Monte Carlo Simulation in Random Coefficient Logit Models Involving Large Sums

Zsolt Sándor

Sapientia Hungarian University of Transylvania
Faculty of Economic and Human Sciences,
Miercurea Ciuc
email: sandorzsolt@sapientia.siculorum.ro

Abstract. We study Monte Carlo simulation in some recent versions of random coefficient logit models that contain large sums of expressions involving multivariate integrals. Such large sums occur in the random coefficient logit with demographic characteristics, the random coefficient logit with limited consumer information and the design of choice experiments for the panel mixed logit. We show that certain quasi-Monte Carlo methods, that is, so-called (t, m, s)-nets, provide improved performance over pseudo-Monte Carlo methods in terms of bias, standard deviation and root mean squared error.

1 Introduction

Random coefficient logit models have become the main tools in the study of demand for differentiated products and related problems. Its popularity comes from the flexible modelling of heterogeneous consumer preferences, which has been shown to facilitate the estimation of realistic patterns of substitution between different products (Berry et al. 1995, Nevo 2001).

The model that is at the basis of these studies can only be estimated with a large data set, which requires a large computational effort in the estimation algorithm. One of the sources of difficulty is the presence of the random coefficients, which compared to the logit model without random coefficients has

fewer degrees of freedom due to the standard deviation parameters, while it is estimated based on the same observations. In addition, the random coefficient logit has analytically intractable market share expressions in the form of multivariate integrals. Since within an estimation algorithm these need to be approximated at each iteration, this further increases the computational burden of the estimation.

Due to their usefulness, random coefficient logit models have been extended in several directions. A straightforward extension is the one that accommodates demographic characteristics (Nevo 2001). Another extension is the model with limited consumer information (Goeree 2008). Although these two models are different, they are similar in that they imply market share expressions that are sums of intractable multivariate integrals, and typically these sums have a large number of terms.

Random coefficient logit models are also used to analyze data from choice experiments. These data are obtained from respondents who choose the best from several sets of hypothetical products. Under the assumption that the preferences of a given respondent do not change, the model yields the so-called panel mixed logit. Bliemer and Rose (2010) consider the problem of designing the hypothetical products into several choice sets for this model. In order to do so, one needs to compute a scalar transformation of the information matrix, which is a large sum of expressions that depend on intractable multivariate integrals.

In this paper we study to what extent quasi-Monte Carlo sampling improves the computational efficiency in these models. For this, first we merge the large sum of multivariate integrals into one higher-dimensional integral of a function that is discontinuous is some variables and then apply quasi-Monte Carlo sampling to estimate the integral. The implied integrand functions are still square-integrable, so asymptotically quasi-Monte Carlo sampling is expected to offer computational gains over pseudo-Monte Carlo. We also expect to get finite sample computational improvement for integrals expressed as probabilities, since intuitively quasi-random draws are designed to provide better coverage than pseudo-random draws over the density for which the expectation is defined. This improved coverage usually translates into smaller expected approximation error. When the simulated probabilities are used within nonlinear expressions, the greater precision and reduced bias in the simulated probabilities translates into greater precision and reduced bias in parameter estimates.

From a computational point of view we contribute to the literature by studying the performance of quasi-Monte Carlo sampling applied to the estimation

of the expectation of some discrete random variables. In the literature there seems to be an emphasis on results that show the gains from using quasi-Monte Carlo in the case of sufficiently smooth integrand functions, but applications to cases of discontinuous integrands are rare.

Recently, in the econometric literature some new approaches have appeared for the estimation of random coefficient logit probabilities. One approach uses Laplace approximations of the integrals (Harding and Hausman 2007). Another approach uses Gaussian quadrature based on sparse grids (Heiss and Winschel 2008). These approaches produce improvements in terms of precision and/or computing time with respect to Monte Carlo estimation based on some quasi-Monte Carlo samples (namely, Halton and modified Latin hypercube by Hess et al. 2006). However, on the one hand, these methods have not yet been compared directly to the quasi-Monte Carlo we consider in this paper, and on the other hand, it is not straightforward how to apply them to estimate expectations of discrete random variables.

The next section presents the random coefficient logit with demographic characteristics and the random coefficient logit with limited consumer information. Section 3 presents the experimental design problem for the panel mixed logit. Section 4 describes briefly the quasi-Monte Carlo methods used in the paper. Section 5 presents the results, and Section 6 concludes.

2   Two demand models involving large sums

We discuss two models whose econometric analysis involves the computation of large sums. Another common feature of these models is that they are used with aggregate market-level data. In both models the market share expressions involve large sums of analytically intractable multi-dimensional integrals. Approximating precisely the market shares is crucial for the estimation of the model parameters. For both models we express the large sums as integrals on the unit hypercube of discontinuous functions, and employ Monte Carlo integration methods to estimate the market shares. Both models are computationally difficult in that they require a very high number of simulation draws, so we believe it is crucial to come up with methods that reduce the computational burden and/or improve the precision of the involved estimates.

2.1   Random coefficient logit with demographic characteristics

We define this model by the utility function. Suppose that there are J products available in the market. The utility of an individual i who chooses j out of J

products is
$$u_{ij} = x_j' \left(\beta_i + \Gamma d_i \right) + \xi_j + \varepsilon_{ij},$$

where x_j is a K-vector of observed characteristics, which may also contain endogenous characteristics like price; it also contains 1 for intercept, β_i is a random parameter whose distribution is assumed to be $N(b, \Sigma)$ with $\Sigma = \mathrm{diag}(\sigma_1^2, ..., \sigma_K^2)$, d_i is a D-vector of demographic characteristics, Γ is a $K \times D$ matrix of parameters, ξ_j is a product-specific error, also called the unobserved characteristic of product j, ε_{ij} is assumed to be iid type I extreme value random variable independent of the rest of the variables. The utility of consumer i who does not purchase any product is

$$u_{i0} = \varepsilon_{i0},$$

where ε_{i0} is also assumed to be type I extreme value random variable independent of the rest of the variables.

We assume that the econometrician observes the product characteristics x_j, $j = 1, ..., J$, the market shares of all products and the distribution of the demographic characteristics. We assume that the demographic characteristics have a discrete distribution with $N_D = n_1 \times ... \times n_D$ support points. This class of models was pioneered by Nevo (2001), who extended the seminal work by Berry et al. (1995) by including demographic characteristics in the demand in an application on measuring market power in the ready-to-eat cereals market.

The market share of product j predicted by the model, that is, the probability that j is chosen, conditional on x_j and ξ_j, $j = 1, ..., J$, as well as the unknown parameters, is

$$s_j = \int \frac{\exp \left(x_j' \left(\beta + \Gamma d \right) + \xi_j \right)}{1 + \sum_{r=1}^{J} \exp \left(x_r' \left(\beta + \Gamma d \right) + \xi_r \right)} dF \left(\beta \right) dG \left(d \right),$$

where $F \left(\beta \right)$ denotes the cumulative distribution function of $\beta \sim N(b, \Sigma)$ and $G \left(d \right)$ denotes the cumulative distribution function of the discrete random variable d. The demographic characteristics need to be integrated out because we assume that individual choices are not observed.

