Who's got the coupon? Estimating Consumer Preferences and Coupon Usage from Aggregate Information

Andrés Musalem Eric T. Bradlow Jagmohan S. Raju

Web Appendix

NUMERICAL EXAMPLE – BASIC MODEL

In order to demonstrate the efficacy of the approach, a numerical example with J=3 brands, T=50 periods and N=500 consumers is considered. The utility function of each of these consumers includes four explanatory variables. The first two correspond to brand dummies for the first two brands, the third is generated from a standard normal distribution and the fourth variable is a coupon indicator (c_{ijt}). The true mean and variance of the individual coefficients (θ_i) are chosen as $\bar{\theta} = (1\ 1\ -1\ 1)'$ and $D = I_4$, respectively, where I_4 denotes the identity matrix with four rows and columns. In addition, coupons are randomly generated and the probability of a consumer receiving a coupon in a given period is equal to .1, .2 and .3 for brands 1, 2 and 3, respectively.

Using only aggregate information (i.e., market shares, number of redeemed coupons and number of coupons available for each brand in each period), $\bar{\theta}$ and D are estimated according to the procedure described in this section. The starting values for $\bar{\theta}$ and D correspond to $\bar{\theta}=(0\ 0\ 0\ 0)'$ and $D = .1\ I_4$. The starting values of the MCMC for Z (choices) and C (coupons) are randomly chosen from a distribution that assigns the same probability to any configuration of choices and coupons satisfying constraints (3), (4) and (5). Specifically, this can be implemented by first assigning the choices of brand j in period t to NS_{jt} randomly chosen consumers, then, among those consumers, we randomly distribute N^r_{jt} redeemed coupons and finally, the remaining coupons N^c_{jt}-N^r_{jt} are randomly distributed among the consumers not choosing brand j in period t. Finally, we specify the following hyperprior distributions: $\bar{\theta} \sim N(0,10^5)$ and D ~ Inverse Wishart (6,6 I_4), very weakly informative. The results are presented in Table W1 and they are based on a single run of 200,000 iterations, where only the last 100,000 were used for the estimation of $\bar{\theta}$ and D, the mean and variance of the preference coefficients.

From the results in Table W1 it is observed that the true values of $\bar{\theta}$ and D are covered by their 95% posterior-probability intervals and that the estimated posterior means are very close to their true values, providing numerical support for this method.

NUMERICAL EXAMPLE – LIMITED INFORMATION

A numerical simulation example is constructed based on the same parameter values for $\bar{\theta}$ and D previously used. The true values for the parameters that determine the coupon probabilities are specified as: q=(.4, .5, .6), α=(-2,-1,0) and

$$\Sigma^c = \begin{bmatrix} 2.0 & 1.0 & -1.0 \\ 1.0 & 2.0 & 0.0 \\ -1.0 & 0.0 & 2.0 \end{bmatrix}.$$

A Beta(1,1) hyperprior distribution (i.e, Uniform(0,1)) is used for q_1, q_2 and q_3 (i.e., $a_j = b_j = 1$), a N(0,1000) for each α_j (i.e., σ_j^2=1000) and an Inverse Wishart (5,5 I_3), weakly informative, for Σ^c. As mentioned before, only aggregate data on market shares and number of *redeemed* coupons for each brand in each period are used to estimate the posterior distribution of the parameters of the model (i.e., we do not use data on how many consumers received a coupon for each brand in each period and, of course, any individual-level data). The starting values for $\bar{\theta}$, D and Z are the same as those used in the previous section, while the initial values for α_j and Σ^c correspond to $\alpha_j = 0$ and $\Sigma^c = I_3$. In the case of the coupon variables (C), the initial value for N^c_{jt} is set to be equal to the integer part of N^r_{jt} +0.3 (N-N^r_{jt}), and then these N^c_{jt} coupons are randomly assigned among the N consumers. Using the method described in this section, the results presented in Table W2 were obtained, where the results are again based on a single run of 200,000 iterations with the last 100,000 used for estimation.

From the results it is observed that the true values of $\bar{\theta}$, D, α, q and Σ^c are covered by their 95% posterior-probability intervals and that the posterior means and posterior medians are very close to the true values (all within 1 posterior standard deviation).

NUMERICAL EXAMPLE –STRUCTURAL DEMAND MODEL

A numerical simulation example is constructed where consumers choose among three brands and a no-purchase alternative. The utility function includes three brand intercepts and a covariate ($x_{4,jt}$) generated from a standard normal distribution. The true values for $\bar{\theta}$ and D correspond to (1,1,1,-1,1) and I_5. In terms of the formulation of the coupon probabilities, m_{jt} includes $x_{4,jt}$ (the value of the fourth explanatory variable for brand j in period t) and we set α = (-2,-1,0,2), ρ = 1 and ϕ = (1,0,-1). In addition, we use the same true values for q and Σ^c as in the previous section. We constrain δ_{j1} to be equal to 1 for every brand in order to obtained meaningful estimation results for each of the components of ϕ (in our real data analyses in the next section this is not required). The coefficients capturing the autocorrelation in the demand and price shocks are defined as follows: $\gamma_p = (.2, 0,-.3)$ and $\gamma_d = (0,.3,-.2)$. In addition, υ_1 =

