

# Sampling Weights in a Bayesian Grade of Membership Model

Marianne (Marnie) Bertolet\*  
Department of Statistics  
Carnegie Mellon University

*Abstract:* The Grade of Membership (GoM) model is a hierarchical mixed-membership model used to characterize underlying latent classes based on categorical data. When using GoM models to analyze survey data, the sampling design needs to be appropriately modeled. Linear mixed-effect models (LME's) easily model the stratification and clustering in sampling designs. This paper introduces a modification of the GoM model to include a polytomous logistic mixed-effects regression prior, designed to take sampling design induced dependencies into account. In addition, there is a debate regarding the use of sampling weights in model based analyses. I developed a new type of weighting, *weighting based on the estimated parameter*, to incorporate the sampling weights in the updated GoM model. Finally, simulation studies demonstrate the effect of the sampling weights under different levels of informative sampling.

*Keywords:* Survey Sampling, GoM model, Inverse Probability Weights, Linear Mixed-Effect Models

## 1 Introduction

The grade of Membership (GoM) model is a hierarchical Bayesian mixed-membership model used to analyze a variety of data, including depression-related psychiatric disorders, (Woodbury and Manton, 1989), the number of likely topics published in the *Proceedings of the National Academy of Sciences* in 1997-2000, and the number of underlying latent class disability profiles in the National Long Term Care Survey (NLTCS), (Airoldi et al., 2005). When analyzing survey data such as NLTCS, the stratification and clustering in the sampling design induce dependencies not easily modeled in the GoM model. Mixed-effects

---

\*This work is based on my thesis, completed at Carnegie Mellon University. Current affiliation is with the Epidemiology Data Center, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, bertoletm@edc.pitt.edu

models, described in Searle et al. (1992), can model the dependencies induced by sampling designs. There is a controversy regarding whether or not to include sampling weights when using mixed-effects models to analyze survey data. Pfeiffermann et al. (1998) and Rabe-Hesketh and Skrondal (2006) have all developed methods to insert sampling weights into linear mixed-effects models. In this paper, I propose a modification of the GoM model to incorporate the sampling design when analyzing survey data. This involves two steps; 1) using a polytomous logistic mixed-effects regression as a prior and 2) developing a modification of the previous methods to inserting sampling weights into a Bayesian model. Finally, simulations demonstrate the effect of the sampling weights under different levels of informative sampling.

Section 2 describes the standard unweighted GoM model and its derivation as seen in Erosheva (2002). Section 3 describes the changing to the polytomous logistic mixed-effects regression prior, first deriving the unweighted model. Weighting of the GoM model is discussed, and weighting based on the estimated parameter is derived. Section 4 describes some rotational indeterminacies in the GoM model, along with two known techniques for solving these indeterminacies. Sections 5 and 6 describe the details and descriptions of the simulation study. Section 7 summarizes the report. Section 8 collects together appendices providing further detail on this work. In particular section 8.3 provides a description and web-links for computer code used to conduct the simulations.

The contributions in this paper involve both the GoM model analysis and incorporation of sampling weights. For the GoM model analysis, the polytomous logistic mixed-effects regression prior allows for model-based incorporation of the sampling design. It also provides a framework for the GoM model to be analyzed with longitudinal data (either with or without weights). With respect to sampling weights, we developed a principled way to incorporate sampling weights into a Bayesian model-based analysis, called weighting based on the estimated parameter. The simulation study provides a contribution regarding the actual performance of the weighting of the GoM model with the new prior. These simu-

lations demonstrate the following: 1) When  $\lambda$  is fixed in the simulations, the mean of the posterior distributions is generally similar to the simulations in which  $\lambda$  is unconstrained with an informative prior. However, the simulations in which  $\lambda$  is unconstrained with an informative prior have larger variance. This is true for the unweighted simulations, and mostly true for the weighted simulations. 2) The differences between the unweighted and weighted estimates of parameters of the polytomous logistic mixed-effects regression parameters behave similarly to the analogous differences seen in the simulation studies of weighted linear mixed-effects models under informative sampling, see Bertolet (2009), with a few exception noted in the simulation descriptions. Finally, 3) the estimates of the  $\lambda$  parameters appear robust to the sampling design and the type of weighting used in the estimation.

## 2 Unweighted Derivation of the GoM Model

Following Erosheva (2002), the GoM model is comprised of extreme profiles and their conditional response probabilities. Let the data consist of  $J$  binary questions for  $I$  individuals. Let  $y_i = (y_{i1}, y_{i2}, \dots, y_{iJ})$  be a vector of 0's and 1's representing the response of individual  $i$  on all  $J$  questions,  $i = 1, \dots, N$ . A vector of GoM scores (latent variables) for each individual,  $g_i = (g_{i1}, g_{i2}, \dots, g_{iC})$ , represents the mixture proportion of individual  $i$  in each of  $C$  unobservable latent classes. These GoM scores are non-negative and sum to 1 for each individual,

$$\sum_{c=1}^C g_{ic} = 1, \quad i = 1, \dots, N. \quad (1)$$

Sole membership in a given class defines the pure response probability,  $\lambda_{cj}$ , for each of the  $J$  items of interest,

$$\lambda_{cj} = P(y_{ij} = 1 | g_{ic} = 1). \quad (2)$$

The following assumptions are made for the GoM model;

**Assumption 1:** The conditional probability of response of individual  $i$  to question  $j$ , given the GoM scores, is  $P(y_{ij} = 1|g_i) = \sum_{c=1}^C g_{ic}\lambda_{cj}$ .

**Assumption 2:** Conditional on the GoM scores, the responses  $y_{ij}$  are independent for different values of  $j$ ,  $(y_{ij_1} \perp y_{ij_2})|g_i$ .

**Assumption 3:** The responses  $y_{ij}$  are independent for different values of  $i$ , or  $y_{i_1j} \perp y_{i_2j}$ .

**Assumption 4:** The GoM scores,  $g_i$ , are realizations of a random vector with a Dirichlet distribution.

The GoM model in Erosheva (2002) allows the responses to the  $J$  items to be polytomous. For simplification, this thesis presents dichotomous response data only.

The GoM model is defined as

$$\begin{aligned} y_{ij}|g_i &\sim \text{Bernoulli}\left(\sum_{c=1}^C g_{ic}\lambda_{cj}\right) \\ g_i &\sim \text{Dirichlet}(\alpha_0\xi) \\ \lambda_{cj} &\sim \text{Beta}(\eta_{1cj}, \eta_{2cj}) \\ \alpha_0 &\sim \text{Gamma}(\tau_1, \tau_2) \\ \xi &\sim \text{Dirichlet}(\zeta), \end{aligned}$$

which contain a number of prior parameters and hyperparameters. Let  $\eta_{1cj}$  and  $\eta_{2cj}$  be parameters for the pure response probabilities,  $\lambda_{cj}$ . The prior parameters for the GoM scores are  $\alpha_0$ , the prior sample size, and  $\xi$ , the prior proportions of the population elements in the underlying latent classes. The  $\tau_1, \tau_2$  and  $\zeta$  hyperparameters are set to be non-informative.

Erosheva (2002) augmented this model with latent variables,  $m_{ijc}$  to assign individual  $i$  to class  $c$  for question  $j$ , to ease the computations in the Bayesian MCMC estimation.

Erosheva's fundamental representation theorem proves the equivalence between the GoM model and the data augmented GoM model below,

$$\begin{aligned}
y_{ij}|m_{ijc}, \lambda &\sim \text{Bernoulli}\left(\prod_{c=1}^C \lambda_{cj}^{m_{ijc}}\right) \\
m_{ijc}|g_i &\sim \text{Multinomial}(1, g_{i1}, \dots, g_{iC}) \\
g_i|\alpha_0, \xi &\sim \text{Dirichlet}(\alpha_0\xi) \\
\lambda_{cj} &\sim \text{Beta}(\eta_{1cj}, \eta_{2cj}) \\
\alpha_0 &\sim \text{Gamma}(\tau_1, \tau_2) \\
\xi &\sim \text{Dirichlet}(\zeta).
\end{aligned}$$

To solve for these parameters using a Bayesian MCMC algorithm, the joint distribution is computed as

$$\begin{aligned}
p(y, m, g, \lambda, \alpha_0, \xi) &\propto p(\lambda, \alpha)p(y, m, g|\lambda, \alpha_0, \xi) \\
&= p(\lambda)p(\alpha_0, \xi) \prod_{i=1}^N p(y_i, m_i, g_i|\lambda_i, \alpha_0, \xi) \\
&= p(\lambda)p(\alpha_0)p(\xi) \prod_{i=1}^N p(y_i|m_i, \lambda)p(m_i|g_i)p(g_i|\alpha_0, \xi).
\end{aligned}$$

This formulation assumes that  $y_i|(m_i, \lambda) \perp (g_i, \alpha)$  and  $g_i|m_i \perp (\lambda, \alpha)$  and  $g_i|\alpha \perp \lambda$ .

Inserting the distributional assumptions into the joint distribution provides,

$$\begin{aligned}
p(y, m, g, \lambda, \alpha) &\propto p(\lambda)p(\alpha) \prod_{i=1}^N [p(m_i|g_i)p(y_i|m_i, \lambda)p(g_i|\alpha)] \\
&= p(\lambda)p(\alpha) \left( \prod_{i=1}^N \left[ \frac{\Gamma(\sum_{c=1}^C \alpha_c)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_C)} g_{i1}^{\alpha_1-1} g_{i2}^{\alpha_2-1} \dots g_{iC}^{\alpha_C-1} \right] \right) \\
&\times \prod_{i=1}^N \left( \prod_{j=1}^J \prod_{c=1}^C (g_{ic} \lambda_{cj}^{y_{ij}} (1 - \lambda_{cj})^{1-y_{ij}})^{m_{ick}} \right)
\end{aligned}$$

The complete conditionals are obtained for  $m_i, \lambda_{cj}$  and  $g_i$ ,

$$\begin{aligned}
m_i|- &\sim \text{Multinomial}(1, p_1, \dots, p_C) \quad p_c \propto (g_{ic} \lambda_{cj}^{y_{ij}} (1 - \lambda_{cj})^{(1-y_{ij})}) \\
\lambda_{cj}|- &\sim \text{Beta}(1 + \sum_{i=1}^N y_{ij} m_{ijc}, 1 + \sum_{i=1}^N (m_{ijc} - m_{ijc} y_{ij})) \\
g_i|- &\sim \text{Dirichlet}(\alpha_1 + \sum_{j=1}^J m_{ij1}, \alpha_2 + \sum_{j=1}^J m_{ij2}, \dots, \alpha_C + \sum_{j=1}^J m_{ijC}).
\end{aligned}$$

For  $\alpha_0$  and  $\xi$ , a Metropolis-Hastings step needs to be used. First consider  $\alpha_0$ ,

$$\begin{aligned}
p(\alpha_0|-) &\propto p(\alpha_0) \prod_{i=1}^N \left[ \frac{\Gamma(\alpha_0)}{\Gamma(\xi_1 \alpha_0) \cdots \Gamma(\xi_C \alpha_0)} \prod_{c=1}^C g_{ic}^{\xi_c \alpha_0} \right] \\
&= \alpha_0^{\tau_1 - 1} e^{-\tau_2 \alpha_0} \prod_{i=1}^N \left[ \frac{\Gamma(\alpha_0)}{\Gamma(\xi_1 \alpha_0) \cdots \Gamma(\xi_C \alpha_0)} \prod_{c=1}^C g_{ic}^{\xi_c \alpha_0} \right] \\
&= \alpha_0^{\tau_1 - 1} \exp\left\{-\alpha_0(\tau_2 - \sum_{c=1}^C \sum_{i=1}^N \xi_k \log g_{ic})\right\} \left[ \frac{\Gamma(\alpha_0)}{\Gamma(\xi_1 \alpha_0) \cdots \Gamma(\xi_C \alpha_0)} \right]
\end{aligned}$$

For the Metropolis-Hastings step, first draw a proposal point,  $\alpha_0^*$  from the jumping distribution, and then calculate the proposal ratio. In this case, we draw a candidate point from a Gamma proposal distribution with parameters  $\alpha = \gamma, \beta = \frac{\gamma}{\alpha_0^{(r)}}$ , where  $\alpha_0^{(r)}$  was the last accepted value for  $\alpha_0$ . The candidate point is accepted as the next element in the sample with probability  $\min\{1, r_{\alpha_0}\}$ . The proposal ratio,  $r_{\alpha_0}$ , is

$$r_{\alpha_0} = \frac{p(\alpha_0^*|-) p(\alpha_0^{(r)}|\alpha_0^*)}{p(\alpha_0^{(r)}|-) p(\alpha_0^*|\alpha_0^{(r)})}.$$

Breaking this into two terms,

$$\begin{aligned}
r_{\alpha_0}(H) &= \frac{p(\alpha_0^{(r)}|\alpha_0^*)}{p(\alpha_0^*|\alpha_0^{(r)})} \\
&= \frac{\Gamma(\gamma, \gamma/\alpha_0^*)(\alpha_0^{(r)})}{\Gamma(\gamma, \gamma/\alpha_0^{(m)})(\alpha_0^*)} \\
&= \left(\frac{\alpha_0^{(r)}}{\alpha_0^*}\right)^\gamma \left(\frac{\alpha_0^{(r)}}{\alpha_0^*}\right)^{\gamma-1} \exp\left\{-\gamma\left(\frac{\alpha_0^{(r)}}{\alpha_0^*} - \frac{\alpha_0^*}{\alpha_0^{(r)}}\right)\right\} \\
r_{\alpha_0}(M) &= \frac{p(\alpha_0^*|-)}{p(\alpha_0^{(r)}|-)} \\
&= \left(\frac{\alpha_0^*}{\alpha_0^{(r)}}\right)^{\tau_1-1} \exp\left\{-(\alpha_0^* - \alpha_0^{(r)})(\tau_2 - \sum_{c=1}^C \sum_{i=1}^N \xi_c \log g_{ic})\right\} \\
&\times \left(\frac{\Gamma(\alpha_0^*) \prod_{c=1}^C \Gamma(\xi_c \alpha_0^{(r)})}{\Gamma(\alpha_0^{(r)}) \prod_{c=1}^C \Gamma(\xi_c \alpha_0^*)}\right)^N.
\end{aligned}$$

