Supplemental materials for

Braun and McAuliffe

"Variational inference for large-scale models of discrete choice"

APPENDIX C: A CONVEXITY RESULT

Let $a_1, \ldots, a_d$ be scalars, $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d$ be n-vectors, $p, r > 0$, and $\boldsymbol{Q} \succeq 0$. We show here that the function

$$f(\boldsymbol{B}) = r \log\left|\boldsymbol{B}\boldsymbol{B}^T\right| - p \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{B}\boldsymbol{B}^\top\right) - \log\left(\sum_{j=1}^{d} \exp\left\{a_j + \boldsymbol{c}_j^\top \boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{c}_j\right\}\right) \tag{C.1}$$

is concave on the set of full-rank $n \times n$ matrices.

We argue that each of the three constituent terms, from left to right, is concave. The second differential of $g(\boldsymbol{B}) = r \log\left|\boldsymbol{B}\boldsymbol{B}^T\right|$ is

$$\mathrm{d}^2 g = \mathrm{d} \operatorname{tr}\left\{2r\boldsymbol{B}^{-1}\mathrm{d}\boldsymbol{B}\right\} = \operatorname{tr}\left\{-2r\left[\boldsymbol{B}^{-1}(\mathrm{d}\boldsymbol{B})\right]^2\right\} . \tag{C.2}$$

By Theorem 10.6.1 of (Magnus and Neudecker 2007), the Hessian of g is $-2r\boldsymbol{K}_n\left(\boldsymbol{B}^{-\top} \otimes \boldsymbol{B}^{-1}\right)$, where $\boldsymbol{K}_n$ is the order-n commutation matrix and $\otimes$ denotes the Kronecker product. We now show that $\boldsymbol{K}_n\left(\boldsymbol{B}^{-\top} \otimes \boldsymbol{B}^{-1}\right)$ is (matrix) positive-definite.

$$(\mathrm{d}\operatorname{vec}\boldsymbol{X})^\top \boldsymbol{K}_n\left(\boldsymbol{B}^{-\top} \otimes \boldsymbol{B}^{-1}\right)(\mathrm{d}\operatorname{vec}\boldsymbol{X}) = (\mathrm{d}\operatorname{vec}\boldsymbol{X})^\top \boldsymbol{K}_n \operatorname{vec}\left\{\boldsymbol{B}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{B}^{-1}\right\} \tag{C.3}$$

$$= (\operatorname{vec}\mathrm{d}\boldsymbol{X})^\top \operatorname{vec}\left\{\boldsymbol{B}^{-\top}(\mathrm{d}\boldsymbol{X})^\top\boldsymbol{B}^{-\top}\right\} \tag{C.4}$$

$$= \operatorname{tr}\left\{\left(\boldsymbol{B}^{-1}\mathrm{d}\boldsymbol{X}\right)^2\right\} \geq 0 . \tag{C.5}$$

Equation (C.3) follows from the well-known fact that $\operatorname{vec}\boldsymbol{A}\boldsymbol{B}\boldsymbol{C} = (\boldsymbol{C}^\top \otimes \boldsymbol{A})\operatorname{vec}\boldsymbol{B}$. Thus, the Hessian of g is negative definite, and $r\log\left|\boldsymbol{B}\boldsymbol{B}^T\right|$ is concave.

Concavity of the middle term in (C.1) follows in the usual way from the univariate convexity of the function

$$g(t) := \operatorname{tr}\left(\boldsymbol{Q}(\boldsymbol{M}+t\boldsymbol{P})(\boldsymbol{M}+t\boldsymbol{P})^\top\right) = \sum_{i=1}^{n}(\boldsymbol{m}_i+t\boldsymbol{p}_i)^\top \boldsymbol{Q}(\boldsymbol{m}_i+t\boldsymbol{p}_i) \tag{C.6}$$

for fixed matrices $\boldsymbol{M}$ and $\boldsymbol{P}$, with columns $\boldsymbol{m}_i$ and $\boldsymbol{p}_i$. To see that the rightmost term in (C.1) is concave, define

$$g_j(t) := a_j + \boldsymbol{c}_j^\top(\boldsymbol{M}+t\boldsymbol{Q})(\boldsymbol{M}+t\boldsymbol{Q})^\top \boldsymbol{c}_j$$

for $j = 1, \ldots, d$ and fixed matrices $\boldsymbol{M}$ and $\boldsymbol{Q}$. Each g_j is convex in t, and the rightmost term in (C.1) is (minus) the log-sum-exp function composed with the g_j's. Concavity of this term in t, and hence in $\boldsymbol{B}$, follows from Boyd and Vandenberghe (2004, p. 86).

