Web Appendix

Retrieving Unobserved Consideration Sets from Household Panel Data

Erjen van Nierop, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses

Web Appendix A: Synthetic Data Analysis

In this appendix, we show the results of a synthetic data analysis. We simulate data in a four-brand market, with 2446 observations. The data is generated according to a two-stage model, consisting of a multivariate probit consideration set formulation, followed by a multinomial probit brand choice model. We replicate the simulation using four sets of data-generating parameters. We eliminate the empty consideration sets from the data to mimic what happens in extant consideration sets models, which allow for empty consideration sets but do not use empty sets observations. Averaged across the replications, 1796 observations remain.

On the resulting dataset, we estimate our model, which takes into account that empty consideration sets might occur, and compare it to a benchmark model that uses a multivariate probit model in the consideration set stage that does not take into account that empty consideration sets are present in practice. The goal of this analysis is twofold:

- To show that our model can recover the DGP parameters well;
- To show that our model outperforms competing models when empty consideration sets are present, but not observed as is the case in most scanner panel datasets.

Table W1 shows the result of these simulations.

Table W1

Results of synthetic data analysis. Displayed are recovered parameters for our model (left panel) and a model that does not take empty consideration sets into account (right panel).

	DGP	OUR MODEL		BENCHMARK MODEL	
		takes empty sets into account		does not take empty sets into account	
Variable	**true**	**M**	**SD**	**M**	**SD**
DGP 1					
Intercept brand 1	-0.788	-0.831	0.133	-0.769	0.110
Intercept brand 2	-0.152	-0.392	0.163	-0.596	0.094
Intercept brand 3	-1.120	-1.880	0.181	-2.020	0.137
Intercept brand 4	-0.837	-0.848	0.116	-0.393	0.090
Display	1.321	1.500	0.108	1.020	0.067
Feature	1.821	2.080	0.261	1.280	0.109
DGP 2					
Intercept brand 1	0.000	0.0462	0.116	-0.141	0.0974
Intercept brand 2	-0.100	0.261	0.180	-0.936	0.084
Intercept brand 3	-0.200	-0.234	0.125	-1.752	0.104
Intercept brand 4	-0.300	-0.215	0.132	-0.271	0.087
Display	1.000	1.310	0.123	0.607	0.059
Feature	1.000	1.560	0.241	0.541	0.091
DGP 3					
Intercept brand 1	-0.500	-0.607	0.157	-0.769	0.091
Intercept brand 2	0.100	0.040	0.175	-1.024	0.067
Intercept brand 3	-1.000	-1.401	0.241	-2.37	0.115
Intercept brand 4	0.500	0.238	0.148	0.432	0.092

	Display	2.000	1.750	0.198	0.988	0.069
	Feature	2.000	2.341	0.236	0.646	0.095
DGP 4						
	Intercept brand 1	-2.000	-2.102	0.169	-1.37	0.299
	Intercept brand 2	0.100	0.207	0.231	-0.486	0.161
	Intercept brand 3	-3.000	-3.373	0.296	-2.991	0.326
	Intercept brand 4	0.500	0.465	0.141	0.405	0.319
	Display	2.000	2.111	0.202	1.162	0.063
	Feature	3.000	3.442	0.455	1.663	0.121

The results show that the recovery of the brand consideration parameters is good for our proposed model. All 24 DGP-parameters are in the 99% HPD region of their respective estimates. For the model that does not allow empty consideration sets, this happens for only eight parameters.

Web Appendix B: Full Conditional Posterior Distributions

In this appendix we provide the full conditional posterior distributions and sampling algorithms of the model parameters and the unobserved utilities.