Similarly, the Metropolis-Hastings step for  $\xi$  is derived,

$$\begin{aligned}
p(\xi|-) &\propto p(\xi) \prod_{i=1}^N \left[ \frac{\Gamma(\alpha_0)}{\prod_{c=1}^C \Gamma(\xi_c \alpha_0)} \prod_{c=1}^C g_{ic}^{(\alpha_c-1)} \right] \\
&= \left[ \frac{\Gamma(\alpha_0)}{\prod_{c=1}^C \Gamma(\xi_c \alpha_0)} \right]^N \exp\left\{ \sum_{c=1}^C \sum_{i=1}^N (\alpha_0 \xi_c - 1) \log g_{ic} \right\}. \tag{3}
\end{aligned}$$

For the  $\xi_c$ 's, a candidate point is drawn from a Dirichlet proposal distribution centered at the previous sample value,  $\text{Dirichlet}_{(\delta_C \xi_1^{(r)}, \dots, \delta_C \xi_C^{(r)})}(\xi^*)$ . The candidate point is accepted as the next element in the sample with probability  $\min\{1, r_\xi\}$ . The proposal ratio is  $r_\xi$ ,

$$r_\xi = \frac{p(\xi^*|-)p(\xi^{(r)}|\xi^*)}{p(\xi^{(r)}|-)p(\xi^*|\xi^{(r)})}.$$

Breaking this into two terms,

$$\begin{aligned}
r_{\xi}(M) &= \frac{p(\xi^*|-)}{p(\xi^{(r)}|-)} \\
&= \left( \frac{\prod_{c=1}^C \Gamma(\alpha_0 \xi_c^{(r)})}{\prod_{c=1}^C \Gamma(\alpha_0 \xi_c^*)} \right)^N \exp \left\{ \sum_{c=1}^C \sum_{i=1}^N \alpha_0 (\xi_c^* - \xi_c^{(r)}) \log g_{ic} \right\} \\
r_{\xi}(H) &= \frac{p(\xi^{(r)}|\xi^*)}{p(\xi^*|\xi^{(r)})} \\
&= \left( \frac{\prod_{c=1}^C \Gamma(\delta C \xi_c^{(r)})}{\prod_{c=1}^C \Gamma(\delta C \xi_c^*)} \right) \left( \frac{\prod_{c=1}^C \xi_c^{(r)(\delta C \xi_c^* - 1)}}{\prod_{c=1}^C \xi_c^{*(\delta C \xi_c^{(r)} - 1)}} \right).
\end{aligned}$$

A sample from the posterior is obtained using MCMC with the complete conditionals and the Metropolis-Hastings steps.

### 3 Incorporation of the Sampling Design in the GoM Model

#### 3.1 Polytomous Logistic Regression Prior in the GoM Model

Assuming clustering in the sampling design, all individuals in the population are not independent and the Assumption 3 from Section 2 no longer holds. Recall Assumption 1, that  $P(y_{ij} = 1|g_i) = \sum_c g_{ic} \lambda_{cj}$ . This suggests that the dependency between  $y_{i_1j}$  and  $y_{i_2j}$  is a result of the dependencies between the GoM scores,  $g_i$ 's and/or the pure response probabilities,  $\lambda_{cj}$ . Given that the GoM scores represent individual characteristics and the pure response probabilities represent class characteristics, I will represent dependencies between individuals through dependencies in their GoM scores. Linear mixed-effects models can model the dependencies induced by clusters, which are present in many sampling designs. Given that the GoM scores for an individual are positive and sum to 1, I propose using a polytomous logistic random-effects regression to model the effect of the sampling design on the GoM scores. Let  $y_{kij}$  represent the response of subject  $i$  in cluster  $k$  on question



$j$ . Similar changes in subscript are made on other variables. The assumptions from the original GoM score are now:

**Assumption 1:** The conditional probability of response of individual  $i$  in cluster  $k$  to question  $j$ , given the GoM scores, is  $P(y_{kij} = 1|g_{ki}) = \sum_{c=1}^C g_{kic}\lambda_{cj}$ .

**Assumption 2:** Conditional on the GoM scores, the responses  $y_{kij}$  are independent for different values of  $j$ ,  $(y_{kij_1} \perp y_{kij_2})|g_{ki}$ .

**Assumption 3:** The responses of  $y_{kij}$  for all subjects  $i$  in the same cluster  $k$  are dependent.

**Assumption 4:** The GoM scores,  $g_{ki}$  are realizations of a random vector with a polytomous logistic random-effects distribution.

These assumptions allow GoM model analysis on data from a survey.

The updated GoM model is defined as

$$\begin{aligned}
y_{kij}|m_{kijc}, \lambda &\sim \text{Bernoulli}\left(\prod_c \lambda_{cj}^{m_{kijc}}\right) \\
m_{kijc}|\psi_i &\sim \text{Multinomial}(1, g_{ki1}, \dots, g_{kiC}) \\
g_{kic} &= \frac{\exp\{\psi_{kic}\}}{\sum_{c=1}^C \exp\{\psi_{kic}\}} \\
\psi_{ic}|X, Z, U, \beta &\sim N(X_i\beta_c + Z_iU_c, \sigma_\psi^2), c = 1, \dots, C - 1 \\
\psi_{kiC} &= 0, \psi_{kic} = \log\left(\frac{g_{kic}}{g_{kiC}}\right), c = 1, \dots, C - 1 \\
\lambda_{cj} &\sim \text{Beta}(\eta_{1cj}, \eta_{2cj}) \\
\beta_c &\sim \text{Normal}(\mu_\beta, \Sigma_\beta) \\
U_c &\sim \text{Normal}(0, \Omega) \\
\sigma_\psi^2 &\sim \text{Scaled Inv } \chi^2(\nu, s_\psi^2)
\end{aligned} \tag{4}$$

$$\tag{5}$$

This utilizes the LME framework from Searle et al. (1992) to insert the effect of the sampling design on the GoM model. While the subscript on  $y_{kij}$  denotes a clustered only design,

more complex designs change the above model trivially, by changing the structure of the  $X$  and  $Z$  matrices to incorporate the stratification and clustering information.

Similar to the unweighted GoM model, this model is estimated using MCMC. Before considering sampling weights, we consider estimation of this unweighted model. The joint conditional distribution is

$$\begin{aligned} p(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto p(\beta, U, \sigma_\psi^2) p(\lambda) p(y, m, \psi | \lambda, \beta, U, \sigma_\psi^2, X, Z) \\ &= p(\beta) p(U) p(\sigma_\psi^2) p(\lambda) p(y | m, \lambda) p(m | \psi) p(\psi | \beta, U, \sigma_\psi^2, X, Z) \end{aligned} \quad (6)$$

In the last equation, we assume that  $(y | m, \lambda) \perp (\beta, U, \sigma_\psi^2, \psi)$  and  $(m | \psi) \perp (\beta, U, \sigma_\psi^2)$  and that  $(\psi | \beta, U, \sigma_\psi^2) \perp \lambda$ . Continuing,

$$\begin{aligned} p(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto p(\beta) p(U) p(\sigma_\psi^2) p(\lambda) \\ &\times \left[ \prod_{k=1}^K \prod_{i=1}^{N_k} \left( \prod_{j=1}^J \prod_{c=1}^C p(y_{kij} | m_{kijc}, \lambda_{cj}) \right) \right] \left[ \prod_{k=1}^K \prod_{i=1}^{N_k} \prod_{j=1}^J \prod_{c=1}^C p(m_{kijc} | \psi_{kic}) \right] \\ &\times \prod_{c=1}^{C-1} \prod_{k=1}^K \prod_{i=1}^{N_k} p(\psi_{kic} | \beta, U, \sigma_\psi^2) \end{aligned}$$

Recall that  $k$  (of  $K$ ) indexes clusters,  $i$  (of  $N_k$ ) indexes individuals in clusters,  $j$  (of  $J$ ) indexes questions,  $c$  (of  $C$ ) indexes latent classes. Writing  $p(U) \prod_k p(\psi_k | \beta, U, \sigma_\psi^2) = \prod_k (\prod_i \prod_c p(\psi_{kic} | \beta, U, \sigma_\psi^2)) p(U_k)$  will be useful for the insertion of sampling weights in the

next section. Inserting in the distributional forms provides

$$\begin{aligned}
p(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto \exp \left\{ -\frac{1}{2} \sum_{c=1}^C (\beta_c - \mu_\beta)^T \Sigma_\beta^{-1} (\beta_c - \mu_\beta) \right\} \\
&\times (\sigma_\psi^2)^{-(\frac{\nu}{2}+1)} \exp \left\{ -\frac{\nu s_\psi^2}{2\sigma_\psi^2} \right\} \left[ \prod_{c=1}^C \prod_{j=1}^J \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \right] \\
&\times \prod_{k=1}^K \prod_{i=1}^{N_k} \prod_{j=1}^J \prod_{c=1}^C \left[ \frac{\exp(\psi_{kic})}{\sum_{c_1=1}^C \exp(\psi_{kic_1})} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{1-y_{kij}} \right]^{m_{kijc}} \\
&\times \prod_{k=1}^K \left[ \prod_{i=1}^{N_k} \left( \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\} \right) \right] \\
&\times \exp \left\{ -\frac{1}{2} U_{kc}^T \mathcal{O}^{-1} U_{kc} \right\} \tag{7}
\end{aligned}$$

Note that the prior on  $U_{kc}$  uses the cluster version of the covariance matrix,  $\mathcal{O}$ , instead of the entire covariance of  $U$ ,  $\Omega$ . From this, we can get the complete conditionals for Gibbs steps in the MCMC. The complete conditionals for the parameters associated with the

polytomous logistic regression are

$$\begin{aligned}
p(\beta_c|-) &\propto \exp\left\{-\frac{1}{2}(\beta_c - \mu_c)^T \Sigma_\beta^{-1}(\beta_c - \mu_c)\right\} \prod_{k=1}^K \prod_{i=1}^{N_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\} \\
&\sim \text{Normal}(\mu_1, \Sigma_1) \\
\mu_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} X_{ki}^T X_{ki}\right)^{-1} \left(\Sigma_\beta^{-1} \mu_c + \frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} X_{ki}^T (\psi_{kic} - Z_{ki}U_c)\right) \\
\Sigma_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} X_{ki}^T X_{ki}\right)^{-1} \\
p(U_c|-) &\propto \prod_{k=1}^K \prod_{i=1}^{N_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\} \exp\left\{-\frac{1}{2}U_{kc}^T \Omega^{-1}U_{kc}\right\} \\
&\sim \text{Normal}(\mu_2, \Sigma_2) \\
\mu_2 &= \left(\mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} Z_{ki}^T Z_{ki}\right)^{-1} \left(\frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} Z_{ki}^T (\psi_{kic} - X_{ki}\beta_c)\right) \\
\Sigma_2 &= \left(\mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^K \sum_{i=1}^{N_k} Z_{ki}^T Z_{ki}\right)^{-1} \\
p(\sigma_\psi^2|-) &\propto (\sigma_\psi^2)^{-(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu s_\psi^2}{2\sigma_\psi^2}\right\} \prod_{k=1}^K \prod_{i=1}^{N_k} \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_{kc})^2\right\} \\
&\sim \text{Scaled Inv } \chi^2 \left(N(C-1) + \nu, \frac{\left(\sum_{k=1}^K \sum_{i=1}^{N_k} \sum_{c=1}^{C-1} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_{kc})^2\right) + \nu s_\psi^2}{N(C-1) + \nu}\right)
\end{aligned}$$

The complete conditionals associated with the pure response probabilities and data augmented values are

$$\begin{aligned}
p(\lambda_{cj}|-) &\propto \prod_{k=1}^K \prod_{i=1}^{N_k} \left[ \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right]^{m_{kijc}} \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \\
&\sim \text{Beta} \left( \sum_{k=1}^K \sum_{i=1}^{N_k} y_{kij} m_{kijc} + \eta_{1cj}, \sum_{k=1}^K \sum_{i=1}^{N_k} (1 - y_{kij}) m_{kijc} + \eta_{2cj} \right) \\
p(m_{kij}|-) &\propto \prod_{c=1}^C \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right]^{m_{kijc}} \\
&\sim \text{Multinomial} \left( 1, \frac{\exp\{\psi_{ki1}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{1j}^{y_{kij}} (1 - \lambda_{1j})^{1-y_{kij}}, \dots, \right. \\
&\quad \left. \frac{\exp\{\psi_{kiC}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{Cj}^{y_{kij}} (1 - \lambda_{Cj})^{1-y_{kij}} \right)
\end{aligned}$$

Finally, a Metropolis step is needed for  $\psi$ ,

$$\begin{aligned}
p(\psi_{kic}|-) &\propto \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \right]^{m_{kijc}} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}
\end{aligned}$$

Let a candidate point be drawn from a Normally distribution, with the mean at the previous MCMC value and a variance of  $\sigma_\psi^2$ . The candidate point is accepted as the next element

in the sample with probability  $\min\{1, r_{\psi_{kic}}\}$ , where

$$\begin{aligned}
r_{\psi_{kic}} &= \frac{p(\psi_{kic}^* | -)}{p(\psi_{kic}^{(r)} | -)} \\
&= \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}^*\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^*\}} \frac{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^{(r)}\}}{\exp\{\psi_{kic}^{(r)}\}} \right]^{m_{kijc}} \\
&\times \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\} \\
&= \prod_{c=1}^C \left[ \frac{g_{kic}^*}{g_{kic}^{(r)}} \right]^{\sum_{j=1}^J m_{kijc}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\}
\end{aligned}$$

where  $g_{kic} = \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \psi_{kic_1}}$  where  $\psi_{kic} = 0$ . These complete conditionals and Metropolis-Hastings steps can be implemented using MCMC algorithms.

### 3.2 Weighting the Logistic Regression GoM Model

Given that the GoM model has been modified to incorporate the sampling design, we next evaluate if the sampling weights provide any additional information. Simulations in Bertolet (2009) demonstrate that the sampling weights did help compensate for informative sampling but not model misspecification. We next investigate the effect of the sampling weights on informative sampling with the GoM model. The effect of model misspecification and sampling weights on the GoM model will be discussed in future work.