Writing out the expectation with respect to the discrete distribution of demographic characteristics, we obtain

$$s_j = \sum_{k=1}^{N_D} p_k \int_{\mathbb{R}^K} \frac{\exp \left(x_j' \left(\beta + \Gamma d_k \right) + \xi_j \right)}{1 + \sum_{r=1}^{J} \exp \left(x_r' \left(\beta + \Gamma d_k \right) + \xi_r \right)} dF \left(\beta \right), \tag{1}$$

where d_k and p_k are the values of d and their probabilities, respectively. In order to evaluate the market share (1) we need to compute N_D integrals for each j. It is well known that these types of integrals are analytically intractable, so they need to be approximated numerically. This implies a non-negligible computational effort, which in the case of the market share (1) is multiplied by the number of terms N_D in the sum. Typically, one of the demographic characteristics is income, and for this the number of support points is rather large making the number of terms N_D also large. In this case the computational burden for obtaining a sufficiently precise approximation of the market share becomes substantial.

As we outlined at the beginning of this section, we replace the sum by integral. At this point we assume that the components of the demographic characteristic vector are independent.[1] Then the market share (1) can be rewritten as

$$s_j = \int_{\mathbb{R}^K \times [0,1]^D} \frac{\exp\left(x_j'\left(\beta + \Gamma\varphi\left(\mathbf{u}\right)\right) + \xi_j\right)}{1 + \sum_{r=1}^{J} \exp\left(x_r'\left(\beta + \Gamma\varphi\left(\mathbf{u}\right)\right) + \xi_r\right)} dF\left(\beta\right) d\mathbf{u}, \qquad (2)$$

where $\varphi\left(\mathbf{u}\right) = \left(\varphi_1\left(u_1\right), ..., \varphi_D\left(u_D\right)\right)'$ with the components $\varphi_h\left(u_h\right)$, $h = 1, ..., D$, defined as follows. Let $v_{h,1}, ..., v_{h,n_h}$ and $p_{h,1}, ..., p_{h,n_h}$ be the possible values and the corresponding probabilities, respectively, of component h of the demographic characteristic vector d; let $q_{h,i} = \sum_{\alpha=1}^{i} p_{h,\alpha}$ for $i = 1, ..., n_h$ and $q_{h,0} = 0$. Then define $\varphi_h\left(u_h\right) = v_{h,i}$ if and only if $q_{h,i-1} \leq u_h < q_{h,i}$. We note that each component φ_h of φ is a step function and hence the integrand function in the above market share expression is square-integrable in $\mathbf{u}$.[2] Therefore, estimation of the integral by the Monte Carlo method is valid. We

---

[1]Although this assumption seems to be restrictive at first sight, there are at least two reasons for which it is not unplausible. First, in many cases the econometrician is forced to assume that the demographic characteristics are independent because from the data set in many cases only the marginal distributions of the demographic characteristics can be inferred. Second, if from the data set one can infer the joint distribution of demographic characteristics, then this can be treated as a one-dimensional demographic characteristic that can take N_D values. Therefore, this situation appears as a special case of our treatment. We devote one simulation study to this case.

[2]For an integral $\int_{[0,1]^D} g\left(\mathbf{u}\right) d\mathbf{u}$ the integrand function g is square-integrable if $\int_{[0,1]^D} g^2\left(\mathbf{u}\right) d\mathbf{u} < \infty$. This property allows the computation of the variance of the Monte Carlo estimator $\sum_{i=1}^{N} g\left(\mathbf{u}_i\right)/N$.

form the (quasi-)Monte Carlo estimator by using joint draws $\Delta_N = (\beta_i, \mathbf{u}_i)_{i=1}^N$:

$$\widetilde{s}_j (\xi, \theta, \Delta_N) = \frac{1}{N} \sum_{i=1}^{N} \frac{\exp \left(x'_j (\beta_i + \Gamma \varphi (\mathbf{u}_i)) + \xi_j \right)}{1 + \sum_{r=1}^{J} \exp (x'_r (\beta_i + \Gamma \varphi (\mathbf{u}_i)) + \xi_r)}, \qquad (3)$$

where β_i is a draw from F and $\mathbf{u}_i$ is a draw from the uniform distribution on $[0,1]^D$.

2.2 Random coefficient logit with limited consumer information

In this subsection we discuss a discrete choice model in which a consumer can potentially choose (at most one out) of J products, but she only observes a subset of the J products. Let

$$P_S (\gamma_i) = \prod_{k \in S} \phi_k (\gamma_i) \prod_{h \notin S} (1 - \phi_h (\gamma_i))$$

be the probability that consumer i observes subset S of the products, where γ_i is a consumer-specific L-vector of parameters affecting this probability; we will refer to S as the the choice set of consumer i. The choice set probability $P_S (\gamma_i)$ also depends on product characteristics t_j (suppressed in the notation) that affect the observability of the products through the factor $\phi_j (\gamma_i)$, $j = 1, ..., J$. The probability that product j is chosen, conditional on x_j, ξ_j, t_j, $j = 1, ..., J$, and the unknown parameters, is

$$s_j = \int_{\mathbb{R}^{K+L}} \sum_{S \in \mathcal{S}_j} P_S (\gamma) \frac{\exp \left(x'_j \beta + \xi_j \right)}{1 + \sum_{r \in S} \exp (x'_r \beta + \xi_r)} dF (\beta, \gamma), \qquad (4)$$

where $\mathcal{S}_j$ denotes the set of subsets of $\{1, 2, ..., J\}$ containing j; x_j, β, ξ_j are as in the model from the previous subsection, and $F (\beta, \gamma)$ denotes the joint cumulative distribution function of (β, γ). It will be convenient to assume that the random parameters can be parameterized so that $\beta \equiv \beta (\theta_1, v_1)$ and $\gamma \equiv \gamma (\theta_2, v_2)$ where θ_1 and θ_2 are unknown parameter vectors and v_1 and v_2 are standard normal random vectors of dimension K and L, respectively. In what follows, instead of $\phi_k (\gamma (\theta_2, v_2))$ we use $\phi_k (\theta_2, v_2)$. With this notation the market share of product j is

$$s_j = \int_{\mathbb{R}^{K+L}} \sum_{S \in \mathcal{S}_j} P_S (\theta_2, v_2) \frac{\exp \left(x'_j \beta (\theta_1, v_1) + \xi_j \right)}{1 + \sum_{r \in S} \exp (x'_r \beta (\theta_1, v_1) + \xi_r)} d\Phi (\mathbf{v}), \qquad (5)$$

where $\Phi(\mathbf{v})$ denotes the standard normal cumulative distribution function of $\mathbf{v} = (v_1, v_2)$.

A similar model is used by Goeree (2008), where the choice sets are determined by the amount of advertising of the products. The number of terms in the above sum is equal to 2^{J-1}, which is the number of subsets of $\{1, 2, ..., J\}$ containing some product j. In many markets the number of products J is large, so the sum in the market share expression (4) will have very many terms. This will cause a substantial computational burden, if one wants to compute the market shares directly.

