D A Grades of Measurement Model for Mixed Data

We observe a vector of $j = 1, \ldots, J$ covariates, $\mathbf{x_i} = (x_{i1}, \ldots, x_{iJ})$, for each of the $i \in 1, \ldots, I$ politicians in our dataset. To provide an overall picture of variability in the dataset, we use a grades of measurement (GoM) approach to model politicians' characteristics in terms of a lower-dimensional set of latent groups. As we outline in the main text, we assume each politician $i \in 1, \ldots, I$ exhibits potentially partial membership in K underlying latent classes. We represent subject i's membership profile with a vector of scores, $\mathbf{g_i} = (g_{i1}, \ldots, g_{ik})$, where $0 \leq g_{ik} \leq 1$ and $\sum_{k=1}^{K} g_{ik} = 1$ for $i = 1, \ldots, I$. Thus, each g_{ik} indicates the extent to which subject i is associated with group k and any $g_{ik} = 1$ implies that subject i belongs only to group k. After estimation, these parameters provide a tool for evaluating the extent to which each individual belongs to each of the K ideal types discovered by the model.

While traditional GoM models deal only with polytomous observed items (Erosheva 2002), one might include both nominal and interval-level variables in such an analysis. Therefore, we develop a mixed-mode generalization of the standard GoM setup. Specifically, we describe each x_{ij} by a density function $f_{jk}(x_{ij}|g_{ik})$ and the form of this density depends on the mode of indicator j. Formally, when indicator j is continuous we assume

$$f_{jk}(x_{ij}|g_{ik} = 1) = \mathcal{N}(\mu_{jk}, \sigma^2_{jk}). \tag{1}$$

In other words, each continuous indicator is associated with a set of group-specific normal densities. After fitting the model to the data, one can interpret μ_{jk} as the expected level of observed variable j for a randomly drawn full member of group k. Similarly, σ^2_{jk} is the variance in observed variable j among full members of k. For example, were we to include politician age as an interval variable in a two-group model, we might find that politicians with full membership in group one had an expected age of 40 years with a variance of 10 years, while full members of group two might be expected to be 50 with a variance of 5 years.

Alternatively, when indicator j is a categorical variable with $l_j = 1, \ldots, L_j$ categories, we assume

$$f_{jk}(x_{ij} = l_j | g_{ik} = 1) = p_{jkl_j} \tag{2}$$

where $0 \leq p_{jkl_j} \leq 1$ and $\sum_{l_j=1}^{L_j} p_{jkl_j} = 1$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$. After estimation, p_{jkl_j} is the probability that a randomly drawn full member of group k will exhibit category l_j on observed variable j. So, for example, if variable j is a binary indicator of gender, an estimated value of $p_{j11} = 0.7$ would indicate that a randomly selected full member of group one has a 70 per cent chance of being a woman.

Putting the pieces together, we assume that, conditional on the GoM scores, the probability of observing a given outcome on indicator j for subject i is

$$f(x_{ij} | \mathbf{g_i}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \sum_{k=1}^{K} \left[g_{ik} \cdot f_{jk}(x_{ij} | g_{ik} = 1) \right]. \tag{3}$$

This key assumption maintains that the density of each individual item response is a convex combination of group-specific densities, weighted by subject-specific group membership scores, and establishes the core structure of the GoM model. It is this structure that allows for a soft clustering of the observed data; essentially, we model each individual observation as a mixture over pure types.

Estimation of the model relies on number of other technical assumptions. First, we assume that, after conditioning on $\mathbf{g}$, each x_{ij} is independent across all values of j. In other words, we assume that covariances across observed indicators are completely explained by latent class memberships. Or, mathematically, we assume that

$$f(\mathbf{x_i} | \mathbf{g_i}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{j=1}^{J} \sum_{k=1}^{K} \left[g_{ik} \cdot f_{jk}(x_{ij} | g_{ik} = 1) \right]. \tag{4}$$

We also assume that each vector of observed characteristics, $\mathbf{x_i}$, is independent across all values of i, or, in other words, that individual observations are drawn randomly from the underlying population. Taken together, the above assumptions yield the following sampling density:

$$f(\mathbf{x}|\mathbf{g},\mathbf{p},\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{i=1}^{I}\prod_{j=1}^{J}\sum_{k=1}^{K}\left[g_{ik}\cdot f_{jk}(x_{ij}|g_{ik}=1)\right]. \tag{5}$$