Pfeffermann et al. (1998) and Rabe-Hesketh and Skrondal (2006) used pseudo-maximum likelihood to estimate census likelihood equations with weighted sample likelihood equations. When using PML on the GoM model, similar issues arise as with the LME model, namely, when should weights be inserted, does inserting the weights in different areas affect the results and does scaling the weights reduce bias in the estimates? More specifically, should we

1. Add weights to Equation 7 and have them propagate through to the complete conditionals and Metropolis-Hastings steps?

2. Add weights directly to the complete conditionals?
3. Use an alternate weighting method?
4. Scale of the weights as was done with the LME's?

These issues are explored next.

### 3.2.1 Adding Sampling Weights to Equation 7

To add weights to the GoM model likelihood, consider the methods described in Pfeffermann et al. (1998) and Rabe-Hesketh and Skrondal (2006) for weighting LME models. This PML estimation of the census joint distribution is similar to the method by Rabe-Hesketh and Skrondal (2006). The subscript  $w$  denotes this weighted joint distribution. This provides a weighted joint distribution of

$$\begin{aligned}
p_w(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto p(\beta)p(\sigma_\psi^2)p(\lambda) \\
&\times \prod_{k=1}^K \prod_{i=1}^{N_k} \left( \prod_{j=1}^J \prod_{c=1}^C p(y_{kij} | m_{kijc}, \lambda_{cj}) \right)^{w_{ki}} \\
&\times \prod_{k=1}^K \prod_{i=1}^{N_k} \left[ \prod_{j=1}^J \prod_{c=1}^C p(m_{kijc} | \psi_{kic}) \right]^{w_{ki}} \\
&\times \prod_{c=1}^{C-1} \prod_{k=1}^K \left[ \left( \prod_{i=1}^{N_k} p(\psi_{kic} | \beta, U, \sigma_\psi^2)^{w_{i|k}} \right) p(U_k) \right]^{w_k}
\end{aligned}$$

The derivation of all the complete conditionals and Metropolis Hastings steps are in Section 8.1. An issue with this weighting is that weights propagate to the complete conditionals in unexpected ways.

Consider, the complete conditional for  $m_{kijc}$  is

$$\begin{aligned}
p_w(m_{ki} | -) &\propto \prod_{j=1}^J \prod_{c=1}^C \left( g_{kic} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right)^{m_{kijc} w_{ki}} \\
m_{ki} | - &\sim \text{Multinomial}(1, p_1, \dots, p_C) \quad p_c \propto (g_{kic} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})})^{w_{ki}},
\end{aligned}$$

where  $g_{kic} = \frac{\exp\{\psi_{kic}\}}{\sum_{c_1} \exp\{\psi_{kic_1}\}}$ . The  $m_{kij}$  parameter describes a characteristic of an individual as opposed to being a summary variable for the finite population (such as the  $\lambda_{c_j}$ 's). It is not immediately clear that, for example, the probabilities in the multinomial distribution for  $m_{kij}$  should be raised to the power of the weight so that it can represent more people. If provided the complete conditionals based on the census, the weights do not seem to have a place in the complete conditionals of  $m_{kij}$ . Similar arguments hold for weights in the  $\psi_{kic}$  complete conditionals. This leads to the next option for incorporating the sampling weights.

### 3.2.2 Adding Sampling Weights to the Complete Conditionals

Consider estimating the census complete conditional with weighted sample complete conditionals. The subscript  $wCC$  below denotes the result from weighting the complete conditionals. The complete derivation of these complete conditionals and Metropolis-Hastings steps are in Section 8.2.

Consider the complete conditionals for  $\lambda$  and  $m$  with this weighting scenario,

$$\begin{aligned}
p_{wCC}(\lambda_{c_j}|-) &\propto \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \left[ \lambda_{c_j}^{y_{kij}} (1 - \lambda_{c_j})^{(1-y_{kij})} \right]^{m_{kijc} w_{ki}} \lambda_{c_j}^{\eta_{1c_j}-1} (1 - \lambda_{c_j})^{\eta_{2c_j}-1} \\
&\sim \text{Beta} \left( \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} y_{kij} m_{kijc} w_{ki} + \eta_{1c_j}, \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} (1 - y_{kij}) m_{kijc} w_{ki} + \eta_{2c_j} \right) \\
p_{wCC}(m_{kij}|-) &\propto \prod_{c=1}^C \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{c_j}^{y_{kij}} (1 - \lambda_{c_j})^{(1-y_{kij})} \right]^{m_{kijc}} \\
&\sim \text{Multinomial} \left( 1, \frac{\exp\{\psi_{ki1}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{1j}^{y_{kij}} (1 - \lambda_{1j})^{1-y_{kij}}, \dots, \right. \\
&\quad \left. \frac{\exp\{\psi_{kiC}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{Cj}^{y_{kij}} (1 - \lambda_{Cj})^{1-y_{kij}} \right)
\end{aligned}$$

With this  $wCC$  weighting, some components of the joint distribution are treated differently in different complete conditionals. For example, in  $p_{wCC}(\lambda_{c_j}|-)$ , the  $\left[ \lambda_{c_j}^{y_{kij}} (1 - \lambda_{c_j})^{1-y_{kij}} \right]$  term from the posterior is weighted. However, the same term in  $p_{wCC}(m_{kij}|-)$  is not



weighted. Treating a component from the posterior differently in different complete conditionals appears unprincipled. This leads to the new weighting scheme below.

### 3.2.3 Weighting based on the Estimated Parameter

A more principled way to add sampling weights to the GoM model is to weight the term of the joint distribution based on the parameter upon which that term is used to make an inference. If the term of the joint distribution is making inferences only on individual parameters, then it does not need to be weighted. If the term of the joint distribution is making inferences on any group parameters (or parameters that more than one individual is dependent upon), then it should be weighted. Call this weighting based on the estimated parameter, and subscript the estimates with  $wEP$ .

To understand the reasoning for this, first define two different types of distributions (or conditional distributions); 1) distributions providing information for at least one group parameter and 2) distributions providing information for individual/cluster parameter or priors with no estimable parameters. Consider the unweighted joint distribution as an example

$$\begin{aligned}
p(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto p(\beta)p(U)p(\sigma_\psi^2)p(\lambda) \\
&\times \left[ \prod_{k=1}^K \prod_{i=1}^{N_k} \left( \prod_{j=1}^J \prod_{c=1}^C p(y_{kij} | m_{kijc}, \lambda_{cj}) \right) \right] \\
&\times \left[ \prod_{k=1}^K \prod_{i=1}^{N_k} \prod_{j=1}^J \prod_{c=1}^C p(m_{kijc} | \psi_{kic}) \right] \\
&\times \prod_{c=1}^{C-1} \prod_{k=1}^K \prod_{i=1}^{N_k} p(\psi_{kic} | \beta, U, \sigma_\psi^2) \tag{8}
\end{aligned}$$

The likelihood portion of the joint distribution has three components; 1)  $p(y_{kij} | m_{kijc}, \lambda_{cj})$ , 2)  $p(m_{kijc} | \psi_{kic})$  and 3)  $p(\psi_{kic} | \beta, U, \sigma_\psi^2)$ . Note that  $p(y_{kij} | m_{kijc}, \lambda_{cj})$  uses the data in  $y_{kij}$  to gain more information about both  $m_{kijc}$  and  $\lambda_{cj}$ . Here  $m_{kijc}$  is an individual parameter

which applies only to individual  $ki$ . However,  $\lambda_{cj}$  is a group parameter, affecting more than just the  $ki^{th}$  individual. The  $p(y_{kij}|m_{kijc}, \lambda_{cj})$  terms combine information across many  $y_{kij}$  to estimate the group parameter  $\lambda_{cj}$ . Because of this, the  $p(y_{kij}|m_{kijc}, \lambda_{cj})$  term provides information for a group parameter. Similarly,  $p(\psi_{kic}|\beta, U, \sigma_\psi^2)$  combines information across many  $\psi_{kic}$  to estimate the group parameters  $\beta, U$  and  $\sigma_\psi^2$ . Contrast this to  $p(m_{kijc}|\psi_{kic})$ . The  $\psi_{kic}$  parameter only pertains to the  $ki^{th}$  individual. The  $\psi_{kic} = X_{ki}\beta_c + Z_{ki}U_{kc} + \epsilon_{ki}$ , so it is a function of group parameters. However, as noted just after Equation 6, the assumption is that  $m_{kijc}|\psi_{kic} \perp (\beta_c, U_{kc}, \sigma_\psi^2)$ . Therefore, the distribution  $p(m_{kijc}|\psi_{kic})$  provides information about the individual parameter  $\psi_{kic}$  but not any of the group parameters.

Next consider the prior distributions used to form the joint distribution. I will classify the priors into two different groups, 1) non data-scalable priors that will not be weighted and 2) data-scalable priors that will be weighted. Non data-scalable priors are priors that do not change dimension regardless of the size of the data. The  $p(\beta), p(\sigma_\psi^2)$  and  $p(\lambda)$  do not change dimension if the number of individuals or clusters increase. However,  $p(U)$  is a data-scalable prior, as the dimensions of  $U$  change as the number of clusters changes. The dimension of  $U$  is  $KQ \times 1$ , where  $K$  is the number of clusters. When a sample is taken, then the dimension of the prior is  $K_s Q \times 1$  where  $K_s$  is the number of sampled clusters.

In this  $wEP$  weighting scheme, the distributions providing information for at least one group parameter and the data-scalable priors are weighted. The distributions providing information for individual/cluster parameters and the non data-scalable priors are not weighted.

The  $wEP$  weighted joint distribution becomes

$$\begin{aligned}
p_{wEP}(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto p(\beta)p(\sigma_\psi^2)p(\lambda) \\
&\times \left[ \prod_{k=1}^K \prod_{i=1}^{N_k} \left( \prod_{j=1}^J \prod_{c=1}^C p(y_{kij} | m_{kijc}, \lambda_{cj}) \right)^{w_{ki}} \right] \\
&\times \prod_{k=1}^K \prod_{i=1}^{N_k} \left[ \prod_{j=1}^J \prod_{c=1}^C p(m_{kijc} | \psi_{kic}) \right] \\
&\times \prod_{c=1}^{C-1} \prod_{k=1}^K \left[ \left( \prod_{i=1}^{N_k} p(\psi_{kic} | \beta, U, \sigma_\psi^2)^{w_{i|k}} \right) p(U_k | \Omega) \right]^{w_k}
\end{aligned}$$

Inserting the distributional forms provides

$$\begin{aligned}
p_{wEP}(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto \exp \left\{ -\frac{1}{2} \sum_{c=1}^C (\beta_c - \mu_\beta)^T \Sigma_\beta^{-1} (\beta_c - \mu_\beta) \right\} \\
&\times (\sigma_\psi^2)^{-\left(\frac{\nu}{2}+1\right)} \exp \left\{ -\frac{\nu s_\psi^2}{2\sigma_\psi^2} \right\} \left[ \prod_{c=1}^C \prod_{j=1}^J \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \right] \\
&\times \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \prod_{j=1}^J \prod_{c=1}^C \left[ \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{1-y_{kij}} \right]^{m_{kijc} w_{ki}} \\
&\times \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \prod_{j=1}^J \prod_{c=1}^C \left[ \frac{\exp(\psi_{kic})}{\sum_{c_1=1}^C \exp(\psi_{kic_1})} \right]^{m_{kijc}} \\
&\times \prod_{k=1}^{K_s} \left[ \prod_{i=1}^{n_k} \left( \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\} \right)^{w_{i|k}} \right] \\
&\times \exp \left\{ -\frac{1}{2} U_{kc}^T \mathcal{O}^{-1} U_{kc} \right\}^{w_k}
\end{aligned}$$

This leads to the following complete conditionals for the regression variables

$$\begin{aligned}
p_{wEP}(\beta_c | -) &\propto \exp \left\{ -\frac{1}{2} (\beta_c - \mu_c)^T \Sigma_\beta^{-1} (\beta_c - \mu_c) \right\} \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}^{w_{ki}} \\
&\sim \text{Normal}(\mu_1, \Sigma_1) \\
\mu_1 &= \left( \Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki} \right)^{-1} \left( \Sigma_\beta^{-1} \mu_c + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T (\psi_{kic} - Z_{ki}U_c) \right) \\
\Sigma_1 &= \left( \Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki} \right)^{-1} \\
p_{wEP}(U_c | -) &\propto \prod_{k=1}^K \left[ \prod_{i=1}^{n_k} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_{kc})^2 \right\}^{w_{i|k}} \exp \left\{ -\frac{1}{2} U_{kc}^T \mathcal{O}^{-1} U_{kc} \right\} \right]^{w_k} \\
&\sim \text{Normal}(\mu_2, \Sigma_2) \\
\mu_2 &= \left( \sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki} \right)^{-1} \left( \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T (\psi_{kic} - X_{ki}\beta_c) \right) \\
\Sigma_2 &= \left( \sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki} \right)^{-1} \\
p_{wEP}(\sigma_\psi^2 | -) &\propto (\sigma_\psi^2)^{-(\frac{\nu}{2}+1)} \exp \left\{ -\frac{\nu s_\psi^2}{2\sigma_\psi^2} \right\} \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}^{w_{ki}} \\
&\sim \text{Scaled Inv } \chi^2(\nu_1, s_1^2) \\
\nu_1 &= \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} (C-1) + \nu \\
s_1^2 &= \frac{\left( \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} \sum_{c=1}^{C-1} w_{ki} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right) + \nu s_\psi^2}{\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} (C-1) + \nu}
\end{aligned}$$

The complete conditionals for the augmented data and pure response probabilities are

$$\begin{aligned}
p_{wEP}(\lambda_{cj}|-) &\propto \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \left[ \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right]^{m_{kijc} w_{ki}} \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \\
&\sim \text{Beta} \left( \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} y_{kij} m_{kijc} w_{ki} + \eta_{1cj}, \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} (1 - y_{kij}) m_{kijc} w_{ki} + \eta_{2cj} \right) \\
p_{wEP}(m_{kij}|-) &\propto \prod_{c=1}^C \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \left( \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right)^{w_{ki}} \right]^{m_{kijc}} \\
&\sim \text{Multinomial} \left( 1, \frac{\exp\{\psi_{ki1}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \left[ \lambda_{1j}^{y_{kij}} (1 - \lambda_{1j})^{1-y_{kij}} \right]^{w_{ki}}, \dots, \right. \\
&\quad \left. \frac{\exp\{\psi_{kiC}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \left[ \lambda_{Cj}^{y_{kij}} (1 - \lambda_{Cj})^{1-y_{kij}} \right]^{w_{ki}} \right)
\end{aligned}$$

Finally, the Metropolis-Hastings step for  $\psi$  is

$$\begin{aligned}
p_{wEP}(\psi_{kic}|-) &\propto \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \right]^{m_{kijc}} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}
\end{aligned}$$

Let a candidate point be drawn from a Normal distribution, with the mean at the previous MCMC value and a variance of  $\sigma_\psi^2$ . The candidate point is accepted as the next element in the sample with probability  $\min\{1, r_{\psi_{kic}}\}$ , where

$$\begin{aligned}
r_{(wEP)\psi_{kic}} &= \frac{p(\psi_{kic}^*|-)}{p(\psi_{kic}^{(r)}|-)} \\
&= \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}^*\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^*\}} \frac{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^{(r)}\}}{\exp\{\psi_{kic}^{(r)}\}} \right]^{m_{kijc}} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\} \\
&= \prod_{c=1}^C \left[ \frac{g_{kic}^*}{g_{kic}^{(r)}} \right]^{\sum_{j=1}^J m_{kijc}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\}
\end{aligned}$$

One concern regarding the  $wEP$  weighting involve the weights in the  $p_{wEP}(m_{kij}|-)$  distribution. When the weights are large, then  $\left(\lambda_{cj}^{y_{kij}}(1-\lambda_{cj})^{(1-y_{kij})}\right)^{w_{ki}}$  may become very small, basically zero to machine precision. This will be addressed by scaling the weights.