We follow the literature (e.g., Goeree 2008) in that we approximate the market shares by Monte Carlo integration. Let (v_{1i}, v_{2i}), $i = 1, ..., N$, be a random sample from the standard normal distribution and

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{S \in \mathcal{S}_j} P_S(\theta_2, v_{2i}) \frac{\exp\left(x_j' \beta_i + \xi_j\right)}{1 + \sum_{r \in S} \exp\left(x_r' \beta_i + \xi_r\right)}$$

the corresponding Monte Carlo estimator of (5), where $\beta_i = \beta(\theta_1, v_{1i})$. For each i we can write the sum with respect to $S \in \mathcal{S}_j$ as the expected value of a discrete random variable. In this regard, for given i, note that

$$\sum_{S \in \mathcal{S}_j} P_S(\theta_2, v_{2i}) \frac{\exp\left(x_j' \beta_i + \xi_j\right)}{1 + \sum_{r \in S} \exp\left(x_r' \beta_i + \xi_r\right)}$$

$$= \phi_j(\theta_2, v_{2i}) \sum_{S \in \mathcal{N}_j} P_S(\theta_2, v_{2i}) \frac{\exp\left(x_j' \beta_i + \xi_j\right)}{1 + \exp\left(x_j' \beta_i + \xi_j\right) + \sum_{r \in S} \exp\left(x_r' \beta_i + \xi_r\right)} \quad (6)$$

where $\mathcal{N}_j$ is the set of subsets of $\{1, 2, ..., J\}$ that do not contain j. Since $\sum_{S \in \mathcal{N}_j} P_S(\theta_2, v_{2i}) = 1$, the sum in (6) can be interpreted as the expected value of a discrete random variable with values

$$P_{j|S}(\theta_1, v_{1i}) = \frac{\exp\left(x_j' \beta_i + \xi_j\right)}{1 + \exp\left(x_j' \beta_i + \xi_j\right) + \sum_{r \in S} \exp\left(x_r' \beta_i + \xi_r\right)}, \quad S \in \mathcal{N}_j$$

and probabilities $P_S(\theta_2, v_{2i})$. This can be used to obtain the Monte Carlo estimator of the sum. For this we need to draw a sample of size N from this discrete distribution. This can be done by drawing N uniform random vectors $\mathbf{u}_1, ..., \mathbf{u}_N \in [0, 1]^J$ and then, based on these, constructing the choice

sets $S_1, ..., S_N$ by $S_i = \{k \in \{1, ..., J\} \setminus \{j\} : u_{ik} \leq \phi_k(\theta_2, v_{2i})\}$ for $i = 1, ..., N$, where u_{ik} is the k-th component of $\mathbf{u}_i$. The obtained Monte Carlo estimator of the sum in (6) is

$$\frac{1}{N} \sum_{i=1}^{N} P_{j|S_i}(\theta_1, v_{1i}).$$

(Note that we do not use the j-th component of the u_i's.)

However, this Monte Carlo estimator is not continuous in the unknown parameters in general, because, for a given sample $\mathbf{u}_1, ..., \mathbf{u}_N \in [0, 1]^J$, the choice sets $S_1, ..., S_N$ may change when θ_2 changes, and this modifies the structure of $P_{j|S_i}(\theta_1, v_{1i})$. A solution to this problem is via importance sampling, as suggested by Goeree (2008). Let θ_{20} be a fixed value of the parameter θ_2 and define the importance sampling probabilities $P_S(\theta_{20}, v_{2i})$. Using these, we can rewrite the sum in (6) as

$$\sum_{S \in \mathcal{N}_j} P_S(\theta_{20}, v_{2i}) \frac{P_S(\theta_2, v_{2i})}{P_S(\theta_{20}, v_{2i})} P_{j|S}(\theta_1, v_{1i}).$$

Now, this can be written as the integral

$$\int_{[0,1]^J} \frac{P_j(\theta_2, v_{2i}, \mathbf{u})}{P_j(\theta_{20}, v_{2i}, \mathbf{u})} \frac{\exp(x_j'\beta_i + \xi_j)}{1 + \exp(x_j'\beta_i + \xi_j) + \sum_{\substack{r=1 \\ r \neq j}}^{J} 1(u_r \leq \phi_r(\theta_{20}, v_{2i})) \exp(x_r'\beta_i + \xi_r)} d\mathbf{u},$$

where for $\mathbf{u} = (u_1, ..., u_J)$

$$P_j(\theta_2, v_2, \mathbf{u}) = \prod_{\substack{k=1, \\ k \neq j}}^{J} \phi_k(\theta_2, v_2)^{1(u_k \leq \phi_k(\theta_2, v_2))} (1 - \phi_k(\theta_2, v_2))^{1(u_k > \phi_k(\theta_2, v_2))}$$

and $1(A)$ is the indicator of the event A. Note that $P_j(\theta_2, v_2, \mathbf{u}) = P_S(\theta_2, v_2)$ if $S = \{k \in \{1, ..., J\} \setminus \{j\} : u_k \leq \phi_k(\theta_2, v_2)\}$. Consequently, we obtain that the market share of product j is[3]

$$
\begin{aligned}
s_j &= \int_{\mathbb{R}^{K+L}} \int_{[0,1]^J} \phi_j(\theta_2, v_2) \frac{P_j(\theta_2, v_2, \mathbf{u})}{P_j(\theta_{20}, v_2, \mathbf{u})} \times \\
&\quad \times \frac{\exp(x_j'\beta(\theta_1, v_1) + \xi_j)}{1 + \exp(x_j'\beta(\theta_1, v_1) + \xi_j) + \sum_{\substack{r=1 \\ r \neq j}}^{J} 1(u_r \leq \phi_r(\theta_{20}, v_2)) \exp(x_r'\beta(\theta_1, v_1) + \xi_r)} d\mathbf{u} d\Phi(\mathbf{v}).
\end{aligned}
\tag{7}
$$

---

[3]In order to keep the notation simple we use the whole vector $\mathbf{u}$ in the integrand function above, although it does not depend on u_j. However, this does not change the results.

The integrand function in this integral is square-integrable for the probabilities ϕ_k that we consider (see Section 5.2), so we approximate the integral by (quasi-) Monte Carlo estimation using joint draws $\Delta_N = (\mathbf{u}_i, \mathbf{v}_i)_{i=1}^N$. The estimator is

$$
\widetilde{s}_j\left(\xi, \theta, \Delta_N\right) = \frac{1}{N} \sum_{i=1}^N \phi_j\left(\theta_2, v_{2i}\right) \frac{P_j\left(\theta_2, v_{2i}, \mathbf{u}_i\right)}{P_j\left(\theta_{20}, v_{2i}, \mathbf{u}_i\right)} \times \tag{8}
$$

$$
\times \frac{\exp\left(x_j'\beta\left(\theta_1, v_{1i}\right) + \xi_j\right)}{1 + \exp\left(x_j'\beta\left(\theta_1, v_{1i}\right) + \xi_j\right) + \sum_{\substack{r=1 \\ r \neq j}}^J \mathbf{1}\left(u_{ir} \leq \phi_r\left(\theta_{20}, v_{2i}\right)\right) \exp\left(x_r'\beta\left(\theta_1, v_{1i}\right) + \xi_r\right)},
$$

where $\xi = (\xi_1, ..., \xi_J)$ and $\theta = (\theta_1, \theta_2)'$; from the the arguments of $\widetilde{s}_j$ we omit the rest of the variables.

