Appendices for the Article "Comparing Two Algorithms for Calibrating the Restricted Non-compensatory Multidimensional IRT Model"

Appendix A: Details of the MCMC algorithm

In this appendix, we detail an MCMC algorithm for the restricted conjunctive MIRT (RC-MIRT) model. As in any adaptation of the MLTM, the item/person parameters to be estimated are a, g, $\boldsymbol{b} = \text{vec}([\boldsymbol{b}_1, \ldots, \boldsymbol{b}_J])^T = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_J^T)^T$ and $\boldsymbol{\theta} = \text{vec}([\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N])^T = (\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_N^T)^T$, where vec stands for the "vectorized" operator and forms a vector by stacking the columns of the matrix inside. To start the exposition, given a set of values from the $(r-1)^{\text{th}}$ iteration, the r^{th} iteration of the MCMC algorithm proceeds in the following steps.

Step 1. $\boldsymbol{\Sigma_\theta}|\boldsymbol{\theta}, \boldsymbol{Y}$

Because the discrimination parameters are fixed to 1, the complete covariance matrix of ability, $\boldsymbol{\Sigma_\theta}$, can be freely estimated. Thus, sample $\boldsymbol{\Sigma_\theta}$ conditional on $\boldsymbol{\theta}^{(r-1)}$ and $\boldsymbol{Y}$ by assuming that the prior distribution of $\boldsymbol{\Sigma_\theta}$ is inverse-Wishart distributed because $\boldsymbol{\theta}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$. Given prior parameters $(\boldsymbol{I}_\nu, \nu)$, we thus have

$$p(\boldsymbol{\Sigma_\theta}|\boldsymbol{\theta}^{(r-1)}) \sim \text{InvWish}\Big(\boldsymbol{\theta}^{(r-1)T}\boldsymbol{\theta}^{(r-1)} + \boldsymbol{I}_\nu, N + \nu\Big),$$

where $\boldsymbol{\theta}^{(r-1)}$ is of dimension $N \times K$, N is the examinee sample size, and ν is generally set to K.

Step 2. $\boldsymbol{\theta}|\boldsymbol{Y}, a, g, \boldsymbol{b}, \boldsymbol{\Sigma_\theta}$

Sample $\boldsymbol{\theta}_i^{(r)}$ conditional on $\boldsymbol{Y}_i$, $\boldsymbol{b}^{(r-1)}$, $a^{(r-1)}$, $g^{(r-1)}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(r)}$ via a Metropolis step. For each $\boldsymbol{\theta}_i^{(r)} =$

$(\theta_{i1}, \ldots, \theta_{iK})^T$, $i = (1, \ldots, N)$, generate a candidate $\boldsymbol{\theta}_i^*$ from a multivariate normal distribution with mean $\boldsymbol{\theta}_i^{(r-1)}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(r)}$. Then set $\boldsymbol{\theta}_i^{(r)} = \boldsymbol{\theta}_i^*$ with probability

$$\alpha(\boldsymbol{\theta}_i^{(r-1)}, \boldsymbol{\theta}_i^*) \equiv \min\left\{1, \frac{\phi(\boldsymbol{\theta}_i^*; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(r)}) L(\boldsymbol{Y}_i | \boldsymbol{b}^{(r-1)}, a^{(r-1)}, g^{(r-1)}, \boldsymbol{\theta}_i^*)}{\phi(\boldsymbol{\theta}_i^{(r-1)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(r)}) L(\boldsymbol{Y}_i | \boldsymbol{b}^{(r-1)}, a^{(r-1)}, g^{(r-1)}, \boldsymbol{\theta}_i^{(r-1)})}\right\},$$

where $\phi(\cdot, \cdot)$ is the multivariate normal density, $\boldsymbol{Y}_i$ is the i^{th} row of the response matrix, and $L(\boldsymbol{Y}_i)$ is the likelihood of the response vector, $\boldsymbol{Y}_i$, and computed by means of

$$L(\boldsymbol{Y}_i | \boldsymbol{b}, a, g, \boldsymbol{\theta}_i) = \prod_{j=1}^{J} \left[a\pi_{ij} + g(1 - \pi_{ij})\right]^{y_{ij}} \left[1 - \left(a\pi_{ij} + g(1 - \pi_{ij})\right)\right]^{1-y_{ij}}.$$

Otherwise, set $\boldsymbol{\theta}_i^{(r)} = \boldsymbol{\theta}_i^{(r-1)}$. Note that $\boldsymbol{\mu}$, the prior mean vector of abilities, is chosen to be a vector of 0s for model identification.

