

The Performance of German Firms in the Business-Related Service Sectors Revisited: Differential Evolution Markov Chain Estimation of the Multinomial Probit Model

W. Erno Kuiper · Anton J. Cozijnsen

Accepted: 21 January 2011 / Published online: 12 February 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We outline a new estimation method for the multinomial probit model (MNP). The method is a differential evolution Markov chain algorithm that employs a Metropolis-within-Gibbs sampler with data augmentation and the Geweke–Hajivassiliou–Keane (GHK) probability simulator. The method lifts the curse of dimensionality in numerical integration as it neither requires simulation of the whole likelihood function nor the computation of its analytical or numerical derivatives. The method is applied to an unbalanced panel dataset of firms from the German business-related service sector over the period 1994–2000. In spite of its less restricted character, the MNP model is found not to provide more accurate estimates for explaining the performance of these firms than the multinomial logit model.

Keywords Differential evolution · GHK · MCMC · MNP · Service sector

JEL Classification C13 · C15 · C25 · L25

1 Introduction

Van Nguyen et al. (2004) used (extensions of) the multinomial logit (MNL) model to analyze the performance of firms from the German business-related service sector during the 1994–2000 period. In this article we re-analyze their data using a multinomial probit (MNP) model and a new estimation method based on the differential evolution (DE) algorithm. The popular MNL model (see McFadden 1974) has several restrictions when compared with the more general MNP model. The most famous of these are that unobservable components of utilities should be mutually independent and

W. E. Kuiper (✉)
Wageningen University, Wageningen, The Netherlands
e-mail: Erno.Kuiper@wur.nl

A. J. Cozijnsen
VU University Amsterdam, Amsterdam, The Netherlands

homoskedastic, which is also known as the independence of irrelevant alternatives (IIA) assumption; see, for example, [McFadden \(1984\)](#) and [Cramer \(1991\)](#) for discussion. Nevertheless, [Train \(2009, pp. 35, 36\)](#) argues that these restrictions can be interpreted as a natural outcome of a well-specified model. However, such a well-specified model could also be seen as rather ideal and hence, one may usually wish that the unobserved portion of utility is allowed to be correlated and heteroskedastic over alternatives. And indeed, although the dataset used by [Van Nguyen et al. \(2004\)](#) is quite an extensive one—the sample contains various indicators of 3,750 firms and the data span the time period from the second quarter of 1994 to the end of 2000 (i.e., 27 quarters) constituting a total of 16,064 observations per variable, it still lacks many factors that might also influence firm performance in addition to the variables available: firm age; firm size; legal status; the number of banking relationships; the degree of diversification; the business-related service sector to which the firm belongs; and region, year and quarter dummies. Moreover, although [Van Nguyen et al. \(2004\)](#) could not reject the IIA property of their simple MNL model using the Hausman test statistic, they had serious doubts about the power of the Hausman test as they also applied some LR specification tests that clearly showed the presence of IIA invalidating heterogeneity. Consequently, using the MNP model to allow for dependence and heteroskedasticity of the unexplained components may provide an interesting extension to the MNL analyses in [Van Nguyen et al. \(2004\)](#).

In contrast to the MNL model, estimation of the parameters in the MNP model is not straightforward as it requires the inclusion of simulators as a consequence of the underlying multiple integrals in the choice probabilities. Fortunately, simulation has proven to be very general and useful for approximating probit probabilities ([Train 2009, p. 115](#)). In this paper, the maximum simulated likelihood (MSL) method, that is, the simulated counterpart of the maximum likelihood (ML) method, incorporating the so-called Geweke–Hajivassiliou–Keane (GHK) simulator (as developed in [Geweke 1989, 1991](#); [Geweke et al. 1994](#); [Börsch-Supan and Hajivassiliou 1993](#); [Hajivassiliou and McFadden 1998](#); [Keane 1990, 1994](#); see also [Train 2009, pp. 122–133](#)), is employed since its use seems to be advantageous compared with the use of other combinations of classical estimation methods and simulators. Moreover, [Bolduc \(1999\)](#) outlined MSL based on the GHK probability simulator while deriving analytical instead of using numerical derivatives. With analytical derivatives computation time considerably reduces and, in general, approaches with analytical derivatives are more reliable than those based on numerical derivatives.

However, derivatives are not needed at all when using a Markov Chain Monte Carlo (MCMC) version of the genetic algorithm Differential Evolution (DE) as proposed by [Ter Braak \(2006\)](#). An additional advantage of this Differential Evolution-Markov Chain (DE-MC) estimation procedure is that the DE part of it, first proposed by [Storn and Price \(1997\)](#), provides a simple genetic algorithm for global numerical optimization in real parameter spaces. Such an algorithm robust to local roots is clearly asked for as local maxima are a common problem among many MNP estimators that use simulation techniques ([Ben-Akiva et al. 1997, p. 284](#)). Furthermore, the combination of DE and MCMC solves an important problem in MCMC in real parameter spaces, namely that of choosing an appropriate scale and orientation of the jumping distribution. Finally, by using a DE-MC-within-Gibbs sampler we can apply data aug-

mentation (Tanner and Wong 1987) with the help of the GHK simulator to estimate the MNP model in principle without having to numerically integrate the multiple integrals in the likelihood function.