APPENDIX D: EXISTENCE OF RELATIVE ENTROPIES

Here we argue that the quantity KL $[q\,||\,p]$ in (3.4) exists and is finite for all values of $\boldsymbol{\lambda}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\Omega}$ that we consider in this paper. We always take $\boldsymbol{\zeta} \in \mathbb{R}^K$ and $0 \prec \boldsymbol{\Omega} \in \mathbb{R}^{K\times K}$. From (3.5), we see it is enough that each of the two quantities $\mathcal{L}(\boldsymbol{\lambda};\boldsymbol{\zeta},\boldsymbol{\Omega})$ and $p(\mathcal{D} \mid \boldsymbol{\zeta},\boldsymbol{\Omega})$ exists and is finite.

First consider $p(\mathcal{D} \mid \boldsymbol{\zeta},\boldsymbol{\Omega})$, the likelihood of $\boldsymbol{\zeta}$ and $\boldsymbol{\Omega}$ in the MNL discrete choice model. It is a product of $(T_1 + \cdots + T_H) < \infty$ factors, each of which is the expected value under the $\mathcal{N}_K(\boldsymbol{\zeta},\boldsymbol{\Omega})$ distribution of the expression in (2.1) (with j equal to the observed outcome of agent h's tth choice event). Since each of these expected values exists and lies in $[0, 1]$, it follows that $p(\mathcal{D} \mid \boldsymbol{\zeta},\boldsymbol{\Omega})$ exists and is finite.

Now consider $\mathcal{L}(\boldsymbol{\lambda};\boldsymbol{\zeta},\boldsymbol{\Omega})$ as represented in (A.1). The argument just applied to $p(\mathcal{D} \mid \boldsymbol{\zeta},\boldsymbol{\Omega})$ shows that the third term in (A.1) exists and is finite, provided that $\boldsymbol{\mu}_h \in \mathbb{R}^K$ and $0 \prec \boldsymbol{\Sigma}_h \in \mathbb{R}^{K\times K}$ for all the variational mean and covariance parameters $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$, $h = 1, \ldots, H$. These conditions are guaranteed by our optimization procedure. The first term in (A.1) is a sum of $H < \infty$ normal entropies, each of which takes on the value $\frac{1}{2}\log\left[(2\pi e)^K|\boldsymbol{\Sigma}_h|\right] < \infty$ (again, provided $\boldsymbol{\Sigma}_h \succ 0$). Finally, the second term in (A.1) is a sum of $H < \infty$ cross entropies between two normal distributions, one with parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\Omega}$ and the other with parameters $\boldsymbol{\mu}_h$ and $\boldsymbol{\Sigma}_h$. It is easy to show (by completing the square inside the expectation) that each cross-entropy has the form

$$\frac{1}{2}\left[\log\left((2\pi)^K|\boldsymbol{\Omega}|\right) + \operatorname{tr}\boldsymbol{\Omega}^{-1}\left\{\boldsymbol{\Sigma}_h + (\boldsymbol{\mu}_h - \boldsymbol{\zeta})(\boldsymbol{\mu}_h - \boldsymbol{\zeta})^\top\right\}\right]. \tag{D.1}$$

By inspection, this quantity is finite under the previously stated conditions on the parameters. We have completed the argument that the quantity KL $[q\,||\,p]$ in (3.4) exists and is finite. One can further argue without difficulty that the two forms of the approximation $\tilde{\mathcal{L}}(\boldsymbol{\lambda};\boldsymbol{\zeta},\boldsymbol{\Omega})$ exist and are finite, as is the quantity KL $[q\,||\,p]$ for variational hierarchical Bayes (with its accompanying approximations $\tilde{\mathcal{L}}$).

APPENDIX E: MONTE CARLO VARIABILITY

Let $p(\boldsymbol{y} \mid \boldsymbol{x})$ be the true predictive choice distribution at $\boldsymbol{x}$. Let $q(\boldsymbol{y} \mid \boldsymbol{x})$ denote some fixed estimate of $p(\boldsymbol{y} \mid \boldsymbol{x})$, such as the one derived from ($\hat{\boldsymbol{\zeta}}_{\text{VEB}}$, $\hat{\boldsymbol{\Omega}}_{\text{VEB}}$). We cannot compute TV $[p, q]$ exactly, because we cannot compute the integral (4.1) to evaluate $p(\boldsymbol{y} \mid \boldsymbol{x})$. Instead, we compute TV $[\hat{p}_M, q]$, where $\hat{p}_M(\boldsymbol{y} \mid \boldsymbol{x})$ is a Monte Carlo approximation to (4.1) based on M iid draws of $\boldsymbol{\beta} \sim N_K(\boldsymbol{\zeta}, \boldsymbol{\Omega})$:

$$\hat{p}_M(\boldsymbol{y} \mid \boldsymbol{x}) := \frac{1}{M}\sum_{m=1}^{M} p\big(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\beta}^{(m)}\big). \tag{E.1}$$