(.5,1)', υ_2 = (.5,2)' and υ_3 = (.5,.5)' and Σ^d is specified as follows:

$$\Sigma^d = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.30 & 0.20 & -0.20 \\ 0.00 & 1.00 & 0.00 & -0.20 & 0.30 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.20 & 0.30 \\ 0.30 & -0.20 & 0.00 & 0.25 & 0.00 & 0.00 \\ 0.20 & 0.30 & 0.20 & 0.00 & 0.25 & 0.00 \\ -0.20 & 0.00 & 0.30 & 0.00 & 0.00 & 0.25 \end{bmatrix}.$$

Finally, the hyperprior distributions for these parameters are $\alpha_{J+1} \sim N(0,10)$, $\alpha_{J+1} \sim N(0,10)$, $\rho \sim N(0,1000)$, $\phi_j \sim N(0,10)$, $\upsilon_j \sim N(0_2, 10^5 I_2)$, $\gamma_{pj} \sim N(0,1)$, $\gamma_{dj} \sim N(0,1)$ and $\Sigma^d \sim$ Inverse Wishart (5, 5 (.1 I_6)), while all other parameters have the same hyperprior distributions as in the numerical experiment in the previous section. Tables W3 and W4 present the results, which are based on a single run of 600,000 iterations with the last 200,000 used for estimation, a longer run than before as the model is more complicated. From the results it is observed that the true values of $\bar{\theta}$, D, q, α, ϕ, ρ, Σ^c, υ, γ and Σ^d are covered by their 95% posterior-probability intervals (except for υ_{32}, $\Sigma^d{}_{16}$, $\Sigma^d{}_{23}$ and $\Sigma^d{}_{34}$) and that the true values are in most cases within one posterior standard deviation around their posterior means. In addition, the posterior mean and median of all of the non-zero off-diagonal elements of Σ^d and the non-zero elements of γ have the right sign.

GIBBS SAMPLER FOR STRUCTURAL DEMAND MODEL

In this section, we describe the procedure to sample each of the parameters of the most general version of the model from their full-conditional posterior distribution. We note that in the case of the full-conditional posterior simulation of θ_i, ξ_t and ν_t we use a MH step with candidate vectors generated from a normal distribution with mean equal to the current value (random walk) and a variance matrix proportional to the current value of D, Σ^d and Σ^c, respectively (see C.3, C.4 and C.5 for details).

a) Sampling Choices (y_{it}):

1. In each iteration (k) randomly select $N/2$ pairs of consumers without replacement and enumerate these pairs. Let (i_{1p}, i_{2p}) be the indices of consumers in pair p and $(z_{i_{1p}t}^{(k)}, z_{i_{2p}t}^{(k)})$ their choices in period t in the current iteration k.

2. For each period t and starting from the first pair, successively and jointly draw the choices of each pair of consumers $(z_{i_{1p}t}^{(k+1)}, z_{i_{2p}t}^{(k+1)})$ from their full-conditional posterior distribution. The pair (p) subscript is dropped for notational convenience. In addition, define $\omega_i = 1$ if $\psi_i > 0$ and, otherwise $\omega_i = 0$. We proceed by assigning $(z_{i_1t}^{(k+1)}, z_{i_2t}^{(k+1)}) = (z_{i_2t}^{(k)}, z_{i_1t}^{(k)})$ according to the following probability:

$$\text{(W1)}\quad f\left((z_{i_1t}^{(k+1)}, z_{i_2t}^{(k+1)}) = (z_{i_2t}^{(k)}, z_{i_1t}^{(k)}) \mid *\right) = \begin{cases} \dfrac{\mathrm{I}_{\left\{\omega_{i_1} c_{i_1j} z_{i_1j}^{(k)} + \omega_{i_2} c_{i_2j} z_{i_2j}^{(k)} = \omega_{i_1} c_{i_1j} z_{i_2j}^{(k)} + \omega_{i_2} c_{i_2j} z_{i_1j}^{(k)}\right\}}}{1 + \dfrac{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_2jt}^{(k)}} p_{i_2jt}^{z_{i_1jt}^{(k)}} h\left(c_{i_1jt+1} | c_{i_1jt}, z_{i_2jt}^{(k)}\right) h\left(c_{i_2jt+1} | c_{i_2jt}, z_{i_1jt}^{(k)}\right)}{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_1jt}^{(k)}} p_{i_2jt}^{z_{i_2jt}^{(k)}} h\left(c_{i_1jt+1} | c_{i_1jt}, z_{i_1jt}^{(k)}\right) h\left(c_{i_2jt+1} | c_{i_2jt}, z_{i_2jt}^{(k)}\right)}}, & 1 \le t \le T-1; \\[4ex] \dfrac{\mathrm{I}_{\left\{\omega_{i_1} c_{i_1j} z_{i_1j}^{(k)} + \omega_{i_2} c_{i_2j} z_{i_2j}^{(k)} = \omega_{i_1} c_{i_1j} z_{i_2j}^{(k)} + \omega_{i_2} c_{i_2j} z_{i_1j}^{(k)}\right\}}}{1 + \dfrac{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_2jt}^{(k)}} p_{i_2jt}^{z_{i_1jt}^{(k)}}}{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_1jt}^{(k)}} p_{i_2jt}^{z_{i_2jt}^{(k)}}}}, & \mathrm{t} = \mathrm{T}, \end{cases}$$

where $h(\cdot\,|\,\cdot,\cdot)$ is the likelihood contribution of next-period coupons based on current coupons and choices. This function is defined as follows:

$$h(c_{ijt+1}\,|\,c_{ijt},z_{ijt}) = \left(r_{ijt+1}(c_{ijt},z_{ijt})\right)^{c_{ijt+1}}\left(1-r_{ijt+1}(c_{ijt},z_{ijt})\right)^{1-c_{ijt+1}},\ t=1,..,T-1.$$