### 3.2.4 Scaling of the Weights

In Bertolet (2009), the scaling of the weights played a role in the estimation of parameters, especially the variance components. The scaling in LME models was introduced to reduce the bias in the variance components. The scaled 1 weightings from adjust the conditional weights,  $w_{i|k}$ , so  $\sum_i w_{i|k}^{s1}$  equals the effective sample size for cluster  $k$ , as defined in Potthoff et al. (1992). The scaled 2 weights adjust the conditional weights so that  $\sum_i w_{i|k}^{s2}$  equals the cluster sample size for cluster  $k$ ,  $n_k$ . The scaling of the weights will also be used in the simulations in this report.

The posterior variances of the parameters are affected by the scaling of the weights. By weighting the data, the sample size affectively becomes  $\sum_k \sum_i w_{ki}$ . When the weights are unscaled, this is an estimate of the size of the *population*, which is larger than the sample size and will create smaller posterior variances. For the simulations in this report, the weights are scaled analogous to the scaled 2 weights so that  $\sum_{i=1}^{n_k} w_{i|k}^s = n_k$ , where  $w_{i|k}^s$  represents the scaled conditional weight from cluster  $k$  and  $n_k$ . Unlike the LME models, it is not clear that the scaling of the cluster weights,  $w_k$ , will have no affect on the estimates. The cluster weights are scaled so that  $\sum_{k=1}^{K_s} w_k^s = K_s$  where  $w_k^s$  represents the scaled cluster weight and  $K_s$  represents the number of sampled clusters. The effect of the scalings of the weights is an area for further investigation.

## 4 Indeterminacies in the GoM model

### 4.1 GoM model, Factor Analysis and Rotations

The GoM model and factor analysis models both contain latent class structures designed to find factors to explain interrelationships among observable variables. Woodbury and Manton (1989), Marini et al. (1996) and Erosheva (2002) compare and contrast the latent structures of the GoM models and the factor analytic models. Unfortunately, these models both have rotational indeterminacies.

Rotational indeterminacies in factor analysis are well known and researched; see any statistical multivariate analysis text such as Johnson and Wichern (1992). Rotational indeterminacies in the GoM model have not been previously documented. To see where these rotations are inserted in the GoM model, consider again Assumption 1 from Section 2 and Assumption 1 from Section 3.1,

$$P_{kij} = P(y_{kij} = 1 | g_{ki}, \lambda_{cj}) = \sum_{c=1}^C g_{kic} \lambda_{cj} = g_{ki}^T \lambda_j$$

where  $g_{ki} = (g_{ki1}, g_{ki2}, \dots, g_{kiC})^T$  are the GoM scores for individual  $ki$  and the pure response probabilities for item  $j$  are  $\lambda_j = (\lambda_{1j}, \lambda_{2j}, \dots, \lambda_{Cj})^T$ . Collecting all the  $P_{kij}$  into a matrix, see that

$$P = G\Lambda$$

where  $G = (g_1^T, \dots, g_N^T)^T$  and  $\Lambda = (\lambda_1, \dots, \lambda_J)$ . If  $R$  is an invertible matrix subject to suitable restrictions discussed below, define,

$$\begin{aligned} G^* &= GR \\ \Lambda^* &= R^{-1}\Lambda \end{aligned}$$

Now  $G^*$  and  $R^*$  are a new rotation for  $G$  and  $R$ . The definition of the probabilities of response remains the same,

$$P = G^* \Lambda^*$$

Restrictions on the matrix  $R$  come from Equations 1 and 2. Namely,

$$GR1_{C \times 1} = 1$$

$$GR \geq 0$$

$$R^{-1}\Lambda \geq 0$$

$$R^{-1}\Lambda \leq 1$$

where all inequalities are element-wise. The set of matrices  $R$  satisfying these conditions usually has positive Lebesgue measure in the space of invertible matrices  $R$ , hence finding these rotations in an MCMC algorithm is possible. Described next are two ways to work with the rotational indeterminacies in the GoM model, using informative priors and fixing  $\lambda$  parameters.

## 4.2 Informative Priors

In frequentist analysis, a number of specific rotations are defined for factor analysis. The purpose of these rotations is to find factor loadings that are easily interpretable. There are a variety of standard rotations that are used, such as orthogonal rotations (including varimax, quartimax and equimax) and oblique rotations (including promax).

In Bayesian factor analysis, using informative (or subjective) priors uniquely determines the rotation, see Kaufman and Press (1973) and Rowe (2001). The formulation of the GoM model in Sections 3.1, and 3.2 allows for informative priors, especially the prior for  $\lambda_{cj}$  and the specification  $\eta_{1cj}$  and  $\eta_{2cj}$ . Whenever informative priors are used in the remainder of



this report, it is clearly stated.

### 4.3 Fix $\lambda$ Parameters

GoM models and item response theory (IRT) models contain similar latent class structures, Erosheva (2005). As an example, the National Assessment of Educational Progress (NAEP) uses IRT models to estimate proficiency scores (analogous to the GoM scores or  $g_i$ 's) of students on different skills, see the special issue of the Journal of Educational Measurement (1992). In estimating the proficiency scores, the NAEP model first estimates item parameters (analogous to the pure response probabilities,  $\lambda$ 's in the GoM model) ignoring the survey design completely, then assumes the item parameters are fixed and produces random draws of the proficiency scores accounting for the survey design, see von Davier et al. (2007). In fixing the  $\lambda$ 's when estimating the  $g$ 's, the NAEP estimation avoids the rotational indeterminacy and provides a precedent for this approach. An explanation of the estimation of the proficiency scores and item parameters in NAEP based upon Mislevy and Sheehan (1989b) follows.

#### 4.3.1 Informative Stratified Sampling and GoM Models

Mislevy and Sheehan (1989a,b) argue that differential probabilities of sampling in a stratified sampling model do not affect estimation of item parameters in an IRT model, but may affect estimation of proficiency scores. Their argument is shown below in the context of the GoM model using the Dirichlet prior from Section 2. Similar results hold for the GoM model with logistic regression prior from Section 3.1.

Suppose that the GoM scores,  $g_i$  and the augmented data inclusion variables,  $m_{ijc}$  are observed. Consider the likelihood portion of Equation 3,

$$p(y, m, g | \lambda, \alpha_0, \xi) = \prod_{i=1}^N p(y_i, m_i, g_i | \lambda, \alpha_0, \xi).$$

Suppose the census data are stratified, where the distribution of the GoM scores differs in each stratum. Let  $h_i$  represent the stratum indicator for element  $i$ . Let  $f(g_i|\alpha_{h_i}, \xi_{h_i}, h_i)$  be the distribution of the GoM scores in stratum  $h_i$ . Let  $\pi_s$  be the proportion of the *population* in stratum  $s$  and assume that

$$f(g|\alpha, \xi) = \sum_{s=1}^H \pi_s f(g|\alpha_s, \xi_s, h_i = s).$$

The likelihood becomes

$$\begin{aligned} p(y, m, g, H|\lambda, \alpha_0, \xi) &= \prod_{i=1}^N f(H = h_i|\lambda, \pi, \alpha, \xi) f(y_i, m_i, g_i|h_i, \lambda, \pi, \alpha, \xi) \\ &= \prod_{i=1}^N \Pr(H = h_i|\pi) f(y_i|m_i, \lambda) p(m_i|g_i) f(g_i|h_i, \xi_{h_i}, \alpha_{h_i}) \\ &= \prod_{s=1}^H \pi_s^{N_s} \times \prod_{s=1}^H f(y_s|m_s, \lambda) p(m_s|g_s) f(g_h|h_s, \xi_{h_s}, \alpha_{h_s}) \end{aligned}$$

where the  $y_s$  represent all responses from people in stratum  $s$ . There are corresponding definitions for  $m_s$  and  $g_s$ . Note that the  $\pi_s$  is distinct from the rest of the likelihood and consistency for these parameters is derived using standard results on the multinomial distribution, as the  $N_s$  grow. The second term is a product of  $H$  likelihoods with a common  $\lambda$  parameter. Bradley and Gart (1962) show conditions for consistency when the likelihood is made up of separate populations that have distinct population parameters (such as  $g_s, m_s, \xi_{h_s}$  and  $\alpha_{h_s}$ .) and a few common parameters, such as  $\lambda$ .

Suppose a sample is taken and that the proportion of sampled elements within a stratum does not equal the proportion of population elements in the stratum. Now define

$$f^*(g|\alpha, \xi) = \sum_{s=1}^H \pi_s^* f(g|\alpha_s, \xi_s, h_i = s),$$

where  $\pi_s^*$  is the *sample* proportion of the elements in stratum  $s$ . Then the likelihood

becomes

$$\begin{aligned}
p(y, m, g, H|\lambda, \alpha_0, \xi) &= f(H|\lambda, \pi, \alpha, \xi)f(y, m, g|H, \lambda, \pi, \alpha, \xi) \\
&= \prod_{i=1}^n \Pr(H = h_i|\pi^*)f(y_i|m_i, \lambda)p(m_i|g_i)f(g_i|h_i, \xi_{h_i}, \alpha_{h_i}) \\
&= \prod_{s=1}^H \pi_s^{*N_s} \times \prod_{s=1}^H f(y_s|m_s, \lambda)p(m_s|g_s)f(g_s|h_s, \xi_{h_s}, \alpha_{h_s}).
\end{aligned}$$

Similar to the case where the census was taken, consistent estimates of  $\lambda, \alpha_h$ 's,  $\xi_h$ 's can be obtained, but we can not reconstruct  $f(g|\alpha, \xi) = \sum_s \pi_s f(g|\alpha_s, \xi_s, h_i = s)$  because of our inability to estimate  $\pi$ , the population proportions when the sampling design uses biased  $\pi^*$ 's.

This argument is used in NAEP to show that the item response parameters (or  $\lambda$ 's in the GoM model) are not affected by the sampling design, whereas the achievement scores (or  $g$ 's in the GoM model) are affected by the sampling design. Next is a brief description of the implementation of the Mislevy and Sheehan (1989b) results in NAEP.

#### 4.3.2 National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) is a national assessment program that regularly tests students in grades 4, 8 and 12 on a variety of academic subjects. NAEP provides an important operational example of the methodology advocated by Mislevy and Sheehan (1989a,b). The data from NAEP are analyzed and published in the National's Report Card (see <http://nces.ed.gov/nationsreportcard/>) for comparative analysis across years. There are many complexities in NAEP's design that are derived from the limited time to administer tests (often about one hour) while at the same time producing reliable and valid assessments. The goal of NAEP is to produce reliable estimates of proficiency for specific population subgroups in various academic subjects.

Group proficiencies are estimated using three stages as described in von Davier et al. (2007). In the *Scaling* stage, an IRT model is fit to the data to estimate the item response

parameters (equivalent to the  $\lambda$ 's in the GoM model). These IRT models do not account for the sampling design. After estimation of the item response parameters, they are considered fixed in the remaining analysis. The justification for this is from Mislevy and Sheehan (1989b) as stated in Thomas (2000). In the *Conditioning* stage, marginal maximum likelihood is used to estimate the mean and variance of the proficiency for students in the population given the students individual covariate values (i.e. the  $g$ 's in the GoM model). Using that mean and variance, random draws (also called multiple imputations or plausible values) are obtained from the examinees posterior latent variable and are used to create subgroup estimates. In the *Variance Estimation* stage, multiple imputation and jackknife approaches are used to estimate subgroup variances.

The simulation study in this report provides results of weighting on the GoM model when there are informative priors on the  $\lambda$  parameters, and when the  $\lambda$ 's are fixed.

## 5 GoM Simulation Study Set-Up

The role of the sampling weights in the GoM model is analyzed using a simulation study. These simulations are designed for a number of comparisons; 1) the difference between the polytomous logistic mixed-effects prior parameters (GoM estimates) when  $\lambda$  is fixed versus when  $\lambda$  is unconstrained with an informative prior, 2) the difference between unweighted and  $wEP$  weighted estimates, both when  $\lambda$  is fixed and when  $\lambda$  is unconstrained with an informative prior, and 3) the difference between unweighted and  $wEP$  weighted estimates of  $\lambda$ .