Sampling of α

To obtain the full conditional posterior distribution of α we rewrite (1) as

(1) $$\Sigma^{-\frac{1}{2}}(C_{it}^{*} - X_{it}\alpha_i) = \Sigma^{-\frac{1}{2}}X_{it}\alpha + \Sigma^{-\frac{1}{2}}\varepsilon_{it},$$

where $X_{it} = (X_{i1t},...,X_{iJt})'$, for i=1,...,I and t=1,...,T_i. We can interpret Equation 1 as J regression equations with regression coefficient α and uncorrelated normally distributed error terms with unit variance. In total we have $J \times \sum_{i=1}^{I} T_i$ of such regression equations. Hence, the full conditional posterior distribution of α given $\{\alpha_i\}_{i=1}^{I}$, Σ and C^* is normal. The mean and variance result from the OLS estimator of α in (1), see Zellner (1971, Chapter VIII).

Sampling of α_i

To sample α_i for i=1,...I we can follow a similar approach as for α. We rewrite Equation 1 as

(2) $$\begin{aligned} \Sigma^{-\frac{1}{2}}(C_{it}^{*} - X_{it}\alpha) &= \Sigma^{-\frac{1}{2}}X_{it}\alpha_i + \Sigma^{-\frac{1}{2}}\varepsilon_{it} \quad \text{for } t = 1,\ldots,T_i \\ \mathbf{0} &= \Sigma_{\alpha}^{-\frac{1}{2}}\alpha_i + \Sigma_{\alpha}^{-\frac{1}{2}}v_i. \end{aligned}$$

The last line follows from the fact that $\alpha_i \sim N(\mathbf{0},\Sigma_\alpha)$ can be written as $v_i = (\alpha_i - 0) \sim N(\mathbf{0},\Sigma_\alpha)$. This represents k_X+JT_i regression equations with regression coefficient α_i and uncorrelated normally distributed error terms with unit variance. Hence, the full conditional posterior distribution of α_i given α, Σ_α, Σ and C^* is normal. The mean and variance result from the OLS estimator of α_i in Equation 2.

Sampling of Σ_α

For the diagonal elements of Σ_α it holds that

(3) $$p(\sigma_{\alpha,kk}^{2} \mid \cdot) \propto (\sigma_{\alpha,kk}^{2})^{-(I+1)} \exp\left(-\frac{1}{2\sigma_{\alpha,kk}^{2}} \sum_{i=1}^{I} \alpha_{ik}^{2} \right),$$

for k = 1,...,k_X, where α_{ik} is the k^{th} element of α_i. Hence, the diagonal elements of Σ_α can be sampled according to

$$(4) \qquad \frac{\sum_{i=1}^{I} \alpha_{ik}^2}{\sigma_{\alpha,kk}^2} \sim \chi^2(I) \quad \text{for } k = 1, \ldots, k_X.$$

Sampling of Σ

To sample Σ we note that

$$(5) \qquad p(\Sigma \mid \cdot) \propto \pi(\Sigma) = |\Sigma|^{-\frac{1}{2}\sum_{i=1}^{I} T_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T_i}(C_{it}^* - X_{it}(\alpha + \alpha_i))'\Sigma^{-1}(C_{it}^* - X_{it}(\alpha + \alpha_i))\right),$$

for t=1,...,T_i and i=1,...,I.

As Σ is not a free covariance matrix (the diagonal elements are 1), the full conditional distribution is not inverted Wishart. In fact the full conditional posterior distribution of Σ is not standard. To sample Σ we propose a sampler based on Besag and Green (1993) and Damien, Wakefield and Walker (1999). Loosely speaking, this sampler interchanges the two steps in the Metropolis-Hastings sampler. A possible Metropolis-Hastings sampler for Σ is:

Step 1 Draw the elements of the matrix Σ from a uniform distribution on the interval (-1,1) under the restriction of positive definiteness, resulting in Σ^{new}.

Step 2 Draw u from a uniform distribution on the interval (0,1) and accept Σ^{new} if $\pi(\Sigma^{new})/\pi(\Sigma^{old}) > u$. Otherwise take $\Sigma^{new}=\Sigma^{old}$.