Finally, with the complement of the probability defined in equation (W1), the pair of choices remain at their current values by assigning: $(z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_1}^{(k)}, z_{i_2}^{(k)})$.

b) Sampling Coupons (c_{it}):

In every iteration k, for each period t and for every consumer i, successively draw $c_{it}^{(k+1)}$ as follows:

1. Let b_i the brand chosen by consumer i in period t (i.e., $z_{ib_it}=1$).
2. Let c_{it}^* be such that:
    (a) If $\omega_i=1$, $c_{ib_it}^* = c_{ib_it}^{(k)}$ (this condition is required in order to satisfy condition (14)), and define the set $Q_i=\{1,...,J\}\setminus\{b_i\}$, otherwise, if $\omega_i=0$ define $Q_i=\{1,...,J\}$.
    (b) If $\delta_{b't}=1$, generate $c_{ib't}^*$ from a Bernoulli distribution with probability 0.5, for all $b\in Q_i$; otherwise, set $c_{ib't}^*=0$.
3. Accept c_{it}^*, according to the following MH probability:

$$P\left(c_{it}^{(k+1)} = c_{it}^*\right) = \begin{cases} \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^*)^{z_{ijt}}\ h(c_{ijt+1}|c_{ijt}^*,z_{ijt})\, r_{ij1}^{c_{ijt}^*}\,(1-r_{ij1})^{(1-c_{ijt}^*)}}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}}\ h(c_{ijt+1}|c_{ijt}^{(k)},z_{ijt})\, r_{ij1}^{c_{ijt}^{(k)}}\,(1-r_{ij1})^{(1-c_{ijt}^{(k)})}}, & t=1; \\[2ex] \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^*)^{z_{ijt}}\ h(c_{ijt+1}|c_{ijt}^*,z_{ijt})\, h(c_{ijt}^*|c_{ijt-1},z_{ijt-1})}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}}\ h(c_{ijt+1}|c_{ijt}^{(k)},z_{ijt})\, h(c_{ijt}^{(k)}|c_{ijt-1},z_{ijt-1})}, & 2\le t\le T-1; \\[2ex] \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^*)^{z_{ijt}}\ h(c_{ijt}^*|c_{ijt-1},z_{ijt-1})}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}}\ h(c_{ijt}^{(k)}|c_{ijt-1},z_{ijt-1})}, & t=T, \end{cases}$$

otherwise, assign $c_{it}^{(k+1)} = c_{it}^{(k)}$.

Also note that the acceptance rate of candidate vectors of coupons can be increased if some elements of Q_i are removed at random in each iteration (i.e., if the value of c_{ijt} is left constant for some components in a given iteration). For example, for the data set used in the empirical application, if four elements are removed at random from Q_i, then the acceptance rate exhibits values closer to 25% instead of values below 10%.

c) Sampling θ_i:

Define $c_{ijt}^r = c_{ijt} z_{ijt} \mathrm{I}_{\{\psi_i > 0\}}$. In every iteration k, for each consumer i, successively draw θ_i using a Metropolis-Hastings step specified as follows:

1. Let $l_i = 0$, if $\max\limits_{j,t} c_{ijt}^r = 1$; otherwise, set $l_i = -\infty$.
2. Let $u_i = 0$, if $\max\limits_{j,t} c_{ijt}^r = 0$ and $\max\limits_{j,t} z_{ijt} c_{ijt} = 1$; otherwise, set $u_i = +\infty$.
3. Generate a candidate value ψ_i^* from a normal distribution with mean equal to $\psi_i^{(k)}$ (the value of ψ_i in the current iteration), variance equal to $a_\psi \cdot D_\psi^{(k)}$ and truncated in the interval (l_i, u_i). We note that D_ψ is the diagonal element of D that corresponds to the coupon redemption utility coefficient (ψ_i), while a_ψ is a scalar tuning parameter for the jumping kernel used in the MH step.
4. Generate a candidate value φ_i^* from a normal distribution centered on the current value (φ_i^k) and with variance matrix equal to $a_\varphi D_\varphi^{(k)}$. We note that $D_\varphi^{(k)}$ is the current value of the variance-covariance matrix of φ_i and a_φ is a scalar tuning parameter for this jumping kernel. In our simulation experiment we used $a_\varphi = 0.28$ and $a_\psi = 6 \cdot 0.28$ in order to obtain a MH acceptance rate for θ_i between 20% and 30%.
5. Define $\theta_i^* = (\varphi_i^{*'}, \psi_i^*)'$.

