Online Appendix for "Benefit-Based Conjoint Analysis"

Dong Soo Kim, Roger A. Bailey, Nino Hardt, and Greg M. Allenby

1 The proof of the conditions for the monotonic and subadditive part-worth function

The proof for monotonicity is straightforward. Since y is monotonically increasing with respect to a_1, a_2, $\cdots$, a_M, b is monotonic if v is monotonically increasing. The proof for subadditivity is as follows: First, we show the first inequality in equation (10) in the main text. Since v is concave for $y \geq 0$ and $v(0) = 0$,

$$v(t \cdot y) \geq t \cdot v(y), \quad \text{for all } t \in [0,1] \text{ and } y \geq 0.$$

This is a well known property of concave functions. Since $\boldsymbol{a}_{B^+} \geq \mathbf{0}$, $v(y(\boldsymbol{a}_{B^+})) \geq 0$ and

$$|b(\boldsymbol{a}_{B^+})| = v\left(\sum_{l=1}^{L^+} \sum_{m \in S_l^+} a_m\right).$$

Let $a_l^* = \sum_{m \in S_l^+} a_m$. Then we obtain the first inequality in equation (10) in the main text as follows:

$$\begin{aligned}
|b(\boldsymbol{a}_{B^+})| = v\left(\sum_{l=1}^{L^+} a_l^*\right) &= \left(\frac{a_1^*}{\sum_{l=1}^{L^+} a_l^*} + \cdots + \frac{a_{L^+}^*}{\sum_{l=1}^{L^+} a_l^*}\right) \cdot v\left(\sum_{l=1}^{L^+} a_l^*\right) \\
&\leq v\left(\frac{a_1^*}{\sum_{l=1}^{L^+} a_l^*} \cdot \sum_{l=1}^{L^+} a_l^*\right) + \cdots + v\left(\frac{a_{L^+}^*}{\sum_{l=1}^{L^+} a_l^*} \cdot \sum_{l=1}^{L^+} a_l^*\right) \\
&= v(a_1^*) + \cdots + v(a_{L^+}^*) \\
&= \sum_{l=1}^{L^+} v\left(\sum_{m \in S_l^+} a_m\right) = \sum_{l=1}^{L^+} \left|b\left(\boldsymbol{a}_{S_l^+}\right)\right|.
\end{aligned}$$

Next, we show the second inequality in equation (10) in the main text. Since $v(y)$ is convex for $y<0$, $|v(-|y|)|$ is concave with respect to $|y|$ for $y<0$ and its value is zero at $y=0$. Therefore,

$$|v(-t\cdot|y|)| \geq t\cdot|v(-|y|)|, \quad \text{for all } t\in[0,1] \text{ and } y<0.$$

Since $\boldsymbol{a}_{B^-}\leq\mathbf{0}$,

$$|b(\boldsymbol{a}_{B^-})| = \left| v\left(-\sum_{l=1}^{L^-}\left|\sum_{m\in S_l^-} a_m\right| \right)\right|.$$

Let $a_l^{**}=\sum_{m\in S_l^-} a_m$. Then we obtain the second inequality in equation (10) in the main text as follows:

$$\begin{aligned}
|b(\boldsymbol{a}_{B^-})| = \left| v\left(-\sum_{l=1}^{L^-}|a_l^{**}|\right)\right| &= \left(\frac{|a_1^{**}|}{\sum_{l=1}^{L^-}|a_l^{**}|}+\cdots+\frac{|a_{L^-}^{**}|}{\sum_{l=1}^{L^-}|a_l^{**}|}\right)\cdot\left|v\left(-\sum_{l=1}^{L^-}|a_l^{**}|\right)\right| \\
&\leq \left|v\left(-\frac{|a_1^{**}|}{\sum_{l=1}^{L^-}|a_l^{**}|}\cdot\sum_{l=1}^{L^-}|a_l^{**}|\right)\right|+\cdots+\left|v\left(-\frac{|a_{L^-}^{**}|}{\sum_{l=1}^{L^-}|a_l^{**}|}\cdot\sum_{l=1}^{L^-}|a_l^{**}|\right)\right| \\
&= |v(-|a_1^{**}|)|+\cdots+|v(-|a_{L^-}^{**}|)| \\
&= \sum_{l=1}^{L^-}\left|v\left(-\left|\sum_{m\in S_l^-}a_m\right|\right)\right| = \sum_{l=1}^{L^-}\left|b\left(\boldsymbol{a}_{S_l^-}\right)\right|.
\end{aligned}$$

2 Simulation study

A simulation study is conducted to verify the ability to recover model parameters including the assignment probabilities under a typical condition in practice. We use the conjoint experiment design of our second empirical application to generate synthetic choice data: three inside options and an outside option with 18 variables (17 binary variables of 10 different attributes and a continuous price variable). The model specification for generating data is the same as equation (25) in the main text with a simple logarithmic form, $f(y)=\log(y+1)$. The true number of benefits is set to two, i.e., $K=2$, identical to the number of revealed benefits in the empirical analysis. With this setup, we generated two different datasets: one (large dataset) as three times great as the empirical data, 843 individuals and 60 observations per individual, and the other (small dataset) as the same as our empirical data, 281 individuals and 20 observations per individual.

We fit both datasets to the proposed model with various values of K, from zero (identical to the standard MNL model as in equation (26) in the main text) to three. For each value of K, we ran 20,000 MCMC iterations and found that the likelihood value converges after the first 5,000 iterations. The first 10,000 draws are dropped as a burn-in period and the last 10,000 draws are used for calculating statistics and estimating the model parameters. Details of the estimation procedure are provided in section 3 in this online appendix.