The goal in this appendix is to choose an M large enough that the Monte Carlo variability in TV $[\hat{p}_M, q]$ is always substantially smaller than the standard errors reported in our simulation results (Tables 1, 2, and 4). However, because TV $[\hat{p}_M, q]$ is a nonlinear, non-separable function of all M draws,

$$\text{TV}\,[\hat{p}_M, q] = \frac{1}{2}\sum_{j=1}^{J}\left|\hat{p}_M(y^j = 1 \mid \boldsymbol{x}) - q(y^j = 1 \mid \boldsymbol{x})\right|, \tag{E.2}$$

computing its Monte Carlo variance is not straightforward. We will now derive an upper bound on Var(TV $[\hat{p}_M, q]$) in terms of quantities we can directly compute. To emphasize, the variance here is with respect to the distribution of $(\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(M)})$. We will need the following lemma.

Lemma 1. *Let the random variable Z have two finite moments. Then Var$(|Z - c|) \leq$ Var(Z) for all $c \in \mathbb{R}$.*

Proof. Since the function $x \mapsto |x|$ is convex, Jensen's inequality shows that $|\mathbb{E}Z - c| \leq \mathbb{E}|Z - c|$. The result follows upon squaring both sides and rearranging terms. □

Now let $\hat{p}_M^j := \hat{p}_M(y^j = 1 \mid \boldsymbol{x})$ and $q^j := q(y^j = 1 \mid \boldsymbol{x})$, for $j = 1, \ldots, J$. We have that

$$\operatorname{Var}(\mathrm{TV}\,[\hat{p}_M, q]) = \frac{1}{4}\operatorname{Var}\left(\sum_{j=1}^{J} \left|\hat{p}_M^j - q^j\right|\right) \tag{E.3}$$

$$= \frac{1}{4}\sum_{j=1}^{J}\sum_{k=1}^{J} \operatorname{Cov}\left(\left|\hat{p}_M^j - q^j\right|, \left|\hat{p}_M^k - q^k\right|\right) \tag{E.4}$$

$$\leq \frac{1}{4}\sum_{j=1}^{J}\sum_{k=1}^{J} \left\{\operatorname{Var}\left(\left|\hat{p}_M^j - q^j\right|\right)\operatorname{Var}\left(\left|\hat{p}_M^k - q^k\right|\right)\right\}^{1/2} \tag{E.5}$$

$$= \left[\frac{1}{2}\sum_{j=1}^{J} \left\{\operatorname{Var}\left(\left|\hat{p}_M^j - q^j\right|\right)\right\}^{1/2}\right]^2 \tag{E.6}$$

$$\leq \left[\frac{1}{2}\sum_{j=1}^{J} \left\{\operatorname{Var}\left(\hat{p}_M^j\right)\right\}^{1/2}\right]^2, \tag{E.7}$$

where (E.5) used Cauchy-Schwarz and (E.7) used Lemma 1. We therefore have the bound

$$\operatorname{SD}(\mathrm{TV}\,[\hat{p}_M, q]) \leq \frac{1}{2}\sum_{j=1}^{J} \operatorname{SD}\left(\hat{p}_M^j\right). \tag{E.8}$$

To estimate the quantity $\operatorname{SD}\left(\hat{p}_M^j\right)$, we plug in $(1/\sqrt{M})$ times the empirical standard deviation of the M Monte Carlo values $p\left(y^j = 1 \mid \boldsymbol{x}, \boldsymbol{\beta}^{(m)}\right)$. (This is just the usual estimator for the standard error of a sample mean.)

We found that when we used $M = 1{,}000{,}000$, the upper bound (E.8) for each entry in Tables 1, 2, and 4 was always less than half of the entry's standard error (which is computed across the 10 replications of the simulation scenario). In every entry, the root-sum-square combination of the SE with the upper bound was equal to the SE, out to the four decimal places we report. Thus, the variability contributed by Monte Carlo integration of the true predictive choice distribution has no practical impact on our results.