For the sampler used in this paper we interchange these two steps. We first draw u from a uniform distribution on the interval (0,1). In the second step we keep sampling candidate draws of the elements of Σ from a uniform distribution on the interval (-1,1) until Σ^{new} is positive definite and $\pi(\Sigma^{new})/\pi(\Sigma^{old}) > u$. The advantage of the latter approach is that it always results in a new draw, which is not the case for the Metropolis-Hastings sampler, see Damien, Wakefield and Walker (1999) for details. The disadvantage is that the sampler is slower as one has to draw new candidates until acceptance. Another possibility to generate Σ based on the Metropolis-Hastings sampler is given in Chib and Greenberg

(1998) or the hit-and-run algorithm in Manchanda et al. (1999).

Sampling of β

In the brand choice model, β is sampled in a similar way as α. We rewrite Equation 4 as

$$\omega_j^{-1}(U_{ijt} - W_{ijt}'\beta_i) = \omega_j^{-1}W_{ijt}'\beta + \omega_j^{-1}\eta_{ijt}, \tag{6}$$

for j=1,...,J, i=1,...,I and t=1,...,T_i. This represents $J \times \sum_{i=1}^{I} T_i$ regression equations with regression coefficient β and uncorrelated normally distributed error terms with unit variance. Hence, the full conditional posterior distribution of β given $\{\beta_i\}_{i=1}^{I}$, Ω and U is normal. The mean and variance result from the OLS estimator of β in (6), see again Zellner (1971, Chapter VIII).

Sampling of β_i

To sample β_i for i=1,...,I we can follow a similar approach as for β. We rewrite Equation 4 as

$$\begin{aligned} \omega_j^{-1}(U_{ijt} - W_{ijt}'\beta) &= \omega_j^{-1}W_{ijt}'\beta_i + \omega_j^{-1}\eta_{ijt} \\ \mathbf{0} &= \Sigma_\beta^{-\frac{1}{2}}\beta_i + \Sigma_\beta^{-\frac{1}{2}}v_i, \end{aligned} \tag{7}$$

for j=1,...,J, i=1,...,I and t=1,...,T_i. The last line follows from the fact that $\beta_i \sim N(\mathbf{0},\Sigma_\beta)$ can be written as $v_i = (\beta_i - 0) \sim N(\mathbf{0},\Sigma_\beta)$.

This represents k_W+JT_i regression equations with regression coefficient β_i and uncorrelated normally distributed error terms with unit variance. Hence, the full conditional posterior distribution of β_i given β, Σ_β, Ω, and U is normal. The mean and variance result from the OLS estimator of β_i in (7).

Sampling of Σ_β

For the diagonal elements of Σ_β it holds that

$$p(\sigma_{\beta,kk}^2 \mid \cdot) \propto (\sigma_{\beta,kk}^2)^{-(I+1)} \exp\left(-\frac{1}{2\sigma_{\beta,kk}^2} \sum_{i=1}^{I} \beta_{ik}^2 \right), \tag{8}$$

for k=1,...,k_W, where β_{ik} is the k^{th} element of β_i. Hence, the diagonal elements of Σ_β can be sampled according to

(9) $$\frac{\sum_{i=1}^{I}\beta_{ik}^2}{\sigma_{\beta,kk}^2} \sim \chi^2(I) \quad \text{for } k=1,\ldots,k_W.$$

Sampling of Ω

To sample the elements of the covariance matrix Ω we use that

(10) $$p(\omega_j^2 \mid \cdot) \propto (\omega_j^2)^{-(v+1)} \exp\left(-\frac{1}{2\omega_j^2}\sum_{i=1}^{I}\sum_{t=1}^{T_i}(U_{ijt} - W_{ijt}'(\beta+\beta_i))^2\right)$$

with $\nu = \sum_{i=1}^{I} T_i$ and hence

(11) $$\frac{\sum_{i=1}^{I}\sum_{t=1}^{T_i}(U_{ijt} - W_{ijt}'(\beta+\beta_i))^2}{\omega_j^2} \sim \chi^2(v)$$

for j=1,...,J-1.

