Taylor & Francis
Taylor & Francis Group

Analytic approximations for computing probit choice probabilities

Richard D. Connors[a]*, Stephane Hess[a] and Andrew Daly[ab]

[a]*Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK;* [b]*RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK*

The multinomial probit model has long been used in transport applications as the basis for mode- and route-choice in computing network flows, and in other choice contexts when estimating preference parameters. It is well known that computation of the probit choice probabilities presents a significant computational burden, since they are based on multivariate normal integrals. Various methods exist for computing these choice probabilities, though standard Monte Carlo is most commonly used. In this article we compare two analytical approximation methods (Mendell–Elston and Solow–Joe) with three Monte Carlo approaches for computing probit choice probabilities. We systematically investigate a wide range of parameter settings and report on the accuracy and computational efficiency of each method. The findings suggest that the accuracy and efficiency of an optimally ordered Mendell–Elston analytic approximation method offers great potential for wider use.

Keywords: multinomial probit; multivariate normal integral; analytical approximation; choice probabilities

1. Introduction

The random utility paradigm models are chosen by identifying the chosen alternative as the one with the highest utility. With the recognition that only part of this utility can be modelled deterministically, a key assumption made in the estimation as well as the application of such models concerns the distribution used to represent the remaining random part of utility. This assumption has major implications in terms of the ability to represent core phenomena, such as unexplained taste heterogeneity across respondents and across choices for the same respondent, correlation across choices for the same respondent and heteroskedasticity across respondents and/or alternatives. Crucially, it also has major implications in terms of the computational cost of both model estimation and application. The choice probabilities are given by multivariate integrals over the distribution of the error terms, and these multivariate integrals have a closed form solution only for certain choice of distribution, such as the family of GEV models (McFadden 1978a, and also see overview in Train 2009).

While the standard closed form GEV models are appealing due to their low-computational cost in estimation and application, there are limits to the degree of complexity that can be conveniently represented by making use of (generalised) only

---

*Corresponding author. Email: R.D.Connors@its.leeds.ac.uk

extreme value distributions for the error term. This justifies moving away from such assumptions in certain contexts. The case where the error terms have a joint multivariate normal (MVN) distribution is referred to as the (multinomial) probit model (dating back to Thurstone 1927). Network modellers, in particular, use the probit specification across a host of application areas, for example, for modelling demand (Zito *et al.* 2010), parking (Teng *et al.* 2008), accidents (Rifaat and Chin 2007) and mode and route choice (Connors *et al.* 2007, Sumalee *et al.* 2009). Similarly, the choice-modelling community is well aware of the theoretical appeal of the probit structure, allowing them to address the various core behavioural phenomena highlighted above. Nevertheless, in the particular context of choice modelling, interest in the probit model waned with the increasing use of mixed logit (see, e.g. Dalal and Klein 1988, McFadden and Train 2000, Train 2009), which has a number of advantages in terms of distributional assumptions and ease of estimation. However, the error structure of the probit model remains appealing and interest has been reinvigorated by Bhat's recent work (Bhat 2010, Bhat and Sidharthan 2011, Bhat *et al.* 2010) that highlights the benefit of adopting the probit choice model within a composite marginal likelihood (CML) framework for the estimation of choice preferences based on survey data.[1]

Notwithstanding the identification issues arising from the number of parameters that may need to be estimated (Dansie 1985, Bunch 1991, Keane 1992), the key constraint in using the probit model remains the cost of computing the choice probabilities that do not have a closed form expression. This is an issue for both prediction and model estimation, where there is a need to evaluate the choice probabilities efficiently and accurately. Numerical integration methods have been developed to accurately compute MVN integrals (Drezner and Wesolowsky 1990, Genz 1992, 2004, Drezner 1994, Genz and Bretz 1999, 2002). Indeed, the works by Genz and Bretz underlie the MVN distribution function currently implemented in MATLAB (as used in this article) that offers off-the-shelf accuracy of at least 10^{-4} for computing each choice probability in situations having up to 25 alternatives. However, the execution time of such methods increases very rapidly with both problem dimension and the accuracy demanded from them. For practical applications that require calculation of very many choice probabilities, these methods are not fast enough.

The standard approach seen in most applications is to evaluate probit choice probabilities via simple frequency simulation (Manski and Lerman 1981): the utilities of the alternatives are drawn from their distribution, the alternative with the highest utility is chosen. This is repeated a number of times, and the choice probability of each alternative is approximated by its choice frequency. This frequency-based (FB) approach faces several difficulties: capturing low-probability alternatives; the results depend on the seed used for the random number generator (i.e. 'noise' is present) and the frequencies are not continuous. The last issue was tackled by McFadden (1989), who processed the simulated frequencies via a logit function. Possibly, the most sophisticated simulation-based approach is the Geweke-Hajivassiliou-Keane (GHK) probability simulator (Börsch-Supan and Hajivassiliou 1993). This approach rewrites the MVN integral as a product of marginal conditional probabilities, and evaluates these marginal probabilities in turn by Monte-Carlo simulation. As it is a probability simulator, GHK avoids the problems of estimating low probabilities, and the probabilities generated are continuous in the parameters. There remains the issue of noise common to all simulation approaches.

In this article we investigate an alternative family of approaches for computing the probit choice probability MVN integrals, that of analytic approximation. Here the MVN integral of interest is transformed into an approximating integral that can be evaluated more easily and without simulation. Several analytic approximations have been proposed in the literature, including the approximations of Clark (1961) and Mendell and Elston (1974), and Langdon's (1984a, 1984b) separated split procedure; all of which were compared by Kamakura (1989). A method proposed by Solow (1990) was extended by Joe (1995), who compared this approach with those of Clark (1961), Mendell and Elston (1974) and Drezner (1990). Within the particular context of algorithms for solving a probit stochastic user equilibrium, Rosa (2003) compared the above-mentioned approximations. Both Kamakura and Rosa recommended the Mendell and Elston (ME) approach as providing the best compromise between speed and accuracy. In contrast, Joe (1995) found the Solow–Joe (SJ) method to perform best. These contradictory conclusions leave the analyst without a clear recommendation on the most appropriate approach to adopt. Additionally, tests reported to date have tended to be rather restrictive in both the range of parameter values considered and the methods compared, hampering the ability to reach general conclusions. The contribution made in this article is to conduct a systematic empirical comparison between key methods. In light of the conclusions from those comparisons listed above, we restrict the attention in this article to the two preferred analytic approximation methods: ME and SJ, and compare their performance with three Monte Carlo-based methods. Our empirical results offer clear evidence that the ME approach outperforms the SJ approach across a wide range of settings. Additionally, ME is significantly faster than using Monte Carlo approaches for choice problems with up to 15 alternatives while attaining comparable accuracy; details of the trade-offs between computation time and accuracy as the problem dimension increases are illustrated and discussed.

The remainder of this article is organised as follows. The next section presents multinomial probit methodology and discusses the specific analytical approximation techniques used in this work. This is followed by the empirical work in Section 3. Finally, Section 4 presents the conclusions from our work.

