plots) provide an important tool for assessing mixing.

In Table B.1 for each condition, we categorize parameters based on whether their ACF values are smaller than .10 at a series of different lag (k) values. Note that the autocorrelation functions (ACF) in Table B.1 are constructed based on 4,000 deviates for each parameter generated over iterations 2,001 to 6,000 of the Gibbs sampler.

First, for condition 1, (i.e., no centering at levels 2 or 3), ACF values for γ_{000} , γ_{001} , γ_{010} , γ_{020} , γ_{030} , γ_{020} , σ^2 , and $\tau_{\beta_{00}}$ are smaller than .10 at $k = 5$, while the ACF values for Bb, Bw_0, Bw_1, and γ_{100} are very high even at $k = 300$. Note that γ_{100} and Bb are fixed effects in the level-3 equation for β_{10j} (school rates of change) and Bw_0 and Bw_1 are fixed effects in the level-3 equations for Bw $_j$. The ACF values at $k = 300$ for those four parameters are .59, .63, .63, and .62, respectively. Thus for γ_{000} , for example, the autocorrelations are decreasing very quickly and are close to 0 before $k = 5$. In contrast, for γ_{100} , the autocorrelations hardly decrease as k increases.

With centering at level 2, we see that there is still a set of fixed effects in the level-3 equations for β_{10j} and Bw $_j$ for which mixing is extremely poor. With centering at level 3, autocorrelations for all parameters are less than .10 by lag 100. Finally, we see a further improvement in mixing for fixed effects in the level-3 equations for β_{10j} and Bw $_j$ when centering is employed at levels 2 and 3. As can be seen, $\rho < .10$ by $k = 50$. We have found that this general pattern tends to hold in fitting numerous LVR-HM3s to the LSAY data (Choi, 2002).

Table B.1. Comparison of the performance of the Gibbs Sampler employing different centerings in Model 2: Parameters with Autocorrelation Function (AFC) values below .10 at lag = k ($\rho(k) < .10$).

	No Centereing	Level-2 centering	Level-3 centering	Centering at levels 2 and 3
$\rho(k) < .10$ $K = 5$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030}, \sigma^2, \tau_{\beta 00}$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030}, \gamma_{102}, \sigma^2, \tau_{\beta 00}$	$\gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030}, \sigma^2, \tau_{\beta 00}$	$\gamma_{000}, \gamma_{001}, \gamma_{010}, \gamma_{020}, \gamma_{030}, \tau_{\beta 00}$
$\rho(k) < .10$ $k = 10$		$\gamma_{110}, \tau_{\beta 10}$	γ_{000}	$\gamma_{102}, \gamma_{100}, \tau_{\beta 10}$
$\rho(k) < .10$ $k = 50$	$\gamma_{110}, \gamma_{120}, \gamma_{130}, \tau_{Bw}$	Bw_3 , $\gamma_{120}, \gamma_{130}, \tau_{\beta 10}$	$\gamma_{110}, \gamma_{120}, \gamma_{130}, \tau_{Bw}$	Bb, Bw_0, Bw_1, Bw_2, Bw_3 , $\gamma_{101}, \gamma_{110}, \gamma_{120}, \gamma_{130}, \sigma^2, \tau_{Bw}$
$\rho(k) < .10$ $k = 100$	Bw_2, Bw_3 , $\gamma_{101}, \gamma_{102}, \tau_{\beta 10}$		Bb, Bw_0, Bw_1, Bw_2, Bw_3 , $\gamma_{100}, \gamma_{101}, \gamma_{102}, \tau_{\beta 10}$	
$\rho(k) < .10$ $k = 150$				
$\rho(k) < .10$ $k = 200$				
$\rho(k) < .10$ $k = 250$				
$\rho(k) > .50$ $k = 300$	Bb, Bw_0, Bw_1 , γ_{100}	Bb, Bw_0, Bw_1, Bw_2 , $\gamma_{100}, \gamma_{101}$		