Table 1 shows the log-marginal density (LMD) estimates. As in the main text, we employ two methods to approximate the LMD estimate: (i) Newton and Raftery's (1994) harmonic mean approximation (NR) and (ii) Gelfand and Dey's (1994) approximation (GD). The NR method has been widely used because of its simplicity for calculation. However, Lenk (2009) pointed out that the NR approximation possibly overestimates the marginal density and reported that GD approximation is significantly accurate. We thus calculated both approximations (see section 4 in this online appendix for the detail) and found that the NR approximation successfully recovered the number of benefits ($K = 2$) but the GD approximation failed. This does not mean that the NR approximation performs better than the GD approximation because the difference between the models with $K = 2$ and $K = 3$ in both approximations is very marginal. Since the model with large K nests the model with small K, the marginal difference in the fit measures can be naturally expected. But, the result implies that the potential overestimation by the NR approximation may not depend on the number of benefits. Table 1 also shows the ratio of the absolute values of two approximations where we found a consistent pattern: the ratio for the MNL model is relatively smaller than that for the benefit-based conjoint (BBC) models but the ratio is almost identical across the BBC models. This pattern is also consistently found in our empirical applications. We thus conclude that the NR approximation is good enough to examine the relative difference in fit performance between the models.

Table 2 shows the posterior estimates of the part-worth parameters when $K = 2$. We generated individual parameters from the same distribution in both datasets, but the empirical mean values differ between the datasets due to the difference in the number of individuals. When $K \geq 1$, i.e., the BBC model is assumed, the 95% confidence interval of the posterior distribution contains the true value for all parameters except some cases. Even in the exceptions, the true value is marginally outside of the interval. However, if the number of benefits is equal to zero, i.e., the standard MNL model is assumed, the true values are not well recovered: the MNL model yields smaller absolute values of

Table 1: In-sample log-marginal density (LMD) in the simulation study

	Large dataset			Small dataset		
Models	LMD-NR	LMD-GD	NR/GD*	LMD-NR	LMD-GD	NR/GD*
MNL, $K = 0$	-39446.00	-48286.36	0.817	-4316.167	-5763.854	0.749
BBC, $K = 1$	-38682.43	-46788.59	0.827	-4240.435	-5222.786	0.812
BBC, $K = 2$	-38286.07	-46210.53	0.829	-4208.915	-5178.071	0.813
BBC, $K = 3$	-38327.77	-46135.50	0.831	-4230.771	-5057.074	0.837

* This indicates the ratio of the absolute values of the NR and GD approximations.

the posterior estimates than those of the true values. This result is consistent with our expectation. A part-worth parameter in the BBC model is independent from the benefit function, i.e., it captures the part-worth of an attribute separately from satiation. Thus, the true values are well recovered regardless of the number of benefits. In contrast, a part-worth parameter in the MNL model may capture the mixed effects of the part-worth and satiation, leading to the downward biases in the absolute values of the parameter estimates.

A label-switching problem (e.g., Stephens 2000) can exist when we are estimating the assignment probabilities. However, switching the labels of the benefits does not affect interpretations of the result. Figure 1 shows the trace plots of the initial 20,000 MCMC draws associated with a selected attribute when the true number of benefits is assumed ($K = 2$). Label-switching occurs only at the beginning of the MCMC iterations and the labels become very stable over the iterations with the large dataset. Even with the small dataset, label-switching does not occur frequently: it occurs at the beginning and end of the iterations. This stability is due to conditional independence. When the assignment variable for an attribute is drawn, how the other attributes are grouped into benefits is given and the likelihood value differs between benefits where the attribute is assigned. The labels therefore become steady over the iterations, once they are fixed with convergence of the model parameters. However, as in the trace plot for the small dataset, label-switching rarely but possibly occurs over MCMC iterations, leading to efficiency loss of the posterior estimates. We thus set the order of the assignment probabilities based on the initial draws as in our empirical application and ran extra MCMC iterations with the fixed order of the assignment probabilities to reduce the efficiency loss. We found that the model likelihood and part-worth parameters do not change in the extra iterations.

Table 3 displays the posterior estimates of the assignment probabilities when $K = 2$. The posterior estimates with the small dataset are calculated from the last 10,000 draws of

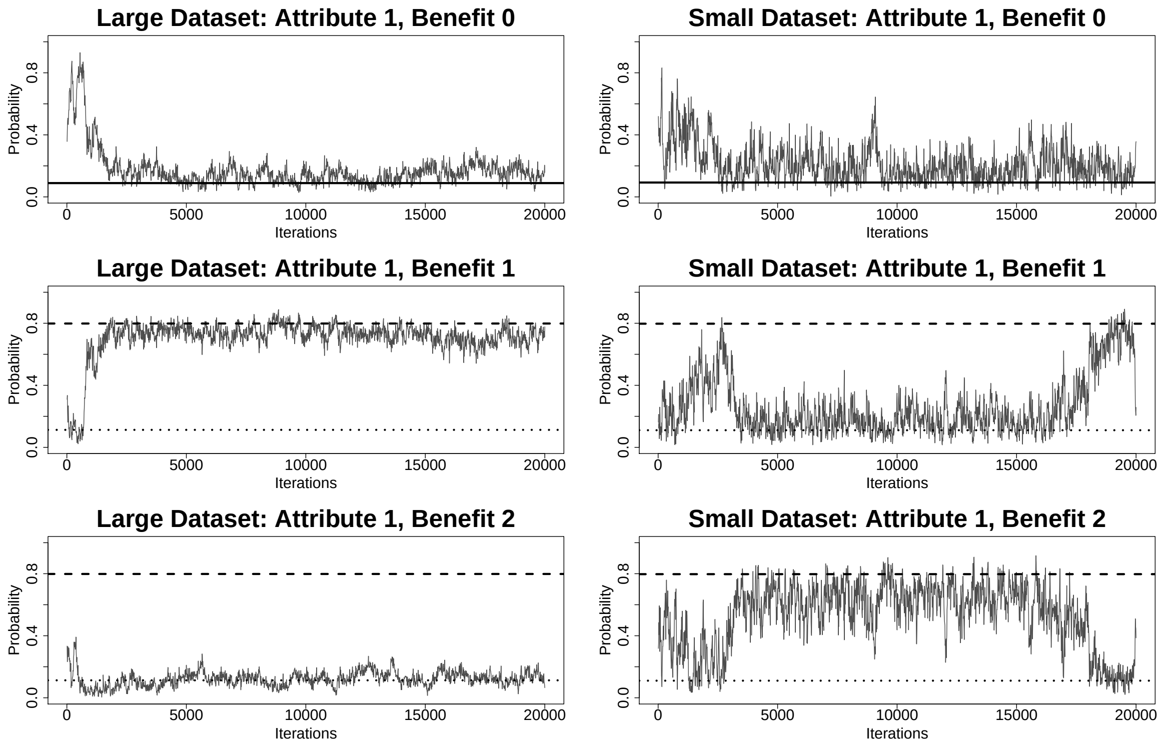
Table 2: Posterior estimates of the part-worth parameters ($\bar{\boldsymbol{\beta}}$) in the simulation study