2.3 Estimation of the model parameters

In this subsection we discuss the essentials on estimation of the model parameters needed for understanding the simulation results from Section 5. In both models the market share of a product j can be expressed as a function of the model ingredients (see equations (1) and (4)). In both models the system of market shares can be inverted, as shown by Berry (1994), to obtain the variables $\xi_1, ..., \xi_J$ as a function of the market shares and the other ingredients of the models. In practice we need to solve in $\xi_1, ..., \xi_J$ the nonlinear systems made up of (3) and (8) for $j = 1, ..., J$. In order to get the solution we appeal to Berry et al. (1995), who have introduced the operator $T\left(\cdot, \mathbf{s}^0, \theta, \Delta_N\right) : \mathbb{R}^J \to \mathbb{R}^J$ such that its component j is defined by

$$
T_j\left(\xi, \mathbf{s}^0, \theta, \Delta_N\right) = \xi_j + \ln s_j^0 - \ln \widetilde{s}_j\left(\xi, \theta, \Delta_N\right),
$$

and whose fixed point is the solution of our nonlinear system, where $\mathbf{s}^0$ is the vector of observed market shares. These authors have established that in the random coefficient logit model $T\left(\cdot, \mathbf{s}^0, \theta, \Delta_N\right)$ is a contraction for any feasible $\left(\mathbf{s}^0, \theta, \Delta_N\right)$. Goeree (2008) has extended this result to the random coefficient logit with limited consumer information.

Let the solution be denoted by $\xi\left(\mathbf{s}^0, \theta, \Delta_N\right)$. The identification of the unknown parameters is based on the moment condition that

$$
E\left[\xi\left(\mathbf{s}^0, \theta, \Delta_N\right)\right] = 0 \text{ implies } \theta = \theta^0,
$$

where θ^0 is the true value of the parameter vector. The method for estimating θ is nonlinear least squares by

$$
\min_\theta \xi\left(\mathbf{s}^0, \theta, \Delta_N\right)' \xi\left(\mathbf{s}^0, \theta, \Delta_N\right).
$$

Clearly, the obtained estimator depends on the simulation draws Δ_N used for estimating the market shares.[4] In the simulation study (Section 5) we evaluate the performance of different samples Δ_N by comparing the mean squared error of the estimator of the objective function evaluated at the true parameter value and normalized by the number of observations, that is, $\xi\left(\mathbf{s}^0, \theta^0, \Delta_N\right)' \xi\left(\mathbf{s}^0, \theta^0, \Delta_N\right)/J$.

3 Experimental design for the panel mixed logit

Suppose that each consumer $i = 1, ..., n$ chooses one alternative from the same S choice sets. Each choice sets has J alternatives. Let $y_{js}^i \in \{0, 1\}$ denote the choice of consumer i in choice set s in the sense that $y_{js}^i = 1$ iff alternative j is chosen. Suppose that the utility of i corresponding to alternative j in s is

$$u_{ijs} = x_{js}\left(\beta + V_i\sigma\right) + \varepsilon_{ijs},$$

where x_{js} is the vector of attributes of alternative j in choice set s, $\beta + V_i\sigma$ is the random coefficient with V_i being a diagonal matrix having standard normal random variables v_i on its main diagonal, and ε_{ijs} is an extreme value distributed error term.

Next we derive the likelihood for this model. For a consumer i the likelihood that a given $y^i = \left(y_{js}^i\right)_{\substack{j=1,...,J \\ s=1,...,S}}$ is chosen, given v_i, is proportional to

$$l_i\left(y^i|\theta, v_i\right) = \prod_{s=1}^{S}\prod_{j=1}^{J} p_{js}^{y_{js}^i}\left(\theta|v_i\right),$$

where $\theta = (\beta', \sigma')'$ and

$$p_{js}\left(\theta|v_i\right) = \frac{\exp\left(x_{js}\left(\beta + V_i\sigma\right)\right)}{\sum_{h=1}^{J}\exp\left(x_{hs}\left(\beta + V_i\sigma\right)\right)} \tag{9}$$

is the probability that alternative j is chosen in choice set s, given v_i. The unconditional likelihood for consumer i is (proportional to)

$$l_i\left(y^i|\theta\right) = \int \prod_{s=1}^{S}\prod_{j=1}^{J} p_{js}^{y_{js}^i}\left(\theta|v\right) d\Phi\left(v\right), \tag{10}$$

---

[4]We use this framework in our simulation study, but we note that in these classes of models in the literature typically one product characteristic considered is price, which is treated as endogenous. Due to this, the estimation method followed in the literature is the method of instrumental variables, which we wanted to avoid due to its potentially poor small sample properties.

where Φ is the cumulative distribution function of the standard normal random vector v. This yields the log-likelihood for all the n consumers

$$L\left(y|\theta\right) = \sum_{i=1}^{n} \ln \left(\int \prod_{s=1}^{S} \prod_{j=1}^{J} p_{js}^{y_{js}^{i}} \left(\theta|v\right) d\Phi\left(v\right) \right).$$

In order to make the expressions more similar to the cross-sectional mixed logit, let

$$L_{i}\left(y^{i}|\theta, v_{i}\right) = \ln \left(\prod_{s=1}^{S} \prod_{j=1}^{J} p_{js}^{y_{js}^{i}} \left(\theta|v\right) \right) = \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js}^{i} \ln p_{js} \left(\theta|v_{i}\right) = \left(y^{i}\right)' \ln p\left(\theta|v_{i}\right),$$

where

$$y^{i} = \left(y_{11}^{i}, y_{21}^{i}, ..., y_{J1}^{i}, y_{12}^{i}, ..., y_{J2}^{i}, ..., y_{1S}^{i}, ..., y_{JS}^{i} \right)'$$

$$p\left(\theta|v\right) = \left(p_{11}\left(\theta|v\right), p_{21}\left(\theta|v\right), ..., p_{J1}\left(\theta|v\right), p_{12}\left(\theta|v\right), ..., p_{J2}\left(\theta|v\right), ..., p_{1S}\left(\theta|v\right), ..., p_{JS}\left(\theta|v\right) \right)'$$

So

$$L\left(y|\theta\right) = \sum_{i=1}^{n} \ln \left(\int \exp L_{i}\left(y^{i}|\theta, v\right) d\Phi\left(v\right) \right).$$

For deriving the design criterion below we use this formula.

3.1 Design criterion

The design criterion is typically based on local D-optimality and defined as

$$D\left(X, \theta\right) = \left[\det I\left(X, \theta\right)\right]^{1/d}, \tag{12}$$

where d denotes the dimension of the information matrix $I\left(X, \theta\right)$. This criterion is called local because it depends on the true parameter value θ and "D" comes from determinant. Other criteria can also be considered; see Kessels et al. (2006).

The design criterion depends on the information matrix, so next we derive this. The information matrix is

$$I\left(X, \theta\right) = E\left[\frac{\partial L\left(y|\theta\right)}{\partial \theta} \frac{\partial L\left(y|\theta\right)}{\partial \theta'} \right] = n \cdot E\left[\frac{\partial \mathcal{L}\left(y^{i}|\theta\right)}{\partial \theta} \frac{\partial \mathcal{L}\left(y^{i}|\theta\right)}{\partial \theta'} \right]$$

because $y^{i}, \; i = 1, ..., n$, are assumed to be iid, where

$$\mathcal{L}\left(y^{i}|\theta\right) = \ln \left(\int \exp L_{i}\left(y^{i}|\theta, v\right) d\Phi\left(v\right) \right)$$

is the log-likelihood corresponding to i. Therefore,

$$I\left(X,\theta\right)=n\cdot\sum_{y^i}\Pr\left(y^i\right)\left[\frac{\partial\mathcal{L}\left(y^i|\theta\right)}{\partial\theta}\frac{\partial\mathcal{L}\left(y^i|\theta\right)}{\partial\theta'}\right],$$

where the summation is according to all possible J^S values of y^i; $\Pr\left(y^i\right)$ is the probability that the choice vector y^i occurs, that is,

$$\Pr\left(y^i\right)=\int\prod_{s=1}^{S}\prod_{j=1}^{J}p_{js}^{y_{js}^i}\left(\theta|v\right)d\Phi\left(v\right).$$

Note that this is identical to $l_i\left(y^i|\theta\right)$ from (10).