6. Accept θ_i^* with the following MH probability:

$$\alpha_{MH,\theta_i} = \min\left\{ \frac{\dfrac{\phi(\theta_i^*;\bar{\theta},D)\prod_{j=1}^{J}\prod_{t=1}^{T} p_{ijt}(\theta_i^*)^{z_{ijt}}}{\phi(\psi_i^*;\psi^{(k)},a_\psi D_\psi,l_i,u_i)}}{\dfrac{\phi(\theta_i^{(k)};\bar{\theta},D)\prod_{j=1}^{J}\prod_{t=1}^{T} p_{ijt}(\theta_i^{(k)})^{z_{ijt}}}{\phi(\psi^{(k)};\psi_i^*,a_\psi D_\psi,l_i,u_i)}},1 \right\},$$

otherwise assign $\theta_i^{(k+1)} = \theta_i^{(k)}$, where $\phi(\,\cdot\,;\psi_i,a_\psi D_\psi,l_i,u_i)$ denotes the density of a normal distribution with mean ψ_i and variance $a_\psi D_\psi$ truncated in (l_i,u_i); and $p_{ijt}(\theta_i)$ denotes the probability that consumer i chooses brand j in period t when her vector of preference coefficients is equal to θ_i.

d) Sampling $\tilde{\xi}_t$ and ξ_t:

1. Let $\bar{\xi}_t = \Sigma^d_{\xi,\eta}(\Sigma^d_{\eta,\eta})^{-1}\eta_t$ and let $\Sigma^d_{\xi|\eta} = \Sigma^d_{\xi,\xi} - \Sigma^d_{\xi,\eta}(\Sigma^d_{\eta,\eta})^{-1}\Sigma^d_{\eta,\xi}$, where $\eta_t = \text{price}_{jt} - w_{jt}\upsilon_j$.
2. Generate $\tilde{\xi}_t^*$ from a normal distribution with mean $\tilde{\xi}_t^{(k)}$ and variance matrix $a_\xi \Sigma^d_{\xi|\eta}$, where a_ξ is a one-dimensional MH tuning parameter (in our numerical experiment we used $a_\xi = 0.55$).
3. Let $\xi_{j1}^* = \tilde{\xi}_{j1}^* / \sqrt{1-\gamma_{dj}^2}$ and $\xi_{jt}^* = \gamma_{dj}\xi_{jt-1}^* + \tilde{\xi}_{jt}^*$.
4. Accept ξ_t^* and $\tilde{\xi}_t^*$ with the following MH probability:

$$\alpha_{MH,\xi_t} = \min\left\{ \frac{\phi(\tilde{\xi}_t^*;\bar{\xi}_t,\Sigma^d_{\xi|\eta})\prod_{i=1}^{N}\prod_{j=1}^{J} p_{ijt}(\xi_t^*)^{z_{ijt}}}{\phi(\tilde{\xi}_t^{(k)};\bar{\xi}_t,\Sigma^d_{\xi|\eta})\prod_{i=1}^{N}\prod_{j=1}^{J} p_{ijt}(\xi_t^{(k)})^{z_{ijt}}},1 \right\},$$

otherwise assign $\xi_t^{(k+1)} = \xi_t^{(k)}$ and $\tilde{\xi}_t^{(k+1)} = \tilde{\xi}_t^{(k)}$, where $p_{ijt}(\xi_t)$ denotes the probability that consumer i chooses brand j in period t when the vector of common demand shocks is equal to ξ_t.

e) Sampling ν_t:

1. Generate ν_t^* from a normal distribution centered on $\nu_t^{(k)}$ and with variance matrix $a_\nu \Sigma^c$, where Σ^c is the current value of the covariance matrix of ν_t (we used $a_\nu = 0.4$ in the numerical experiment).
2. Accept ν_t^* with the following MH probability:

$$\alpha_{MH,\nu_t} = \min\left\{ \frac{\phi(\nu_t^*;0,\Sigma^c)\prod_{i=1}^{N}\prod_{j=1}^{J} r_{ijt}(\nu_t^*)^{c_{ijt}}\left(1-r_{ijt}(\nu_t^*)\right)^{1-c_{ijt}}}{\phi(\nu_t^{(k)};0,\Sigma^c)\prod_{i=1}^{N}\prod_{j=1}^{J} r_{ijt}(\nu_t^{(k)})^{c_{ijt}}\left(1-r_{ijt}(\nu_t^{(k)})\right)^{1-c_{ijt}}}, 1 \right\},$$

otherwise assign $\nu_t^{(k+1)} = \nu_t^{(k)}$, where $r_{ijt}(\cdot)$ denotes the probability that consumer i has a coupon available for brand j in period t as a function of the value of ν_t.

f) Sampling δ_t:

1. If in iteration k there is a positive number of coupons assigned to brand j in period t (i.e., if $\sum_{i=1}^{N} c_{ijt} > 0$), then set $\delta_{jt}^{(k+1)} = 1$ (see condition (11)).
2. If no coupons are assigned, set $\delta_{jt}^{(k+1)} = 1$ with the following probability:

$$P(\delta_{jt}^{(k+1)} = 1 \mid *) = \frac{\left(\frac{1}{1+e^{\alpha_j+\nu_{jt}+\rho m_{jt}}}\right)^N q_j}{\left(\frac{1}{1+e^{\alpha_j+\nu_{jt}+\rho m_{jt}}}\right)^N q_j + (1-q_j)},$$

otherwise, set $\delta_{jt}^{(k+1)} = 0$.

g) Sampling $\bar{\theta}$:

1. Define $A_{\bar{\theta}}$ and $B_{\bar{\theta}}$ as follows:

$$A_{\bar{\theta}} = B_{\bar{\theta}}(D^{-1}\sum_{i=1}^{N}\theta_i)$$

$$B_{\bar{\theta}} = \left(V_0^{-1} + ND^{-1}\right)^{-1}$$

where $V_{0,\bar{\theta}}^{-1}$ is the inverse of the prior variance for $\bar{\theta}$ ($V_{0,\bar{\theta}} = 10^5 \mathrm{I}_5$ in the simulation experiment).