Variables	Large dataset					Small dataset				
	True value	MNL $K=0$	BBC $K=1$	BBC $K=2$	BBC $K=3$	True value	MNL $K=0$	BBC $K=1$	BBC $K=2$	BBC $K=3$
A1M	0.58	**0.63**	**0.57**	**0.61**	**0.60**	0.52	**0.54**	**0.53**	**0.55**	**0.60**
		(0.04)	(0.05)	(0.05)	(0.05)		(0.09)	(0.13)	(0.14)	(0.14)
A1H	1.54	1.23	1.34	**1.52**	**1.50**	1.59	1.24	**1.51**	**1.55**	**1.50**
		(0.05)	(0.06)	(0.07)	(0.06)		(0.12)	(0.17)	(0.15)	(0.16)
A2M	-1.47	-1.24	**-1.48**	**-1.48**	**-1.46**	-1.55	**-1.52**	-1.95	**-1.81**	**-1.77**
		(0.05)	(0.06)	(0.05)	(0.06)		(0.11)	(0.14)	(0.15)	(0.15)
A2H	-0.59	-0.38	**-0.62**	-0.69	-0.68	-0.63	-0.29	**-0.55**	**-0.48**	**-0.45**
		(0.04)	(0.04)	(0.05)	(0.04)		(0.11)	(0.12)	(0.12)	(0.10)
A3M	1.73	1.31	**1.80**	**1.79**	**1.83**	1.72	1.26	**1.55**	**1.57**	**1.52**
		(0.05)	(0.06)	(0.09)	(0.06)		(0.10)	(0.17)	(0.12)	(0.16)
A3H	2.56	2.00	**2.64**	**2.59**	**2.61**	2.53	2.12	**2.62**	**2.48**	**2.47**
		(0.04)	(0.06)	(0.08)	(0.04)		(0.11)	(0.20)	(0.16)	(0.15)
A4M	1.50	1.21	1.38	**1.52**	**1.50**	1.37	1.06	**1.22**	**1.27**	**1.24**
		(0.04)	(0.04)	(0.05)	(0.05)		(0.12)	(0.17)	(0.14)	(0.13)
A4H	2.22	1.84	2.04	**2.27**	**2.24**	2.14	**1.84**	**2.16**	**2.31**	**2.12**
		(0.05)	(0.05)	(0.07)	(0.06)		(0.13)	(0.19)	(0.18)	(0.14)
A5M	0.80	0.50	**0.72**	**0.79**	**0.76**	0.67	**0.53**	**0.63**	**0.64**	**0.63**
		(0.04)	(0.06)	(0.06)	(0.07)		(0.11)	(0.15)	(0.16)	(0.15)
A5H	1.20	0.86	**1.21**	**1.24**	**1.20**	1.23	**1.17**	**1.48**	**1.33**	**1.36**
		(0.04)	(0.07)	(0.07)	(0.07)		(0.10)	(0.22)	(0.16)	(0.12)
A6M	0.79	**0.80**	**0.74**	**0.77**	**0.80**	0.71	0.99	0.98	1.03	1.04
		(0.04)	(0.05)	(0.06)	(0.06)		(0.12)	(0.14)	(0.13)	(0.13)
A6H	1.55	**1.48**	**1.49**	**1.55**	**1.58**	1.56	**1.71**	1.86	1.87	**1.81**
		(0.04)	(0.06)	(0.07)	(0.07)		(0.13)	(0.13)	(0.14)	(0.14)
A7M	0.52	**0.48**	**0.59**	**0.56**	**0.55**	0.49	**0.32**	**0.43**	**0.37**	**0.43**
		(0.04)	(0.06)	(0.07)	(0.07)		(0.10)	(0.19)	(0.14)	(0.14)
A7H	0.97	0.78	1.07	**1.02**	**1.02**	1.05	0.80	**1.20**	**0.97**	**1.02**
		(0.04)	(0.05)	(0.06)	(0.05)		(0.11)	(0.15)	(0.12)	(0.12)
A8	1.09	0.98	1.28	**1.21**	1.20	1.15	0.89	**1.22**	**1.03**	**1.09**
		(0.05)	(0.06)	(0.06)	(0.05)		(0.12)	(0.20)	(0.17)	(0.14)
A9	1.44	1.11	**1.53**	**1.48**	**1.48**	1.39	1.15	**1.65**	**1.41**	**1.37**
		(0.04)	(0.05)	(0.06)	(0.05)		(0.10)	(0.17)	(0.17)	(0.15)
A10	1.60	1.37	**1.66**	**1.63**	**1.65**	1.64	**1.64**	2.02	1.90	1.89
		(0.04)	(0.06)	(0.05)	(0.05)		(0.09)	(0.15)	(0.13)	(0.11)
Price	-1.49	-1.39	**-1.48**	**-1.50**	**-1.51**	-1.50	**-1.52**	**-1.56**	**-1.57**	**-1.59**
		(0.04)	(0.04)	(0.04)	(0.04)		(0.08)	(0.09)	(0.08)	(0.09)

* Posterior standard deviations are given in parentheses ().

* Numbers in boldface indicate that the 95% confidence interval contains the true value.