D.1 Estimating the Mixed-Mode GoM Model

D.1.1 Prior Distribution

We use a Bayesian estimation approach and, therefore, must specify prior distributions for the model parameters. We assume, *a priori*, that the subject-level membership parameters are independent of the structural parameters and that the categorical and continuous structural parameters are independent of one another: $p(\mathbf{g},\mathbf{p},\boldsymbol{\mu},\boldsymbol{\sigma}) = p(\mathbf{g})p(\mathbf{p})p(\boldsymbol{\mu},\boldsymbol{\sigma})$. First, we adopt a conjugate Dirichlet prior on the GoM parameters, assuming that

$$\mathbf{g_i} \sim \mathcal{D}_k(\boldsymbol{\alpha}), \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_k)$ and $\boldsymbol{\alpha}$ is known. Similarly, we use Dirichlet priors for the categorical structural parameters, and assume that categorical response probabilities are independent across items and groups, such that

$$p(\mathbf{p}) \propto \prod_{k=1}^{K}\prod_{j\in J_P}\prod_{l_j=1}^{L_j} p_{jkl_j}^{\beta_{jl_j}-1} \tag{7}$$

where $\boldsymbol{\beta}_\mathbf{j} = (\beta_{j1},\ldots,\beta_{jL_j})$ is known and J_P is the subset of $j=1,\ldots,J$ for which item j is polytomous. For the normal parameters describing the continuous item responses, we use standard, semi-conjugate priors, again assuming prior independence across both groups and parameters. Specifically, we adopt normal priors on the group means and inverse gamma priors on the group variances, such that,

$$p(\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{k=1}^{K}\prod_{j\in J_C} p(\mu_{jk})p(\sigma_{jk}^2), \tag{8}$$

each

$$\mu_{jk} \sim \mathcal{N}(\mu_{j0}, \sigma_{j0}^2), \tag{9}$$

and each

$$\sigma_{jk}^2 \sim \Gamma^{-1}(c_{j0}, d_{j0}), \tag{10}$$

where $\Gamma^{-1}(\cdot)$ is the inverse gamma density function and J_C is the subset of $j = 1, \ldots, J$ for which item j is continuous. In sum, we assume the prior distribution:

$$p(\mathbf{g}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) \propto \prod_{k=1}^{K} \left(\left[\prod_{i=1}^{I} g_{ik}^{\alpha_k - 1} \right] \left[\prod_{j \in J_P} \prod_{l_j=1}^{L_j} p_{jkl_j}^{\beta_{jl_j} - 1} \right] \left[\prod_{j \in J_C} \frac{1}{\sigma_{j0}} \phi \left(\frac{\mu_{jk} - \mu_{j0}}{\sigma_{j0}} \right) (\sigma_{jk}^2)^{-(c_{j0}+1)} e^{-d_{j0}/\sigma_{jk}^2} \right] \right), \tag{11}$$

where $\phi(\cdot)$ is the standard normal density function.

For actual estimation we use the flat priors $\alpha_k = 1 \ \forall k$ for the GoM parameters and $\beta_{jl} = 1 \ \forall j, l$ for the categorical structural parameters. For the continuous parameters we set each μ_{j0} to the sample mean for variable j and each $\sigma_{j0}^2 = R_j^2$, where R_j is the observed range of variable j. Moreover, we set each $c_{j0} = 2$ and each $d_{j0} = \frac{2}{R^2}$. In general, these priors are quite uninformative.

D.1.2 A Gibbs Sampler for the Mixed-Mode GoM Model

Erosheva (2002) shows that the GoM model is equivalent to a parameterization of latent class model proposed by Haberman (1995). One obtains this latent class representation by augmenting the data with latent binary vectors, $\mathbf{z_{ij}} = (z_{ij1}, \ldots, z_{ijK})$, where $\mathbf{z_{ij}} \sim \mathcal{M}_K(1, \mathbf{g_i})$, where $\mathcal{M}_K(\cdot)$ is the multinomial distribution with K dimensions. Thus, the GoM model where each individual i is considered a partial member of K groups is equivalent to a pure latent class model where each individual i belongs to one, and only one, of K^J latent groups. This equivalence allows one to rewrite the GoM sampling density in equation 5 as the augmented likelihood

$$f(\mathbf{x}, \mathbf{z}|\mathbf{g}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\left[g_{ik} \cdot f_{jk}(x_{ij}|g_{ik}=1)\right]^{z_{ijk}}, \tag{12}$$