2. Generate $\bar{\theta}$ from a normal distribution with mean $A_{\bar{\theta}}$ and variance $B_{\bar{\theta}}$.

h) Sampling D:

1. Let K be equal to the number of columns of D.
2. Generate D from an IW $(K+2+N,(K+2)I_K+\sum_{i=1}^{N}(\theta_i-\bar{\theta})(\theta_i-\bar{\theta})')$.

i) Sampling $\alpha_1,\ldots,\alpha_J$ and ρ:

1. Let $\overset{\square}{\alpha}=(\alpha_1,\ldots,\alpha_J,\rho)'$.
2. Let $X_{c,t}$ be a matrix such that $X_{c,t}=[\mathrm{I}_J\ m_t]$, where I_J denotes an identity matrix with J rows and columns, while m_t is a matrix that contains the values in period t of marketing variables assumed to be coordinated with coupon promotion.
3. Let $E_t=x_{c,t}\overset{\square}{\alpha}+v_t$.
4. Let $A_{\overset{\square}{\alpha}}=B_{\overset{\square}{\alpha}}\left(\sum_{t=1}^{T}X_{c,t}(\Sigma^c)^{-1}E_t\right)$ and $B_{\overset{\square}{\alpha}}=\left(V_{0,\overset{\square}{\alpha}}^{-1}+\sum_{t=1}^{T}X_{c,t}(\Sigma^c)^{-1}X_{c,t}\right)^{-1}$, where $V_{0,\overset{\square}{\alpha}}$ is the prior variance of $\overset{\square}{\alpha}$ ($V_{0,\overset{\square}{\alpha}}=10^4\mathrm{I}_4$ in the simulation experiment).
5. Generate $\overset{\square}{\alpha}^{(k+1)}$ from a normal distribution with mean $A_{\overset{\square}{\alpha}}$ and variance $B_{\overset{\square}{\alpha}}$ and set $v_t=E_t-x_{c,t}\overset{\square}{\alpha}^{(k+1)}$.

j) Sampling α_{J+1}:

1. Generate α_{J+1}^* from a normal distribution with mean $\alpha_{J+1}^{(k)}$ and variance $\sigma_{\alpha_{J+1}}^2$.
2. Accept α_{J+1}^* with the following MH probability:

$$\alpha_{MH,\alpha_{J+1}}=\min\left\{\frac{\phi(\alpha_{J+1}^*;0,V_{0,\alpha_{J+1}})\prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t=2}^{T}r_{ijt}(\alpha_{J+1}^*)^{c_{ijt}}(1-r_{ijt}(\alpha_{J+1}^*))^{1-c_{ijt}}}{\phi(\alpha_{J+1}^{(k)};0,V_{0,\alpha_{J+1}})\prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{t=2}^{T}r_{ijt}(\alpha_{J+1}^{(k)})^{c_{ijt}}(1-r_{ijt}(\alpha_{J+1}^{(k)}))^{1-c_{ijt}}},1\right\},$$

3.

otherwise, let $\alpha_{J+1}^{(k+1)} = \alpha_{J+1}^{(k)}$ (in the numerical experiment, we set $V_{0,\alpha_{J+1}} = 10$).

k) Sampling q:

Generate q_j from the following beta distribution: $\text{Beta}\left(1+\sum_{t=1}^{T}\delta_{jt}, 1+T-\sum_{t=1}^{T}\delta_{jt}\right)$.

l) Sampling υ:

1. Let $\upsilon = (\upsilon_1, .., \upsilon_J)'$.
2. Let $zp_{j1} = \sqrt{1-\gamma_{pj}}\,(\text{price}_{j1} - \Sigma^d_{\text{price},\xi}(\Sigma^d_{\xi,\xi})^{-1}\xi_1)$ and let

