Web Appendix

Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation

MARTIJN G. DE JONG, JAN-BENEDICT E.M. STEENKAMP, JEAN-PAUL FOX, and HANS BAUMGARTNER*

*Martijn G. de Jong is Assistant Professor of Marketing, Erasmus University (e-mail: MJong@rsm.nl), Jan-Benedict E.M. Steenkamp is C. Knox Massey Distinguished Professor of Marketing and Marketing Area Chair, University of North Carolina at Chapel Hill, NC 27599-3490 (e-mail: JBS@unc.edu). Jean-Paul Fox is Assistant Professor at the University of Twente, Department of Research Methodology, Measurement and Data Analysis (e-mail: G.J.A.Fox@edte.utwente.nl). Hans Baumgartner is Professor of Marketing and Charles & Lillian Binder Faculty Fellow, Smeal College of Business, Pennsylvania State University (e-mail: jxb14@psu.edu). The article is based on the first author's doctoral dissertation. The authors thank the editors, Dick Wittink, Russel Winer, and Joel Huber, and the anonymous reviewers for their extremely useful and constructive comments. In addition, they gratefully acknowledge AiMark for providing the data.

In this Web appendix, we provide estimation details for the hierarchical IRT model.

ESTIMATION DETAILS

This section presents the full MCMC algorithm for the multilevel testlet IRT model with varying item parameters. The model without covariates is a special case of the model that is developed here. Let the observed data be (**EXTR**, **X**, **W**) measuring the item responses **EXTR** for the latent trait ERS, **X** the individual-level explanatory variables, and **W** the country-level variables. The Gibbs sampler draws stepwise from the full conditional distributions. The first step is to augment the observed data with latent data **Z**. By defining a continuous latent variable **Z** that underlies the dichotomous responses contained in **EXTR,** it is easier to sample from the conditional distributions of the parameters of interest. Data augmentation has been widely applied. To identify the model, we use the restriction $\prod_k a_{kj} = 1,$ and $\sum_k b_{kj} = 0,$ in each country *j*.

1) Sample from $[Z_{ijk} \mid EXTR_{ijk}, ERS_{ij}, a_{kj}, b_{kj}, \psi_{ij,r_k}]$, for $k = 1,\ldots,K,\ i = 1,\ldots,n_j,$ and $j = 1,\ldots,J.$ Given the parameters $ERS_{ij}, a_{kj}, b_{kj}, \psi_{ij,r_k}$ the variables Z_{ijk} are independent and normally distributed, that is,

$$Z_{ijk} \Big| EXTR_{ijk}, ERS_{ij}, a_{kj}, b_{kj}, \psi_{ij,r_k} \sim$$
$$\begin{cases} N\left(a_{kj}\left(ERS_{ij} - \psi_{ij,r_k}\right) - b_{kj}, 1\right) \text{ truncated right by 0 if } EXTR_{ijk} = 0. \\ N\left(a_{kj}\left(ERS_{ij} - \psi_{ij,r_k}\right) - b_{kj}, 1\right) \text{ truncated left by 0 if } EXTR_{ijk} = 1. \end{cases}$$

2) Sample from $\left[ERS_{ij} \mid \mathbf{Z}_{ij}, \mathbf{a}, \mathbf{b}, \mathbf{EXTR}, \boldsymbol{\psi}, \boldsymbol{\beta}_j, \sigma^2\right]$ for $i = 1,\ldots,n_j,\ j = 1,\ldots,J.$ This full conditional distribution is a product of two normal distributions and from standard properties of normal distributions it follows that

$$ERS_{ij} \mid \mathbf{Z}_{ij}, \mathbf{a}, \mathbf{b}, \boldsymbol{\psi}, \boldsymbol{\beta}_j, \sigma^2 \sim N\left(\frac{\sum_{k=1}^{K} a_{kj}(Z_{ijk} + b_{kj} + a_{kj}\psi_{ij,r_k}) + \mathbf{X}_{ij}\boldsymbol{\beta}_j / \sigma^2}{1/\sigma^2 + \sum_{k=1}^{K} a_{kj}^2}, \frac{1}{1/\sigma^2 + \sum_{k=1}^{K} a_{kj}^2} \right)$$

3) Sample from $\left[\psi_{ij,r} \mid \mathbf{Z}_{ij}, \mathbf{a}, \mathbf{b}, \sigma^2_{\psi_r}\right]$, for $i = 1,\ldots,n_j,\ j = 1,\ldots,J$ and $r = 1,\ldots,R$ where R is the total number of testlets. Consider testlet r of size N_r, and let r_k denote the testlet of item k. If N_r=1 then $\psi_{ij,r} = 0$ for all i and j. The full conditional distribution for $\psi_{ij,r}$ is normal with parameters

$$E\left(\psi_{ij,r} \mid \mathbf{Z}_{ij}, \mathbf{a}, \mathbf{b}, \sigma^2_{\psi_r}\right) = \frac{\sum_{\{k:r_k=r\}} a_{kj}\left(a_{kj}ERS_{ij} - b_{kj} - Z_{ijk}\right)}{\sum_{\{k:r_k=r\}} a^2_{kj} + 1/\sigma^2_{\psi_r}}$$

$$\mathrm{Var}\left(\psi_{ij,r} \mid \mathbf{Z}_{ij}, \mathbf{a}, \mathbf{b}, \sigma^2_{\psi_r}\right) = \frac{1}{\sum_{\{k:r_k=r\}} a^2_{kj} + 1/\sigma^2_{\psi_r}},$$