Step 3. $\boldsymbol{b} | \boldsymbol{Y}, \boldsymbol{\theta}, a, g$

Sample $\boldsymbol{b}_j^{(r)}$ conditional on $\boldsymbol{Y}_j$, $\boldsymbol{\theta}^{(r)}$, $a^{(r-1)}$, and $g^{(r-1)}$ via a Metropolis step. For item j, generate a candidate $\boldsymbol{b}_j^*$ from a multivariate normal distribution with mean $\boldsymbol{b}_j^{(r-1)}$ and covariance matrix $\boldsymbol{C}$, where $\boldsymbol{C}$ is chosen to restrain the acceptance rate to an adequate level, usually between 20% to 40%. If $q_{jk} = 0$, then $b_{jk} = 0$ and will not be updated. Otherwise, accept $\boldsymbol{b}_i^{(r)} = \boldsymbol{b}_i^*$ with probability

$$\alpha(\boldsymbol{b}_j^{(r-1)}, \boldsymbol{b}_j^*) \equiv \min\left\{1, \frac{L(\boldsymbol{Y}_j | \boldsymbol{b}_j^*, \boldsymbol{\theta}^{(r)}, a^{(r-1)}, g^{(r-1)}) \phi(\boldsymbol{b}_j^*; \boldsymbol{\mu}_{\mathbf{b}}, \boldsymbol{\Sigma}_{\mathbf{b}})}{L(\boldsymbol{Y}_j | \boldsymbol{b}_j^{(r-1)}, \boldsymbol{\theta}^{(r)}, a^{(r-1)}, g^{(r-1)}) \phi(\boldsymbol{b}_j^{(r-1)}; \boldsymbol{\mu}_{\mathbf{b}}, \boldsymbol{\Sigma}_{\mathbf{b}})}\right\}, \tag{A1}$$

where $\boldsymbol{Y}_j$ is the j^{th} column of the response matrix, and $L(\boldsymbol{Y}_j)$ is computed via

$$L(\boldsymbol{Y}_j | \boldsymbol{b}, a, g, \boldsymbol{\theta}_i) = \prod_{i=1}^{N} \left[a\pi_{ij} + g(1 - \pi_{ij})\right]^{y_{ij}} \left[1 - \left(a\pi_{ij} + g(1 - \pi_{ij})\right)\right]^{1-y_{ij}}.$$

The parameters $\boldsymbol{\mu}_{\mathbf{b}}$ and $\boldsymbol{\Sigma}_{\mathbf{b}}$ can take on any form, but they are typically fixed to a vector of zeros and an identity matrix, respectively.

Note that the proposed RC-MIRT model has one (and only one) item parameter for each item on each dimension on which that item loads. If using a two-parameter version of the conjunctive MIRT model, such as in Babcock (2011), then one must specify the diagonal elements of $\boldsymbol{\Sigma_\theta}$ to establish model identifiability. Rather than estimating the covariance matrix, $\boldsymbol{\Sigma_\theta}$, one would instead proceed by estimating the correlation matrix, $\boldsymbol{R_\theta}$. Edwards (2010) recommended estimating correlation matrices by means of a Metropolis hit-and-run algorithm. In that algorithm, one should specify the prior distribution on the correlation terms as a standard normal distribution truncated between -1 and 1 to avoid using complicated priors, such as those introduced in Babcock (2011).

Initial values. The MCMC algorithm can only proceed after specifying appropriate, initial values for several of the unknown parameters. The initial values play an important role in any MCMC algorithm, as they help determine the performance of a chain. We propose slightly informative initial values. Specifically, to determine the initial value of $\boldsymbol{b}_j^{(0)}$, assume $b_{j1} = b_{j2} = \cdots = b_{jk} = b_j$ for $q_{jk} \neq 0$ and $\boldsymbol{\theta} = \mathbf{0}$. Then, $\left(\frac{1}{1+\exp(b_j)}\right)^{K_j} = p_j$, where K_j and p_j are the total number of attributes measured by and the proportion of examinees who correctly responded to item j, respectively. After solving the above equation for b_j, replace $b_{jk} = b_j$ in $\sum_{j=1}^{J} \left(\frac{\exp(\theta_{ik}-b_{jk})}{1+\exp(\theta_i - b_{jk})}\right)^{q_{jk}} = J \times p_i$, and solve for each θ_{ik}, where p_i is the proportion of items correctly answered by examinee i and $\theta_{i1} = \theta_{i2} = \cdots = \theta_{ik}$.

Estimating a and g parameters.

To estimate RC-MIRT a and g parameters using MCMC, we need to define a binary variable, W_{ij}, such that

$$W_{ij} = \begin{cases} 1 & \text{if person } i \text{ knows the correct answer to item } j \\ 0 & \text{if person } i \text{ doesn't know the correct answer to item } j \end{cases}$$

Then the conditional probability of $W_{ij} = w_{ij}$ given $Y_{ij} = y_{ij}$ is

$$P(W_{ij} = 1 | Y_{ij} = 1) \propto a\pi_{ij} \tag{A2}$$

$$P(W_{ij} = 0 | Y_{ij} = 1) \propto g(1 - \pi_{ij}) \tag{A3}$$

$$P(W_{ij} = 1 | Y_{ij} = 0) \propto (1 - a)\pi_{ij} \tag{A4}$$

$$P(W_{ij} = 0 | Y_{ij} = 0) \propto (1 - g)(1 - \pi_{ij}) \tag{A5}$$

where

$$\pi_{ij} = \prod_{k=1}^{K} \left(\frac{\exp(\theta_{ik} - b_{jk})}{1 + \exp(\theta_{ik} - b_{jk})} \right)^{q_{jk}}. \tag{A6}$$

If estimating a and g parameters, also add the following two steps to the Markov chain update:

Step 1. $\boldsymbol{W} | \boldsymbol{Y}, a, g$

Sample $W_{ij}^{(r)}$ conditional on Y_{ij}, $\pi_{ij}^{(r-1)}$, $a^{(r-1)}$, and $g^{(r-1)}$ based on Equations (A2)–(A5).