Figure 1: Trace plots of the initial MCMC draws for the assignment probability of the first attribute when $K = 2$ in the simulation study



* The solid, dashed, and dotted lines indicate the true values of the probabilities of assignment into the null group, first benefit, and second benefit, respectively.

the extra MCMC iterations. We report the empirical mode of the posterior distribution along with the mean and standard deviation. We found that the true values of the assignment probabilities are well recovered. Figures 2 and 3 show the empirical posterior distributions of the assignment probability. The empirical posterior mode of each attribute within a benefit is consistently close to one of different line styles (solid, dashed, and dotted) associated with different benefits. Therefore, the parameter recovery is clearly confirmed in this simulation study.

Table 3: Posterior estimates (means, modes, and standard deviations) of the assignment probabilities (θ) when $K = 2$ in the simulation study

	Large dataset						Small dataset***					
	True value			Estimates			True value			Estimates		
Attributes	Null	B1	B2	Null	B1	B2	Null	B1	B2	Null	B1**	B2**
				0.15	**0.72**	**0.13**				**0.20**	**0.19**	**0.61**
A1	0.09	0.80	0.11	*0.14*	*0.74*	*0.13*	0.09	0.80	0.11	*0.16*	*0.15*	*0.62*
				(0.05)	(0.05)	(0.04)				(0.10)	(0.09)	(0.12)
				0.21	**0.67**	**0.11**				**0.34**	**0.18**	**0.48**
A2	0.19	0.70	0.12	*0.22*	*0.67*	*0.10*	0.21	0.69	0.11	*0.27*	*0.17*	*0.50*
				(0.06)	(0.07)	(0.04)				(0.15)	(0.08)	(0.15)
				0.15	**0.13**	**0.72**				**0.20**	**0.64**	**0.16**
A3	0.19	0.11	0.70	*0.14*	*0.12*	*0.73*	0.20	0.09	0.71	*0.16*	*0.68*	*0.14*
				(0.04)	(0.03)	(0.04)				(0.09)	(0.11)	(0.07)
				0.12	**0.72**	**0.17**				**0.19**	**0.18**	**0.63**
A4	0.10	0.80	0.10	*0.10*	*0.72*	*0.17*	0.09	0.81	0.10	*0.17*	*0.14*	*0.65*
				(0.05)	(0.05)	(0.04)				(0.09)	(0.09)	(0.13)
				0.14	**0.13**	**0.74**				**0.19**	**0.62**	**0.19**
A5	0.19	0.10	0.71	*0.13*	*0.13*	*0.75*	0.26	0.10	0.64	*0.15*	*0.63*	*0.16*
				(0.04)	(0.04)	(0.05)				(0.09)	(0.13)	(0.09)
				0.61	**0.27**	**0.12**				**0.44**	**0.22**	**0.33**
A6	0.61	0.29	0.10	*0.59*	*0.27*	*0.13*	0.64	0.27	0.09	*0.41*	*0.18*	*0.36*
				(0.07)	(0.05)	(0.05)				(0.12)	(0.11)	(0.12)
				0.19	**0.18**	**0.63**				**0.19**	**0.55**	**0.26**
A7	0.14	0.20	0.67	*0.17*	*0.18*	*0.63*	0.07	0.23	0.70	*0.15*	*0.56*	*0.24*
				(0.05)	(0.05)	(0.05)				(0.08)	(0.12)	(0.12)
				0.12	**0.57**	**0.32**				**0.18**	**0.25**	**0.57**
A8	0.10	0.69	0.21	*0.10*	*0.56*	*0.32*	0.09	0.73	0.17	*0.13*	*0.22*	*0.65*
				(0.05)	(0.07)	(0.07)				(0.09)	(0.12)	(0.15)
				0.25	**0.21**	**0.54**				**0.21**	**0.48**	**0.32**
A9	0.20	0.20	0.60	*0.25*	*0.20*	*0.54*	0.23	0.20	0.58	*0.17*	*0.50*	*0.30*
				(0.05)	(0.04)	(0.05)				(0.10)	(0.13)	(0.11)
				0.25	**0.13**	**0.62**				**0.21**	**0.62**	**0.17**
A10	0.21	0.11	0.67	*0.24*	*0.11*	*0.61*	0.22	0.10	0.68	*0.18*	*0.63*	*0.13*
				(0.05)	(0.04)	(0.04)				(0.10)	(0.12)	(0.08)

* Italic numbers are the empirical posterior mode values.