4) Sampling $[\boldsymbol{\xi}_{kj} \mid \tilde{\boldsymbol{\xi}}_k, \mathbf{ERS}^*, \boldsymbol{\psi}, \sigma^2_a, \sigma^2_b]$

Consider the augmented likelihood. Then,

$$\boldsymbol{Z}_{kj} = \mathbf{H}_j\boldsymbol{\xi}_{kj} + \boldsymbol{\varepsilon}_{kj},\ \mathbf{H}_j = [\mathbf{ERS}^*_j, -\mathbf{1}]$$

$$\boldsymbol{\xi}_{kj} \sim N\left(\tilde{\boldsymbol{\xi}}_k, \boldsymbol{\Psi}\right) = N\left(\begin{bmatrix}\tilde{a}_k \\ \tilde{b}_k\end{bmatrix}, \begin{bmatrix}\sigma^2_a & 0 \\ 0 & \sigma^2_b\end{bmatrix}\right)$$

$$\boldsymbol{\xi}_{kj} \mid \tilde{\boldsymbol{\xi}}_k, \mathbf{Z}_{kj}, \mathbf{ERS}^*_j, \boldsymbol{\Psi} \sim N\left(\boldsymbol{\Omega}\boldsymbol{\mu}_{kj}, \boldsymbol{\Omega}\right) I(a_{kj} \in A)$$

$$\boldsymbol{\mu}_{kj} = \mathbf{H}^t_j\mathbf{Z}_{kj} + \boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{\xi}}_k$$

$$\boldsymbol{\Omega}^{-1} = \left(\mathbf{H}^t_j\mathbf{H}_j\right)^{-1} + \boldsymbol{\Psi}^{-1}.$$

5) Sample from $[\tilde{\boldsymbol{\xi}}_k \mid \boldsymbol{\xi}_k, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b}]$.

Given the prior $\begin{bmatrix}\tilde{a}_k \\ \tilde{b}_k\end{bmatrix} \sim N\left(\begin{bmatrix}\mu_{\tilde{a}} \\ \mu_{\tilde{b}}\end{bmatrix}, \begin{bmatrix}\sigma^2_{\tilde{a}} & \sigma^2_{\tilde{a}\tilde{b}} \\ \sigma^2_{\tilde{a}\tilde{b}} & \sigma^2_{\tilde{b}}\end{bmatrix}\right) = N\left(\boldsymbol{\mu}_{\tilde{\xi}}, \boldsymbol{\Sigma}\right) I\left(\tilde{a}_k \in A\right)$,

we have that

$$\tilde{\boldsymbol{\xi}}_k \left| \boldsymbol{\xi}_k, \boldsymbol{\Sigma} \sim N\left(\Xi_k\boldsymbol{\zeta}_k, \Xi_k\right) I\left(\tilde{a}_k \in A\right)\right.$$

$$\boldsymbol{\zeta}_k = \boldsymbol{\Lambda}^{-1}\bar{\boldsymbol{\xi}}_k + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\tilde{\xi}} \qquad \text{where } \bar{\boldsymbol{\xi}}_k = \begin{bmatrix}\sum_j^J a_{kj}/J \\ \sum_j^J b_{kj}/J\end{bmatrix}$$

$$\Xi^{-1}_k = \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Sigma}^{-1}$$

6) Sample from $[\sigma^2_a \mid \mathbf{a}, \tilde{\mathbf{a}}]$, $[\sigma^2_b \mid \mathbf{b}, \tilde{\mathbf{b}}]$, $[\sigma^2_{\psi_r}]$. For each variance parameter an inverse gamma prior is specified with parameters g_1 and g_2. As a result, the full conditionals are inverse gamma distributions with shape parameters $KJ/2 + g_1$, $KJ/2 + g_1$, and $N/2 + g_1$, respectively, and scale parameters

$$g_2 + \sum_{j=1}^J\sum_{k=1}^K\left(a_{kj} - \tilde{a}_k\right)^2/2,\ g_2 + \sum_{j=1}^J\sum_{k=1}^K\left(b_{kj} - \tilde{b}_k\right)^2/2,\quad g_2 + \frac{1}{2}\sum_{j=1}^J\sum_{i=1}^{n_i}\psi^2_{ij,r_k} \text{ respectively.}$$

In this article, we specified a noninformative proper prior with $g_1 = g_2 = 1$.

7) The conditional distribution of the covariance matrix $\boldsymbol{\Sigma}$ has an inverse-Wishart distribution,

$$\boldsymbol{\Sigma} \mid \tilde{\boldsymbol{\xi}} \sim \mathrm{Inv-W}\left(K + n_0, \left(\sum_{k=1}^{K} (\tilde{\boldsymbol{\xi}}_k - \mu_{\tilde{\xi}})(\tilde{\boldsymbol{\xi}}_k - \mu_{\tilde{\xi}})^t + \mathbf{S} \right)^{-1} \right).$$ We used **S**=diag(100,100), n_0=2.