Sampling of U

To sample U_{it} for i=1,...,I and t=1,...,T_i we consider

(12) $$U_{it} = W_{it}(\beta + \beta_i) + \eta_{it}$$

and hence U_{it} is normally distributed with mean $W_{it}(\beta + \beta_i)$ and covariance matrix Ω. The conditional distribution of U_{ijt} given $(U_{i1t},\ldots,U_{i,j-1,t}, U_{i,j+1,t},\ldots,U_{iJt})$ is of course also normal with, let's say, mean m_j and variance s_j^2. Hence, U_{ijt} can be sampled from truncated normal distributions in the following way

(13) $$U_{ijt} \mid \cdot \sim \begin{cases} N(m_j, s_j^2) \times I(-\infty, U_{i,d_{it},t}) & \text{if } j \neq d_{it} \wedge c_{ijt} = 1 \\ N(m_j, s_j^2) \times I(\max_{k|k\neq j}(U_{ikt} \mid c_{ikt} = 1), \infty) & \text{if } j = d_{it} \\ N(m_j, s_j^2) \times I(-\infty, \infty) & \text{if } c_{ijt} = 0, \end{cases}$$

for j=1,...,J, see Geweke (1991) for details. The value of c_{ijt} is of course determined by the values of the draws of C_{ijt}^*.

Sampling of C^*

To sample C_{it}^* for i=1,...,I and t=1,...,T_i we consider

(14) $$C^*_{it} = X_{it}(\alpha + \alpha_i) + \varepsilon_{it}$$

and hence C^*_{it} is normally distributed with mean $X_{it}(\alpha + \alpha_i)$ and covariance matrix Σ. The conditional distribution of C^*_{ijt} given $(C^*_{1it}, \ldots, C^*_{i,j-1,t}, C^*_{i,j+1,t}, \ldots, C^*_{iJT})$ is in this case also normal with, let say, mean m_j and variance s_j^2, for j=1,...,J. We need to distinguish four situations for the sampling of C^*_{ijt} for j=1,...,J:

1. The first situation corresponds to $j = d_{it}$. In this case we sample

(15) $$C^*_{ijt} \mid \cdot \sim \begin{cases} N(m_j, s_j^2) \times I(0, \infty) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} > 0 \\ N(m_j, s_j^2) \times I(\max_{k|k\neq j}(C^*_{ikt}), \infty) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} = 0. \end{cases}$$

2. The second situation corresponds to $j \neq d_{it}$ and $U_{ijt} > U_{i,d_{it},t}$, in which case we sample

(16) $$C^*_{ijt} \mid \cdot \sim \begin{cases} N(m_j, s_j^2) \times I(-\infty, 0) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} > 1 \\ N(m_j, s_j^2) \times I(-\infty, \min(0, C^*_{i,d_{it},t})) & \text{if } \sum_{j=1,k\neq j}^{J} c_{ijt} = 1. \end{cases}$$

3. The third situation corresponds to $j \neq d_{it}$ and $U_{ijt} < U_{i,d_{it},t}$ with $C^*_{i,d_{it},t} < 0$. In this case we sample

(17) $$C^*_{ijt} \mid \cdot \sim N(m_j, s_j^2) \times I(-\infty, C^*_{i,d_{it},t}).$$

4. In all other cases we use the following approach:

(18) $$C^*_{ijt} \mid \cdot \sim N(m_j, s_j^2).$$

REFERENCES

Besag, Julian and Peter J. Green (1993), “Spatial statistics and bayesian computation,” *Journal of the Royal Statistical Society B*, 55 (1), 25-37.

Chib, Siddhartha and Edward Greenberg (1998), “Analysis of multivariate probit models,” *Biometrika*, 85 (2), 347-361.

Damien, Paul, Jon Wakefield and Stephen Walker (1999), “Gibbs sampling for bayesian non-conjugate and hierarchical models by using auxiliary variables,” *Journal of the Royal Statistical Society B*, 61 (2), 331-344.

Geweke, John F. (1991), “Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints,” in Computer Science and Statistics: Proceedings of the 23rd Symposium on Interface, Alexandria, VA, American Statistical Association, 571-578.

Manchanda, P., A. Ansari and S. Gupta (1999), “The shopping basket: A model for multicategory purchase incidence decisions,” *Marketing Science*, 18 (2), 95-115.

Zellner, Arnold (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.