2. Methodology

Consider the issue of calculating the MVN integral necessary to determine choice probabilities for a probit model. The choice probability for a given alternative j from a choice set containing K elements $[j = 1, 2, \ldots, K]$ is given by

$$P_j = \Pr\big[U_j > U_k; k = 1, \ldots, K, k \neq j\big] \tag{1}$$

where the vector of random utilities $\big[U_j\big] = \boldsymbol{U}$ is MVN distributed with mean $\boldsymbol{V}$ and covariance Σ.

$$\boldsymbol{U} \sim \mathrm{MVN}(\boldsymbol{V}, \Sigma) \tag{2}$$

where utility comprises the deterministic utility given by $\boldsymbol{V}$, and, without loss of generality, a zero-mean random vector $\boldsymbol{e}$

$$U_j = V_j + e_j \quad \text{where } \boldsymbol{e} \sim \mathrm{MVN}(0, \Sigma) \tag{3}$$

Thus

$$P_j = \Pr\big[e_j - e_k > V_k - V_j; k \in C; k \neq j\big] \tag{4}$$

and the differences $Y_{jk} = e_j - e_k$ are also normally distributed. Define

$$\Delta^j = [\delta_{lm}] \text{ where } \delta_{lm} = \begin{cases} -1 & \text{for } m = l \\ 1 & \text{for } l = j \qquad l = 1, 2, \ldots, K; m = 1, 2, \ldots, K-1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The Δ^j matrix needed to give the differences for computing P_j is generated from a $K \times K$ zero matrix by putting -1 on the diagonal, setting the j-th row to 1s and deleting the j-th column. The $K-1$ vector of differences $\mathbf{Y}^j = \Delta^j T\mathbf{e}$ have zero mean and covariance $\Omega^j = \Delta^j T\Sigma\Delta^j$. Then with $\mathbf{y}^j = \Delta^j T\mathbf{V}$ we can write the j-th choice probability

$$P_j = \Pr[\mathbf{Y}^j < \mathbf{y}^j] = \iiint_{-\infty\ldots-\infty}^{y_1^j\ldots y_{K-1}^j} \frac{1}{\sqrt{(2\pi)^{K-1}|\Omega^j|}} \exp\left[-\frac{1}{2}\mathbf{Y}^T(\Omega^j)^{-1}\mathbf{Y}\right] \mathrm{d}Y_1 \mathrm{d}Y_2 \ldots \mathrm{d}Y_{K-1} \tag{6}$$

where the upper limits are given by the $\mathbf{y}^j$. For simplicity of notation, we will now begin to drop the index j. Finally, we transform to standard (multivariate) normal by setting all variances to unity; writing the diagonal matrix

$$S = \mathrm{diag}\left[\frac{1}{\sqrt{\Omega_{1,1}^j}}, \ldots, \frac{1}{\sqrt{\Omega_{K-1,K-1}^j}}\right] \tag{7}$$

then defining

$$Z = S \cdot \mathbf{Y}^j \tag{8}$$

The covariance matrix for Z is $\Sigma_Z = S\Omega^j S$, which then has ones along its diagonal. With $\mathbf{z} = S \cdot \mathbf{y}^j$, the choice probability we wish to compute is now

$$P_j = \Phi_{\Sigma_Z}(\mathbf{z}) \tag{9}$$

with Φ_Σ the standard MVN CDF with covariances given by Σ, i.e.

$$P_j = \iiint_{-\infty\ldots-\infty}^{z_1\ldots z_{K-1}} \frac{1}{\sqrt{(2\pi)^{K-1}|\Sigma_Z|}} \exp\left[-\frac{1}{2}Z^T\Sigma_Z^{-1}Z\right] \mathrm{d}Z_1 \mathrm{d}Z_2 \ldots \mathrm{d}Z_{K-1} \tag{10}$$

Calculating the choice probability for each alternative is hence equivalent to the evaluation of the CDF for the $(K-1)$-dimensional standard MVN with mean $\mathbf{0}$, unit variances and covariance Σ_Z (comprising correlations ρ_{ij}). All lower limits of integration are $-\infty$ and the upper limits are finite but may take positive or negative values.

The choice probability for each of the K alternatives requires one evaluation of the CDF for the $(K-1)$-dimensional standard MVN. We therefore restrict our attention to

the case of an n-dimensional random variable $\boldsymbol{Z} \sim \text{MVN}(0, \Sigma)$ with unit variances and dimension n, (for ease of notation, set $n = K - 1$). We wish to compute the probability

$$P = \Pr[Z_1 < z_1, Z_2 < z_2, \ldots, Z_n < z_n] \tag{11}$$

This probability can be evaluated by Monte Carlo simulation. We draw many values of $\boldsymbol{Z} = [Z_1, \ldots, Z_n]$ from its distribution, and count the proportion of these samples that satisfy (11). This gives an estimate of the true probability, which will improve as the number of draws increases.

Mendell and Elston (1974) employ a result of Aitken (1935) to extend the bivariate results of Pearson (1903) to the multivariate case. Rewriting the MVN integral (11) in terms of conditional probabilities:

$$P = \Pr[Z_1 < z_1] \cdot \Pr[Z_2 < z_2 | Z_1 < z_1] \cdot \Pr[Z_3 < z_3 | \{Z_1 < z_1, Z_2 < z_2\}] \cdot \cdots$$
$$\cdot \Pr[Z_n < z_n | \{Z_1 < z_1, \ldots, Z_{n-1} < z_{n-1}\}] \tag{12}$$

$$P = \Pr[Z_1 < z_1] \cdot \prod_{i=2}^{n} \Pr[Z_i < z_i | \{Z_1 < z_1, \ldots, Z_{i-1} < z_{i-1}\}] \tag{13}$$

The ME approach is to approximate each univariate conditional distribution (which is not normal) with a normal distribution matching the mean and variance. Kamakura (1989) gives a clear account of the ME approximation for the evaluation of probit choice probabilities.[2]

Joe (1995) presents several closely related methods, extending the work of Solow (1990) to the more general problem of computing probabilities for 'rectangular areas', i.e.

$$\Pr(w_1 < Z_1 < z_1, w_2 < Z_2 < z_2, \ldots, w_n < Z_n < z_n) \quad \text{where } w_i, z_i \in R \tag{14}$$

We consider only the case of evaluating orthant probabilities, having infinite lower limits. For such cases, the simplest SJ approach is to write

$$P = \Pr[Z_2 < z_2, Z_1 < z_1] \cdot \prod_{i=3}^{n} \Pr[Z_i < z_i | \{Z_1 < z_1, Z_2 < z_2, \ldots, Z_{i-1} < z_{i-1}\}] \tag{15}$$

The conditional probabilities in the product are then approximated using indicator functions and the following result, which is exact only in the bivariate case, extended to higher dimensions:

$$\text{If} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right) \text{then } E(Y_2 | Y_1 = y_1) = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1}(y_1 - \mu_1). \tag{16}$$

Extended versions of this approach were also proposed, founded on the evaluation of an exact trivariate or quadrivariate probability, in place of the bivariate used in (15), with remaining terms similarly collected as a product of conditional probabilities. These extended versions were tested as part of our investigations. Clearly, they offer more accuracy, though they are very much slower than the bivariate version due to the multiple 'exact' evaluations of three- or four-dimensional integrals required.

In both ME and SJ approaches, the order of conditional terms may influence the result. For ME, Kamakura (1989) suggests ordering the terms *within* the calculation of each choice probability according to their covariances. Joe suggests that (for both ME and SJ approaches) an average is taken over many permutations of the variate order. Since the number of available permutations is $n!/2$, Joe suggests 100–10,000 should be sufficient for high-dimensional problems. Averaging over multiple re-orderings increases the computation time in proportion to the number of repetitions comprising the average.

Comparisons of analytic approximation methods have appeared in published articles, although Joe (1995) appears to be the only comparison including both ME and SJ. The tests in Joe (1995) considered problems of dimension 5 and 9, with constant positive correlation of 0.1 and then 0.4, computing the specific choice probabilities

$$P = \Pr[Z_1 < z_1, Z_2 < z_2, \ldots, Z_n < z_n] \quad \text{with } z_1 = z_2 = \cdots = z_n = w \qquad (17)$$

for the cases $w = 0, 0.2, 0.4, 0.6, 0.8$. Therefore, all the tests comparing ME and SJ comprise a total of just 20 individual choice probabilities. The results indicate that SJ is more accurate than ME. Computation time is not explicitly compared, though ME is noted to be faster.