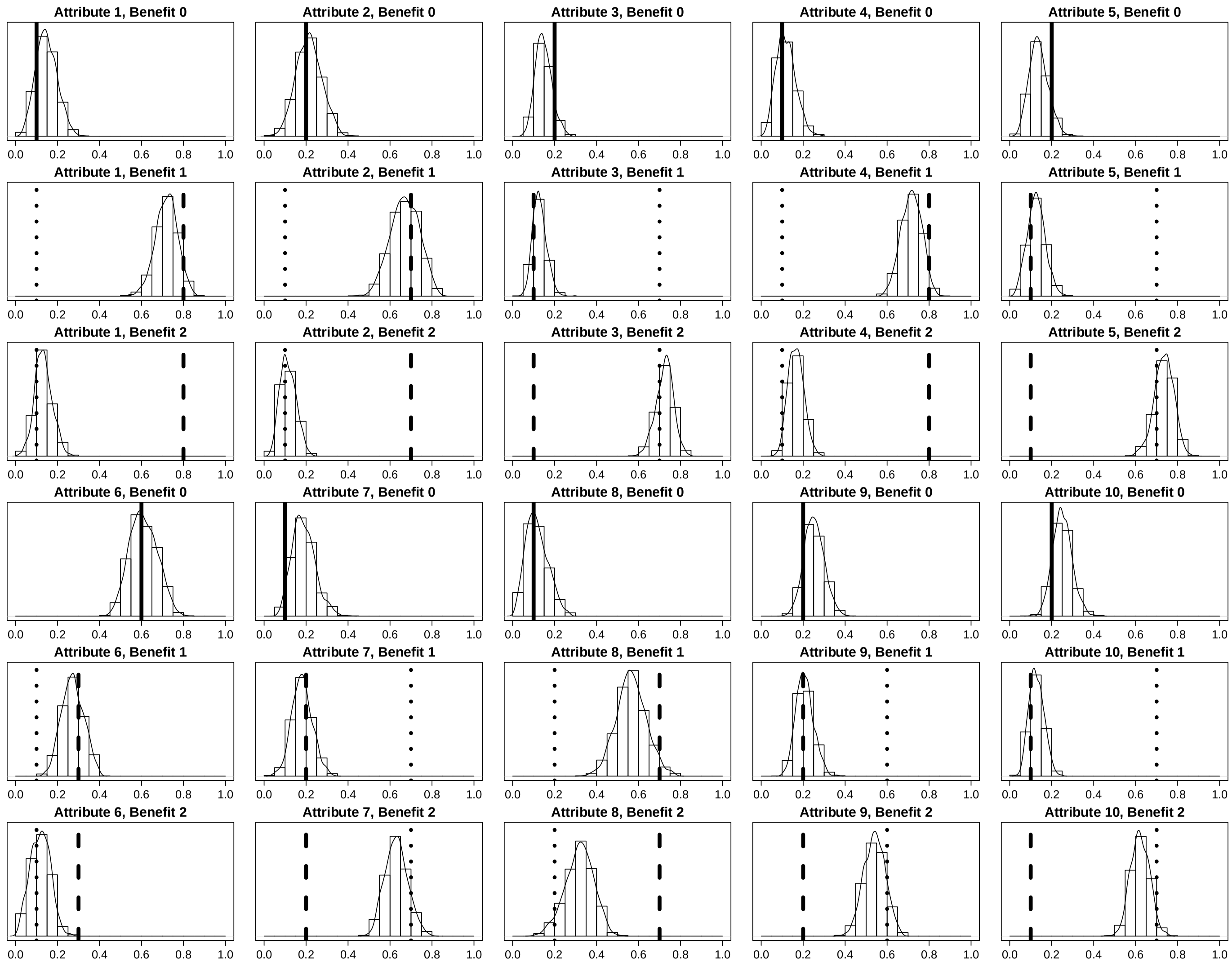
* Posterior standard deviations are given in parentheses ().

* Numbers in boldface indicate that the 95% confidence interval contains the true value.

** The labels are switched, i.e., the estimates corresponding to B1 (B2) should be compared to the true values of B2 (B1).

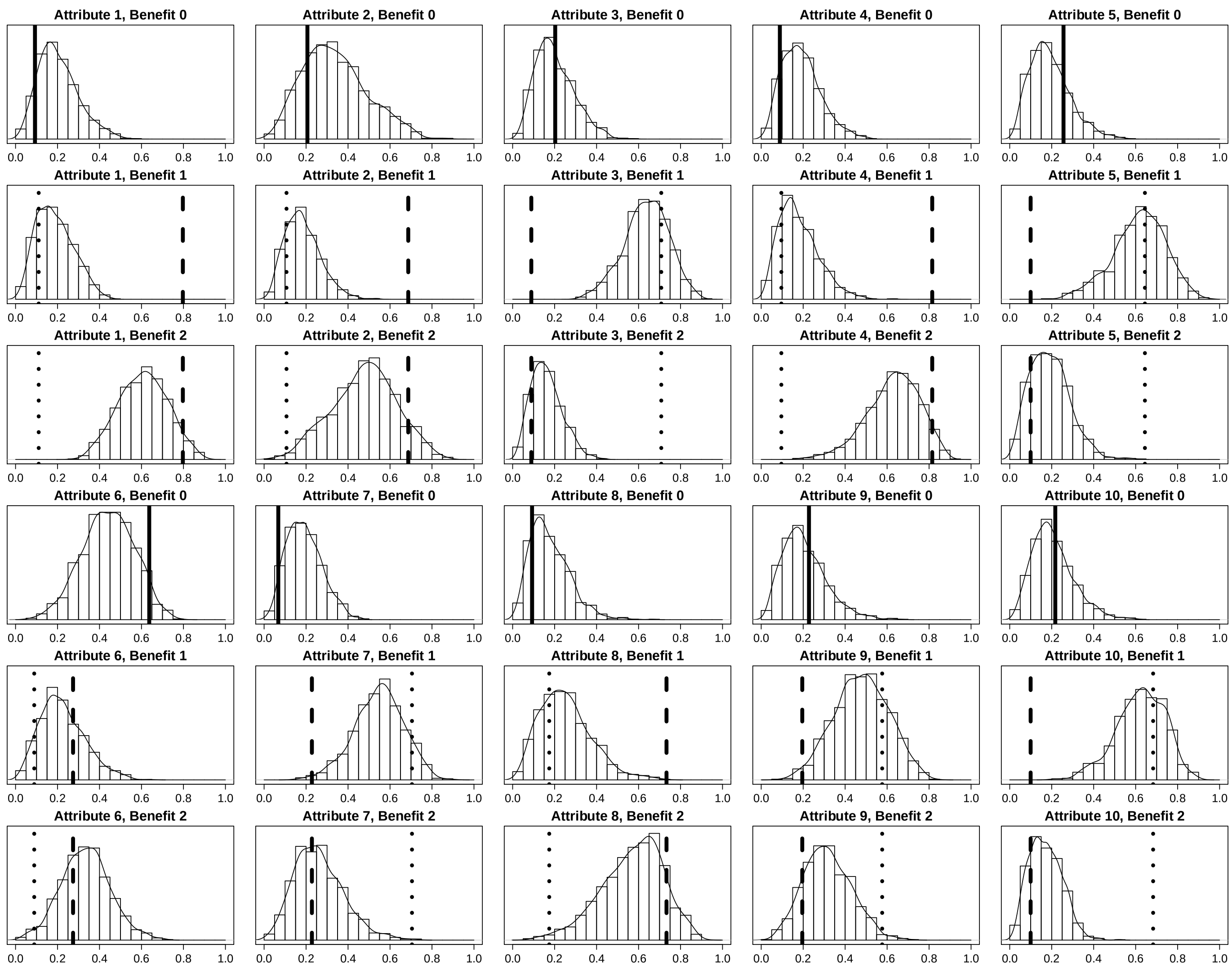
*** The posterior estimates with the small dataset are calculated from the last 10,000 draws of the extra MCMC iterations.