Step 2. $a, g | \boldsymbol{Y}, \boldsymbol{W}$

Sample $a^{(r)}$ and $g^{(r)}$ conditional on W_{ij} and Y_{ij}. To determine the appropriate posterior distribution of $a^{(r)}$ and $g^{(r)}$ given the r^{th} iteration of the algorithm, let $R^{(r)} = \sum_{j=1}^{J} \sum_{i=1}^{N} I(w_{ij}^{(r)} = 0)$ be the estimated number of guesses for all people across all items, where $I(w_{ij}^{(r)} = 0)$ is an indicator function that equals 1 if person i guessed on item j. Moreover, let $S^{(r)} = \sum_{j=1}^{J} \sum_{i=1}^{N} I(y_{ij} = 1) I(w_{ij} = 0)$ be the number of correct responses obtained by guessing. Because $P(Y_{ij} = 1 | W_{ij} = 0) = g$ for all persons and items due to a strict assumption of the MLTM, $S \sim \text{Binom}(R, g)$. Assuming a conjugate Beta prior on g with prior parameters α_g and β_g, then

$$g^{(r)} \sim \text{Beta}(S^{(r)} + \alpha_g, R^{(r)} - S^{(r)} + \beta_g). \tag{A7}$$

Similarly, let $U^{(r)} = \sum_{j=1}^{J} \sum_{i=1}^{N} I(w_{ij}^{(r)} = 1)$ be the estimated number of items that are known across all people. If $V^{(r)} = \sum_{j=1}^{J} \sum_{i=1}^{N} I(y_{ij} = 1) I(w_{ij} = 1)$ is the number of correct responses obtained without guessing, then $V \sim \text{Binom}(U, a)$. Assuming a conjugate Beta prior on a with prior parameters α_a and β_a, then

$$a^{(r)} \sim \text{Beta}(v^{(r)} + \alpha_a, u^{(r)} - v^{(r)} + \beta_a). \tag{A8}$$

Appendix B: Details of the MH-RM algorithm

Assume that a researcher wants to estimate item parameters for the RC-MIRT model given a set of data and using the MH-RM algorithm, and that this researcher acquires reasonable initial values. If $\boldsymbol{b}^{(r-1)}$ is the vector of item parameter estimates after the $(r-1)^{\text{th}}$ loop, then the r^{th} iteration proceeds as follows.

Step 1. Stochastic imputation.

Draw m_r sets of missing data $\{\boldsymbol{\theta}_s^{(r)},\, \boldsymbol{\Sigma_\theta};\; s = 1, \ldots, m_r\}$, where m_r is chosen by the researcher and allowed to depend on the iteration number. These missing data are combined with the observed responses, $\boldsymbol{Y}$, to form m_r arrays of complete data $\left(\boldsymbol{Z}_s^{(r)} = (\boldsymbol{Y},\, \boldsymbol{\theta}_s^{(r)},\, \boldsymbol{\Sigma_\theta});\; s = 1, \ldots, m_r\right)$. One should use the MH sampler described in the previous section to mechanically update $\boldsymbol{\theta}$ (conditional on $\boldsymbol{Y}$, $\boldsymbol{b}^{(r-1)}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(r-1)}$) and $\boldsymbol{\Sigma_\theta}$ (conditional on $\boldsymbol{\theta}^{(r-1)}$ and $\boldsymbol{Y}$). A general rule in choosing the number of parameters to update via MCMC is that one should impute the least missing data needed to simplify complete data calculations. Using this logic, one might argue that $\boldsymbol{\theta}$ form such a minimal set. However, we decided to impute both $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma_\theta}$ for two reasons. First, $\boldsymbol{\Sigma_\theta}$ is required only for determining the distribution used to sample $\boldsymbol{\theta}$ via the Metropolis-Hastings step and does not affect futher analyses. Second, as will be clarified shortly, including $\boldsymbol{\Sigma_\theta}$ as an estimation component of the Robbins-Monro step results in a full-rank Hessian matrix and thus, a more difficult and costly matrix inversion. In a pilot study, we found that imputing only $\boldsymbol{\theta}$ (as compared to imputing both $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma_\theta}$) does not improve the estimation accuracy of $\boldsymbol{b}$ enough to justify the extra complexity of adding $\boldsymbol{\Sigma_\theta}$ to the estimation step.

Step 2. Stochastic approximation.