The above discussion has shown that the existing comparison between the two methods is not comprehensive enough to allow us to reach general conclusions. Additionally, the question of optimal reordering remains. This article addresses these two issues. We present a wide range of numerical tests to investigate the accuracy and efficiency of both the ME and SJ methods, implemented using the algorithms described in Kamakura (1989) and Joe (1995), respectively. We also investigate the impact of reordering the terms in (13) and (15).

To give a more complete picture, we also present the same computations made using three Monte Carlo methods. First, using a simple FB approach, second using the GHK method (implemented as described in Bolduc 1999) and finally using a quasi-Monte Carlo approach (GZ) (Genz, 1992; using results from Niederreiter 1972, Cranley and Patterson 1976) for which MATLAB code is readily available.[3]

3. Empirical work

At this stage, it is worth justifying our focus on the accuracy of the computation of probabilities alone, rather than quantities that depend on these parameters. Choice models have two primary uses, namely the estimation of parameters that are used to explain the choices observed in data, and the forecasting of choices in hypothetical or future scenarios. Clearly, forecasts themselves are just functions of probabilities (e.g. the forecast share for a given mode is obtained by weighted summation across individuals), and as such, the importance lies in the accurate calculation of the underlying probabilities. Similarly, parameter estimates are obtained through maximisation of the likelihood function, and any error in the computation of probabilities that the likelihood function is based on may translate into error in these estimates. This has been illustrated extensively for the case of maximum simulated likelihood in the context of using quasi-Monte Carlo methods (Train 2000, Bhat 2001, Hess *et al.* 2006), and is also reflected in the theoretical properties of maximum simulated likelihood estimates (Lee 1992, 1995).

3.1. *Test methodology*

In the tests presented here, we consider MVN-distributed utilities whose differences will then be transformed into standard MVN variates (as outlined in Section 2) to calculate the choice probability for each alternative. Each test therefore comprises the calculation of choice probabilities for MVN-distributed utilities $U \sim \text{MVN}(V, \Sigma)$. The choice probabilities themselves, and the accuracy of each approximation method, will depend on the mean and the variance–covariance matrix. Specifically, it may be that some approximation methods do better or worse when faced with correlations, or large differences in the variances or in the means of the alternatives.

In the tests below we generate many MVN distributions; each requires a mean vector, V, and a covariance matrix, Σ. We need to consider how to generate these means and covariances in order to systematically investigate the space of all possible MVN distributions, with the aim of uncovering the accuracy (or lack thereof) of the methods being tested.

With a fixed covariance matrix, we illustrate the impact of generating the mean, V, from different ranges. Figure 1 shows three bivariate utility distributions (A, B, C). Sampling each component of the mean from $U([-1, 1])$, as marked by the dashed square, would result in distributions such as A. The distribution straddles the diagonal $U_1 = U_2$, and hence both alternatives have significant (non-zero) choice probabilities. With the same covariance matrix, randomly sampling the components of the mean from a larger range such as $U([-10, 10])$ often results in a distribution located like B, where the choice probabilities will be 0/1 even to the maximum accuracy we might hope to attain. Such cases do not rely on highly accurate estimation of that part of the probability density function (PDF) above or below the diagonal. Correlations between alternatives are also influential, and may raise or lower the chance of selecting the alternative with lower mean utility. The strong positive correlation between alternatives in distribution C increases the probability of choosing alternative 1, in comparison with having independent alternatives.

In our tests, we sample the mean for the MVN distribution from $U([-L, L]^N)$, where the number of alternatives is N. Initial tests indicated that the accuracy of the ME and SJ

Distributions plotted

$$A = \text{MVN}\left([0.9, 0.5], \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right)$$

$$B = \text{MVN}\left([1.0, 7.0], \begin{bmatrix} 2 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right)$$

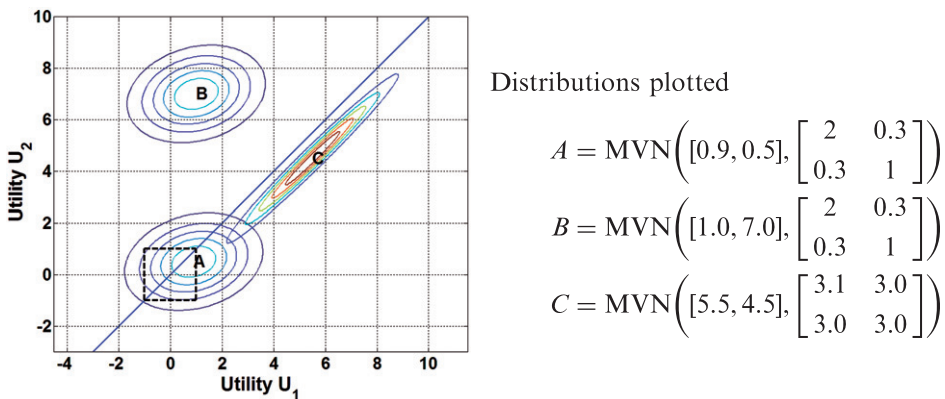$$C = \text{MVN}\left([5.5, 4.5], \begin{bmatrix} 3.1 & 3.0 \\ 3.0 & 3.0 \end{bmatrix}\right)$$

Figure 1. Bivariate distributions illustrating the region of integration for choice probabilities at different settings for the mean and covariance.

approximation methods may depend on the (relative) size of the off-diagonal terms in the covariance matrix. With this in mind, we test covariance matrices $\{\Sigma_r\}$ of the form $\Sigma_r = \alpha I + C_r$ where I is the identity matrix and $\{C_r\}$ are randomly generated correlation matrices (built using eigenvalues randomly drawn from a uniform distribution). The scalar α therefore parameterises the relative magnitude of the variances compared with the (off-diagonal) correlations. Increasing α increases the magnitude of the diagonal terms, without changing the off-diagonal/correlation terms.

We test the values of α and L labelled $\{\alpha_j : j = 1, \ldots, J\}$ and $\{L_k : k = 1, \ldots, K\}$. Increasing L leads to the occurrence of more near 0/1 probabilities, as described above. Increasing α amplifies the diagonal terms, which dominate the covariance matrix, *i.e.* larger α corresponds to effectively smaller correlations. The tests are built from several components:

 (I) The number of alternatives (for each choice) determines the problem dimension N.
 (II) Given N, we generate a set of $N \times N$ correlation matrices $\{C_r, r = 1, \ldots, R\}$.
 (III) Given L_k, we draw R mean vectors V_r from the uniform distribution $U\big([-L_k, L_k]^N\big)$.
 (IV) Given α_j, we compute the choice probabilities for $U \sim \text{MVN}\big(V_r, \alpha_j I + C_r\big)$ by each method, for each of the pairs $V_r, C_r \; r = 1, \ldots, R$.
 (V) For each method, we compute the error in each individual choice probability (there are $N \times R$).

Hence for each problem dimension N, at each setting of L_k and α_j we compute the choice probabilities for R MVN distributions. With N alternatives, the choice probabilities for a given $U \sim \text{MVN}\big(V_r, \alpha_j I + C_r\big)$ give an N-vector: P_{ML} from 'exact' computation (using MATLAB's mvncdf function), P_{SJ} from the SJ approximation, P_{ME} via ME and the three Monte Carlo methods P_{FB}, P_{GHK} and P_{GZ}.

The 'exact' choice probabilities were computed to an accuracy of 1E−4 in MATLAB. These values are used to calculate errors in the choice probabilities computed by other methods, for example the ME errors are $E_{\text{ME}} = P_{\text{ML}} - P_{\text{ME}}$. Errors below 1E−4 therefore cannot be regarded as precise.

3.2. *Role of internal averaging in analytic approximation methods*

The first set of tests consider the improvement gained by averaging over multiple reordering of the variates (as suggested by Joe 1995). These re-orderings are random permutations of the variates in (13) and (15). We also establish a number of draws for each Monte Carlo method that offers reasonable comparison with the analytical approximation methods.