Figure 2: Empirical posterior distributions of the assignment probabilities when $K = 2$ with the large dataset



* The solid, dashed, and dotted lines indicate the true values of the probabilities of assignment into the null group, first benefit, and second benefit, respectively.

Figure 3: Empirical posterior distributions of the assignment probabilities when $K = 2$ with the small dataset



* The solid, dashed, and dotted lines indicate the true values of the probabilities of assignment into the null group, first benefit, and second benefit, respectively.

3 Estimation procedure

This section provides the estimation procedure for the models in the empirical analysis. The benefit-based conjoint (BBC) model used in the empirical analysis is as follows:

$$u_{hjt} = \sum_{n=1}^{N} \tau_{hn0} \cdot \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn} + \sum_{k=1}^{K} g\left(\sum_{n=1}^{N} \tau_{hnk} \cdot g^{-1}\left(\boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn}\right)\right) + \epsilon_{hjt},$$

where $g(y) = \operatorname{sgn}(y)\log(|y|+1)$. We employ the following standard MNL model as the benchmark model:

$$u_{hjt} = \sum_{n=1}^{N} \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn} + \epsilon_{hjt}.$$

We estimate the individual parameters, $\boldsymbol{\beta}_h$, and the hyperparameters for heterogeneity on the individual parameters, $\bar{\boldsymbol{\beta}}$ and $\mathbf{V}_\beta$, in both models. For the BBC model, we additionally infer about the individual attribute-benefit mapping, $\{\tau^*_{hn}\}$, and the hyperparameters for heterogeneity on the individual attribute-benefit mapping, $\{\theta_{nk}\}$.

Let z_{hjt} denote the response of individual h to choice task t, which has 1 if option j is chosen, and 0 otherwise. The likelihood of the responses of individual h given $\boldsymbol{\beta}_h$ and $\{\tau^*_{hn}\}$ is

$$\ell_h\left(\boldsymbol{\beta}_h, \{\tau^*_{hn}\} \mid \{z_{hjt}\}\right) = \prod_{t=1}^{T}\prod_{j=1}^{J}\left[\frac{\exp\left(\bar{u}_{hjt}\left(\boldsymbol{\beta}_h, \{\tau^*_{hn}\}\right)\right)}{\sum_{j'=1}^{J}\exp\left(\bar{u}_{hj't}\left(\boldsymbol{\beta}_h, \{\tau^*_{hn}\}\right)\right)}\right]^{z_{hjt}},$$

where, for the BBC model,

$$\begin{aligned}\bar{u}_{hjt}\left(\boldsymbol{\beta}_h, \{\tau^*_{hn}\}\right) &= u_{hjt} - \epsilon_{hjt}\\ &= \sum_{n=1}^{N} \tau_{hn0} \cdot \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn} + \sum_{k=1}^{K} g\left(\sum_{n=1}^{N} \tau_{hnk} \cdot g^{-1}\left(\boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn}\right)\right),\end{aligned}$$

and, for the standard MNL model,

$$\bar{u}_{hjt}\left(\boldsymbol{\beta}_h, \{\tau^*_{hn}\}\right) = u_{hjt} - \epsilon_{hjt} = \sum_{n=1}^{N} \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hn}.$$

Given this notation, the estimation procedure is as follows:

Step 1. Set initial values for all variables to be inferred: $\boldsymbol{\beta}_h$, $\{\tau^*_{hn}\}$, $\bar{\boldsymbol{\beta}}$, $\mathbf{V}_\beta$, and $\{\theta_{nk}\}$.

Step 2. Generate $\boldsymbol{\beta}_h$ for $h = 1, 2, \cdots, H$ given $\{\tau^*_{hn}\}$, $\bar{\boldsymbol{\beta}}$, and $\mathbf{V}_\beta$ via the random-walk Metropolis-Hastings (henceforth, RWMH) algorithm:

(a) Draw candidate $\boldsymbol{\beta}^{\text{new}}_h$ from $N\left(\boldsymbol{\beta}^{\text{old}}_h, d^2 \cdot \mathbf{V}_\beta\right)$, where $\boldsymbol{\beta}^{\text{old}}_h$ is the previous value of $\boldsymbol{\beta}_h$ and d is the step size of the RWMH algorithm. In our analysis, $d = 0.3$.

(b) Accept $\boldsymbol{\beta}^{\text{new}}_h$ with the following probability:

$$\Pr(\text{accept}) = \min\left[1, \frac{\ell_h\left(\boldsymbol{\beta}^{\text{new}}_h, \{\tau^*_{hn}\} \mid \{z_{hjt}\}\right) \cdot \phi\left(\boldsymbol{\beta}^{\text{new}}_h \mid \bar{\boldsymbol{\beta}}, \mathbf{V}_\beta\right)}{\ell_h\left(\boldsymbol{\beta}^{\text{old}}_h, \{\tau^*_{hn}\} \mid \{z_{hjt}\}\right) \cdot \phi\left(\boldsymbol{\beta}^{\text{old}}_h \mid \bar{\boldsymbol{\beta}}, \mathbf{V}_\beta\right)}\right],$$

where $\phi\left(\cdot \mid \bar{\boldsymbol{\beta}}, \mathbf{V}_\beta\right)$ is the density of the normal distribution with mean $\bar{\boldsymbol{\beta}}$ and variance $\mathbf{V}_\beta$.