Estimate the gradient and Hessian of the complete data log-likelihood with respect to $\boldsymbol{b}$ by averaging gradients and Hessians across the m_r sets of complete data. Specifically, calculate

$$\tilde{\boldsymbol{s}}^{(r)} = \frac{1}{m_r}\sum_{s=1}^{m_r} \boldsymbol{s}(\boldsymbol{b}^{(r-1)}|\boldsymbol{Z}_s^{(r)}),$$

where $\boldsymbol{s}(\boldsymbol{b}^{(r-1)}|\boldsymbol{Z}_s^{(r)}) \approx \frac{\partial l(\boldsymbol{b}|\boldsymbol{Z})}{\partial \boldsymbol{b}}$ is a JK-by-1 vector with the $[(j-1)K+k]^{\text{th}}$ element computed by

$$s_{jk} = \sum_{i=1}^{N} \left(\frac{y_{ij}-p_{ij}}{1-p_{ij}}\right)\left(\frac{-1}{1+\exp(\theta_{ik}^{(r)} - b_{jk}^{(r-1)})}\right), \qquad \text{(B1)}$$

with $p_{ij} = (a-g)\prod_{k=1}^{K}\left(\frac{\exp(\theta_{ik}-b_{jk})}{1+\exp(\theta_{ik}-b_{jk})}\right)^{q_{jk}} + g.$

Next, estimate the $JK \times JK$ Hessian matrix of $\boldsymbol{b}$, $\boldsymbol{H}(\boldsymbol{b}|\boldsymbol{Z}) = -\frac{\partial^2 l(\boldsymbol{b}|\boldsymbol{Z})}{\partial \boldsymbol{b}\partial \boldsymbol{b}^t}$. Most IRT models assume local item independence, so that the partial derivatives with respect to parameters for two different items are all 0, and consequently, the Hessian matrix for these models takes block diagonal form. (Inverting a very sparse matrix is much easier than inverting a dense matrix.) Unfortunately, the a and g parameters from Embretson's MLTM model (Equation 2) are shared across *all* items, so that the rows and columns of $\boldsymbol{H}$ corresponding to a and g will all be non-zero. For the sake of estimation efficiency, assume that $a = 1$ and $g = 0$, so that $\boldsymbol{H}$ is block diagonal. Then the k^{th} diagonal term for the j^{th} item of the Hessian matrix would be

$$H_{j,kk} = \frac{d^2 l(\mathbf{b}|\mathbf{Z})}{db_{jk}^2} = \sum_{i=1}^{N}\left\{\left(\frac{y_{ij}-p_{ij}}{1-p_{ij}}\right)\left(\frac{\exp(\theta_{ik}-b_{jk})}{[1+\exp(\theta_{ik}-b_{jk})]^2}\right) - \left(\frac{p_{ij}y_{ij}-p_{ij}}{[1-p_{ij}]^2}\right)\left(\frac{1}{[1+\exp(\theta_{ik}-b_{jk})]^2}\right)\right\}, \qquad \text{(B2)}$$

and the k_1, k_2^{th} off-diagonal term for the j^{th} item of the Hessian matrix would be

$$H_{j,k_1k_2} = \frac{d^2 l(\mathbf{b}|\mathbf{Z})}{db_{jk_1}db_{jk_2}} = -\sum_{i=1}^{N}\left(\frac{p_{ij}y_{ij}-p_{ij}}{[1-p_{ij}]^2}\right)\left(\frac{1}{1+\exp(\theta_{ik_1}-b_{jk_1})}\right)\left(\frac{1}{1+\exp(\theta_{ik_2}-b_{jk_2})}\right). \qquad \text{(B3)}$$

As in estimating the gradient, the Hessian is also estimated using the complete data sets, $\boldsymbol{Z}_s^{(r)}$,

$s = (1, \ldots, m_r)$, and the proximate parameter estimates, $\boldsymbol{b}^{(r-1)}$. The estimated Hessian is then used to approximate the conditional expectation of the complete data information matrix by calculating

$$\boldsymbol{\Gamma}^{(r)} = \boldsymbol{\Gamma}^{(r-1)} + \gamma_r \left(\frac{1}{m_r} \sum_{s=1}^{m_r} [\boldsymbol{H}(\boldsymbol{b}^{(r-1)} | \boldsymbol{Z}_s^{(r+1)})] - \boldsymbol{\Gamma}^{(r-1)} \right),$$

where $\gamma_r \in (0, 1]$ is a small constant such that $\sum_{r=1}^{\infty} \gamma_r = \infty$ and $\sum_{r=1}^{\infty} \gamma_r^2 < \infty$ (Cai, 2010a). A final adjustment to each $\tilde{\boldsymbol{s}}_s$ and $\boldsymbol{H}_s$ must also be made due to the Q-matrix, which restricts many entries in $\boldsymbol{b}$ to be 0. Let $\boldsymbol{b}_c = \boldsymbol{L}\boldsymbol{b}$ denote the constrained item parameter matrix, where $\boldsymbol{L}$ is a $JK \times JK$ diagonal matrix with the j^{th} set of K diagonal elements representing the j^{th} row of the Q-matrix. Treating $\boldsymbol{b}_c$ as the target matrix of unknown parameters, one must modify the estimated gradient and Hessian matrices to be $\tilde{\boldsymbol{s}}_c = (\tilde{\boldsymbol{s}}^T \boldsymbol{L})^T$ and $\boldsymbol{H}_c = \boldsymbol{L}^T \boldsymbol{H} \boldsymbol{L}$, respectively, in all of the above equations.