The simulated sampling design is the same in all the simulations; two top-level strata, and within each stratum there are clusters. Elements from the clusters are then sampled. There are three different levels of informativeness,

**Non-Informative (Non):** Clusters and elements are sampled according to the size of an independently generated random variable.

**Informative Clusters (Clust):** Clusters with larger random effects are over-sampled.

Elements are sampled according to the size of an independently generated random variable.

**Informative Individuals (Indiv):** Clusters are sampled according to the size of an

independently generated random variable. Elements with larger random errors,  $\epsilon_{ik}$  are over-sampled.

### 5.1 True Values in the Simulated Model

The simulated model contains  $C = 2$  underlying classes and  $J = 5$  questions. The *population* has 2 strata, with 20 clusters per stratum ( $Q = 40$  population clusters) and 250 elements per cluster for a population size of  $N = 10,000$ . The polytomous regression prior has two  $X$  covariates which are indicators of stratum inclusion and no intercept. The population  $Z$  matrix is of dimension  $(10000 \times 80)$  or  $(N \times KQ)$ . The regression function for a given element is

$$\begin{aligned}\psi_{hki1} &= -1I_{h==1} + 0.5I_{h==2} + U_{1k}I_{h==1} + U_{2k}I_{h==2} + \epsilon_{hki1} \\ U_{01k} &\sim N(0, 0.04), \quad U_{02k} \sim N(0, 0.64), \quad \epsilon_{hki1} \sim N(0, 0.25)\end{aligned}$$

where  $h = 1$  or  $2$  denotes stratum inclusion and  $k$  represents the cluster. This LME is similar to the generating model in Simulation Set 11 in Bertolet (2009) modified for only one level of stratification. This LME contains two random slopes, each on a stratum inclusion indicator variable. There is no data to estimate the values of  $U_{2k}$  when the element is in stratum 1. As such, the posterior of  $U_{2k}$  for values of  $k$  in stratum 1 matches the prior. Those values are not reported in the simulation. The reverse holds true for  $U_{1k}$  and stratum 2.

In this model,  $C = 2$  so the  $\psi_{hki1}$  is only defined for  $c = 1$ , as the baseline class for the polytomous logistic regression is  $c = 2$ . In other words, this is a logistic re-

	class 1	class 2
Question 1	0.765	0.050
Question 2	0.723	0.407
Question 3	0.447	0.410
Question 4	0.642	0.483
Question 5	0.950	0.250

Table 1: True Value of Simulated  $\lambda$

gression as there are only two classes. The mean GoM score for someone in stratum 1 is  $E(g_{1ki}) = (\frac{\exp\{-1\}}{1+\exp\{-1\}}, \frac{1}{1+\exp\{-1\}}) = (0.27, 0.73)$ , and for stratum 2 is  $E(g_{2ki}) = (\frac{\exp\{0.5\}}{1+\exp\{0.5\}}, \frac{1}{1+\exp\{0.5\}}) = (0.62, 0.38)$ .

The true value of  $\lambda_{cj} = P(y_{hki} = 1 | g_{ki} = c)$  for  $C = 2$  classes and  $J = 5$  questions is in Table 1. These correspond roughly to a "sick" class and a "healthy" class. Prior parameters are set at  $\mu_\beta = 0I_{2 \times 1}$ ,  $\Sigma_\beta = 10I_{2 \times 2}$ . The value of  $\Omega$  is

$$\Omega = \begin{bmatrix} 0.01 & 0.001 \\ 0.001 & 0.01 \end{bmatrix} \otimes I_{K_s \times K_s}$$

Throughout the simulations, the estimates of  $\sigma_\psi^2$  tended to drift. To control this, an informative prior was used, with prior degrees of freedom  $\nu = 200$ , and prior mean equalling the true value of  $s_\psi^2 = 0.25$ . Changing this back to a non-informative prior is discussed in the future work. For the simulations below that use informative priors on  $\lambda$ , the values of  $\eta_{1cj}, \eta_{2cj}$  are listed in the descriptions below.

## 5.2 MCMC Notes

Details of the twelve MCMC simulations presented in the graphs below are in Tables 2 and 3 and discussed next.

For the implementation of the MCMC, there is a Metropolis-Hastings step for the  $\psi_{hki}$ . The acceptance ratio is computed for each  $\psi_{hki}$  for the 1000 elements in the sample and the average acceptance ratio over the sampled elements is in Table 2 and 3. The target

	UNnon FixL	UNclust FixL	Unindiv FixL	wEPnon FixL	wEPclust FixL	wEPindiv FixL
Acceptance Ratio of $\psi$ 's	45.5%	45.5%	45.5%	40.1%	42.3%	44.2%
Number of Iterations	200K	200K	200K	400K	300K	300K
Number of Burn-In Iterations	10K	10K	10K	10K	10K	10K
Amount of Thinning	20	20	20	20	20	20

Table 2: Notes for the MCMC simulation when  $\lambda$  is Fixed

	UNnon Uncon- strL	UNclust Uncon- strL	Unindiv Uncon- strL	wEPnon Uncon- strL	wEPclust Uncon- strL	wEPindiv Uncon- strL
Acceptance Ratio of $\psi$ 's	45.4%	45.5%	45.5%	40.8%	43.1%	44.4%
Value of $\eta_{1c_j} + \eta_{2c_j}$	100	100	100	200	300	150
Number of Iterations	300K	200K	200K	500K	500K	300K
Number of Burn-In Iterations	10K	10K	10K	10K	30K	10K
Amount of Thinning	20	20	20	20	20	20

Table 3: Notes for the Informative Prior (Unconstrained)  $\lambda$  MCMC simulation

acceptance ratio is about 40%. The jumping distribution for the  $\psi$  is normal, centered at the previous  $\psi$  value with variance 1.

It is not possible to state exactly when a sample from the MCMC algorithm represents a random sample from the posterior distribution. Assessing convergence of an MCMC chain is delicate and I used a number of tools. The MCMC chains have a burn-in period to remove effects of initial values and are thinned (to remove iteration to iteration dependencies and save computing resources). These values were chosen by running each simulations for an initial 5000 iterations (no burn-in and no thinning) and using the Raftery & Lewis convergence diagnostics in the R package `boa` (Raftery and Lewis, 1992). The Raftery and Lewis convergence diagnostics provides guidelines on MCMC burn-in, thinning and number of runs to achieve an estimate of the 0.025 and 0.975 percentiles with an accuracy of  $\pm 0.005$  with a probability of 95%. Convergence of the MCMC runs after the burn-in and thinning was monitored using the Heidelberger & Welch convergence diagnostics, also in `boa` (Heidelberger and Welch, 1983). The Raftery & Lewis and Heidelberger & Welch diagnostics are summarized in Cowles and Carlin (1996). The Raftery & Lewis

total number of iterations, burn-in and thinning were altered based on the results of the Heidelberger & Welch diagnostics, if necessary. In addition, I visually examine the plots for any evidence of nonstationarity. Finally, I also examine the trace of the log likelihood using the same set of methods. For comparison, Erosheva (2002) analyzed a GoM model with 2 underlying classes and the Dirichlet prior, using 100,000 MCMC iterations with 10,000 iterations of burn-in and thinned by taking every 10<sup>th</sup> iteration for the sample.

Most of the parameters passed convergence tests, however each of the simulations had non-convergence in some of the the random effects (i.e. the  $U_{01k}$  and  $U_{02k}$ ). Most of the failed tests were the Heidelberger & Welch halfwidth tests, though two of the  $U$ 's failed both the stationary test and the halfwidth test. However in mixed-effects regressions, the specific values of the  $U_{0k}$ 's are usually not of interest. It is the variance of the random effect that is of interest, and that value is reported in the simulations below. The variance of the random effects was computed for each MCMC iteration. The chain of variances of the  $U$ 's did converge for all of the simulations.

### 5.3 Presentation Format

The format of the presentation of the results is similar to that of Bertolet (2009). The differences from that format are described here.

For each parameter there are two vertical lines, one grey and one light blue. The grey line is the simulated parameter value. Because there is known shrinkage towards the prior mean in Bayesian analyses, the light blue line indicates the mean of the unweighted non-informative mean. Comparisons of the effects of informative sampling and weights are compared to the unweighted non-informative estimates. Each line is labeled above. The labels are in Table 4.



Label	Meaning
UN	Unweighted analyses
wEP	Weighted analysis using $wEP$ weights
non	Non-informative sampling
clust	Clusters with large random effects, $U_{01k}$ or $U_{02k}$ , are over-sampled
indiv	Individuals with large random error, $\epsilon_{hki}$ , are over-sampled
UnconstrL	Unconstrained $\lambda$ 's
FixedL	Fixed $\lambda$ 's

Table 4: Labels on GoM Simulations

## 6 GoM Simulation Results

Five simulation plots are shown below to evaluate the comparisons noted above; 1) the difference between unweighted and  $wEP$  weighted polytomous logistic mixed-effects regression prior parameters when  $\lambda$  is fixed versus when  $\lambda$  is unconstrained with an informative prior, 2) the difference between unweighted and  $wEP$  weighted estimates, both when  $\lambda$  is fixed and when  $\lambda$  is unconstrained with an informative prior, and 3) the difference between unweighted and  $wEP$  weighted estimates of  $\lambda$ .

### 6.1 Unweighted Results of Sampling Design Parameters (GoM Scores)

The results from this simulation are in Figure 1. First, examine the difference in the estimates where  $\lambda$  is fixed versus where  $\lambda$  is unconstrained with an informative prior. The unweighted estimates under the non-informative sampling scheme when  $\lambda$  is fixed versus when  $\lambda$  is unconstrained are very similar. For the unweighted estimates under informative cluster sampling scheme, the unconstrained  $\lambda$  estimates have larger posterior spread for the  $\beta$ 's. The posterior spread on the variance components are similar for the fixed versus the unconstrained  $\lambda$  estimates. For the unweighted estimates under the informative individual sampling scheme, the unconstrained  $\lambda$  estimates also have larger posterior spread for the  $\beta$ 's. The posterior spread on the variance components are similar for the fixed versus the unconstrained  $\lambda$  estimates.

Next, examine the difference in the fixed-effects under the different sampling schemes.

As mentioned earlier, the  $\beta$  estimates exhibit shrinkage towards their prior mean of zero. When the sampling design over-over-samples clusters with large random effects ( $U_{01k}$  or  $U_{02k}$ ) then the estimates of the  $\beta$ 's increase, as expected. Similarly, when the sampling design over-samples individuals with large random errors ( $\epsilon_{hki}$ ) the estimates of  $\beta$  increase.

Finally, examine the difference in the variance components under the different sampling schemes. The estimates of  $\sigma_\psi^2$  are consistent across sampling schemes. I expect that the random error variance would decrease when the individuals are informative sampled based on  $\epsilon_{hki}$ . I believe that the underestimation of  $\sigma_\psi^2$  is not seen in these simulations because of the informative prior placed on the parameter, as discussed in Section 5.1. The loosening of this informative prior is discussed in the future work. The estimates of  $\text{Var}(U_k)$  for Stratum 1 and Stratum 2 are smaller for the informative cluster sampling than the non-informative sampling, as expected. Under the informative individual sampling scheme, the estimates of  $\text{Var}(U_k)$  for Stratum 1 and Stratum 2 are larger than the informative cluster sampling scheme, as expected. There appears to be some positive bias for the estimates of  $\text{Var}(U_k)$  Stratum 1 and some negative bias for the estimates of  $\text{Var}(U_k)$  Stratum 2 when compared to the non-informative sampling scheme. This bias is not well understood, however it may be due to the lack of bias on the  $\sigma_\psi^2$  estimate.

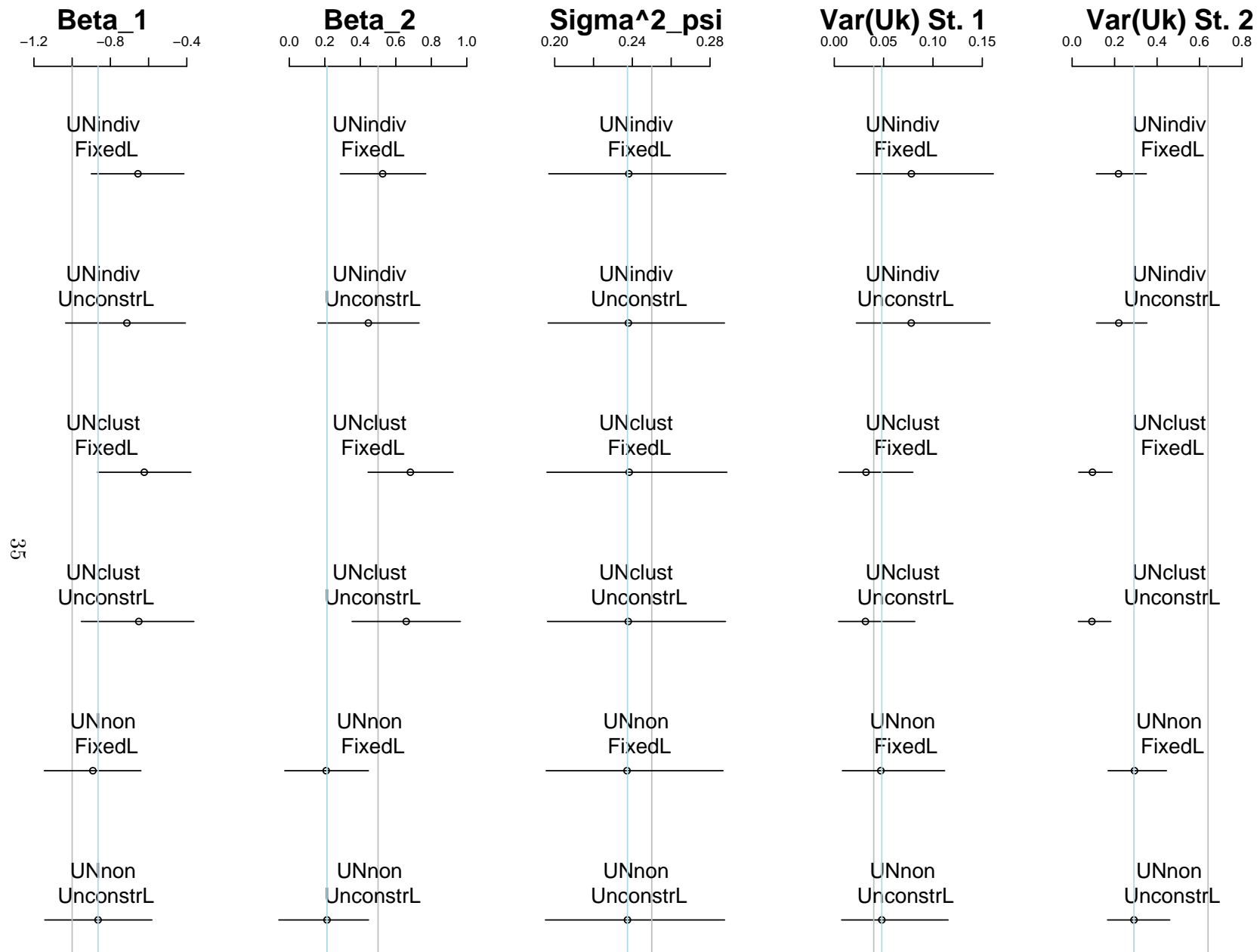


Figure 1: Unweighted GoM Results of Sampling Design Parameters (GoM Scores)