Note that the bound (E.8) is very loose: the covariances in (E.4) which we upper-bound by Cauchy-Schwarz are almost always observed to be close to zero or negative for $j \neq k$.

APPENDIX F: CONVERGENCE DIAGNOSTICS AND PARAMETER ESTIMATES

F.1 Simulation study A

We present summary statistics, MCMC standard error estimates, and Gelman-Rubin scale reduction factors for the conditional log likelihood (LL) and ζ for all models. Each model is identified by the degree of heterogeneity (Het), number of simulated households (N), number of coefficients (K) and number of choice alternatives (J). Further definitions and details are in the main body of the manuscript. Results are averages over dataset replicates and chains.

The posterior mean and standard deviation are simple summary statistics of the retained draws. The Gelman-Rubin scale reduction factor, effective sample size, and MCMC error for the posterior mean were computed using the CODA package in R. The method for computing MCMC error is the same as described in Chapter 7 of Brockwell and Davis (1991).

F.1.1 Posterior mean

Het	N	K	J	LL	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
Low	250	3	3	-2873	-2.00	0.02	2.00							
Low	250	3	12	-6652	-2.00	0.01	2.02							
Low	250	10	3	-2167	-1.84	-1.42	-1.02	-0.62	-0.25	0.21	0.62	1.05	1.43	1.85
Low	250	10	12	-4665	-1.91	-1.46	-1.07	-0.64	-0.21	0.21	0.64	1.05	1.50	1.91
High	250	3	3	-2757	-1.95	-0.01	2.03							
High	250	3	12	-6449	-1.99	-0.01	2.01							
High	250	10	3	-1988	-1.66	-1.25	-0.90	-0.56	-0.19	0.20	0.52	0.94	1.31	1.65
High	250	10	12	-3970	-1.90	-1.45	-1.09	-0.61	-0.25	0.18	0.60	1.06	1.47	1.92
Low	1000	3	3	-11493	-1.99	-0.00	1.99							
Low	1000	3	12	-26849	-2.00	-0.01	2.00							
Low	1000	10	3	-8249	-1.92	-1.50	-1.06	-0.65	-0.21	0.20	0.66	1.06	1.50	1.93
Low	1000	10	12	-18123	-1.98	-1.54	-1.09	-0.67	-0.22	0.22	0.65	1.11	1.54	1.97
High	1000	3	3	-11240	-1.97	0.00	1.96							
High	1000	3	12	-25691	-2.00	0.00	2.01							
High	1000	10	3	-7370	-1.88	-1.46	-1.07	-0.62	-0.20	0.19	0.63	1.05	1.44	1.90
High	1000	10	12	-15693	-1.98	-1.53	-1.10	-0.65	-0.23	0.22	0.66	1.10	1.54	1.96
Low	5000	3	3	-57572	-2.00	0.00	2.00							
Low	5000	3	12	-134005	-2.00	0.00	2.00							
Low	5000	10	3	-40695	-1.97	-1.54	-1.10	-0.66	-0.22	0.21	0.66	1.09	1.54	1.98
Low	5000	10	12	-90342	-2.00	-1.55	-1.11	-0.66	-0.23	0.22	0.66	1.11	1.55	1.99
High	5000	3	3	-55559	-2.00	-0.01	2.00							
High	5000	3	12	-129295	-2.01	-0.00	2.00							
High	5000	10	3	-36034	-1.95	-1.51	-1.08	-0.65	-0.22	0.22	0.65	1.08	1.52	1.94
High	5000	10	12	-78290	-1.98	-1.54	-1.10	-0.65	-0.21	0.22	0.66	1.11	1.54	1.98