Practitioners could additionally modify the gradient and Hessian by assuming a prior distribution on the unknown item parameters, $\boldsymbol{b}$. Imposing a prior distribution on the item parameters would allow a fairer comparison between the MH-RM and MCMC algorithms. Moreover, the empirical likelihood function used to estimate parameters in MH-RM is strangely shaped (see Figure 2), and a moderately informative prior distribution should improved the peakedness of the resulting posterior distribution and better allow the MH-RM algorithm to find the corresponding maximum. In particular, assume that $\boldsymbol{b}$ is multivariate normally distributed with mean vector $\boldsymbol{\mu_b}$ and covariance matrix $\boldsymbol{\Sigma_b}$. Then we would simply need to modify $s_{jk}^B = s_{jk} + (b_{jk}^{(r-1)} - \boldsymbol{\mu}_{bk}(\boldsymbol{\Sigma}_b^{-1})_{kk})$, $H_{j,kk}^B = H_{j,kk} + (\boldsymbol{\Sigma}_b^{-1})_{kk}$, and $H_{j,k_1k_2}^B = H_{j,k_1k_2} + (\boldsymbol{\Sigma}_b^{-1})_{k_1k_2}$. This modified, Bayesian version of MH-RM will be referred to as Bayesian MH-RM in any subsequent simulations and discussion.

Step 3. Robbins-Monro update

Refine the estimate of $\boldsymbol{b}$ with a Robbins-Monro update, where

$$\boldsymbol{b}_c^{(r)} = \boldsymbol{b}_c^{(r-1)} + \gamma_r (\boldsymbol{\Gamma}^{(r)})^{-1} \tilde{\boldsymbol{s}}_c^{(r)}.$$

Note that a common choice of γ_r is $1/r$, and then $\boldsymbol{\Gamma}^{(0)}$ can be arbitrarily picked. Convergence of the MH-RM algorithm is monitored by computing a window of successive differences, such as $\{\max|\boldsymbol{b}_c^{(r)} - \boldsymbol{b}_c^{(r-1)}|, \max|\boldsymbol{b}_c^{(r-1)} - \boldsymbol{b}_c^{(r-2)}|, \ldots, \max|\boldsymbol{b}_c^{(r-W-1)} - \boldsymbol{b}_c^{(r-W)}|\}$, where W is the predetermined window size (Cai, 2008, recommended setting $W = 3$). The algorithm terminates if and only if all differences in the window are below a pre-specified small number, such as .001. Once the estimates converge, the algorithm terminates and $\boldsymbol{b}$ is taken to be the final parameter estimates.

Appendix C: The Role of the Q-matrix

This appendix shows how the covariance matrix of $\boldsymbol{\theta}$ changes when Q-matrix displays a hierarchical structure (Rupp & Templin, 2008), that is, when each item loads either on the first dimension or on both dimensions. Tests display hierarchical structure when an item loading on one dimension implies that it also loads on the second dimension. If a test manifests hierarchical structure, then $t_3 = 0$, and the information matrix reduces to

$$\mathcal{I}^*(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j1})^2}{1-p_{j1}p_{j2}} + \sum_{j=t_1+1}^{t_1+t_2} p^*_{j1}(1-p^*_{j1}) & \sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j1})(1-p_{j2})}{1-p_{j1}p_{j2}} \\ \sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j1})(1-p_{j2})}{1-p_{j1}p_{j2}} & \sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j2})^2}{1-p_{j1}p_{j2}} \end{pmatrix}.$$

Now assume that $p^*_{j1} \approx p_{j1}p_{j2}$, so that items that measure both dimensions have roughly the same item response function as items that measure a single dimension. Then we have $p^*_{j1}(1-p^*_{j1}) \approx p_{j1}p_{j2}(1-p_{j1}p_{j2}) = \frac{p_{j1}p_{j2}(1-p_{j1}p_{j2})^2}{1-p_{j1}p_{j2}} \geq \frac{p_{j1}p_{j2}(1-p_{j1})^2}{1-p_{j1}p_{j2}}$. As a consequence of that simple inequality, it can be shown that when

$$\frac{\left[\sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j1})(1-p_{j2})}{1-p_{j1}p_{j2}}\right]\left[\sum_{j=t_1+1}^{t_1+t_2} \frac{p_{j1}p_{j2}(1-p_{j2})^2}{1-p_{j1}p_{j2}}\right]}{\left[\sum_{j=t_1+1}^{t_1+t_2} \frac{p_{j1}p_{j2}(1-p_{j1})(1-p_{j2})}{1-p_{j1}p_{j2}}\right]\left[\sum_{j=1}^{t_1} \frac{p_{j1}p_{j2}(1-p_{j2})^2}{1-p_{j1}p_{j2}}\right]} \geq 0.5 \tag{C1}$$

then $[\mathcal{I}^{-1}(\boldsymbol{\theta})]_{11} \leq [\mathcal{I}^{*-1}(\boldsymbol{\theta})]_{11}$. Therefore, replacing t_2 items that load on both dimensions with items that only on the first dimension results in *increased* estimation precision of θ_1. Notice that the inequality in Equation (C1) is sufficient but not necessary for the variance of $\hat{\theta}_1$ to decrease. Moreover, Equation (C1) is only violated when the items corresponding to t_1 are very easy on the first dimension and the items corresponding to t_2 are very difficult on the first dimension. But with a well-balanced item bank, inequality (C1) will normally be satisfied.

Appendix D: Why correlation between ability dimensions (ρ) matters?