In order to derive the information matrix we need the first order derivative of the log-likelihood corresponding to i:

$$\frac{\partial\mathcal{L}\left(y^i|\theta\right)}{\partial\theta}=\frac{\dfrac{\partial\left(\int\exp L_i\left(y^i|\theta,v\right)d\Phi\left(v\right)\right)}{\partial\theta}}{\int\exp L_i\left(y^i|\theta,v\right)d\Phi\left(v\right)}=\frac{\displaystyle\int\frac{\partial\left(\exp L_i\left(y^i|\theta,v\right)\right)}{\partial\theta}d\Phi\left(v\right)}{\int\exp L_i\left(y^i|\theta,v\right)d\Phi\left(v\right)},$$

where recall $L_i\left(y^i|\theta,v_i\right)=\left(y^i\right)'\ln p\left(\theta|v_i\right)$. Of this,

$$\frac{\partial\left(\exp L_i\left(y^i|\theta,v\right)\right)}{\partial\theta}=\frac{\partial L_i\left(y^i|\theta,v\right)}{\partial\theta}\exp L_i\left(y^i|\theta,v\right)=\left(\begin{array}{c}\dfrac{\partial L_i\left(y^i|\theta,v\right)}{\partial\beta}\exp L_i\left(y^i|\theta,v\right)\\[2mm]\dfrac{\partial L_i\left(y^i|\theta,v\right)}{\partial\sigma}\exp L_i\left(y^i|\theta,v\right)\end{array}\right).$$

The first order derivatives

$$\frac{\partial L_i\left(y^i|\theta,v\right)}{\partial\theta}=\left(\frac{\partial L_i\left(y^i|\theta,v\right)}{\partial\theta'}\right)'=\left(\frac{\partial\ln p\left(\theta|v_i\right)}{\partial\theta'}\right)'y^i.$$

Split the probability vector $p\left(\theta|v_i\right)$ into the probability vectors corresponding to the choice sets, so

$$\ln p\left(\theta|v_i\right)=\left(\begin{array}{c}\ln p_{(1)}\left(\theta|v_i\right)\\\vdots\\\ln p_{(S)}\left(\theta|v_i\right)\end{array}\right),$$

where $p_{(s)}\left(\theta|v_i\right),\ s=1,...,S$, are the probability vectors corresponding to choice set s, so

$$\frac{\partial\ln p\left(\theta|v_i\right)}{\partial\theta'}=\left(\begin{array}{c}\frac{\partial\ln p_{(1)}(\theta|v_i)}{\partial\theta'}\\\vdots\\\frac{\partial\ln p_{(S)}(\theta|v_i)}{\partial\theta'}\end{array}\right)$$

and

$$\frac{\partial L_i \left(y^i | \theta, v \right)}{\partial \theta} = \left(\ \left(\frac{\partial \ln p_{(1)} \left(\theta | v_i \right)}{\partial \theta'} \right)' \quad \cdots \quad \left(\frac{\partial \ln p_{(S)} \left(\theta | v_i \right)}{\partial \theta'} \right)' \ \right) y^i.$$

By Sándor (2001, p.9)

$$\frac{\partial \ln p_{(s)} \left(\theta | v_i \right)}{\partial \theta'} = P_{(s)}^{-1} \frac{\partial p_{(s)}}{\partial \theta'},$$

where $P_{(s)}$ is the diagonal matrix of $p_{(s)}$ and for simplicity we omitted the argument $(\theta | v_i)$. Let us particularize to β. Then we obtain,

$$\frac{\partial p_{(s)}}{\partial \beta'} = \left(P_{(s)} - p_{(s)} p'_{(s)} \right) X_{(s)},$$

so

$$\frac{\partial \ln p_{(s)} \left(\theta | v_i \right)}{\partial \beta'} = P_{(s)}^{-1} \left(P_{(s)} - p_{(s)} p'_{(s)} \right) X_{(s)} = \left(I_{(s)} - \iota_{(s)} p'_{(s)} \right) X_{(s)},$$

hence

$$\left(\frac{\partial \ln p_{(s)} \left(\theta | v_i \right)}{\partial \beta'} \right)' = X'_{(s)} \left(I_{(s)} - p_{(s)} \iota'_{(s)} \right),$$

and

$$
\begin{aligned}
\frac{\partial L_i \left(y^i | \theta, v \right)}{\partial \beta} &= \left(\ X'_{(1)} \left(I_{(1)} - p_{(1)} \iota'_{(1)} \right) \quad \cdots \quad X'_{(S)} \left(I_{(S)} - p_{(S)} \iota'_{(S)} \right) \ \right) y^i \\
&= \left(\ X'_{(1)} \quad \cdots \quad X'_{(S)} \ \right) \begin{pmatrix} I_{(1)} - p_{(1)} \iota'_{(1)} & & 0 \\ & \ddots & \\ 0 & & I_{(S)} - p_{(S)} \iota'_{(S)} \end{pmatrix} y^i \\
&= X' \left(I - B_i \right) y^i,
\end{aligned}
$$

where

$$B_i \equiv B \left(\theta | v_i \right) = \begin{pmatrix} p_{(1)} \iota'_{(1)} & & 0 \\ & \ddots & \\ 0 & & p_{(S)} \iota'_{(S)} \end{pmatrix}.$$

In a similar way, we can derive

$$\frac{\partial L_i \left(y^i | \theta, v \right)}{\partial \sigma} = V X' \left(I - B_i \right) y^i.$$

Consequently,

$$
\frac{\partial \left(\exp L_i \left(y^i | \theta, v \right) \right)}{\partial \theta} = \left(\begin{array}{c} X' \left(I - B_i \right) y^i \exp L_i \left(y^i | \theta, v \right) \\ V X' \left(I - B_i \right) y^i \exp L_i \left(y^i | \theta, v \right) \end{array} \right)
$$

$$
= \left(\begin{array}{c} I_V \\ V \end{array} \right) X' \left(I - B \left(\theta | v \right) \right) y^i \exp L_i \left(y^i | \theta, v \right),
$$

and

$$
\frac{\partial \mathcal{L} \left(y^i | \theta \right)}{\partial \theta} = \frac{\int \left(\begin{array}{c} I_V \\ V \end{array} \right) X' \left(I - B \left(\theta | v \right) \right) y^i \exp L_i \left(y^i | \theta, v \right) d\Phi \left(v \right)}{\int \exp L_i \left(y^i | \theta, v \right) d\Phi \left(v \right)}. \tag{13}
$$

Since this expression is in vector-matrix form, it is easier to program (in Gauss or Matlab) than the one provided by Bliemer and Rose (2010), which uses sums.