   $zp_t = \text{price}_t - \gamma_p \text{price}_{t-1} - \Sigma^d_{\text{price},\xi}(\Sigma^d_{\xi,\xi})^{-1}\xi_t$ for $t \geq 2$.
3. Let $\Sigma^d_{\text{price},\text{price}|\xi} = \Sigma^d_{\text{price},\text{price}} - \Sigma^d_{\text{price},\xi}(\Sigma^d_{\xi,\xi})^{-1}\Sigma^d_{\xi,\text{price}}$.
4. Let $\tilde{w}_{j1} = \sqrt{1-\gamma_{pj}}\, w_{j1}$ and $\tilde{w}_t = w_{jt} - \gamma_{pj} w_{jt-1}$ for $t \geq 2$.
5. Let W_t a block-diagonal matrix, where the j^{th} block corresponds to $\tilde{w}_{jt}$, for $j = 1, .., J$.
6. Let $A_\upsilon = B_\upsilon\left(\sum_{t=1}^{T} W_t(\Sigma^d_{\text{price},\text{price}|\xi})^{-1} zp_t\right)$ and $B_\upsilon = \left(V_{0,\upsilon}^{-1} + \sum_{t=1}^{T} W_t(\Sigma^d_{\text{price},\text{price}|\xi})^{-1} W_t\right)^{-1}$, where $V_{0,\upsilon}$

   is the prior variance of υ ($V_{0,\upsilon} = 100\text{I}_4$ in the simulation experiment).
7. Generate $\upsilon^{(k+1)}$ from a normal distribution with mean A_υ and variance B_υ.

m) Sampling Σ^d:

1. Let $\tilde{\zeta}_t = (\tilde{\eta}_t, \tilde{\xi}_t)'$.
2. Generate Σ^d from an $\text{IW}(J+2+T, (J+2)\ 0.01\ \text{I}_{2J} + \sum_{t=1}^{T}\tilde{\zeta}_t\tilde{\zeta}_t)$.

n) Sampling Σ^c:

Generate Σ^c from an $\text{IW}(J+2+T-1, (J+2)\ \text{I}_J + \sum_{t=2}^{T}\nu_t\nu_t)$.

o) Sampling γ:

1. Let $\zeta_t = (\eta_t, \xi_t)'$.
2. Let $\tilde{\zeta}_t = (\tilde{\eta}_t, \tilde{\xi}_t)'$.
3. Let D_t a diagonal matrix with diagonal elements equal to the components of ζ_t.
4. Let $A_\gamma = B_\gamma \left(\sum_{t=1}^{T-1} D_t (\Sigma^d)^{-1} \zeta_{t+1} \right)$ and $B_\upsilon = \left(V_{0,\gamma}^{-1} + \sum_{t=1}^{T} D_t (\Sigma^d)^{-1} D_t \right)^{-1}$, where $V_{0,\gamma}$ is the prior variance of γ ($V_{0,\gamma}$ equals the identity matrix in the simulation experiment).
5. Generate γ^* from a normal distribution with mean A_γ and variance B_γ.
6. Let $\tilde{\eta}^*_{j1} = \sqrt{1-(\gamma^*_{pj})^2}\ \eta_{j1}$, $\tilde{\xi}^*_{j1} = \sqrt{1-(\gamma^*_{dj})^2}\ \xi_{j1}$. Let $\tilde{\zeta}^*_1 = (\tilde{\eta}^{*\prime}_1, \tilde{\xi}^{*\prime}_1)'$.
7. Accept γ^* according to the following MH probability:

$$\alpha_{MH,\gamma} = \min \left\{ \frac{\phi(\tilde{\zeta}^*_1; 0, \Sigma^d)}{\phi(\tilde{\zeta}^{(k)}_1; 0, \Sigma^d)}, 1 \right\},$$

otherwise, set $\gamma^{(k+1)} = \gamma^{(k)}$.

ESTIMATION OF THE MARGINAL LIKELIHOOD

In what follows we derive an estimator of the marginal likelihood by generalizing the harmonic mean method proposed by Newton and Raftery (1994). This generalization is needed for the aggregate estimation procedures presented in this paper that are based on augmenting the aggregate data (A) with unobserved sequences of choices (Z) and coupons (C).

As before, define $\omega_i = 1$ if $\psi_i > 0$ and, otherwise $\omega_i = 0$. Let ω define a vector with the values of ω_i for all consumers. In addition, Let Ω_M denote the set of all values of (Z, C, ω) consistent with the aggregate data (A) under model M and let ϕ denote the collection of parameters that determine the likelihood of the augmented choices and coupons (i.e., $\phi = \{\theta, r\}$). We are interested in computing $p(A|M)$, the marginal likelihood of the aggregate data A under model M. For notational convenience, we drop the model subscript (M) and we refer to $p(A|M)$ and Ω_M simply as $p(A)$ and Ω, respectively. By noting that $\int p(\phi)d\phi = 1$, it is straightforward to verify that the marginal likelihood $p(A)$ satisfies the following equation:

(W2)
$$\frac{1}{p(A)} = \frac{1}{|\Omega|} \sum_{(Z,C,\omega)\in\Omega} \int \frac{p(\phi)}{p(A)} d\phi.$$

Using Bayes Law and noting that $p(A|Z,C,\omega,\phi) = 1$ for any pair $(Z,C,\omega) \in \Omega$, the following identity can be easily derived:

(W3)
$$\frac{1}{p(A)} = \frac{p(Z,C,\omega,\phi|A)}{p(Z,C,\omega,\phi)}, \quad \forall (Z,C,\omega) \in \Omega.$$

Using this identity in equation (W2) we obtain:

$$\begin{aligned} \frac{1}{p(A)} &= \frac{1}{|\Omega|} \sum_{(Z,C,\omega)\in\Omega} \int \frac{p(\phi)}{p(Z,C,\omega,\phi)} p(Z,C,\omega,\phi|A)d\phi \\ &= \frac{1}{|\Omega|} \sum_{(Z,C,\omega)\in\Omega} \int \frac{1}{p(Z,C,\omega|\phi)} p(Z,C,\omega,\phi|A)d\phi \\ &= \frac{1}{|\Omega|} \mathrm{E}\left[\frac{1}{p(Z,C,\omega|\phi)} | A \right]. \end{aligned}$$