An oft-documented benefit of using multidimensional models is that one obtains greater precision in estimating dimensional ability by "borrowing strength" from other dimensions (i.e., de la Torre & Patz, 2005; Wang, Chen, & Cheng, 2004). Moreover, as is clearly shown in de la Torre and Patz (2005), a higher correlation among the dimensions results in *improved* estimation accuracy. But whereas moderate-to-large correlations between ability dimensions are helpful in $\boldsymbol{\theta}$ estimation, they impede item parameter estimation when using a non-compensatory MIRT model (Bolt & Lall, 2003, Babcock, 2012) such as the MLTM discussed in this paper. One might wonder why high ability correlations hinder estimating item parameters. A primary reason for the resulting difficulties in item parameter estimation can (of course) be traced back to the Fisher information matrix. The volume of the confidence ellipsoid of $\hat{\boldsymbol{b}}_j$ is proportional to the determinant of the item parameter Fisher information matrix (as is the case in most applications of maximum likelihood). Therefore, we conducted a small simulation to demonstrate the change in estimation precision for different values of ρ. To do this, we generated sets of $N = 2000$ examinees from bivariate normal distributions with varying correlations (ρ) between the dimensions. For each set of examinees, we calculated the determinant of the item parameter Fisher information matrix for three two-dimensional items and plotted the results in Figure C1.

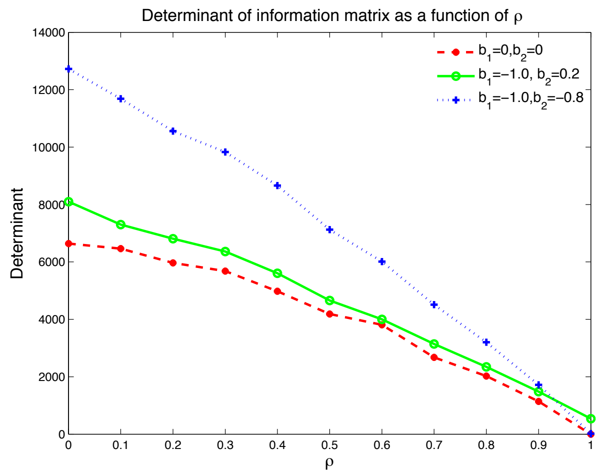

Figure C1: Determinant of the item information matrix as a function of ρ

As is obvious from Figure C1, the determinant of the item parameter Fisher information matrix monotonically decreases as the correlation between the dimensions is increased regardless of the actual item parameter values. For all three items, a larger correlation between the dimensions results in a flatter likelihood surface and a more difficult item parameter estimation problem. Intuitively, if an item measures multiple dimensions, then large correlations between those dimensions implies nearly indistinguishable item parameters. This problem is identical to the well-known headache of near-multicolinearity in estimating parameters of the linear regression model.

Appendix E: Additional Results

Table E1: An illustration of a complex Q-matrix structure

item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
d1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	1	1	1	0	0
d2	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
d3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
sum	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	2	2	2	3	2	2	3	2	3	3	3	2	2

Table E2: Bias, MSE, Correlation, computation time (in seconds), and number of iterations (for both MH-RM algorithm and Bayesian MH-RM (denoted as B MH-RM) algorithm) for the item parameter b_{jk} estimates given a $J = 30$ item test.