3.2   Estimation of the design criterion

The design criterion cannot be computed exactly, so it needs to be estimated/approximated. Recall that the design criterion is $D \left(X, \theta \right) = \left[\det I \left(X, \theta \right) \right]^{1/d}$, where

$$
I \left(X, \theta \right) = n \cdot \sum_{y^i} \Pr \left(y^i \right) \left[\frac{\partial \mathcal{L} \left(y^i | \theta \right)}{\partial \theta} \frac{\partial \mathcal{L} \left(y^i | \theta \right)}{\partial \theta'} \right]. \tag{14}
$$

Estimation of $I \left(X, \theta \right)$ involves estimation of the integrals in $\frac{\partial \mathcal{L} \left(y^i | \theta \right)}{\partial \theta}$ and the sum over all possible realizations of y^i. The number of these possible realizations is J^S, which in typical practical examples has the magnitude of several thousands (e.g., $2^{15} = 32768$ in the case of 15 choice sets with 2 alternatives or $3^{10} = 59049$ in the case of 10 choice sets with 3 alternatives).

The integrals involved in $\frac{\partial \mathcal{L} \left(y^i | \theta \right)}{\partial \theta}$ can be estimated by one of the commonly used methods (Monte Carlo, quasi-Monte Carlo or Gaussian quadrature using sparse grids). The sum over y^i typically contains excessively many terms, and therefore, we also estimate it by Monte Carlo simulations. To do this, we draw from the discrete distribution of the y^i's. In order to accomplish this, first for each consumer i we draw v_i, the standard normally distributed random vector that determines the random coefficient of respondent i. Once we know v_i, the choices of i across the choice sets become independent; in choice set s the choice (i.e., $y^i_{js}, \ j = 1, ..., J$) can be obtained as one multinomial draw with

probabilities $p_{js}(\theta|v_i)$, $j = 1, ..., J$ defined in (9). If we denote these draws by $\widetilde{y}^i_{js}$ and let $\widetilde{y}^i = \left(\widetilde{y}^i_{js}\right)_{\substack{j=1,...,J \\ s=1,...,S}}$, and obtain a sample $\left(\widetilde{y}^{i,r}\right)_{r=1,...,R}$ of such replicated draws, then a Monte Carlo estimate of $I(X,\theta)$ is

$$\widetilde{I}(X,\theta) = \frac{n}{R} \sum_{r=1}^{R} \left[\frac{\partial \mathcal{L}\left(\widetilde{y}^{i,r}|\theta\right)}{\partial \theta} \frac{\partial \mathcal{L}\left(\widetilde{y}^{i,r}|\theta\right)}{\partial \theta'} \right].$$

4 Quasi-Monte Carlo sampling

We use two different types of quasi-Monte Carlo samples. The first type is a randomized (t, m, s)-net and the second type is a randomly shifted lattice. They are both samples from $[0, 1)^s$ and they are meant to replace in the Monte Carlo estimator pseudo-random draws from the uniform distribution on $[0, 1)^s$ in order to estimate integrals on $[0, 1)^s$. These samples are randomized versions of deterministic sequences. There are several quasi-Monte Carlo samples available in the literature; the reason for using exactly these is that they perform relatively well in the estimation of multivariate normal probabilities (Sándor and András 2004). We mention below only the most essential features of these; for more details we refer to Sándor and András (2004) and the references therein.

4.1 Randomized (t, m, s)-nets

A (t, m, s)-net in base b is a set of b^m points from $[0, 1)^s$ that satisfy certain equidistribution property, namely that all subintervals of $[0, 1)^s$ of a certain type contain a given number of points of the sequence. This equidistribution property ensures that the sequence approximates closely the continuous uniform distribution on $[0, 1)^s$. (t, m, s)-nets are deterministic, so we need to randomize them. We use a method that randomizes these numbers from the s-dimensional unit hypercube by permuting their digits in base b representation such that the permutation depends on the coordinate $j \in \{1, ..., s\}$ and the order of the digits in base b representation. By this randomization one obtains vectors whose elements are from the set $\left\{0, \frac{1}{b^m}, \frac{2}{b^m}, ..., \frac{b^m-1}{b^m}\right\}$. In order to make these coordinates uniform random on $[0, 1)$, we add to them $\frac{u_{ij}}{b^m}$, where u_{ij} is uniform random on $[0, 1)$. This way we obtain randomized (t, m, s)-nets that inherit the equidistribution property of the original nets and contain points that are uniformly distributed.

4.2 Randomized good lattice points

The simplest type of lattice points s-dimensional vectors

$$x_i = \left(\left\{ \frac{i}{n} \right\}, \left\{ \frac{iq}{n} \right\}, ..., \left\{ \frac{iq^{s-1}}{n} \right\} \right) \text{ for } i = 0, 1, ..., n-1, \qquad (15)$$

where n is the number of lattice points, q is a positive integer relative prime with n, and the symbol $\{x\}$ means the fractional part of the number x. Lattice points are easy to generate, provided that we know q. Procedures for obtaining appropriate q values are based on minimizing the worst-case integral error. This criterion is based on the Fourier series representation of the integrand. Fourier series are helpful in expressing the integration error of lattice points because lattice points are especially suited for integrating periodic functions, or, more precisely, functions that admit a continuous periodic extension.

Randomization of lattice points is typically done by random shifting. A randomly shifted lattice has the points

$$x_i = \left(\left\{ \frac{i}{n} + u_1 \right\}, \left\{ \frac{iq}{n} + u_1 \right\}, ..., \left\{ \frac{iq^{s-1}}{n} + u_s \right\} \right) \text{ for } i = 0, 1, ..., n-1, \ (16)$$

where $u_1, ..., u_s$ are independent random uniform numbers on $[0, 1)$. These randomly shifted lattice points can be regarded as having been drawn from the uniform distribution if we replace i by $\pi(i)$ in (16), where π is a random uniform permutation of $0, 1, ..., n-1$.

The lattice points perform better if they are transformed by the so-called baker's transformation $\varphi(z) = |2z - 1|$ (that is, each coordinate of each point is transformed). This in fact transforms the integrand function into a function that has a continuous periodic extension, so it is expected to improve the performance of the lattice points.[5]

5 Simulation studies

The motivation of the simulation studies, as mentioned in Section 4, is to find out if quasi-Monte Carlo sampling can improve the precision of the objective function simulator, and if so, what the gains are in terms of precision and computing time. In the simulation design we intend to come as close as possible to reality.

---

[5]Gauss-codes for generating the samples are available from the author.

5.1 Random coefficient logit with demographic characteristics

There are some common ingredients in the data generating processes (DGP's) that we use for the two models. In both models $K = 5$, which is the number of random coefficients, the characteristics are generated as $x_{j1} = 1$, $x_{j2}, ..., x_{j5} \sim U[0,1]$, $\xi_j \sim N(0, 0.25)$. The true parameters are $b = (-6, -1, 1, 1, 1)'$, $\sigma_1 = ... = \sigma_5 = 1$. The true market shares are computed by MC with 10,000 pseudo-random draws.