F.1.2 Posterior standard deviation

Het	N	K	J	LL	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
Low	250	3	3	19.72	0.06	0.04	0.06							
Low	250	3	12	20.06	0.04	0.04	0.04							
Low	250	10	3	52.73	0.07	0.06	0.05	0.05	0.04	0.04	0.05	0.05	0.06	0.07
Low	250	10	12	52.97	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05
High	250	3	3	20.21	0.08	0.07	0.08							
High	250	3	12	19.90	0.07	0.07	0.07							
High	250	10	3	60.47	0.10	0.08	0.08	0.07	0.06	0.07	0.07	0.07	0.09	0.10
High	250	10	12	49.30	0.09	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.09
Low	1000	3	3	43.49	0.03	0.02	0.03							
Low	1000	3	12	43.18	0.02	0.02	0.02							
Low	1000	10	3	150.72	0.04	0.03	0.03	0.03	0.02	0.02	0.03	0.03	0.03	0.04
Low	1000	10	12	153.70	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03
High	1000	3	3	40.83	0.04	0.04	0.04							
High	1000	3	12	40.21	0.04	0.03	0.04							
High	1000	10	3	218.90	0.10	0.08	0.07	0.05	0.04	0.04	0.05	0.06	0.08	0.10
High	1000	10	12	131.78	0.07	0.06	0.05	0.04	0.03	0.03	0.04	0.05	0.06	0.07
Low	5000	3	3	131.79	0.02	0.01	0.02							
Low	5000	3	12	126.14	0.01	0.01	0.01							
Low	5000	10	3	597.80	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
Low	5000	10	12	666.87	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
High	5000	3	3	100.52	0.02	0.02	0.02							
High	5000	3	12	98.07	0.02	0.01	0.02							
High	5000	10	3	1068.45	0.10	0.08	0.06	0.04	0.02	0.02	0.04	0.06	0.08	0.10
High	5000	10	12	545.23	0.06	0.05	0.03	0.02	0.02	0.02	0.02	0.03	0.05	0.06

F.1.3 MCMC error for posterior mean

Het	N	K	J	LL	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
Low	250	3	3	0.540	0.002	0.001	0.002							
Low	250	3	12	0.372	0.001	0.000	0.001							
Low	250	10	3	2.206	0.004	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.004
Low	250	10	12	1.818	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
High	250	3	3	0.370	0.002	0.001	0.002							
High	250	3	12	0.301	0.001	0.001	0.001							
High	250	10	3	2.910	0.005	0.004	0.003	0.002	0.002	0.002	0.002	0.003	0.004	0.005
High	250	10	12	2.028	0.003	0.003	0.002	0.001	0.001	0.001	0.001	0.002	0.003	0.004
Low	1000	3	3	1.194	0.001	0.000	0.001							
Low	1000	3	12	0.783	0.000	0.000	0.000							
Low	1000	10	3	5.724	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
Low	1000	10	12	4.597	0.001	0.001	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.001
High	1000	3	3	0.726	0.001	0.000	0.001							
High	1000	3	12	0.602	0.000	0.000	0.000							
High	1000	10	3	12.724	0.006	0.005	0.003	0.002	0.001	0.001	0.002	0.003	0.004	0.006
High	1000	10	12	6.419	0.003	0.003	0.002	0.001	0.001	0.001	0.001	0.002	0.003	0.003
Low	5000	3	3	3.549	0.001	0.000	0.001							
Low	5000	3	12	2.292	0.000	0.000	0.000							
Low	5000	10	3	18.632	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.001
Low	5000	10	12	17.601	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
High	5000	3	3	1.523	0.000	0.000	0.000							
High	5000	3	12	1.419	0.000	0.000	0.000							
High	5000	10	3	70.836	0.006	0.005	0.003	0.002	0.001	0.001	0.002	0.003	0.005	0.006
High	5000	10	12	32.193	0.003	0.002	0.002	0.001	0.000	0.000	0.001	0.002	0.002	0.003

F.1.4 Gelman-Rubin scale reduction factor

Het	N	K	J	LL	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
Low	250	3	3	1.004	1.007	1.001	1.008							
Low	250	3	12	1.001	1.001	1.000	1.001							
Low	250	10	3	1.022	1.045	1.039	1.023	1.012	1.005	1.006	1.012	1.026	1.033	1.053
Low	250	10	12	1.009	1.015	1.010	1.007	1.004	1.002	1.002	1.003	1.006	1.008	1.015
High	250	3	3	1.001	1.002	1.000	1.001							
High	250	3	12	1.001	1.000	1.000	1.000							
High	250	10	3	1.057	1.052	1.041	1.020	1.009	1.002	1.004	1.010	1.024	1.041	1.058
High	250	10	12	1.005	1.003	1.002	1.001	1.000	1.001	1.001	1.001	1.001	1.002	1.002
Low	1000	3	3	1.003	1.004	1.001	1.004							
Low	1000	3	12	1.001	1.001	1.001	1.002							
Low	1000	10	3	1.058	1.103	1.078	1.052	1.021	1.007	1.011	1.023	1.046	1.073	1.107
Low	1000	10	12	1.013	1.018	1.013	1.010	1.005	1.002	1.002	1.004	1.010	1.012	1.018
High	1000	3	3	1.001	1.001	1.000	1.002							
High	1000	3	12	1.001	1.000	1.000	1.001							
High	1000	10	3	1.022	1.023	1.016	1.009	1.005	1.002	1.001	1.004	1.010	1.019	1.023
High	1000	10	12	1.004	1.004	1.003	1.002	1.001	1.001	1.001	1.002	1.001	1.003	1.004
Low	5000	3	3	1.005	1.008	1.001	1.007							
Low	5000	3	12	1.001	1.002	1.001	1.001							
Low	5000	10	3	1.081	1.156	1.111	1.090	1.036	1.009	1.011	1.037	1.066	1.119	1.143
Low	5000	10	12	1.012	1.019	1.015	1.011	1.005	1.002	1.003	1.005	1.009	1.014	1.017
High	5000	3	3	1.001	1.002	1.000	1.002							
High	5000	3	12	1.001	1.000	1.000	1.001							
High	5000	10	3	1.035	1.031	1.024	1.016	1.008	1.003	1.002	1.007	1.017	1.024	1.033
High	5000	10	12	1.005	1.006	1.005	1.002	1.001	1.001	1.001	1.002	1.003	1.005	1.005