We generated a set of 25 choice situations: $U^i \sim \text{MVN}(V_i, \alpha_i I + C_i)$, $i = 1, \ldots, 25$. Each V_i drawn from $U\big([-1.5, 1.5]^N\big)$, each α_i drawn from $U([0, 4])$ and $\{C_i\}$ are generated as described above. For these 25 choice problems, the choice probabilities were computed using each method and mean absolute error across all 25 situations calculated. This was repeated for the number of choice alternatives $N = 5, 7, 9, 15$. Figures 2 and 3 show the mean absolute error in the choice probabilities computed by ME and SJ as the number of
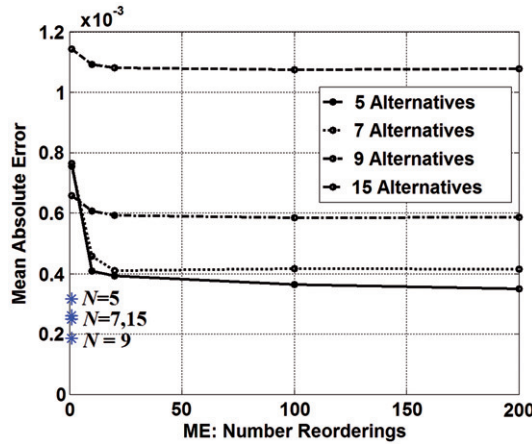
Figure 2. Mean absolute error in choice probabilities computed by ME, with increasing number of terms included in the internal average. Asterisks indicate mean absolute error for the optimal ordering MEO (for $N = 5, 7, 9, 15$).
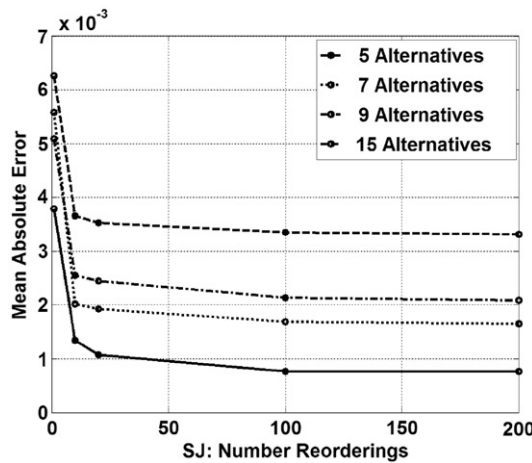

Figure 3. Mean absolute error in choice probabilities computed by SJ with increasing number of terms included in the internal average.

re-orderings included in the average increases. For both ME and SJ, $[1, 10, 20, 100, 200]$ random re-orderings were tested.

For both SJ and ME methods, various non-random orderings were also tested (prompted by Langdon 1984b); specifically, the terms in (11) were ordered by increasing or decreasing $\{z_i\}$ before writing expressions (13) and (15). For the ME method, the order of decreasing $\{z_i\}$ was found to be very effective. ME using this single optimal order of terms is labelled 'MEO'. For the SJ method, decreasing $\{z_i\}$ made a slight improvement, but averaging over multiple random orderings was much more effective than any single ordering.
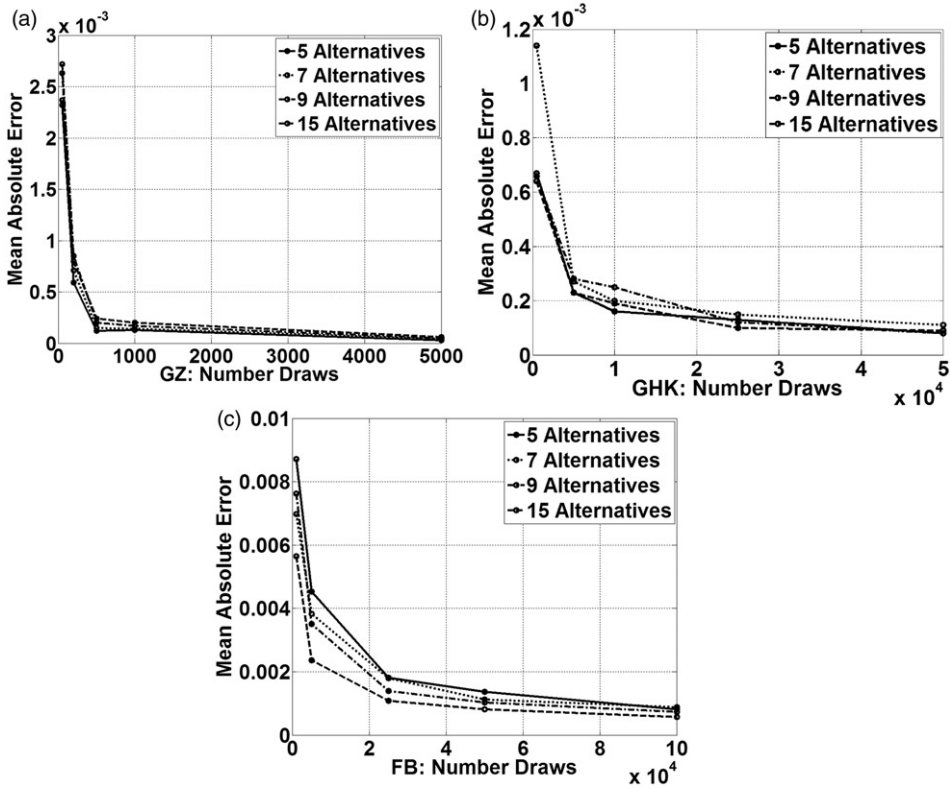
Figure 4. (a) Mean absolute error in choice probabilities computed by GZ with increasing number of draws. (b) Mean absolute error in choice probabilities computed by GHK with increasing number of draws. (c) Mean absolute error in choice probabilities computed by FB with increasing number of draws.

Many more tests were performed that confirmed the trends shown here: for all problem dimensions internal averaging significantly improves the accuracy of SJ, much less so for ME. The ME outperforms SJ (note *y*-scale in Figures 2 and 3); in fact, the ME method with no averaging (any random single internal ordering of terms) is often more accurate than the SJ method with all permutations included. MEO with the single optimal ordering of terms (asterisks in Figure 2) significantly outperforms the unordered ME.

Similar tests were run to determine appropriate numbers of draws for the three Monte Carlo-based methods (Figure 4). Note that different numbers of draws were tested for each method, hence both axes have different scales in each of the graphs. For the GZ method the tests examined [50, 200, 500, 1000, 5000] draws. For GHK [500, 5000, 10,000, 25,000, 50,000] draws were tested and for the FB approach [1000, 5000, 25,000, 50,000, 100,000] draws were considered.

Based on these findings, the tests that follow in the next section consider only the single optimal ordering of terms for ME, labelled MEO. For SJ we report tests using 10 internal re-orderings, labelled SJ10; averaging over 10 random re-orderings improves accuracy appreciably. No more re-orderings were included as execution time increases with each additional reordering. For the Monte Carlo methods, the simple FB MC is run using

50,000 draws (labelled FB50k), GHK with 10,000 draws (labelled GHK10k) and the quasi-MC method of Genz is run with 500 draws (labelled GZ500). These selections are a compromise between accuracy and computation time and were chosen to give the best comparison with MEO and SJ in the tests below.