Step 3. (For the BBC model only) Generate τ^*_{hn} for $n = 1, 2, \cdots, N$ and $h = 1, 2, \cdots, H$ given $\tau^*_{h,-n}$, $\boldsymbol{\beta}_h$, and $\{\theta_{nk}\}$ via the following posterior multinomial distribution:

$$\tau^*_{hn} \mid \tau^*_{h,-n}, \boldsymbol{\beta}_h, \{\theta_{nk}\} \sim \text{Multinomial}_{K+1}\left(\tilde{\theta}_{n0}, \tilde{\theta}_{n1}, \cdots, \tilde{\theta}_{nK}\right),$$

where

$$\tilde{\theta}_{nk} = \frac{\ell_h\left(\boldsymbol{\beta}_h, \tau^*_{hn} = k, \tau^*_{h,-n}\right) \cdot \theta_{nk}}{\sum_{k'=0}^{K} \ell_h\left(\boldsymbol{\beta}_h, \tau^*_{hn} = k', \tau^*_{h,-n}\right) \cdot \theta_{nk'}}.$$

Step 4. Generate $\bar{\boldsymbol{\beta}}$ and $\mathbf{V}_\beta$ given $\{\boldsymbol{\beta}_h\}$ via the following Bayesian multivariate regression (Rossi et al. 2005):

$$\boldsymbol{\beta}_h = \bar{\boldsymbol{\beta}} + \boldsymbol{\zeta}_h, \quad \boldsymbol{\zeta}_h \sim N\left(\mathbf{0}, \mathbf{V}_\beta\right).$$

In our analysis, the prior distributions are $\bar{\boldsymbol{\beta}} \mid \mathbf{V}_\beta \sim N\left(\mathbf{0}, 100 \cdot \mathbf{V}_\beta\right)$ and $\mathbf{V}_\beta \sim \text{IW}\left(\text{nvar} + 3, (\text{nvar} + 3) \cdot \mathbf{I}_{\text{nvar}}\right)$, where nvar denotes the number of the individual parameters, i.e., the number of elements in $\bar{\boldsymbol{\beta}}$, and $\mathbf{I}_{\text{nvar}}$ is an nvar$\times$nvar identity matrix.

Step 5. (For the BBC model only) Generate $\{\theta_{nk}\}$ given τ^*_{hn} via the following posterior Dirichlet distribution:

$$\theta_{n0}, \theta_{n1}, \cdots, \theta_{nK} \sim \text{Dirichlet}\left(\eta_{n0} + \sum_{h=1}^{H} \tau_{hn0}, \eta_{n1} + \sum_{h=1}^{H} \tau_{hn1}, \cdots, \eta_{nK} + \sum_{h=1}^{H} \tau_{hnK}\right),$$

where η_{nk} is the prior. $\eta_{nk} = 3$ for all $n = 1, 2, \cdots, N$ and $k = 0, 1, 2, \cdots, K$ in our analysis. With the order restriction, we use an ordered Dirichlet distribution rather than the Dirichlet distribution with the same distribution parameters (see Lenk and DeSarbo (2000) for the method of generating the ordered Dirichlet distribution).

Step 6. Repeat step 2 through 5 at each iteration of the MCMC.

4 Calculating fit measures

The marginal density of data y given model $\mathcal{M}$, $p(y \mid \mathcal{M})$, can be estimated from the following identity (Gelfand and Dey 1994):

$$\int_{\Omega_{\mathcal{M}}} \frac{q(\Theta)}{\ell(\Theta \mid \mathcal{M})\, p(\Theta \mid \mathcal{M})} p(\Theta \mid y, \mathcal{M})\, d\Theta = \frac{1}{p(y \mid \mathcal{M})},$$

where Θ is the set of model parameters, $\Omega_{\mathcal{M}}$ is the parameter space for model $\mathcal{M}$, and $q(\cdot)$ is an auxiliary density function. The MCMC approximation, $\hat{p}(y \mid \mathcal{M})$, is given by

$$\hat{p}(y \mid \mathcal{M}) = \left[\frac{1}{R} \sum_{r=1}^{R} \frac{q(\Theta_r)}{\ell(\Theta_r \mid \mathcal{M})\, p(\Theta_r \mid \mathcal{M})} \right]^{-1},$$

where R is the number of MCMC iterations and Θ_r is the r-th posterior draw of the parameters. This approximation is the Gelfand and Dey's (GD) estimator of the marginal density. If q is the posterior density, i.e., $q(\Theta) = p(\Theta \mid y, \mathcal{M})$, this approximation becomes exact. The performance of the GD approximation depends on the choice of the auxiliary density to approximate the posterior density of model parameters. We slightly modified the auxiliary density proposed by Lenk and DeSarbo (2000) for model selection for finite mixtures of generalized linear models because the posterior distributions of their model parameters are the same as those of our model parameters. q consists of the product of the following densities:

1. $q_{\boldsymbol{\beta}_h}(\boldsymbol{\beta}_h)$ is a multivariate normal density and its mean and covariance matrix are given by the sample mean and covariance of the posterior MCMC draws of $\boldsymbol{\beta}_h$.

2. $q_{\bar{\boldsymbol{\beta}}}\left(\bar{\boldsymbol{\beta}}\right)$ is a multivariate normal density and its mean and covariance matrix are given by the sample mean and covariance of the posterior MCMC draws of $\bar{\boldsymbol{\beta}}$.

3. $q_{\mathbf{V}_\beta}(\mathbf{V}_\beta)$ is an inverted Wishart density with ν_q degrees of freedom and scale matrix $\mathbf{V}_q$ so that $E\left(\mathbf{V}_\beta^{-1}\right) = \nu_q \mathbf{V}_q^{-1}$. ν_q is given by $\nu_0 + H$, where ν_0 is the prior degrees

of freedom and H is the number of individuals. $\mathbf{V}_q$ is given by a method of moment estimator, $\nu_q \left[\frac{1}{R}\sum_{r=1}^{R}\mathbf{V}_{\beta,r}^{-1}\right]^{-1}$, where $\mathbf{V}_{\beta,r}$ is the r-th posterior draw of $\mathbf{V}_\beta$.