					bias			MSE			Correlation			time	n.of iter
					b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3		
$J = 30$	Simple Q	$\rho = 0.2$	$N = 1000$	MCMC	-0.01	0.01	0.00	0.01	0.01	0.01	1.00	1.00	1.00	492.02	
				MH-RM	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	79.50	84
				B MH-RM	0.00	0.00	0.01	0.01	0.01	0.01	1.00	1.00	1.00	80.83	84
			$N = 2000$	MCMC	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	966.64	
				MH-RM	-0.01	0.00	-0.01	0.00	0.01	0.00	1.00	1.00	1.00	163.95	94
				B MH-RM	0.00	0.00	-0.01	0.00	0.01	0.01	1.00	1.00	1.00	161.49	94
		$\rho = 0.5$	$N = 1000$	MCMC	0.01	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	491.95	
				MH-RM	0.02	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	89.05	60
				B MH-RM	0.01	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	91.08	62
			$N = 2000$	MCMC	0.00	-0.01	-0.01	0.00	0.00	0.00	1.00	1.00	1.00	965.69	
				MH-RM	0.00	0.00	-0.01	0.00	0.00	0.00	1.00	1.00	1.00	179.50	66
				B MH-RM	0.00	-0.01	0.00	0.00	0.00	0.00	1.00	1.00	1.00	180.77	66
		$\rho = 0.75$	$N = 1000$	MCMC	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	491.98	
				MH-RM	-0.01	-0.01	0.00	0.01	0.01	0.01	1.00	1.00	1.00	97.03	48
				B MH-RM	-0.01	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	87.21	44
			$N = 2000$	MCMC	-0.01	0.00	-0.01	0.00	0.00	0.00	1.00	1.00	1.00	965.70	
				MH-RM	-0.01	0.00	-0.01	0.00	0.00	0.00	1.00	1.00	1.00	179.80	48
				B MH-RM	-0.01	0.00	-0.01	0.00	0.00	0.00	1.00	1.00	1.00	177.71	47
	Complex Q	$\rho = 0.2$	$N = 1000$	MCMC	-0.05	-0.08	-0.07	0.12	0.10	0.13	0.96	0.97	0.97	552.77	
				MH-RM	-0.12	-0.12	-0.15	0.22	0.13	0.18	0.94	0.96	0.96	557.12	550
				B MH-RM	-0.14	-0.15	-0.16	0.18	0.17	0.19	0.95	0.96	0.95	440.81	436
			$N = 2000$	MCMC	-0.04	-0.01	-0.04	0.06	0.09	0.08	0.98	0.97	0.98	1087.74	
				MH-RM	-0.13	-0.13	-0.16	0.19	0.20	0.17	0.96	0.95	0.96	891.39	476
				B MH-RM	-0.14	-0.14	-0.16	0.20	0.22	0.20	0.96	0.95	0.96	812.32	442
		$\rho = 0.5$	$N = 1000$	MCMC	-0.08	-0.09	-0.06	0.22	0.23	0.18	0.95	0.94	0.94	551.44	
				MH-RM	-0.12	-0.14	-0.11	0.33	0.29	0.19	0.92	0.92	0.93	913.78	571
				B MH-RM	-0.20	-0.17	-0.14	0.36	0.33	0.25	0.92	0.92	0.91	752.88	468
			$N = 2000$	MCMC	-0.07	-0.09	-0.06	0.14	0.13	0.11	0.96	0.97	0.97	1088.72	
				MH-RM	-0.15	-0.18	-0.16	0.24	0.26	0.25	0.94	0.94	0.95	1439.10	482
				B MH-RM	-0.18	-0.20	-0.17	0.29	0.31	0.28	0.92	0.93	0.93	1305.82	440
		$\rho = 0.75$	$N = 1000$	MCMC	-0.09	-0.08	-0.16	0.32	0.34	0.43	0.91	0.91	0.90	554.18	
				MH-RM	-0.12	-0.13	-0.21	0.41	0.44	0.51	0.89	0.87	0.85	1497.51	682
				B MH-RM	-0.14	-0.14	-0.26	0.39	0.40	0.56	0.89	0.89	0.85	1164.89	531
			$N = 2000$	MCMC	-0.03	-0.11	-0.02	0.21	0.25	0.22	0.94	0.93	0.93	1091.40	
				MH-RM	-0.11	-0.17	-0.17	0.32	0.45	0.44	0.89	0.87	0.89	2274.41	555
				B MH-RM	-0.14	-0.19	-0.18	0.35	0.39	0.42	0.90	0.87	0.88	2123.96	514

Table E3: Bias, MSE, Correlation, computation time (in seconds), and number of iterations (for both MH-RM algorithm and Bayesian MH-RM (denoted as B MH-RM) algorithm) for the item parameter b_{jk} estimates given a $J = 45$ item test.

					bias			MSE			Correlation			time	n.of iter
					b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3		
$J = 45$	Simple Q	$\rho = 0.2$	$N = 1000$	MCMC	0.00	0.01	0.00	0.01	0.01	0.01	1.00	1.00	1.00	717.87	
				MH-RM	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	108.22	78
				B MH-RM	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	106.09	76
			$N = 2000$	MCMC	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1420.63	
				MH-RM	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	210.95	87
				B MH-RM	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	205.53	85
		$\rho = 0.5$	$N = 1000$	MCMC	0.01	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	718.08	
				MH-RM	0.01	0.00	0.01	0.01	0.01	0.01	1.00	1.00	1.00	115.78	55
				B MH-RM	0.01	0.00	0.01	0.01	0.01	0.01	1.00	1.00	1.00	119.23	56
			$N = 2000$	MCMC	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1419.71	
				MH-RM	0.00	0.01	0.00	0.00	0.00	0.00	1.00	1.00	1.00	236.57	60
				B MH-RM	0.00	0.01	0.00	0.00	0.00	0.00	1.00	1.00	1.00	236.00	60
		$\rho = 0.75$	$N = 1000$	MCMC	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	717.28	
				MH-RM	0.01	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	122.24	42
				B MH-RM	0.00	0.00	0.00	0.01	0.01	0.01	1.00	1.00	1.00	118.53	41
			$N = 2000$	MCMC	0.01	0.01	0.01	0.00	0.00	0.00	1.00	1.00	1.00	1420.27	
				MH-RM	0.01	0.01	0.01	0.00	0.00	0.00	1.00	1.00	1.00	252.04	47
				B MH-RM	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	245.56	46
	Complex Q	$\rho = 0.2$	$N = 1000$	MCMC	-0.09	-0.03	-0.06	0.13	0.12	0.12	0.96	0.97	0.97	815.87	
				MH-RM	-0.15	-0.10	-0.14	0.21	0.16	0.17	0.95	0.95	0.96	915.15	611
				B MH-RM	-0.20	-0.13	-0.17	0.23	0.19	0.19	0.95	0.95	0.95	705.39	471
			$N = 2000$	MCMC	-0.03	-0.04	-0.04	0.06	0.08	0.06	0.98	0.98	0.98	1618.52	
				MH-RM	-0.16	-0.12	-0.12	0.13	0.14	0.13	0.97	0.98	0.97	1397.23	527
				B MH-RM	-0.18	-0.16	-0.16	0.14	0.18	0.15	0.97	0.97	0.97	1225.45	463
		$\rho = 0.5$	$N = 1000$	MCMC	-0.07	-0.09	-0.11	0.17	0.18	0.19	0.96	0.95	0.95	815.59	
				MH-RM	-0.14	-0.12	-0.16	0.20	0.21	0.26	0.95	0.94	0.94	1526.35	654
				B MH-RM	-0.18	-0.18	-0.22	0.25	0.25	0.31	0.94	0.94	0.93	1203.88	515
			$N = 2000$	MCMC	-0.05	-0.05	-0.05	0.08	0.10	0.10	0.97	0.97	0.97	1612.69	
				MH-RM	-0.13	-0.12	-0.16	0.17	0.17	0.16	0.96	0.96	0.95	2355.77	547
				B MH-RM	-0.16	-0.16	-0.15	0.18	0.18	0.20	0.95	0.95	0.94	2028.03	469
		$\rho = 0.75$	$N = 1000$	MCMC	-0.11	-0.15	-0.11	0.34	0.41	0.34	0.91	0.89	0.90	816.04	
				MH-RM	-0.17	-0.16	-0.12	0.39	0.47	0.39	0.88	0.87	0.88	2551.96	799
				B MH-RM	-0.21	-0.23	-0.19	0.46	0.56	0.42	0.88	0.87	0.88	2000.09	626
			$N = 2000$	MCMC	-0.08	-0.08	-0.09	0.21	0.21	0.21	0.94	0.94	0.93	1620.53	
				MH-RM	-0.14	-0.15	-0.12	0.38	0.41	0.37	0.90	0.91	0.88	3992.68	668
				B MH-RM	-0.14	-0.19	-0.18	0.38	0.42	0.41	0.91	0.90	0.89	3361.92	565