3.3. *Comparison of computation speed with precision*

We now test the trade off between execution speed and accuracy across a wide range of settings. In turn, $N = 3, 5, 7, 9, 15$ are considered. For each problem dimension N, we consider $K = 17$ settings of the utility range L_k, and $J = 19$ values for the (relative) magnitude of variance α_j:

$$L_k \in \{0.1, 0.72, 1.34, 1.96, 2.58, 3.19, 3.81, 4.43, 5.05, 5.67, 6.29, 6.91, 7.53, 8.14, 8.7, 9.38, 10.0\}$$

$$\alpha_j \in \{0.1, 0.76, 1.42, 2.08, 2.74, 3.41, 4.07, 4.73, 5.39, 6.05, 6.71, 7.37, 8.03, 8.69, 9.36, 10.02,$$
$$10.68, 11.34, 12.0\}$$

At each setting (L_k, α_j), we draw R values for the deterministic utility $V \sim U(-L_k, L_k)^N$ and generate R covariance matrices (having diagonal elements $= \alpha_j + 1$); the N choice probabilities are then computed for these R MVN distributions via each method. We use $R = 25$, hence for each problem dimension, N, we compute the N choice probabilities for $R \times J \times K = 8075$ MVN distributions. Tests extending beyond these parameter ranges did not reveal any substantial new insights.

We present a summary of results in Table 1. The average time for evaluating one choice with N alternatives is recorded in seconds, and also in multiples of the time taken for the MEO computation. The proportion of errors $> 1E-4$ (the error used for computing 'exact' values) and $> 1E-3$ is recorded.

Note that for choice situations with three alternatives, there are only two differences in utility, and hence the integral (10) has only two variates. In this case, due to the bivariate term in (15), SJ10 is not an approximation method and evaluates the bivariate integral 'exactly'. Nevertheless, even in this case MEO offers a significant improvement in speed and gives 99.4% of choice probabilities accurate to $1E-3$. From hereon we do not consider the case of $N = 3$ where exact methods are applicable.

Throughout the tests reported ($N = 5, 7, 9, 15$), MEO with optimal ordering is always the quickest method, and is more than 20 times faster than any of the methods that give comparable accuracy. As the problem dimension increases, FB50k improves in its speed relative to MEO, though it is still four times slower at $N = 15$ and five times less precise.

The execution time of every method increases with problem dimension. The execution times of methods ML, SJ10, GHK10k not only increase, but also worsen significantly in comparison with MEO. The relative speed of GZ500 also gets slightly worse compared to MEO as the problem dimension increases. Not surprisingly the execution time of FB scales best with N and its accuracy does not degrade appreciably, though it is one of the least accurate methods throughout (with this number of draws).

The most accurate methods are MEO, GHK10k and GZ500, though the latter is consistently $20\times$–$30\times$ slower than MEO. For $N = 5$, GZ500 offers the highest level of accuracy, with more than half of all points having errors within the range of the 'exact' ML method. For $N = 7$, the MEO and GZ500 methods are comparable in accuracy, and at $N = 9$ and 15 we see that MEO has eclipsed the accuracy of GZ500. MEO is more than

Table 1. Summary of computation times and errors.

nAlt		MEO	GZ 500	GHK 10 k	FB 50 k	SJ10	ML
3	Time per choice (s)	0.0005	0.0073	0.0250	0.0135	0.0061	0.0058
	Time per choice ($\times$ MEO)	1	14	49	26	12	11
	Errors $>1E-4$	54	7	32	94	0	0
	Errors >0.001	0.6	0	1.1	56	0	0
5	Time per choice (s)	0.0014	0.0303	0.0871	0.0226	0.0292	0.6580
	Time per choice ($\times$MEO)	1	22	63	16	21	477
	Errors $>1E-4$ (%)	69	38	47	92	86	0
	Errors >0.001 (%)	0.9	0.0	2.3	48.8	39.3	0
7	Time per choice (s)	0.0029	0.0703	0.1974	0.0316	0.0846	2.2461
	Time per choice ($\times$MEO)	1	25	69	11	30	788
	Errors $>1E-4$ (%)	55	51	47	90	91	0
	Errors >0.001 (%)	0.8	0.1	2.4	42.6	55.8	0
9	Time per choice (s)	0.0048	0.1265	0.3487	0.0367	0.1811	4.3376
	Time per choice ($\times$MEO)	1	26	73	8	38	906
	Errors $>1E-4$ (%)	50	56	46	88	92	0
	Errors >0.001 (%)	0.8	0.8	2.5	37.9	62.3	0
15	Time per choice (s)	0.014	0.42	1.24	0.064	0.878	22.224
	Time per choice ($\times$MEO)	1	30	88	5	63	1584
	Errors $>1E-4$ (%)	61	62	42	84	92	0
	Errors >0.001 (%)	0.3	3.6	1.8	28.3	68.0	0

25 times faster in both cases. GHK10k also competes well in terms of accuracy, especially at higher dimensions, but is 60–80 times slower than MEO. Note that (as with all Monte Carlo methods) GZ could be run with more draws, and may then match GHK in terms of both increased accuracy and reduced speed.

Using the same test data, we plot the cumulative distributions of absolute errors in the $8075 \times N$ individual choice probabilities in Figures 5–8 for $N = 5, 7, 9, 15$, respectively. The lines showing the distribution of errors for MEO, GZ500 and GHK10k are close together with more than half the errors below $2E-3$, while the absolute errors for SJ10 and FB50k are substantially larger. Note the errors for MEO are greater than GZ500 and GHK10k for $N = 5$, these three methods are almost equivalent for $N = 5, 9$ while for $N = 15$ GHK10k appears more accurate (though 88 times slower) than MEO.

Next we examine whether or not the errors are dependent on the values of α and L. Recall that the mean is sampled from $U\big([-L, L]^N\big)$ with larger L increasing the occurrence of 0/1 probabilities, while α controls dominance of the diagonal terms in the covariance matrix. At each setting $\big(L_k, \alpha_j\big)$, we compute the mean absolute error across the $25 \times N$ individual choice probabilities. For $N = 7$, contours of these mean absolute errors are plotted in Figure 9. The tests with 5, 9 and 15 alternatives reveal a similar pattern in the errors for each method, although the magnitude of the errors changes as shown above.

The upper limit of the displayed errors is taken from the maximum of the mean absolute errors from all methods, excluding SJ10. The large white area in SJ10 therefore denotes errors of at least 0.0014. SJ10 is the least accurate method across a wide range of settings, though does well along the α axis where the mean of the underlying MVN PDF is
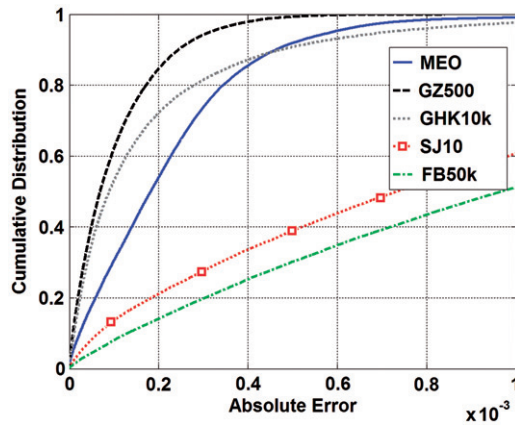
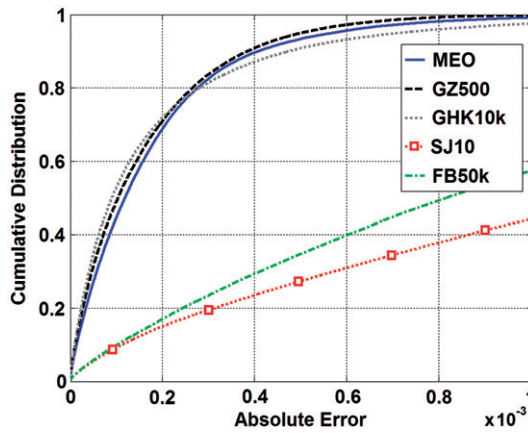Figure 5. Distribution of absolute errors in choices with five alternatives.