4. $q_{\tau_{hn}^*}\left(\tau_{hn}^*\right)$ is a multinomial distribution with probabilities $\bar{\tau}_{hn0}$, $\bar{\tau}_{hn1}$, $\cdots$, $\bar{\tau}_{hnK}$, where $\bar{\tau}_{hnk} = \frac{1}{R}\sum_{r=1}^{R} I\left(\tau_{hnr}^* = k\right)$, τ_{hnr}^* is the r-th posterior draw of τ_{hn}^*, and $I(\cdot)$ is an indicator function.

5. $q_{\boldsymbol{\theta}_n}\left(\theta_{n0}, \theta_{n1}, \cdots, \theta_{nK}\right)$ is a Dirichlet density with parameters $\bar{\eta}_{n0}$, $\bar{\eta}_{n1}$, $\cdots$, $\bar{\eta}_{nK}$. Define $\bar{\eta}_n = \sum_{k=0}^{K}\bar{\eta}_{nk}$. Then, $E\left(\theta_{nk}\right) = \bar{\eta}_{nk}/\bar{\eta}_n$, $\operatorname{var}\left(\theta_{nk}\right) = \left(\bar{\eta}_n + 1\right)^{-1}\left[E\left(\theta_{nk}\right) - E\left(\theta_{nk}\right)^2\right]$, and $\sum_{k=0}^{K}\operatorname{var}\left(\theta_{nk}\right) = \left(\bar{\eta}_n + 1\right)^{-1}\left[1 - \sum_{k=0}^{K} E\left(\theta_{nk}\right)^2\right]$. Based on this, $\bar{\eta}_{nk}$ is given by a method of moment estimator,

$$\hat{E}\left(\theta_{nk}\right)\left[\frac{1 - \sum_{k=0}^{K}\hat{E}\left(\theta_{nk}\right)^2}{\sum_{k=0}^{K}\widehat{\operatorname{var}}\left(\theta_{nk}\right)} - 1\right],$$

where $\hat{E}\left(\theta_{nk}\right)$ and $\widehat{\operatorname{var}}\left(\theta_{nk}\right)$ are the sample mean and variance of the posterior MCMC draws of θ_{nk}.

As a special case, if $q(\Theta) = p(\Theta \mid \mathcal{M})$, this approximation becomes the Newton and Raftery's (1994) estimator of the marginal density:

$$\hat{p}_{\mathrm{NR}}(y \mid \mathcal{M}) = \left[\frac{1}{R}\sum_{r=1}^{R}\frac{1}{\ell\left(\Theta_r \mid \mathcal{M}\right)}\right]^{-1}.$$

This is the harmonic mean of the likelihood values and the computation is straightforward.

We additionally computed hit ratio and hit probability to evaluate both in-sample and holdout sample fit. Hit ratio (HR) measures how many of actual choices are expected conditional on parameter estimates and hit probability (HP) measures the expected probability of the actual choices conditional on parameter estimates. They are calculated as follows:

$$HR = \frac{1}{H \times T \times R}\sum_{h=1}^{H}\sum_{t=1}^{T}\sum_{r=1}^{R} I\left(y_{ht} = \operatorname{argmax}_j\left[\bar{u}_{hjt}\left(\boldsymbol{\beta}_{hr}, \{\tau_{hn}^*\}_r\right)\right]\right),$$
$$HP = \frac{1}{H \times T \times R}\sum_{h=1}^{H}\sum_{t=1}^{T}\sum_{r=1}^{R}\prod_{j=1}^{J}\left[\frac{\exp\left(\bar{u}_{hjt}\left(\boldsymbol{\beta}_{hr}, \{\tau_{hn}^*\}_r\right)\right)}{\sum_{j'=1}^{J}\exp\left(\bar{u}_{hj't}\left(\boldsymbol{\beta}_{hr}, \{\tau_{hn}^*\}_r\right)\right)}\right]^{z_{hjt}},$$

where H is the number of individuals, T is the number of observations per individual, R is the number of MCMC iterations, J is the number of choice options in a choice task, y_{ht} indicates which option individual h chose at choice task t among options 1 through J, z_{hjt}

denotes the response of individual h to choice task t, which has 1 if option j is chosen, and 0 otherwise, $\boldsymbol{\beta}_{hr}$ is the r-th draw of individual h's part-worth parameter vector, $\{\tau^*_{hn}\}_r$ denotes the r-th draw of individual h's assignment probabilities for all attributes, $I(\cdot)$ is an indicator function, and

$$\begin{aligned}\bar{u}_{hjt}\left(\boldsymbol{\beta}_{hr}, \{\tau^*_{hn}\}_r\right) &= u_{hjt} - \epsilon_{hjt} \\ &= \sum_{n=1}^{N} \tau_{hn0r} \cdot \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hnr} + \sum_{k=1}^{K} g\left(\sum_{n=1}^{N} \tau_{hnkr} \cdot g^{-1}\left(\boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hnr}\right)\right),\end{aligned}$$

for the benefit-based conjoint (BBC) model, and

$$\bar{u}_{hjt}\left(\boldsymbol{\beta}_{hr}, \{\tau^*_{hn}\}_r\right) = u_{hjt} - \epsilon_{hjt} = \sum_{n=1}^{N} \boldsymbol{x}'_{hjnt}\boldsymbol{\beta}_{hnr},$$

for the standard MNL model.

References

Gelfand, Alan E, Dipak K Dey. 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(3) 501–514.

Lenk, Peter. 2009. Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *Journal of Computational and Graphical Statistics* **18**(4) 941–960.

Lenk, Peter J., Wayne S. DeSarbo. 2000. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* **65**(1) 93–119.

Newton, Michael A, Adrian E Raftery. 1994. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**(1) 3–48.

Rossi, Peter E., Greg M. Allenby, Robert McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley and Sons Ltd.

Stephens, Matthew. 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4) 795–809.