(W4)

Consequently, using equation (W4) we can estimate $p(A)$ as follows:

(W5)
$$\hat{p}(A) = \frac{|\Omega|}{\frac{1}{m}\sum_{l=1}^{m} \frac{1}{p(Z^{(l)},C^{(l)},\omega^{(l)}|\phi^{(l)})}},$$

where each quadruplet $\left(Z^{(l)}, C^{(l)}, \omega^{(l)}, \phi^{(l)}\right)$ is drawn from the posterior distribution $p(Z, C, \omega, \phi \mid A)$. Therefore, this estimator corresponds to the harmonic mean of the likelihood of the augmented choices and coupons amplified by $|\Omega|$, where the values for $\left(Z^{(l)}, C^{(l)}, \omega^{(l)}, \phi^{(l)}\right)$ can be obtained from the MCMC output.

Finally, we note that if two models M_1 and M_2 share the same set of feasible combinations of choices and coupons (i.e., $\Omega_{M_1} = \Omega_{M_2} = \Omega$), then for the purposes of model selection, it is not necessary to compute $|\Omega|$, which is constant for these two models and, thus, it is not needed to compute the corresponding Bayes factors.

REFERENCES

Newton, Michael A. and Adrian E. Raftery (1994), “Approximating Bayesian Inference with the Weighted Likelihood Bootstrap,” *Journal of the Royal Statistical Society (B)*, 56: 3-48.

TABLE W1: NUMERICAL EXAMPLE RESULTS - BASIC MODEL

Estimated posterior mean, standard deviation and
quantiles for $\bar{\theta}$ and D.

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	D_{11}	D_{22}	D_{33}	D_{44}	D_{12}	D_{13}	D_{14}	D_{23}	D_{24}	D_{34}
M	1.000	1.005	-1.036	1.020	.996	1.159	1.003	1.006	.227	.112	.178	-.013	.001	-.117
SD	.076	.077	.069	.076	.434	.436	.145	.413	.248	.094	.231	.097	.255	.153
2.5%	.868	.869	-1.190	.881	.423	.508	.773	.459	-.150	-.069	-.228	-.206	-.440	-.431
50.0%	.993	1.000	-1.031	1.014	.915	1.095	.983	.910	.193	.111	.158	-.012	-.006	-.117
97.5%	1.165	1.171	-.917	1.186	2.167	2.272	1.344	2.058	.833	.306	.676	.176	.586	.182
True Values	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	.000	.000	.000	.000	.000	.000

TABLE W2: NUMERICAL EXAMPLE RESULTS - LIMITED INFORMATION

Estimated posterior mean, standard deviation and

quantiles for $\bar{\theta}$, D, q, α and Σ^c.

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	D_{11}	D_{22}	D_{33}	D_{44}	D_{12}	D_{13}	D_{14}	D_{23}	D_{24}	D_{34}
M	1.064	1.008	-.969	.925	1.075	1.093	.917	1.180	-.011	.057	.109	-.006	.148	.186
SD	.072	.072	.063	.084	.416	.374	.128	.409	.197	.103	.249	.107	.202	.201
2.5%	.933	.876	-1.100	.765	.496	.511	.695	.594	-.373	-.149	-.388	-.222	-.259	-.200
50.0%	1.061	1.005	-.966	.924	.982	1.050	.906	1.123	-.024	.058	.117	-.005	.141	.176
97.5%	1.218	1.161	-.854	1.094	2.126	1.933	1.196	2.161	.441	.257	.585	.196	.560	.597
True Values	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	.000	.000	.000	.000	.000	.000

	q_1	q_2	q_3	α_1	α_2	α_3	Σ^c_{11}	Σ^c_{22}	Σ^c_{33}	Σ^c_{12}	Σ^c_{13}	Σ^c_{23}
M	.408	.560	.655	-2.049	-1.095	.182	1.704	2.397	1.584	.846	-.278	-.691
SD	.068	.069	.065	.284	.301	.219	.627	.754	.450	.619	.380	.524
2.5%	.280	.424	.522	-2.597	-1.637	-.264	.864	1.336	.939	-.309	-1.081	-1.776
50.0%	.408	.560	.657	-2.058	-1.105	.174	1.576	2.261	1.508	.816	-.259	-.681
97.5%	.543	.692	.779	-1.477	-.483	.628	3.288	4.247	2.677	2.154	.437	.321
True Values	.400	.500	.600	-2.000	-1.000	.000	2.000	2.000	2.000	1.000	.000	-1.000

TABLE W3: NUMERICAL EXAMPLE RESULTS I - STRUCTURAL DEMAND MODEL

Estimated posterior mean, standard deviation and

quantiles for $\bar{\theta}$, D, α, ρ, ϕ, q and Σ^c.