8) Sample from $\left[\boldsymbol{\beta}_j \mid \mathbf{ERS}_j, \sigma^2, \boldsymbol{\gamma}, \mathbf{T}\right]$, for $j = 1, \ldots, J$. Define $\mathbf{X}_j = \left(\mathbf{X}_{1j}, \ldots, \mathbf{X}_{ij}, \ldots, \mathbf{X}_{n_j j}\right)$, with $\mathbf{X}_{ij} = \left(X_{0ij}, ..., X_{Qij}\right)'$. So, $\mathbf{X}_j$ is of dimension $n_j \times (Q+1)$. The Level 2 explanatory variables for goup j are stored in $\mathbf{W}_j$. This matrix is the direct product of $\left(W_{0qj}, \ldots, W_{Sqj}\right)$ and a Q+1 identity matrix. The prior of the random regression coefficients at Level 1 is $\boldsymbol{\beta}_j \sim N\left(\mathbf{W}_j \boldsymbol{\gamma}, \mathbf{T}\right)$. Then, the full conditional posterior of $\boldsymbol{\beta}_j$ is given by,

$$\boldsymbol{\beta}_j \mid \mathbf{ERS}_j, \sigma^2, \boldsymbol{\gamma}, \mathbf{T} \sim N\left(\left(\boldsymbol{\Upsilon}_j^{-1} + \mathbf{T}^{-1}\right)^{-1} \left(\boldsymbol{\Upsilon}_j^{-1} \hat{\boldsymbol{\beta}}_j + \mathbf{T}^{-1} \mathbf{W}_j \boldsymbol{\gamma}\right), \left(\boldsymbol{\Upsilon}_j^{-1} + \mathbf{T}^{-1}\right)^{-1} \right)$$

where $\hat{\boldsymbol{\beta}}_j = \left(\mathbf{X}_j' \mathbf{X}_j\right)^{-1} \mathbf{X}_j' \mathbf{ERS}_j$ and $\boldsymbol{\Upsilon}_j = \sigma^2 \left(\mathbf{X}_j' \mathbf{X}_j\right)^{-1}$. In case of fixed Level 1 regression coefficients, a noninformative prior is used, and the resulting conditional distribution is also normal with mean $\hat{\beta}_F = \left(\mathbf{X}_F' \mathbf{X}_F\right)^{-1} \mathbf{X}_F' \left(\mathbf{ERS} - \mathbf{X}_R \boldsymbol{\beta}_R\right)$ and variance $\sigma^2 \left(\mathbf{X}_F' \mathbf{X}_F\right)^{-1}$, where $\mathbf{X}_F$ and $\mathbf{X}_R$ are the explanatory variables related to the fixed and random regression coefficents, respectively.

9) Sample from $\left[\boldsymbol{\gamma} \mid \boldsymbol{\beta}, \mathbf{T}\right]$. The full conditional for the Level 2 regression coefficients $\boldsymbol{\gamma}$ with a noninformative prior equals

$$\boldsymbol{\gamma} \mid \boldsymbol{\beta}, \mathbf{T} \sim N\left(\left(\sum_{j=1}^{J} \mathbf{W}_j' \mathbf{T}^{-1} \mathbf{W}_j \right)^{-1} \sum_{j=1}^{J} \mathbf{W}_j' \mathbf{T}^{-1} \boldsymbol{\beta}_j, \left(\sum_{j=1}^{J} \mathbf{W}_j' \mathbf{T}^{-1} \mathbf{W}_j \right)^{-1} \right)$$

10) Sample from $\left[\sigma^2 \mid \mathbf{ERS}, \boldsymbol{\beta}\right]$. A prior for the variance is specified in the form of an inverse gamma distribution with shape and scale parameters g_1 and g_2, respectively. It follows that

$$\sigma^2 \mid \mathbf{ERS}, \boldsymbol{\beta} \sim IG\left(N/2 + g_1, \frac{N}{2} \sum_j \left(\mathbf{ERS}_j - \mathbf{X}_j \boldsymbol{\beta}_j\right)^2 + g_2 \right).$$

A noninformative but proper prior is specified with $g_1 = g_2 = 1$.

11) Sample from $\left[\mathbf{T} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}\right]$. An inverse-Wishart distribution with small degrees of freedom, but greater than the dimension of $\boldsymbol{\beta}_j$, n_0 and unity matrix $\mathbf{S}_0$ can be used to specify a diffuse proper prior for **T**. It follows that

$$\mathbf{T} \mid \boldsymbol{\beta}, \boldsymbol{\gamma} \sim \mathrm{Inv-W}\left(J + n_0, \left(\sum_j \left(\boldsymbol{\beta}_j - \mathbf{W}_j \boldsymbol{\gamma}\right)\left(\boldsymbol{\beta}_j - \mathbf{W}_j \boldsymbol{\gamma}\right)' + \mathbf{S}_0 \right)^{-1} \right).$$