Table E4: Standard error of parameter estimates

				MCMC			MH-RM			B MH-RM		
				b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
$J = 30$	Simple Q	$\rho = 0.2$	$N = 1000$	0.076	0.075	0.078	0.141	0.161	0.144	0.140	0.159	0.148
			$N = 2000$	0.055	0.059	0.058	0.096	0.126	0.105	0.101	0.125	0.106
		$\rho = 0.5$	$N = 1000$	0.087	0.078	0.081	0.210	0.183	0.177	0.201	0.180	0.172
			$N = 2000$	0.058	0.059	0.055	0.149	0.126	0.154	0.143	0.127	0.154
		$\rho = 0.75$	$N = 1000$	0.083	0.082	0.078	0.267	0.259	0.303	0.277	0.252	0.278
			$N = 2000$	0.061	0.060	0.061	0.216	0.202	0.203	0.187	0.202	0.178
	Complex Q	$\rho = 0.2$	$N = 1000$	0.277	0.161	0.242	0.272	0.188	0.285	0.276	0.168	0.267
			$N = 2000$	0.158	0.200	0.175	0.189	0.213	0.191	0.191	0.202	0.189
		$\rho = 0.5$	$N = 1000$	0.326	0.326	0.305	0.297	0.309	0.276	0.319	0.310	0.246
			$N = 2000$	0.185	0.186	0.180	0.267	0.222	0.234	0.261	0.214	0.238
		$\rho = 0.75$	$N = 1000$	0.432	0.393	0.462	0.471	0.379	0.365	0.443	0.372	0.452
			$N = 2000$	0.397	0.314	0.395	0.272	0.466	0.578	0.525	0.477	0.499
$J = 45$	Simple Q	$\rho = 0.2$	$N = 1000$	0.081	0.081	0.079	0.146	0.150	0.166	0.145	0.145	0.165
			$N = 2000$	0.059	0.058	0.059	0.105	0.094	0.105	0.103	0.099	0.103
		$\rho = 0.5$	$N = 1000$	0.085	0.078	0.088	0.179	0.167	0.204	0.176	0.171	0.200
			$N = 2000$	0.054	0.057	0.058	0.133	0.124	0.143	0.129	0.122	0.137
		$\rho = 0.75$	$N = 1000$	0.083	0.081	0.078	0.217	0.270	0.220	0.217	0.251	0.223
			$N = 2000$	0.059	0.057	0.056	0.207	0.204	0.189	0.192	0.200	0.169
	Complex Q	$\rho = 0.2$	$N = 1000$	0.252	0.258	0.296	0.235	0.229	0.281	0.222	0.228	0.261
			$N = 2000$	0.186	0.207	0.188	0.186	0.225	0.194	0.176	0.221	0.192
		$\rho = 0.5$	$N = 1000$	0.321	0.205	0.250	0.306	0.238	0.305	0.280	0.223	0.288
			$N = 2000$	0.161	0.206	0.237	0.234	0.196	0.251	0.233	0.182	0.242
		$\rho = 0.75$	$N = 1000$	0.348	0.379	0.335	0.400	0.380	0.332	0.319	0.377	0.317
			$N = 2000$	0.296	0.293	0.311	0.386	0.384	0.461	0.384	0.326	0.379