In all cases considered we generate randomly 11 data sets and across all the different simulation studies these are constructed by the same random number generating seeds. We compare the performance of four samples: two pseudo-random samples of size N and $r \times N$ (denoted MC and MCr), a digital $(0, m, s)$-net (denoted Net) and lattice points with the bakers's transformation (found to be the best in estimating normal probabilities by Sándor and András 2004, denoted Lat in the tables) of size (approximately) N. In order to measure the performance we estimate the bias, standard deviation (SD) and root mean squared error (RMSE) of the objective function estimates by using 100 replications of the objective function estimates. Across all the different simulation studies we construct the samples based on the same 100 random number generating seeds. Out of the 11 cases we only present three cases for each simulation study. These correspond to the minimum, median and maximum of the ratios of $RMSE_{Net}/RMSE_{MCr}$. In all cases when we use lattice points we chose the parameter value $q = 1571$ from those given by Hickernell et al. (2000). We use this value for all sample sizes and dimension values.

For this model we consider two different data generating processes (DGP's) that differ in the dimension L and number of support points n_h, $h = 1, ..., L$, of the demographic characteristics. In both DGP's we use $J = 100$, we assume that d_{ih} has support of n_h values generated from $N(0, 4)$, where n_h will be specified below. For the sample size we take $N = 512$.

In the first DGP we take $L = 5$, $n_1 = 30$ and $n_2 = ... = n_5 = 3$. For the coefficients of the demographic characteristics we take $\Gamma = \text{diag}(1, 1, 2, 2, 2)$. The results are summarized in Table 1. In this case the sample sizes differ slightly due to the fact that there is no $(0, m, s)$-net of size exactly 512 and dimension 10. Instead we use a $(0, 2, 10)$-net in base 23, which is of size 529. From the results we can see that the estimates of the objective functions obtained by the samples Lat, Net and $MC6$ appear to be fairly precise. In general, the estimates of the objective functions tend to be upward biased. We can also see that a smaller SD corresponds to smaller bias.

Table 1. Random coefficient logit with multiple demographic characteristic

Sample	Size	$\text{Min}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC6}}\right\}$			$\text{Median}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC6}}\right\}$			$\text{Max}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC6}}\right\}$		
		True value=0.2587			True value=0.3040			True value=0.2835		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
MC	529	0.3234	0.4705	0.5709	0.0509	0.0619	0.0801	0.0478	0.0511	0.0700
Lat	512	0.0152	0.0167	0.0226	0.0124	0.0121	0.0173	0.0223	0.0207	0.0304
Net	529	0.0187	0.0180	0.0260	0.0093	0.0114	0.0147	0.0261	0.0190	0.0323
MC6	3072	0.0377	0.0501	0.0627	0.0095	0.0134	0.0164	0.0186	0.0173	0.0254

Notes. (i) The minimum, median and maximum of the RMSE ratios are computed from 11 randomly generated realizations.
(ii) The $RMSE_{Net}/RMSE_{MC6}$ values are 0.41, 0.53, 0.84, 0.86, 0.87, 0.90, 0.98, 1.21, 1.22, 1.23, 1.27.
(iii) The $RMSE_{Lat}/RMSE_{MC6}$ values are 0.36, 0.75, 0.78, 0.81, 0.90, 1.06, 1.07, 1.20, 1.26, 1.42, 1.45.

Regarding the performance of the samples, *Net* has a better performance in the median case than the other samples by having both SD and RMSE the smallest. When the ratio $RMSE_{Net}/RMSE_{MC6}$ is minimal, *Lat* is the best, while when it is maximal, *MC6* is the best. In all cases presented, the other three samples clearly outperform *MC*. The overall relative performances can be assessed from the values of the ratios $RMSE_{Net}/RMSE_{MC6}$ and $RMSE_{Lat}/RMSE_{MC6}$ in the *Notes* at the bottom of the table. According to these, *Net* outperforms *MC6* seven times, so we can say that *Net* performs at least as good as *MC6* in this example. This implies that by using a $(0, 2, 10)$-net in base 23 instead of a pseudo-random sample of size 3072, in order to achieve the same precision of the objective function estimate, we can reduce the computing time by about 83%. According to the values of the ratios, we can also conclude that the lattice sample has a performance comparable to *MC6*.

In the second DGP we take $L = 1$, $n_1 = 2000$ and $\Gamma = 1$. Such a case, in which there is one demographic characteristic with very many possible values, covers the situation when the demographic characteristics are dependent and the researcher has information about their joint discrete distribution (as we mention in a footnote on page 89).

The results are summarized in Table 2. In this example we use a $(0, 3, 6)$-net in base 8, which is of size 512. We note that a sample of this type was found to be the best by Sándor and Train (2004) among several different samples. The larger pseudo-random sample has size 1536. The estimates of the objective functions appear to be more precise than in the previous example.

Table 2. Random coefficient logit with one demographic characteristic

Sample	Size	$\text{Min}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC3}}\right\}$			$\text{Median}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC3}}\right\}$			$\text{Max}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC3}}\right\}$		
		True value=0.2616			True value=0.2009			True value=0.3362		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
MC	512	0.0065	0.0074	0.0098	0.0073	0.0110	0.0132	0.0068	0.0113	0.0131
Lat	512	0.0024	0.0041	0.0047	0.0028	0.0061	0.0067	0.0081	0.0144	0.0165
Net	512	0.0016	0.0020	0.0026	0.0015	0.0046	0.0049	0.0023	0.0077	0.0080
MC3	1536	0.0025	0.0038	0.0045	0.0023	0.0043	0.0048	0.0026	0.0056	0.0062

Notes. (i) The minimum, median and maximum of the RMSE ratios are computed from 11 randomly generated DGP realizations.
(ii) The $RMSE_{Net}/RMSE_{MC3}$ values are 0.57, 0.73, 0.79, 0.85, 0.89, 1.01, 1.02, 1.06, 1.15, 1.21, 1.30.
(iii) The $RMSE_{Lat}/RMSE_{MC3}$ values are 0.46, 0.78, 0.84, 0.84, 1.04, 1.13, 1.21, 1.29, 1.38, 2.68, 3.09.

The results in Table 2 imply that *Net* has a performance much better than *MC*. In the median case its performance is very similar to that of *MC3* and better than *Lat*. Based on the $RMSE_{Net}/RMSE_{MC3}$ ratios, the two samples have rather similar performances. This implies a computing time reduction of about 66% by using a $(0, 3, 6)$-net in base 8 instead of a pseudo-random sample of size 1536. According to the values of the ratios $RMSE_{Lat}/RMSE_{MC3}$, we can conclude that the lattice sample has a performance just slightly poorer than *MC3*.

5.2 Random coefficient logit with limited consumer information

For this model, besides the variables specified at the beginning of this section, we take $J = 20$, $K = 5$, $L = 1$. The probability that affects whether consumer i observes product j is specified as

$$\phi_j(\gamma_i) = \frac{\exp(\gamma_i t_j)}{1 + \exp(\gamma_i t_j)},$$

where $\gamma_i \sim N(-1.5, 1)$, so $\theta_2 = (-1.5, 1)$ and t_j are generated as $U[0, 1]$, $j = 1, ..., J$. We take $\theta_{20} = (-1, 1.5)$ for the fixed value of the parameter used for the importance sampling probabilities. We use a $(0, 2, 26)$-net in base 32, which has size 1024, while the larger pseudo-random sample has size 2024. The results are presented in Table 3.