Figure 6. Distribution of absolute errors in choices with seven alternatives.
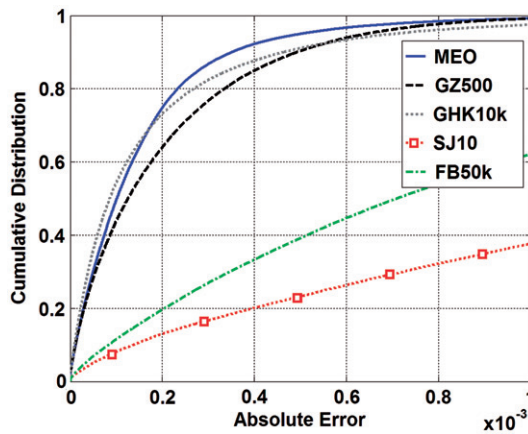

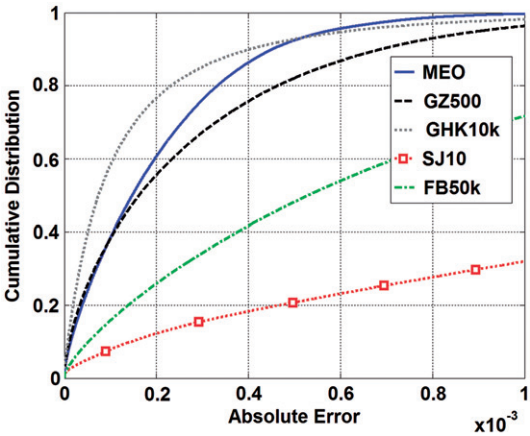Figure 7. Distribution of absolute errors in choices with nine alternatives.

Figure 8. Distribution of absolute errors in choices with 15 alternatives.
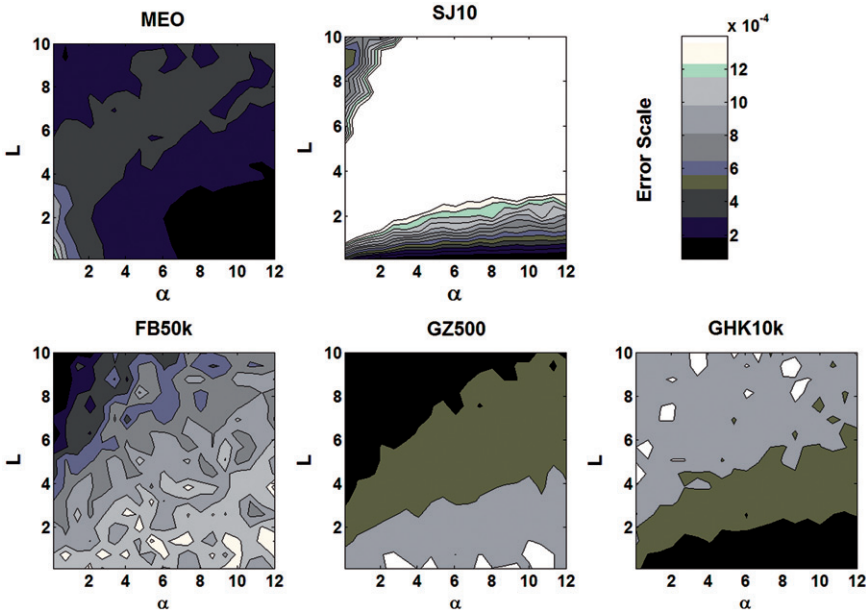

Figure 9. Mean absolute error for choices with seven alternatives, sampling the mean from $[-L, L]^N$ and setting the variances to be α in the underlying MVN distribution of utilities (Section 3.1).

constrained to be near zero. MEO is accurate throughout the range of different choice situations tested, doing less well only in a region near to the origin, where $\alpha = 0.1$ so that correlations dominate the covariance matrix. In this specific region, GZ500 and GHK10k perform better; this was similarly the case for $N = 5, 9, 15$. Methods GZ500 and GHK10k display a distinct overall trend in errors, which was noticeable across all problem dimensions tested. This indicates the values of α and L where one of these methods might be preferred over the other, though computation time also should be considered.

Plots (not displayed here) of the maximum error for any individual choice probability, reveal similar patterns of performance to those shown in Figure 9.

It is worth noting a potential weakness of the SJ method that arises from the need to invert a matrix relating to the covariance. Computing the probabilities via SJ (implemented here following the algorithm described in Joe 1995) can lead to ill-conditioning of this matrix and hence inaccurate results. It may be possible to design a different algorithmic implementation of the SJ approach to guard against this problem, but this is outside the scope of this article.

3.4. *Distribution of error with individual probability value*

We now investigate whether there are systematic errors with probability value, with a view to determine, for example, if small probabilities are always under or over-estimated. The errors in every individual choice probability from the above tests are plotted in Figures 10 and 11. There are $8075 \times N$ points for each method on each plot.

The distribution of errors may not be clear due to the concentration of the plotted points, hence also refer to Figures 5 and 7. All sets of axes have the same scale to aid comparison. Similar plots of the errors for 7 and 15 alternatives do not reveal any substantial new insights.
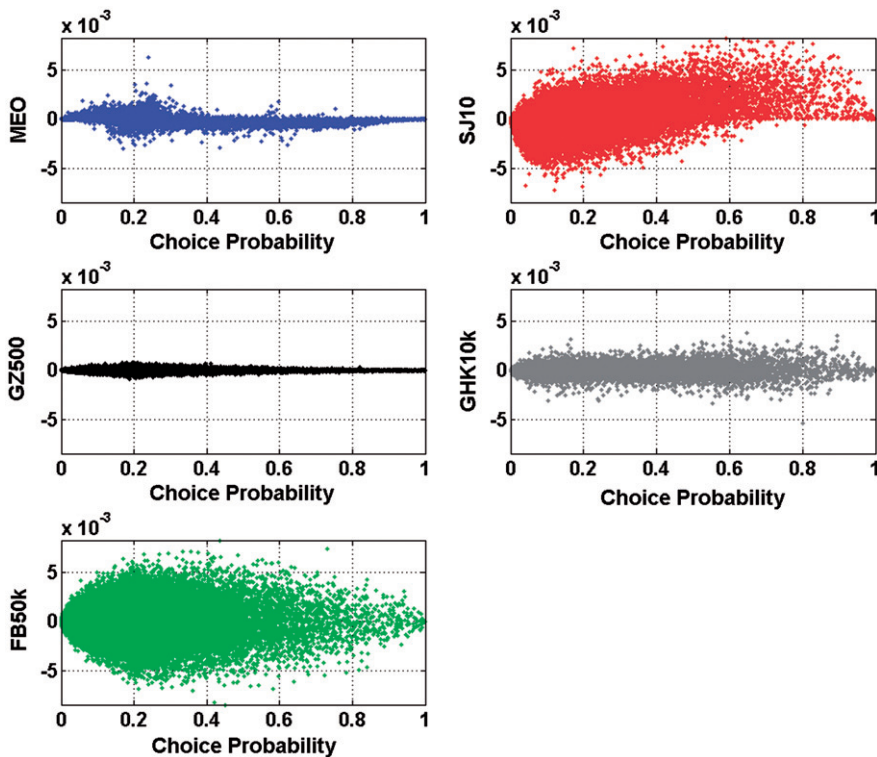