F.1.5 Effective sample size

Het	N	K	J	LL	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
Low	250	3	3	1684	937	4540	960							
Low	250	3	12	3562	3212	7478	3246							
Low	250	10	3	544	151	225	345	678	996	1005	606	332	225	154
Low	250	10	12	1105	433	661	990	1629	2269	2291	1714	962	638	476
High	250	3	3	3711	2969	8243	2952							
High	250	3	12	4651	6990	10250	6596							
High	250	10	3	254	184	270	479	1017	2281	2269	1273	428	253	172
High	250	10	12	745	572	814	1251	2795	4755	5334	2645	1315	834	571
Low	1000	3	3	1661	1009	4557	981							
Low	1000	3	12	3711	3158	7433	3194							
Low	1000	10	3	924	118	179	287	627	1093	1040	607	325	181	116
Low	1000	10	12	1177	458	606	1081	1673	2527	2325	1817	985	632	468
High	1000	3	3	3950	2916	8133	2868							
High	1000	3	12	4720	6477	10499	6452							
High	1000	10	3	145	98	115	144	269	1855	1832	296	152	119	100
High	1000	10	12	555	330	402	588	1176	4178	4217	1127	590	409	326
Low	5000	3	3	1840	826	4434	817							
Low	5000	3	12	3813	3093	7487	3083							
Low	5000	10	3	1192	140	191	346	622	1126	1109	632	358	193	148
Low	5000	10	12	1212	411	588	885	1480	2270	2480	1544	865	562	431
High	5000	3	3	4569	2177	8145	2183							
High	5000	3	12	4843	4990	10449	5069							
High	5000	10	3	110	74	78	89	117	474	441	114	88	79	74
High	5000	10	12	354	230	250	291	442	2060	1994	450	287	249	229

F.2 Convergence statistics for stated choice experiment

Parameter	Effective size	Gelman-Rubin	Posterior Mean	Posterior SD	MCMC error
LL	237	1.035	-13138	265.6	9.317
ζ_1	413	1.009	-0.672	0.028	0.001
ζ_2	427	1.051	0.212	0.028	0.001
ζ_3	344	1.110	0.036	0.029	0.001
ζ_4	280	1.022	-0.098	0.030	0.001
ζ_5	441	1.009	0.909	0.047	0.002
ζ_6	568	1.047	0.458	0.034	0.001
ζ_7	648	1.021	-0.509	0.042	0.001
ζ_8	625	1.038	0.405	0.032	0.001
ζ_9	713	1.020	0.049	0.035	0.001
ζ_{10}	780	1.019	-0.348	0.033	0.001

F.3 Convergence statistics for simulation study B

Parameter	Effective size	Gelman-Rubin	Posterior Mean	Posterior SD	MCMC error
LL	289	1.033	-14861	326.3	11.652
ζ_1	268	1.02	-0.60	0.028	0.001
ζ_2	457	1.03	0.20	0.026	0.001
ζ_3	450	1.04	0.03	0.025	0.001
ζ_4	411	1.05	-0.08	0.027	0.001
ζ_5	240	1.04	0.77	0.041	0.002
ζ_6	379	1.04	0.39	0.029	0.001
ζ_7	305	1.04	-0.40	0.032	0.001
ζ_8	365	1.03	0.38	0.029	0.001
ζ_9	438	1.04	0.08	0.026	0.001
ζ_{10}	325	1.04	-0.29	0.029	0.001