## 6.2 Weighted Results of Sampling Design Parameters (GoM Scores)

The weighted estimates of the sampling design parameters (GoM Scores) are in Figure 2. This figure is included for completeness, though it is difficult to interpret. There are two confounding effects in this comparison. The first difference is the difference between the estimates of the parameters of the polytomous logistic mixed-effects regression prior when  $\lambda$  is fixed versus when  $\lambda$  is unconstrained with an informative prior. The second difference is that the  $wEP$  weighting is different when  $\lambda$  is fixed and when  $\lambda$  is unconstrained. To see this consider the unweighted joint distribution from Equation 8. As discussed in the *Weighting based on the Estimated Parameter* subsection of Section 3.2.3, the  $p(y_{kij}|m_{kijc}, \lambda_{cj})$  entry of the likelihood gets weights because the estimated parameter  $\lambda_{cj}$  is a group parameter. However, when  $\lambda_{cj}$  is fixed, then the term  $p(y_{kij}|m_{kijc})$  no longer contains an estimated group parameter and does not get weights.

Given the confounding described above, there are a few items of note in Figure 2. First consider the fixed effects, or  $\beta$  parameter estimates. The parameter estimates when  $\lambda$  is unconstrained have a larger posterior spread than  $\lambda$  is fixed. This is reasonable since unconstrained  $\lambda$ 's will contribute variability and the unconstrained  $\lambda$   $wEP$  estimates contain weights in more places in the analysis (it is well known that weighted estimates have larger variances). The estimates of  $\beta_1$  when  $\lambda$  is unconstrained have a lower mean than the estimates when  $\lambda$  is fixed. The means  $\beta_2$  when  $\lambda$  is fixed and the estimates when  $\lambda$  is unconstrained with an informative prior are much closer to each other than for  $\beta_1$ . The basic trend of the estimates when  $\lambda$  is fixed is as expected, with the informative cluster and individual sampling scheme estimates larger than the non-informative sampling scheme. The same is true with the estimates when  $\lambda$  is unconstrained. I would expect the weighted estimates to have less bias in the  $wEP$  estimates than the unweighted estimates in Figure 1. This comparison is done in Figures 3 and 4.

Next consider the variance components. The estimates of  $\sigma_\psi^2$  when  $\lambda$  is unconstrained have similar posterior variance than the estimates when  $\lambda$  is fixed. The means of the esti-

mates when  $\lambda$  is fixed are very close to the means of the estimates when  $\lambda$  is unconstrained. The general trend on the  $\sigma_\psi^2$  estimates is not expected. The estimates under informative cluster sampling are larger than the estimates under non-informative sampling and the estimates under informative individual sampling are larger than under informative cluster sampling. Consider the estimates of  $\text{Var}(U_k)$  Stratum 1 and Stratum 2. The estimates when  $\lambda$  is unconstrained have larger posterior variance than the estimates when  $\lambda$  is fixed, as expected. The mean of the estimates of  $\text{Var}(U_k)$  Stratum 1 when  $\lambda$  is fixed are close to the means when  $\lambda$  is unconstrained when the informative sampling does not affect this parameter (for the informative individual and non-informative sampling schemes). When the informative sampling does affect this parameter (informative cluster sampling), the estimates when  $\lambda$  is unconstrained are closer to the true value. The same holds true for  $\text{Var}(U_k)$  Stratum 2, except there is some difference between the estimates when  $\lambda$  is unconstrained and when  $\lambda$  is fixed in the non-informative sampling scheme, with the estimates when  $\lambda$  is fixed being closer to the true value.

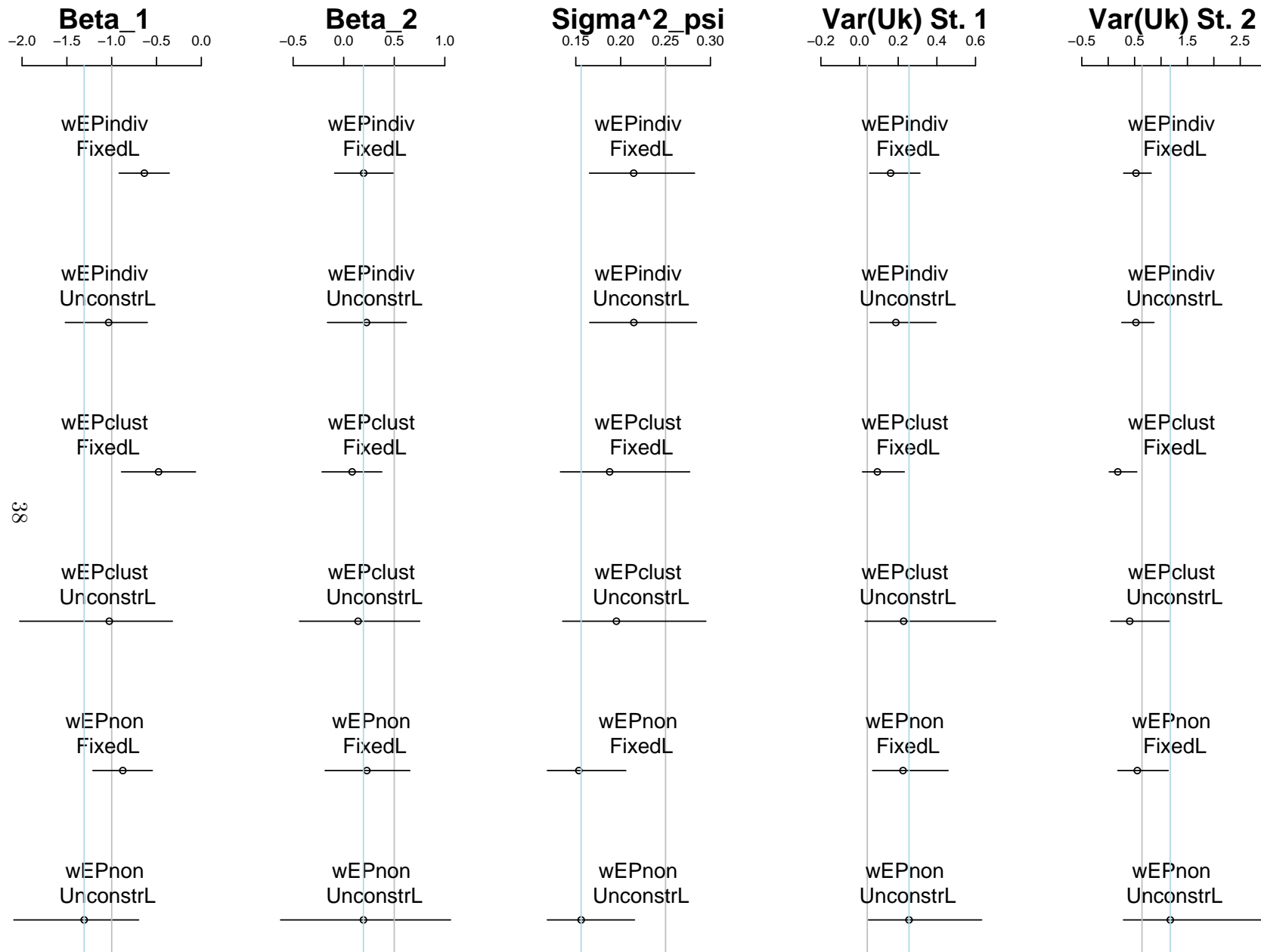


Figure 2: Weighted Results of Sampling Design Parameters (GoM Scores)

### 6.3 Results of Sampling Design Parameters (GoM Scores) when $\lambda$ is Fixed

Figure 3 compares the weighted and  $wEP$  weighted estimates of the sampling design parameters (GoM scores) when  $\lambda$  is fixed.

First consider the fixed-effects or  $\beta$  estimates. The general trend of the unweighted estimates is as expected, as noted in the description of Figure 1. The weighted estimate under the non-informative sampling scheme match the unweighted estimate, as expected. For the  $\beta_1$  estimates under the informative cluster sampling scheme, the weighted estimate has more bias in the same direction as the unweighted estimate, which is not expected as the informative sampling of clusters increases the  $\beta$  estimates and the weighting should reduce the affect of the informative sampling. The weights do not appear to help under informative individual sampling, though they do not increase bias either. For  $\beta_2$ , the weighted estimates adjust the mean of the estimate in the correct direction, with some overcompensation.

Next, consider the variance components. The general trends in the unweighted estimates are in the description of Figure 1. The weighted estimate under non-informative sampling is biased low, as expected from the simulations from Bertolet (2009) simulation set 4. This underestimation of the random error variance is seen in all sampling schemes in Figure 3. The unweighted estimates of  $\sigma_\psi^2$  are similar across all sampling schemes (as discussed with Figure 1). The weighted estimates produce largest bias in the non-informative sampling scheme, and least bias in the informative individual sampling scheme. The weighted estimates of  $\text{Var}(U_k)$  Stratum 1 are larger than the unweighted estimates. This is due to the small intra-class correlation ( $\text{icc} = \frac{0.04}{0.25} = 0.1$ ). This overestimation of the random intercept was also seen simulation set 4 of Bertolet (2009). The weighted estimates of  $\text{Var}(U_k)$  Stratum 2 are also larger than the corresponding unweighted estimates. The reason for this is not known. In simulation Set 12 of Bertolet (2009) there was also positive bias of the scaled 2 weights on the estimate of the random intercept,  $\sigma_{0k}^2$ , when the intra-class correlation was not very small ( $\text{icc} \approx \frac{5}{20} = 0.25$ ).

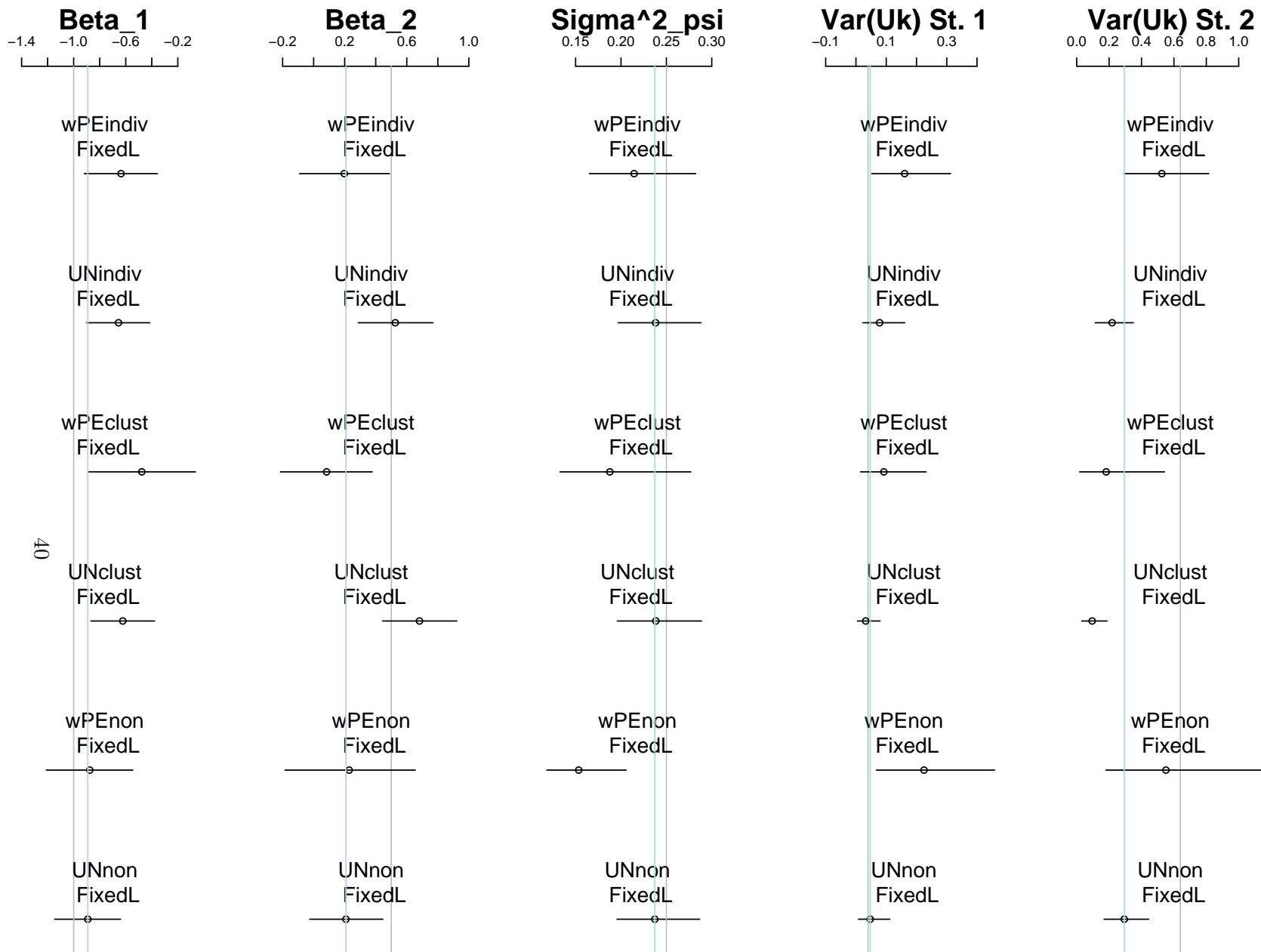


Figure 3: Weighted versus Unweighted GoM Results of Sampling Design Parameters (GoM Scores) when  $\lambda$  is Fixed



## 6.4 Results of Sampling Design Parameters (GoM Scores) when $\lambda$ has an Informative Prior

Figure 4 compares the unweighted and  $wEP$  weighted estimates of the sampling design parameters (GoM scores) when  $\lambda$  is unconstrained with an informative prior.