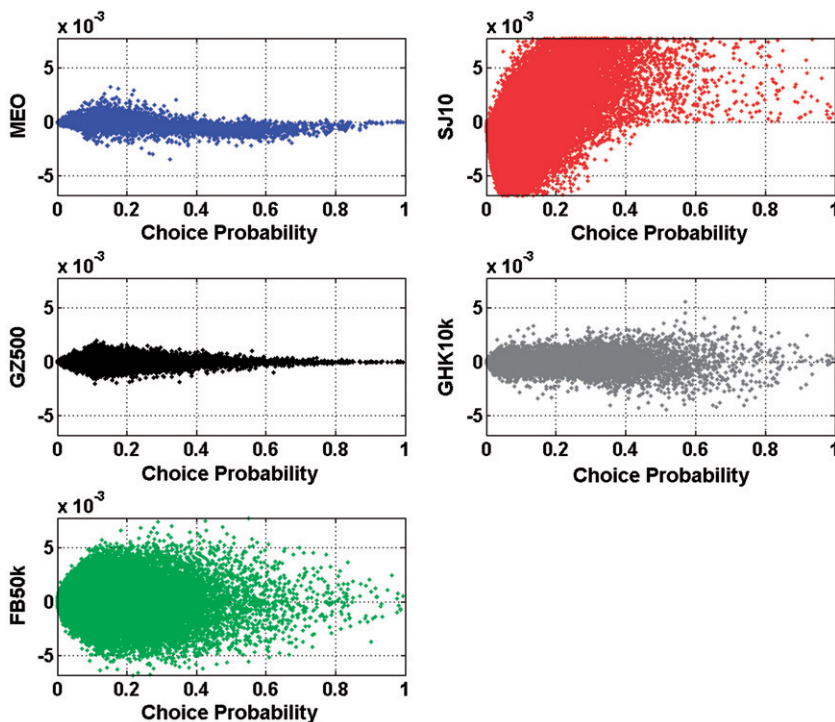Figure 10. Error with choice probability value for choice situations with five alternatives.

Figure 11. Error with choice probability value for choice situations with nine alternatives.

Some errors in SJ10 extend beyond the area plotted in Figures 10 and 11, increasing the *y*-axis to include these points would not allow the detail in the MEO, GZ500 and GHK10k errors to be examined. Our results show the accuracy for MEO and GZ500 across all probability values. All Monte Carlo methods show a symmetric distribution of errors, this is not the case for MEO, and even more markedly so for SJ10.

3.5. *Computing derivatives*

As Daganzo (1979, p. 72) explains (attributed to McFadden 1978b), the (off diagonal) partial derivatives of the choice probabilities can be expressed exactly as the choice probability from a probit model with one fewer alternative, i.e.:

$$\frac{\partial P_i(V, \Sigma)}{\partial V_j} = -\left(\frac{|\Sigma(ij)|}{2\pi|\Sigma|}\right)^{\frac{1}{2}} \exp\left(\frac{1}{2}K(ij)\right) P_{i'}(V(ij), \Sigma(ij)) \tag{18}$$

where $i' = i - 1$ if $j < i$, $i' = i$ if $j > i$ and $\Sigma(ij), K(ij), V(ij)$ are simple transformations of V and Σ with the *j*th rows and columns removed. The diagonal terms $\partial P_i/\partial V_i$ can then be computed immediately from the off-diagonal derivatives. Hence, fast and accurate calculation of the derivatives is equivalent to fast and accurate calculation of the choice probabilities as presented above.

For the estimation of choice models, the derivatives of the log-likelihood function are of interest, therefore requiring first derivatives of the choice probabilities with respect to the preference parameters. These are easily constructed from the above result. Similarly, in traffic networks, sensitivity analysis of the equilibrium flows can be an important tool and this relies on the Jacobian of the choice probabilities exactly as written above (see, e.g. Connors *et al.* 2007).

For completeness, we also briefly present results showing the accuracy of using numerical differences to approximate these partial derivatives, when the choice probabilities being differenced have been calculated using MEO, SJ or via the Monte Carlo approaches. Following the same general approach as above, here we consider 100 distributions $U^i \sim \text{MVN}(V_i, \alpha_i I + C_i)$, which is effectively a subsample of the scenarios investigated in the above tests (Sections 3.3 and 3.4). The matrices $\{C_i\}$ are randomly generated correlation matrices as before, with α drawn from $U([0.1, 12])$. The mean utilities V are drawn from $U([-10, 10]^N)$. For each V, we construct the difference matrix as follows (shown here for $N = 3$ to conserve space):

$$\Delta P = \frac{1}{\delta V} \begin{bmatrix} P_1(V) - P_1(V_1) & P_1(V) - P_1(V_2) & P_1(V) - P_1(V_3) \\ P_2(V) - P_2(V_1) & P_2(V) - P_2(V_2) & P_2(V) - P_2(V_3) \\ P_3(V) - P_3(V_1) & P_3(V) - P_3(V_2) & P_3(V) - P_3(V_3) \end{bmatrix} \qquad (19)$$

where $V_m = V - \delta V \times [0, \ldots, 0, 1, 0, \ldots 0]^T$ with a one in the mth place. This approximate Jacobian is calculated using choice probabilities according to each of the above methods, and using the 'exact' choice probabilities as above. The absolute errors in the elements of the Jacobian are calculated and averaged over all tested choice situations (100 for each of 5, 7 and 9 alternatives). Each entry in Table 2 shows the mean absolute error when differencing using $\delta V = [0.001, 0.01, 0.1]$ in turn.

As noted earlier, errors in the 'exact' method may strongly influence the results when differencing small probabilities. We therefore record how many choice probabilities are sufficiently small that they are within the computational error 1E-4 of the 'exact' method.

All tests show that SJ is the least accurate method. FB50k and MEO are similarly accurate for $\delta V = 0.001$, while MEO offers improved accuracy at $\delta V = 0.01$ and $\delta V = 0.1$. Meanwhile, methods GZ500 and GHK10k are more accurate in computing these derivatives than all other methods tested, it should be noted that any speed penalty is increased by N times in computing these numerical differences. For 5, 7 and 9 alternatives,

Table 2. Approximating derivatives by numerical differencing: mean absolute error in ΔP over 100 computations using interval $\delta V = [0.001, 00A00.01, 00A00.1]$.

nAlt	5	7	9
$P < 1E - 4$ (%)	40	50	62
MEO	[0.0319, 0.0032, 0.0003]	[0.0240, 0.0024, 2.9E−4]	[0.0263, 0.0026, 3.0E−4]
SJ10	[1.1519, 0.1168, 0.0119]	[1.5407, 0.1556, 0.0154]	[1.8384, 0.1823, 0.0182]
FB50k	[0.0369, 0.0066, 0.0019]	[0.0283, 0.0050, 0.0013]	[0.0287, 0.0043, 0.0011]
GZ500	[3.1E−4, 3.1E−5, 3.1E−6]	[2.7E−4, 2.7E−5, 2.5E−6]	[2.7E−4, 2.7E−5, 2.7E−6]
GHK10k	[3.1E−4, 3.1E−5, 3.2E−6]	[2.7E−4, 2.7E−5, 2.7E−6]	[2.7E−4, 2.7E−5, 3.1E−6]

this implies GZ500 will be 110, 175 and 234 times slower than MEO while GHK10k will be 315, 483 and 657 times slower.

4. Conclusions

In this article we have compared two analytical approximations and three Monte Carlo methods for computing the probit choice probabilities in terms of accuracy and computational efficiency. The first of these is the 'SJ' approach, implemented using the algorithm proposed in Joe (1995). The other is the 'ME' approach, using the algorithm described in Kamakura (1989), though significantly improved with the optimal ordering of conditional terms. While it is not possible to be exhaustive, the results presented here give what we believe to be the most thorough and in-depth comparison of these analytical approximation methods to date in the context of multinomial probit choice. The results should be of interest to both network modellers and choice modellers, not least given the renewed interest in probit among the latter, where recent CML developments have adopted the SJ approach.

We investigated the ordering of terms in both SJ and ME methods. For ME there appears to be an optimal ordering of conditional terms (as described) that significantly improves accuracy, and this makes redundant the process of averaging across multiple random orderings. For the case of SJ we did not find an optimal ordering of terms that substantially increased the accuracy of this method. Averaging across random orderings was found to be the most effective way to attain the maximum accuracy available with the SJ approach, with the most important increase in accuracy coming from the inclusion of just 10–20 re-orderings.