Table 3. Random coefficient logit with limited information

Sample	Size	$\text{Min}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=0.1594			$\text{Median}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=0.3209			$\text{Max}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=0.2240		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
MC	1024	0.0280	0.0337	0.0438	0.0252	0.0487	0.0549	0.0116	0.0263	0.0287
Lat	1024	0.0275	0.0884	0.0926	0.0144	0.0419	0.0443	0.0203	0.1028	0.1048
Net	1024	0.0190	0.0208	0.0282	0.0164	0.0339	0.0377	0.0065	0.0248	0.0256
MC2	2048	0.0258	0.0381	0.0460	0.0222	0.0379	0.0440	0.0069	0.0105	0.0125

Notes. (i) The minimum, median and maximum of the RMSE ratios are computed from 11 randomly generated DGP realizations.
(ii) The $RMSE_{Net}/RMSE_{MC4}$ values are 0.61, 0.63, 0.72, 0.73, 0.83, 0.86, 0.97, 1.20, 1.21, 1.35, 2.04.
(iii) The $RMSE_{Lat}/RMSE_{MC4}$ values are 1.01, 1.50, 1.97, 2.01, 2.02, 2.16, 2.43, 2.57, 3.02, 5.26, 8.36.

The results in Table 3 imply that *Net* has a performance clearly better than *MC*. In the median case its performance is very similar to that of *MC2* and better than *Lat*, which has a surprisingly poor performance. In fact, according to the values of the ratios $RMSE_{Lat}/RMSE_{MC2}$, the lattice sample has a performance poorer than *MC3* in all cases. Based on the $RMSE_{Net}/RMSE_{MC2}$ ratios, *Net* outperforms *MC2* in 7 out of the the 11 cases considered. This implies a computing time reduction of at least 50% by using a $(0, 2, 26)$-net in base 32 instead of a pseudo-random sample of size 2024.

5.3   Design for the panel mixed logit

We consider a design of type $3^4/2/15$, that is, with $S = 15$ choice sets, $J = 2$ alternatives in each set with four attributes, each attribute having three levels: 1, 2, 3. We use effect coding that assigns [1 0], [0 1], [-1 -1] to the levels 1, 2, 3, respectively. We assume that the true parameters are $\beta = [\text{-1 0 -1 0 -1 0 -1 0}]'$, $\sigma = [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0]'$. We construct nine such designs by the swapping (Huber and Zwerina 1996) and cycling (Sándor and Wedel 2001) algorithms using randomly generated starting designs. For each design we evaluate the design criterion (12) in the way described in Section 3.2. For the integrals (13) involved in the design criterion we use a $(0, 3, s)$-net in base 4. We estimate the sum from (14) by a $(0, 2, s)$-net in base 19, which has size 361, and pseudo-random samples of size 361 and 650. For computing the true values we used a sample of size 100,000. The results are reported in Table 4.

Table 4. Design criterion estimation for the panel mixed logit

Sample	Size	$\mathrm{Min}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=1.497			$\mathrm{Median}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=1.559			$\mathrm{Max}\left\{\dfrac{RMSE_{Net}}{RMSE_{MC2}}\right\}$ True value=1.580		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
MC	361	-0.033	0.066	0.074	-0.035	0.064	0.073	-0.039	0.060	0.072
Net	361	-0.019	0.045	0.049	-0.022	0.047	0.052	-0.028	0.047	0.055
MC2	650	-0.013	0.050	0.052	-0.019	0.049	0.053	-0.022	0.047	0.052

Notes. (i) The minimum, median and maximum of the RMSE ratios are computed from 9 randomly generated DGP realizations.
(ii) The $RMSE_{Net}/RMSE_{MC2}$ values are 0.95, 0.96, 0.97, 0.99, 0.99, 1.02, 1.02, 1.02, 1.06.

The first impression from the results is that all estimates are rather precise and have rather small downward bias. *Net* improves on the performance of *MC* only slightly; its performance is close to that of the pseudo-random sample *MC2* of size slightly less than double. The $RMSE_{Net}/RMSE_{MC2}$ ratios in the Notes below the table show the similarity of *Net* and *MC2*. We can conclude that the reduction in computing time is about 45%.

6   Conclusions

This paper presents results on how quasi-Monte Carlo methods ((t, m, s)-nets and lattice points) perform for estimating large sums. The terms of these large sums typically contain analytically intractable integrals so that computing the sums exactly is computationally unfeasible. The paper concludes that quasi-Monte Carlo methods offer significant advantages in terms of computation. Specifically, $(0, m, s)$-nets used in this paper yield reductions in computing time relative to pseudo-random draws that range between 45-83%.

When the terms of the large sums are integrals, then it is possible to express the sum as one integral having dimension higher than the original integrals. In such cases the performance of the quasi-Monte Carlo methods compared to pseudo-Monte Carlo is better due to the joint treatment of the variables of the obtained integral (see the results reported in Sections 5.1 and 5.2. The design criterion for the panel mixed logit model is a large sum of expressions that are nonlinear in several integrals, so its estimation makes only limited use of the advantages offered by quasi-Monte Carlo samples.

Acknowledgements

References

Berry, S. (1994). Estimating discrete choice models of product differentiation, *RAND Journal of Economics* **25**: 242–262.

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium, *Econometrica* **63**: 841–890.

Bliemer, M.C.J. and J.M. Rose (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations, *Transportation Research B* **44**: 720–734.

Goeree, M.S. (2008). Limited information and advertising in the U.S. personal computer industry, *Econometrica* **76**: 1017–1074.

Harding, M.C. and J. Hausman (2007). Using a Laplace approximation to estimate the random coefficients logit model by nonlinear least squares, *International Economic Review* **48**: 1311–1328.

Heiss, F. and V. Winschel (2008). Likelihood approximation by numerical integration on sparse grids, *Journal of Econometrics* **144**: 62–80.

Hess, S., K.E. Train and J.W. Polak (2006). On the use of a modified latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice, *Transportation Research B* **40**: 147–163.

Hickernell, F.J., H.S. Hong, P. L'Ecuyer and C. Lemieux (2000). Extensible lattice sequences for quasi-Monte Carlo quadrature, *SIAM Journal on Scientific Computing* **22**: 1117–1138.

Huber, J. and K. Zwerina (1996). The importance of utility balance in efficient choice design, *Journal of Marketing Research* **33**: 307–17.

Kessels, R., B. Jones, P. Goos and M. Vandebroek (2009). An efficient algorithm for constructing bayesian optimal choice designs, *Journal of Business and Economic Statistics* **27**: 279–291.

Kessels, R., P. Goos and M. Vandebroek (2006). A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research* **43**: 409–419.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry, *Econometrica* **69**: 307–42.

Sándor, Z. (2001). *Computation, Efficiency and Endogeneity in Discrete Choice Models*, PhD thesis, University of Groningen.

Sándor, Z. and K. Train (2004). Quasi-random simulation of discrete choice models, *Transportation Research B* **38**: 313–327.

Sándor, Z. and M. Wedel (2001). Designing conjoint choice experiments using managers' prior beliefs, *Journal of Marketing Research* **38**: 430–444.

Sándor, Z. and P. András (2004). Alternative sampling methods for multivariate normal probabilities, *Journal of Econometrics* **120**: 207–234.