First consider the fixed-effects or  $\beta$  estimates. The general trend of the unweighted estimates is as expected, as noted in the description of Figure 1. When there is informative cluster or individual sampling, the weighted estimates correctly compensate in the correct direction. For the estimate of  $\beta_1$  in the non-informative sampling scheme, the unweighted estimates produce negative bias. However, for the estimates of  $\beta_2$  under non-informative sampling, the weighted and unweighted estimates have similar means. As expected the weighted estimates have larger posterior spread than the unweighted estimates.

Next consider the variance components. The behavior of the weighted and unweighted estimates of  $\sigma_\psi^2$  is similar to the behavior when  $\lambda$  is fixed as seen in Figure 3. The behavior of the estimates of  $\text{Var}(U_k)$  in Stratum 1 and Stratum 2 is also the same as when  $\lambda$  is fixed as seen in Figure 3, however the posterior spreads are larger than in the fixed  $\lambda$  case.

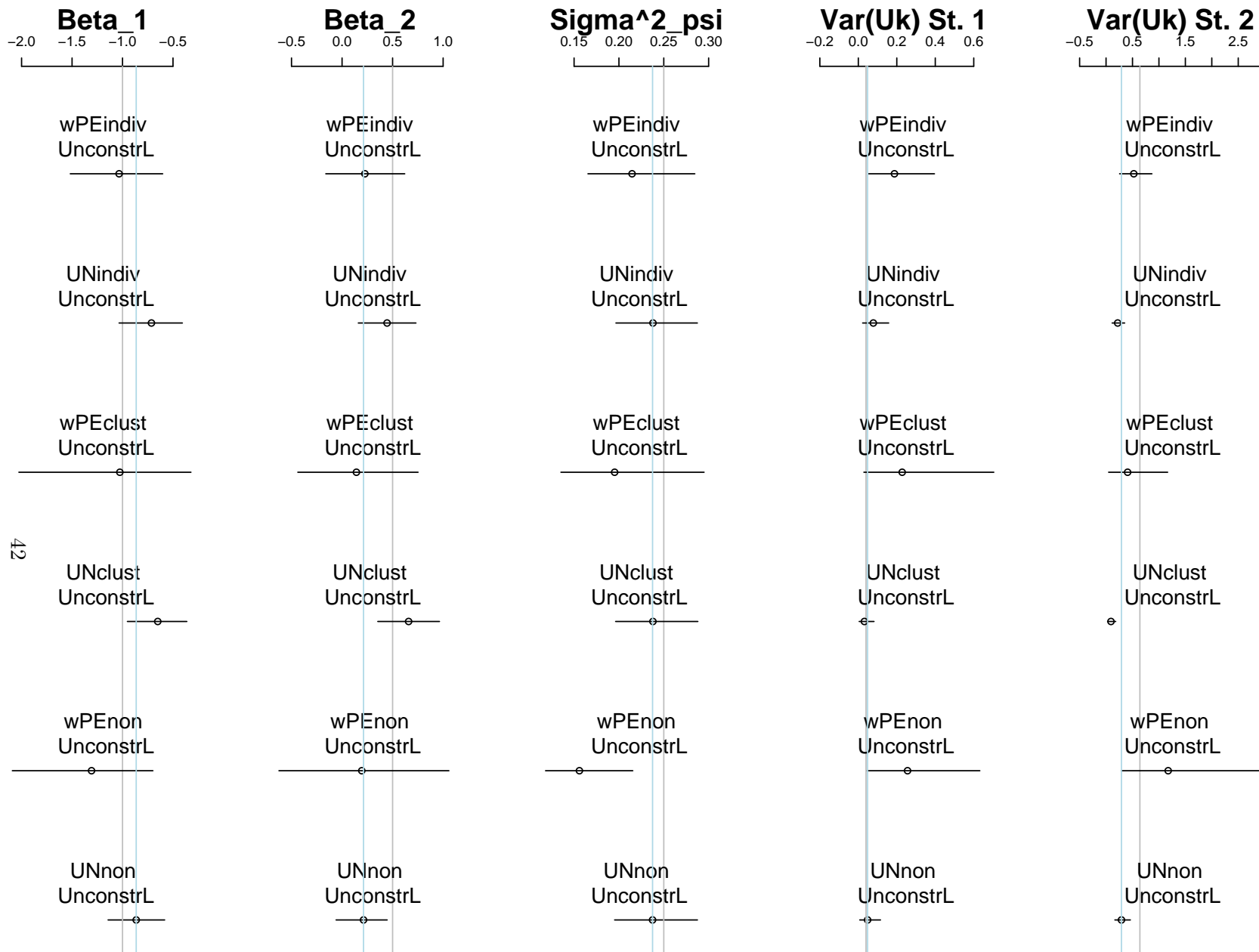


Figure 4: Weighted versus Unweighted GoM Results of Sampling Design Parameters (GoM Scores) with an Informative Prior on  $\lambda$

## 6.5 Comparison of Weighted versus Unweighted Estimates of $\lambda$

Figure 5 compares the unweighted and *wEP* weighted estimates of  $\lambda$ . The scale on all the graphs is 0 to 1 as the  $\lambda$  parameters are probabilities.

The main feature of Figure 5 is the consistency of the means regardless of sampling scheme or type of weighting (unweighted or *wEP* weighted). I next highlight the estimates whose 0.025 and 0.975 quantiles either do not include the true value, or barely include it. These estimates include the unweighted and weighted estimates of  $\lambda_{1,4}$  under the informative individual sampling scheme, the weighted estimate of  $\lambda_{2,3}$  under the informative individual sampling scheme, and the weighted estimate of  $\lambda_{2,4}$  under the informative cluster sampling scheme.

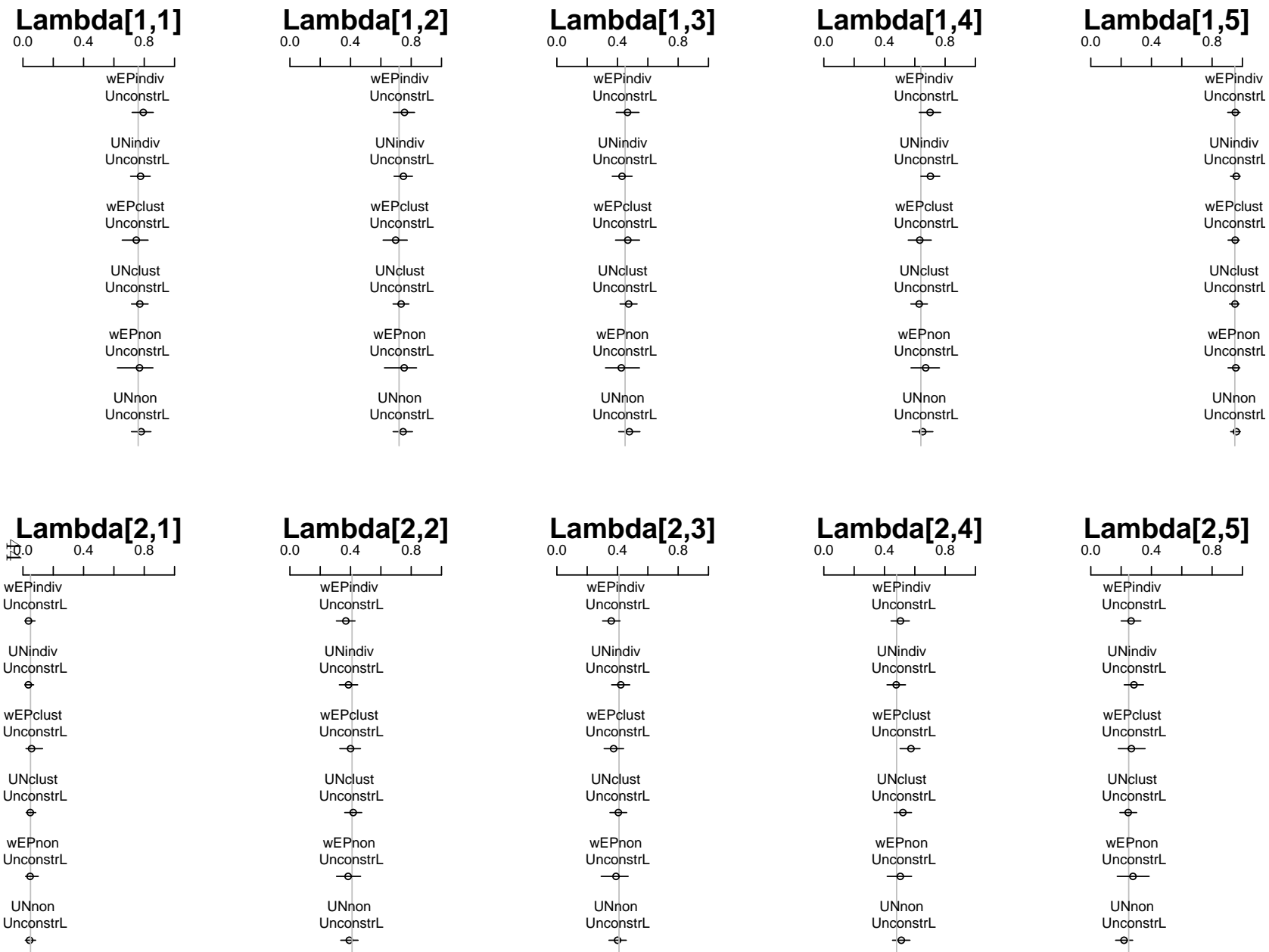


Figure 5: Weighted versus Unweighted Estimates of  $\lambda$

## 7 Summary

The goals of this report are to 1) modify the GoM model to incorporate the sampling design, 2) insert weights into the modified GoM model and 3) analyze the performance of the new unweighted and weighted GoM model through a simulation study. A number of new contributions were made in supporting these three goals.

The original Dirichlet prior GoM model was modified to use a polytomous logistic mixed-effects regression prior. This prior allows incorporation of the dependencies in the GoM scores induced by the sampling design. Another advantage to this prior, as discussed in the future work, is that it can also easily analyze dependencies of longitudinal data.

The insertion of sampling weights expanded upon the PML method from Pfeffermann et al. (1998) and Rabe-Hesketh and Skrondal (2006). In addition, the new method, weighted based on the estimated parameter, introduced a principled type of weighting for complex analyses.

Lastly, the simulation study characterizes the performance of the new polytomous logistic mixed-effects regression prior and the weighting based on the estimated parameter. The simulations indicate that the effect of the sampling design and the effect of adding weights to the analysis strongly parallel the results from the LME simulation study in Bertolet (2009).

## 8 Appendices

### 8.1 PML Weighting of the GoM Model

The weighting method of Pfeffermann et al. (1998) inserts the weights in the process of solving for the estimators. Using the Bayesian modeling and estimation techniques that are different from those used by Pfeffermann et al. (1998), it is difficult to update their method for use on the GoM model. The weighting method of Rabe-Hesketh and Skrondal (2006) creates a weighted likelihood, which is easily incorporated into the Bayesian GoM model structure.

A slight re-parameterization of the prior on the  $U$ 's is needed to incorporate the Rabe-Hesketh and Skrondal (2006) weighting. The variance structure of  $\Omega$  is described in Searle et al. (1992). The model from the GoM description above assumes that the elements of  $U$  are ordered according to random effect. To incorporate the weighting of Rabe-Hesketh and Skrondal (2006), change the ordering of the  $Z$  and  $U$  matrices to be according to cluster instead of element. Allowing  $U_k$  to be the random effects corresponding to cluster  $k$ , the prior on  $U_{kc}$  for the GoM model becomes

$$U_{kc} \sim \text{Normal}(0, \mathcal{O}), \quad , c = 1, \dots, C - 1 \quad (9)$$
$$U_{k_1c} \perp U_{k_2c}$$

where Equation 9 is a prior on each cluster for class  $c, c = 1, \dots, C - 1$  and  $\mathcal{O}$  is the corresponding covariance matrix.