yields—after multiplying by equation 11—the posterior distribution

$$p(\mathbf{g}, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}|\mathbf{x}, \mathbf{z}) \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\left[g_{ik} \cdot f_{jk}(x_{ij}|g_{ik}=1)\right]^{z_{ijk}} \\ \times \prod_{k=1}^{K}\left(\left[\prod_{i=1}^{I} g_{ik}^{\alpha_k - 1}\right]\left[\prod_{j\in J_P}\prod_{l_j=1}^{L_j} p_{jkl_j}^{\beta_{jl_j}-1}\right]\left[\prod_{j\in J_C}\frac{1}{\sigma_0}\phi\left(\frac{\mu_{jk}-\mu_0}{\sigma_0}\right)(\sigma_{jk}^2)^{-(c0+1)}e^{-d_0/\sigma_{jk}^2}\right]\right), \tag{13}$$

and motivates a Gibbs sampling algorithm for estimating the model parameters.

In its first step, the Gibbs sampler imputes $\mathbf{z}$, drawing each $\mathbf{z_{ij}}$ from its full conditional distribution,

$$\mathbf{z_{ij}} \sim \mathcal{M}_K\left(1, \rho_1, \ldots, \rho_K\right). \tag{14}$$

When item j is nominal,

$$\rho_k \propto g_{ik}\prod_{l_j=1}^{L_j} p_{jkl_j}^{x_{ijl_j}}, \tag{15}$$

where

$$x_{ijl_j} = \begin{cases} 1 & \text{if } x_{ij} = l_j \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, when item j is continuous,

$$\rho_k \propto g_{ik} \cdot \frac{1}{\sigma_{jk}}\phi\left(\frac{x_{ij}-\mu_{jk}}{\sigma_{jk}}\right). \tag{16}$$

After the imputation step for $\mathbf{z}$, the algorithm proceeds to the posterior step, first sampling each $\mathbf{g_i}$ from the full conditional distribution

$$\mathbf{g_i} \sim \mathcal{D}_K\left(\alpha_1 + \sum_{j=1}^{J} z_{ij1}, \ldots, \alpha_K + \sum_{j=1}^{J} z_{ijK}\right). \tag{17}$$

Next, it samples the item parameters, $\mathbf{p}$, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma^2}$. For the nominal item parameters the conditional distribution of each $\mathbf{p_{jk}}$ is

$$\mathbf{p_{jk}} \sim \mathcal{D}_{L_j}\left(\beta_{j1} + \sum_{i=1}^{I} x_{ij1} z_{ijk}, \ldots, \beta_{jL_j} + \sum_{i=1}^{I} x_{ijL_j} z_{ijk}\right). \tag{18}$$

If, on the other hand, item j is continuous, the conditional distributions are

$$\mu_{jk} \sim \mathcal{N}\left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{I} z_{ijk}}{\sigma_{jk}^2} \cdot \frac{\sum_{i=1}^{I} z_{ijk} x_{ij}}{\sum_{i=1}^{I} z_{ijk}}}{\frac{1}{\sigma_0^2} + \frac{\sum_{i=1}^{I} z_{ijk}}{\sigma_{jk}^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{\sum_{i=1}^{I} z_{ijk}}{\sigma_{jk}^2}}\right) \tag{19}$$

and

$$\sigma_{jk}^2 \sim \Gamma^{-1}\left(c_0 + \frac{\sum_{i=1}^{I} z_{ijk}}{2}, d_0 + \frac{\sum_{i=1}^{I} z_{ijk}(x_{ij} - \mu_{jk})^2}{2}\right). \tag{20}$$

D.1.3 Handling Missing Data

We can easily extend the Gibbs sampler to deal with missing data by treating missing values as additional model parameters, and sampling from the conditional posterior distributions of the missing values at each sampler iteration to augment the observed data. Specifically, the sampler draws missing nominal values from

$$x_{ijl_j} \sim \mathcal{M}_{L_j}\left(1, \mathbf{p_{jk'}}\right), \tag{21}$$

where k' is the value of k for which $\mathbf{z_{ij}} = 1$.[1] The sampler draws missing continuous values from

$$x_{ij} \sim \mathcal{N}\left(\mu_{jk'}, \sigma_{jk'}^2\right). \tag{22}$$

[1] By definition $\mathbf{z_{ij}}$ is a K-vector with all values but one set to zero, and the remaining value set to one.

References

Erosheva, Elena A. 2002. Grade of Membership and Latent Structure Models With Application to Disability Survey Data PhD thesis Carnegie Melon University.

Haberman, Shelby J. 1995. "Book review of Statistical Applications Using Fuzzy Sets, by Manton, K. G., Woodbury, M. A. and Corder, L.S." *Journal of the American Statistical Association* 90(431):1131– 1133.