The optimally ordered ME approach substantially outperforms the SJ approach across a wide range of settings. MEO is substantially more accurate and faster than using a standard FB Monte Carlo (FB50k) approach. More sophisticated Monte Carlo approaches can offer equal or greater accuracy than MEO, but at higher computational cost, especially as the problem dimension increases.

Especially in model estimation, chosen alternatives with very small probabilities can have a disproportionate impact, making accurate calculation of such probabilities essential. The results from Section 3.4 show the spread of errors across the computed choice probabilities. We note that MEO is more accurate in computing small probabilities than SJ or MC, and is comparable with GZ and GHK; doing worse for $N = 5$ and 7 and as well or better for $N = 9$ and 15.

It is important to acknowledge that there are many parameter dimensions to test in assessing whether one approximation method is more accurate than another. The tests reported here are an illustrative selection of all tests performed in this research. Each analytic approximation method suffers from the disadvantage that the analyst cannot demand arbitrary accuracy, whereas in principle this is possible with FB, GZ and GHK. Moreover, we cannot offer bounds for the errors of the analytical approximation methods. Finally, for problems with a very small number of alternatives, exact evaluation of choice probabilities is computationally feasible using established methods of numerical integration. For higher dimensional problems, MEO is an accurate and efficient method for computing choice probabilities.

    With the strength of the results presented here, it seems that the wider use of the 'ME' method in probit work and implementation within a CML estimator are thus promising areas for new developments. While the primary motivation for this article is to help with the estimation and application of discrete choice models, there may also be applications in other areas where the computation of MVN integrals is needed.

Notes

1. The CML approach works by replacing the complex multivariate integrals inherent to MVN likelihood functions by a combination of separate univariate or bivariate integrals. In the work by Bhat, this is applied to the case of probit structures, given the inherent suitability of normal distributions to a CML context.
2. Note that the ME algorithm in Kamakura (1989) has a slight error in equation 15a.
3. Uses QSIMVNV available from http://www.math.wsu.edu/faculty/genz/software/software.html

References

Aitken, A.C., 1935. Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, 4, 106–110.

Bolduc, D., 1999. A practical technique to estimate multinomial probit models in transportation. *Transportation Research Part B: Methodological*, 33, 63–79.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research B*, 35, 677–693.

Bhat, C.R., 2010. *The maximum approximated composite marginal likelihood (MACML) estimation of mixed unordered response choice models*. Technical Paper, Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, Austin, Texas.

Bhat, C.R. and Sidharthan, R., 2011. Maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B*, 45 (7), 940–953.

Bhat, C.R., Varin, C., and Ferdous, N., 2010. A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered Response Model. *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, 26, 65–106.

Börsch-Supan, A. and Hajivassiliou, V., 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics*, 58, 347–368.

Bunch, D.S., 1991. Estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, 25, 1–12.

Clark, C.E., 1961. The greatest of a finite set of random variables. *Operations Research*, 9, 145–162.

Connors, R.D., Sumalee, A., and Watling, D.P., 2007. Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes. *Transportation Research Part B: Methodological*, 41, 593–615.

Cranley, R. and Patterson, T.N.L., 1976. Randomization of number theoretic methods for multiple integration. *SIAM Journal on Numerical Analysis*, 13, 904–914.

Daganzo, C., 1979. *Multinomial probit: the theory and its application to demand forecasting*. New York, NY: Academic Press.

Dalal, S.R. and Klein, R.W., 1988. A flexible class of discrete choice models. *Marketing Science*, 7, 232–251.

Dansie, B.R., 1985. Parameter estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, 19, 526–528.

Drezner, Z., 1990. Approximations to the multivariate normal integral. *Communications in Statistics – Simulation and Computation*, 19, 527–534.

Drezner, Z., 1994. Computation of the trivariate normal integral. *Mathematics of Computation*, 62, 289–294.

Drezner, Z. and Wesolowsky, G.O., 1990. On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation*, 35, 101–107.

Genz, A., 1992. Numerical computation of multivariate normal probabilities. *Journal of computational and graphical statistics*, 1, 141–149.

Genz, A., 2004. Numerical computation of rectangular bivariate and trivariate normal and t-probabilities. *Statistics and Computing*, 14, 251–260.

Genz, A. and Bretz, F., 1999. Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation*, 63, 103–117.

Genz, A. and Bretz, F., 2002. Comparison of methods for the computation of multivariate t-probabilities. *Journal of Computational and Graphical Statistics*, 11, 950–971.

Hess, S., Train, K.E., and Polak, J.W., 2006. On the use of a modified latin hypercube sampling (MLHS) approach in the estimation of a mixed logit model for vehicle choice. *Transportation Research Part B*, 40 (2), 147–163.

Joe, H., 1995. Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90, 957–964.

Kamakura, W.A., 1989. The estimation of multinomial probit models: a new calibration algorithm. *Transportation Science*, 23, 253–265.

Keane, M.P., 1992. A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics*, 10, 193–200.

Langdon, M.G., 1984a. Improved algorithms for estimating choice probabilities in the multinomial probit model. *Transportation Science*, 18, 267–299.

Langdon, M.G., 1984b. Methods of determining choice probability in utility maximising multiple alternative models. *Transportation Research Part B: Methodological*, 18, 209–234.

Lee, L., 1992. On the efficiency of methods of simulated moments and simulated likelihood estimation of discrete choice models. *Econometric Theory*, 8, 518–552.

Lee, L., 1995. Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 11, 437–483.

Manski, C. and Lerman, S., 1981. On the use of simulated frequencies to approximate choice probabilities. *In*: C.F. Manski and D. McFadden, eds. *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press, 305–319.

McFadden, D., 1978a. Modeling the choice of residential location. *In*: A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, eds. *Spatial interaction theory and planning models*. Amsterdam: North-Holland, 75–96.

McFadden, D., 1978b. Quantitative methods for analysing travel behaviour of individuals: some recent developments., *In*: D.A. Hensher and P.R. Stopher, eds. Behavioral travel modeling. Croom Helm, London, pp. 279–318.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 57, 995–1026.

McFadden, D. and Train, K.E., 2000. Mixed MNL models of discrete response. *Journal of Applied Econometrics*, 15, 447–470.

Mendell, N. and Elston, R., 1974. Multifactorial qualitative traits – genetic-analysis and prediction of recurrence risks. *Biometrics*, 30, 41–57.

Niederreiter, H., 1972. On a number-theoretical integration method. *Aequationes Mathematicae*, 8, 304–311.

Pearson, K., 1903. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society London*, A200, 1–66.

Rifaat, S. M. and Chin, H. C., 2007. Accident severity analysis using ordered probit model. *Journal of Advanced Transportation*, 41 (1), 91–114.

Rosa, A., 2003. Probit based methods in traffic assignment and discrete choice modelling, (unpublished PhD Thesis), Napier University.

Solow, A.R., 1990. A method for approximating multivariate normal orthant probabilities. *Journal of Statistical Computation and Simulation*, 37, 225–229.

Sumalee, A., Uchida, K. and Lam, WHK. 2009. Stochastic multi-modal transport network under demand uncertainties and adverse weather condition. *In: 18th International symposium on transportation and traffic theory*, 16–18 July 2009, The Hong Kong Polytechnic University, Hong Kong, 19–41.

Teng, H., Qi, Y., and Martinelli, D.R., 2008. Parking difficulty and parking information system technologies and costs. *Journal of Advanced Transportation*, 42 (2), 151–178.

Thurstone, L., 1927. A law of comparative judgment. *Psychological Review*, 34, 273–286.

Train, K. 2000, *Halton sequences for mixed logit*. Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

Train, K.E., 2009. *Discrete choice methods with simulation*. 2nd ed. Cambridge, MA: Cambridge University Press.

Zito, P., *et al*., 2010. The effect of advanced traveller information systems on public transport demand and its uncertainty. *Transportmetrica*, 7 (1), 31–43.