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	$\bar{\theta}_5$	D_{11}	D_{22}	D_{33}	D_{44}	D_{55}	D_{12}	D_{13}	D_{14}
M	.972	.955	.957	-1.086	1.067	1.641	1.287	1.968	1.183	1.137	-.120	-.091	-.003
SD	.153	.191	.157	.107	.171	.783	.572	1.148	.239	.382	.456	.526	.216
2.5%	.664	.538	.620	-1.328	.785	.589	.562	.711	.832	.580	-.931	-1.137	-.449
50.0%	.971	.967	.963	-1.076	1.049	1.464	1.162	1.675	1.145	1.066	-.148	-.098	.013
97.5%	1.276	1.308	1.242	-.906	1.456	3.512	2.826	5.179	1.774	2.009	.919	1.102	.402
True Values	1.000	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	1.000	.000	.000	.000

	D_{15}	D_{23}	D_{24}	D_{25}	D_{34}	D_{35}	D_{45}	α_1	α_2	α_3	α_4	ρ
M	.174	-.085	.009	-.307	-.308	.071	.037	-2.459	-1.014	-.051	1.488	-.837
SD	.349	.412	.215	.331	.252	.401	.223	.594	.379	.358	.976	.113
2.5%	-.490	-1.101	-.430	-1.088	-.907	-1.031	-.437	-3.935	-1.766	-.738	-.363	-1.064
50.0%	.146	-.046	.007	-.298	-.276	.118	.045	-2.399	-1.017	-.060	1.472	-.834
97.5%	.909	.669	.416	.305	.078	.714	.485	-1.483	-.259	.683	3.548	-.621
True Values	.000	.000	.000	.000	.000	.000	.000	-2.000	-1.000	.000	2.000	-1.000

	ϕ_1	ϕ_2	ϕ_3	q_1	q_2	q_3	Σ^c_{11}	Σ^c_{22}	Σ^c_{33}	Σ^c_{12}	Σ^c_{13}	Σ^c_{23}
M	1.311	.444	-.749	.406	.508	.599	2.941	2.774	1.847	1.111	-.500	-1.397
SD	.625	.526	.400	.089	.075	.068	1.644	1.086	.653	.872	.724	.682
2.5%	.238	-.595	-1.539	.252	.362	.464	1.221	1.321	.955	-.303	-1.831	-2.912
50.0%	1.258	.448	-.744	.399	.507	.601	2.517	2.559	1.721	1.010	-.539	-1.330
97.5%	2.792	1.480	.035	.603	.654	.727	7.340	5.479	3.483	3.165	1.136	-.225
True Values	1.000	.000	-1.000	.400	.500	.600	2.000	2.000	2.000	1.000	.000	-1.000

TABLE W4: NUMERICAL EXAMPLE RESULTS II - STRUCTURAL DEMAND MODEL

Estimated posterior mean, standard deviation and
quantiles for υ, Σ^d and γ.

	υ_{11}	υ_{12}	υ_{21}	υ_{22}	υ_{31}	υ_{32}	Σ^d_{11}	Σ^d_{22}	Σ^d_{33}	Σ^d_{44}	Σ^d_{55}	Σ^d_{66}	Σ^d_{12}	Σ^d_{13}
M	.610	1.044	.618	2.093	.377	.306	1.461	1.442	.725	.379	.399	.353	-.216	-.213
SD	.208	.111	.166	.131	.095	.087	.315	.308	.159	.117	.127	.114	.219	.161
2.5%	.198	.826	.287	1.842	.184	.138	.971	.962	.480	.203	.215	.186	-.673	-.562
50.0%	.611	1.044	.617	2.091	.379	.305	1.419	1.402	.703	.361	.377	.334	-.208	-.204
97.5%	1.012	1.262	.952	2.355	.561	.479	2.194	2.158	1.102	.654	.707	.628	.198	.082
True Values	.500	1.000	.500	2.000	.500	.500	1.000	1.000	1.000	.250	.250	.250	.000	.000

	Σ^d_{14}	Σ^d_{15}	Σ^d_{16}	Σ^d_{23}	Σ^d_{24}	Σ^d_{25}	Σ^d_{26}	Σ^d_{34}	Σ^d_{35}	Σ^d_{36}	Σ^d_{45}	Σ^d_{46}	Σ^d_{56}
M	.427	.250	-.456	.340	-.425	.479	.104	-.206	.176	.231	-.033	-.123	-.043
SD	.154	.139	.147	.163	.145	.161	.123	.101	.100	.103	.073	.076	.073
2.5%	.175	.011	-.797	.056	-.755	.220	-.126	-.429	.003	.055	-.184	-.299	-.193
50.0%	.410	.239	-.437	.327	-.407	.459	.099	-.197	.167	.220	-.033	-.115	-.041
97.5%	.781	.559	-.221	.704	-.190	.855	.368	-.029	.397	.462	.112	.006	.102
True Values	.300	.200	-.200	.000	-.200	.300	.000	.000	.200	.300	.000	.000	.000

	γ_{p1}	γ_{p2}	γ_{p3}	γ_{d1}	γ_{d2}	γ_{d3}
M	.175	-.049	-.360	-.014	.303	-.176
SD	.112	.116	.171	.142	.121	.152
2.5%	-.052	-.280	-.708	-.306	.051	-.483
50.0%	.177	-.048	-.356	-.011	.307	-.177
97.5%	.392	.173	-.044	.256	.530	.122
True Values	.200	.000	-.300	.000	.300	-.200