Consider the census joint distribution defined in Equation 7. Estimate the census joint distri-

bution with the sample weighted joint distribution. Let

$$\begin{aligned}
p_w(y, m, \psi, \lambda, \beta, U, \sigma_\psi^2, X, Z) &\propto \exp \left\{ -\frac{1}{2} \sum_{c=1}^C (\beta_c - \mu_\beta)^T \Sigma_\beta^{-1} (\beta_c - \mu_\beta) \right\} \\
&\times (\sigma_\psi^2)^{-\left(\frac{\nu}{2} + 1\right)} \exp \left\{ -\frac{\nu s_\psi^2}{2\sigma_\psi^2} \right\} \left[ \prod_{c=1}^C \prod_{j=1}^J \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \right] \\
&\times \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \prod_{j=1}^J \prod_{c=1}^C \left[ \frac{\exp(\psi_{kic})}{\sum_{c_1=1}^C \exp(\psi_{kic_1})} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{1-y_{kij}} \right]^{m_{kijc} w_{ki}} \\
&\times \prod_{k=1}^{K_s} \left[ \prod_{i=1}^{n_k} \left( \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\} \right) \right]^{w_{i|k}} \\
&\times \exp \left\{ -\frac{1}{2} U_{kc}^T \mathcal{O}^{-1} U_{kc} \right\} \Big]^{w_k} \tag{10}
\end{aligned}$$

where  $K_s$  is the number of sampled clusters and  $n_k$  is the number of sampled individuals in cluster  $k$ . The weights are not inserted on the prior distributions of  $\beta, \sigma_\psi^2$  or  $\lambda$  and are inserted in the likelihood conditional distributions of  $y_{ki}|m_{kijc}, \lambda_{cj}$  and  $m_{kijc}|\psi_{kic}$ . The weighting of the  $\psi$ 's mimics the Rabe-Hesketh and Skrondal (2006) weighting, which weights both the  $\psi_{kic}|\beta, U, \sigma_{psi}^2$  and  $U$  distributions. The incorporation of sampling weights on the prior of  $U$  is reasonable if the prior on  $U$  only contains priors on the random effects for the sampled clusters. This weighting of the sample joint density propagates to the complete conditionals corresponding to the polytomous regression

parameters as follows:

$$\begin{aligned}
p_w(\beta_c|-) &\propto \exp\left\{-\frac{1}{2}(\beta_c - \mu_c)^T \Sigma_\beta^{-1}(\beta_c - \mu_c)\right\} \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\}^{w_{ki}} \\
&\sim \text{Normal}(\mu_1, \Sigma_1) \\
\mu_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki}\right)^{-1} \left(\Sigma_\beta^{-1} \mu_c + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T (\psi_{kic} - Z_{ki}U_c)\right) \\
\Sigma_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki}\right)^{-1} \\
p_w(U_c|-) &\propto \prod_{k=1}^K \left[ \prod_{i=1}^{n_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_{kc})^2\right\}^{w_{ik}} \exp\left\{-\frac{1}{2}U_{kc}^T \mathcal{O}^{-1}U_{kc}\right\} \right]^{w_k} \\
&\sim \text{Normal}(\mu_2, \Sigma_2) \\
\mu_2 &= \left(\sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki}\right)^{-1} \left(\frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T (\psi_{kic} - X_{ki}\beta_c)\right) \\
\Sigma_2 &= \left(\sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki}\right)^{-1} \\
p_w(\sigma_\psi^2|-) &\propto (\sigma_\psi^2)^{-\left(\frac{\nu}{2}+1\right)} \exp\left\{-\frac{\nu S_\psi^2}{2\sigma_\psi^2}\right\} \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\}^{w_{ki}} \\
&\sim \text{Scaled Inv } \chi^2 \left(\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki}(C-1) + \nu, \frac{\left(\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} \sum_{c=1}^{C-1} w_{ki}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right) + \nu S_\psi^2}{\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki}(C-1) + \nu}\right)
\end{aligned}$$

The complete conditionals for the augmented data and the pure response probabilities are,

$$\begin{aligned}
p_w(\lambda_{cj}|-) &\propto \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \left[\lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})}\right]^{m_{kijc} w_{ki}} \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \\
&\sim \text{Beta} \left(\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} y_{kij} m_{kijc} w_{ki} + \eta_{1cj}, \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} (1 - y_{kij}) m_{kijc} w_{ki} + \eta_{2cj}\right) \\
p_w(m_{kij}|-) &\propto \prod_{c=1}^C \left[\frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})}\right]^{m_{kijc} w_{ki}} \\
&\sim \text{Multinomial} \left(1, \left[\frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{1-y_{kij}}\right]^{w_{ki}}, \dots, \right. \\
&\quad \left. \left[\frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{Cj}^{y_{kij}} (1 - \lambda_{Cj})^{1-y_{kij}}\right]^{w_{ki}}\right)
\end{aligned}$$



Finally, a Metropolis step is needed for  $\psi$ ,

$$p_w(\psi_{kic}|-) \propto \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \right]^{m_{kijc}w_{ki}} \\ \times \exp \left\{ -\frac{1}{2\sigma_\psi^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}$$

Let the Jumping distribution be Normally distributed, with the mean at the previous MCMC value and a variance of  $\sigma_{\psi\text{jmp}}^2$ . The acceptance ratio is

$$r_{(w)\psi_{kic}} = \frac{p_w(\psi_{kic}^*|-)}{p_w(\psi_{kic}^{(r)}|-)} \\ = \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}^*\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^*\}} \frac{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^{(r)}\}}{\exp\{\psi_{kic}^{(r)}\}} \right]^{m_{kijc}w_{ki}} \\ \times \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\} \\ = \prod_{c=1}^C \left[ \frac{g_{kic}^*}{g_{kic}^{(r)}} \right]^{\sum_{j=1}^J m_{kijc}w_{ki}} \exp \left\{ -\frac{1}{2\sigma_\psi^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\}$$

where  $g_{kic} = \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \psi_{kic_1}}$  where  $\psi_{kic} = 0$ . These complete conditionals and Metropolis-Hasting steps are implemented using MCMC algorithms.

## 8.2 Complete Conditional Weighting

To add weights to the complete conditionals, take the census complete conditionals from the unweighted GoM derivation, and add sampling weights to estimate them with sample complete conditionals. The subscript  $wCC$  below denotes the result from weighting the complete conditionals. The weighted complete conditionals for the polytomous regression parameters are the same as above,

$$\begin{aligned}
p_{wCC}(\beta_c|-) &\propto \exp\left\{-\frac{1}{2}(\beta_c - \mu_c)^T \Sigma_\beta^{-1}(\beta_c - \mu_c)\right\} \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\}^{w_{ki}} \\
&\sim \text{Normal}(\mu_1, \Sigma_1) \\
\mu_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki}\right)^{-1} \left(\Sigma_\beta^{-1} \mu_c + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T (\psi_{kic} - Z_{ki}U_c)\right) \\
\Sigma_1 &= \left(\Sigma_\beta^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} X_{ki}^T X_{ki}\right)^{-1} \\
p_{wCC}(U_c|-) &\propto \prod_{k=1}^K \left[ \prod_{i=1}^{n_k} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_{kc})^2\right\}^{w_{i|k}} \exp\left\{-\frac{1}{2}U_{kc}^T \mathcal{O}^{-1}U_{kc}\right\} \right]^{w_k} \\
&\sim \text{Normal}(\mu_2, \Sigma_2) \\
\mu_2 &= \left(\sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki}\right)^{-1} \left(\frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T (\psi_{kic} - X_{ki}\beta_c)\right) \\
\Sigma_2 &= \left(\sum_{k=1}^K w_k \mathcal{O}^{-1} + \frac{1}{\sigma_\psi^2} \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki} Z_{ki}^T Z_{ki}\right)^{-1} \\
p_{wCC}(\sigma_\psi^2|-) &\propto (\sigma_\psi^2)^{-(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu s_\psi^2}{2\sigma_\psi^2}\right\} \prod_{k=1}^{K_w} \prod_{i=1}^{n_k} \prod_{c=1}^{C-1} (\sigma_\psi^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\psi^2}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right\}^{w_{ki}} \\
&\sim \text{Scaled Inv } \chi^2(\nu_1, s_1^2) \\
\nu_1 &= \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki}(C-1) + \nu \\
s_1^2 &= \frac{\left(\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} \sum_{c=1}^{C-1} w_{ki}(\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2\right) + \nu s_\psi^2}{\sum_{k=1}^{K_s} \sum_{i=1}^{n_k} w_{ki}(C-1) + \nu}
\end{aligned}$$

The complete conditional for  $\lambda_{cj}$  remains the same, and the complete conditional for  $m_{kijc}$  changes as described above,

$$\begin{aligned}
p_{wCC}(\lambda_{cj}|-) &\propto \prod_{k=1}^{K_s} \prod_{i=1}^{n_k} \left[ \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right]^{m_{kijc} w_{ki}} \lambda_{cj}^{\eta_{1cj}-1} (1 - \lambda_{cj})^{\eta_{2cj}-1} \\
&\sim \text{Beta} \left( \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} y_{kij} m_{kijc} w_{ki} + \eta_{1cj}, \sum_{k=1}^{K_s} \sum_{i=1}^{n_k} (1 - y_{kij}) m_{kijc} w_{ki} + \eta_{2cj} \right) \\
p_{wCC}(m_{kij}|-) &\propto \prod_{c=1}^C \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{cj}^{y_{kij}} (1 - \lambda_{cj})^{(1-y_{kij})} \right]^{m_{kijc}} \\
&\sim \text{Multinomial} \left( 1, \frac{\exp\{\psi_{ki1}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{1j}^{y_{kij}} (1 - \lambda_{1j})^{1-y_{kij}}, \dots, \right. \\
&\quad \left. \frac{\exp\{\psi_{kiC}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \lambda_{Cj}^{y_{kij}} (1 - \lambda_{Cj})^{1-y_{kij}} \right)
\end{aligned}$$

The Metropolis-Hastings step for  $\psi_{kic}$  is the same as the unweighted case,

$$\begin{aligned}
p_{wCC}(\psi_{kic}|-) &\propto \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}\}} \right]^{m_{kijc}} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma_{\psi}^2} (\psi_{kic} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right\}
\end{aligned}$$

Let the Jumping distribution be Normally distributed, with the mean at the previous MCMC value and a variance of  $\sigma_{\psi}^2$ . The acceptance ratio is

$$\begin{aligned}
r_{(wCC)\psi_{kic}} &= \frac{p(\psi_{kic}^*|-)}{p(\psi_{kic}^{(r)}|-)} \\
&= \prod_{c=1}^C \prod_{j=1}^J \left[ \frac{\exp\{\psi_{kic}^*\}}{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^*\}} \frac{\sum_{c_1=1}^C \exp\{\psi_{kic_1}^{(r)}\}}{\exp\{\psi_{kic}^{(r)}\}} \right]^{m_{kijc}} \\
&\quad \times \exp \left\{ -\frac{1}{2\sigma_{\psi}^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\} \\
&= \prod_{c=1}^C \left[ \frac{g_{kic}^*}{g_{kic}^{(r)}} \right]^{\sum_{j=1}^J m_{kijc}} \exp \left\{ -\frac{1}{2\sigma_{\psi}^2} \left[ (\psi_{kic}^* - X_{ki}\beta_c - Z_{ki}U_c)^2 - (\psi_{kic}^{(r)} - X_{ki}\beta_c - Z_{ki}U_c)^2 \right] \right\}
\end{aligned}$$

By construction, these complete conditionals and Metropolis-Hastings steps insert weights only when using sample quantities to estimate finite population quantities.

One problem with this  $wCC$  weighting that some components to the posterior (or log likeli-

hood times the prior) are treated differently in different complete conditionals. For example, in  $p_{wCC}(\lambda_{cj}|-)$ , the  $[\lambda_{cj}^{y_{kij}}(1 - \lambda_{cj})^{1-y_{kij}}]$  term from the posterior is weighted. However, the same term in  $p_{wCC}(m_{kij}|-)$  is not weighted. Treating a component from the posterior differently in different complete conditionals appears unprincipled.

### 8.3 Computer Code

The GoM code used for the results of this report started with the code from Elena Erosheva's thesis, see Erosheva (2002), and was modified by Cyrille Joutard. I then modified that code to include the polytomous logistic random-effects prior and the  $wEP$  weighting scheme. This code may be found at <http://stat.cmu.edu> under the **Recent PhD Theses** link. The c-code uses the VMR library, downloaded from <http://www.stat.cmu.edu/~hseltman/>. It is in the Computer Programming, C/C++ section. The code uses the IMSL library, available from Visual Numerics at <http://www.vni.com> for a fee. The compilation instructions are commented in the beginning of the code. Along with the code are sample input files and the corresponding output file.

## References

- Airoldi, E. M., Fienberg, S. E., Joutard, and C. Love, T. M. (2005). Hierarchical Bayesian mixed-membership models and latent pattern discovery. *International Journal of Parallel, Emergent and Distributed Systems*, 00(00):1–12.
- Bertolet, M. (2009). Sampling weights, model misspecification and informative sampling: A simulation study. Technical Report 872, Carnegie Mellon University, Department of Statistics.
- Bradley, R. and Gart, J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, 49(1/2):205–214.
- Cowles, M. and Carlin, B. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 92(434):883–904.
- Erosheva, E. (2002). *Grade of Membership and Latent Structure Models with Application to Disability Survey Data*. PhD thesis, Carnegie Mellon University, Department of Statistics.
- Erosheva, E. (2005). Comparing latent structures of the grade of membership, rasch and latent class models. *Psychometrika*, 70(4):619–628.
- Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31:1109–1144.
- Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Simon & Schuster.
- Kaufman, G. M. and Press, S. J. (1973). Bayesian factor analysis. Technical Report 7322, Center for Mathematical Studies in Business and Economics, Department of Economics and Graduate School of Business.
- Marini, M., Li, X., and Fan, P.-L. (1996). Characterizing latent structure: Factor analytic and grade of membership models. *Sociological Methodology*, 26:133–164.
- Mislevy, R. J. and Sheehan, K. M. (1989a). Information matrices in latent-variable models. *Journal of Educational Statistics*, 14(4):335–350.

- Mislevy, R. J. and Sheehan, K. M. (1989b). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, 54(4):661–679.
- Pfeffermann, D., Skinner, C., Holmes, D., Goldstein, H., and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B*, 60(1):23 – 40.
- Potthoff, R., Woodbury, M., and Manton, K. (1992). "equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association*, 87(418):838–396.
- Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A*, 169:805–827.
- Raftery, A. E. and Lewis, S. (1992). How many iterations in the gibbs sampler? In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 763–773. WOXford University Press.
- Rowe, D. B. (2001). A model for Bayesian factor analysis with jointly distributed means and loadings. Technical Report 1108, Division of the Humanities and Social Sciences, California Institute of Technology, Social Science Working Paper.
- Searle, S., Casella, G., and McCulloch, C. (1992). *Variance Components*. John Wiley & Sons, Inc.
- Thomas, N. (2000). Assessing model sensitivity of the imputations methods used in the national assessment of educational progress. *Journal of Educational and Behavioral Statistics*, 25(4):351–371.
- von Davier, M., Sinharay, S., Oranje, A., and Beaton, A. (2007). The statistical procedures used in national assessment of educational progress: Recent developments and future directions. In *Handbook of Statistics:Psychometrics*.
- Woodbury, M. and Manton, K. (1989). Grade-of-membership analysis of depression-related psychiatric disorders. *Sociological Methods and Research*, 18:126–163.