The article is organized as follows. Section 2 specifies the model of firm performance from economic and econometric perspectives subject to the data analyzed in Van Nguyen et al. (2004). Section 3 outlines the estimation methods used. Section 4 discusses the empirical results and evaluates the performance of the models and estimators. Section 5 concludes.

2 Model, Data and Variables

Firm performance is represented by the change in profit between subsequent quarters measured on a three-point ordinal scale (increase, decrease, and no change in profit). As many firms are included in the survey during at least two subsequent quarters (waves), the data allow for the estimation of transition probabilities according to a Markov chain of order 1, that is

$$P_{itj'j} = \text{Prob}(S_{it} = j' | S_{i,t-1} = j) \tag{1}$$

where S_{it} represents the state of firm i 's profit change between quarter $t - 1$ and quarter t which takes on three possible values ($j, j' = 1, \dots, J$ with $J = 3$): “1” (or “unchanged”, that is, profit at t remains unchanged compared with quarter $t - 1$); “2” (or “up”, that is, increased profit); and “3” (or “down”, that is, decreased profit). Taking into account the three possible values of the departure state $S_{i,t-1}$ as well, there are $3 \times 3 = 9$ transition possibilities. However, since $P_{itj1} + P_{itj2} + P_{itj3} = 1$, six transition probabilities are left to be estimated.

To specify the transition probabilities, we indicate firm i at time t and departure state $j = 1$ (note that a firm i at time t belongs to only one of the three departure states) simply as “case $it1$ ” and assume that this case assigns the highest utility $U_{it1j'}$ to that arrival state j' that is most likely to occur. The utilities of the arrival states j' depend on the characteristics of case $it1$. To model this relationship, let us consider a simple linear model as follows

$$U_{it1j'} = \mathbf{X}'_{it1} \boldsymbol{\beta}_{1j'} + \varepsilon_{it1j'} \quad \forall j' \tag{2}$$

where \mathbf{X}'_{it1} is a $k \times 1$ vector with a constant and $k - 1$ explanatory variables that vary over the $it1$ cases, $\boldsymbol{\beta}_{1j'}$ is an unknown $k \times 1$ parameter vector that is only allowed to vary over the arrival states j' , and $\varepsilon_{it1j'}$ is an unobserved random variable.

Neither the random terms $\varepsilon_{it1j'}$ nor the perceived utilities $U_{it1j'}$ are observed. In fact, we observe only the actual arrival state. Define the variable $d_{it1s'}$ ($s' = 1, 2, 3$) with $d_{it1s'} = 1$ if arrival state s' occurs regarding case $it1$, and 0 if not. The transition probability is given by

$$\begin{aligned} P_{it1s'} &= \text{Prob}(d_{it1s'} = 1) = \text{Prob}(U_{it1s'} > U_{it1j'} \forall j' \neq s') \\ &= \text{Prob}(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1s'} + \varepsilon_{it1s'} > \mathbf{X}'_{it1} \boldsymbol{\beta}_{1j'} + \varepsilon_{it1j'} \forall j' \neq s') \end{aligned}$$

$$\begin{aligned}
 &= \text{Prob}(\varepsilon_{it1j'} - \varepsilon_{it1s'} < \mathbf{X}'_{it1}(\boldsymbol{\beta}_{1s'} - \boldsymbol{\beta}_{1j'}) \forall j' \neq s') \\
 &= \text{Prob}(\varepsilon_{it1,j's'} < \mathbf{X}'_{it1}\boldsymbol{\beta}_{1,s'j'} \forall j' \neq s') = \int_{-\infty}^{\infty} d_{it1s'} f(\mathbf{e}_{it1,s'}) d\mathbf{e}_{it1,s'} \quad (3)
 \end{aligned}$$

where $\varepsilon_{it1,j's'} = \varepsilon_{it1j'} - \varepsilon_{it1s'}$, $\boldsymbol{\beta}_{1,s'j'} = \boldsymbol{\beta}_{1s'} - \boldsymbol{\beta}_{1j'}$ and $\mathbf{e}_{it1,s'} = (\varepsilon_{it1,1s'}, \varepsilon_{it1,2s'})'$ if $s' = 3$, $\mathbf{e}_{it1,s'} = (\varepsilon_{it1,1s'}, \varepsilon'_{it1,3s'})'$ if $s' = 2$, and $\mathbf{e}_{it1,s'} = (\varepsilon_{it1,2s'}, \varepsilon_{it1,3s'})'$ if $s' = 1$. Hence, $\mathbf{e}_{it1,s'}$ has dimension $J - 1 = 3 - 1 = 2$ and the probability $P_{it1s'}$ depends on the assumptions about the distribution $f(\cdot)$ of $\mathbf{e}_{it1,s'}$.

The MNL model (Nerlove and Press 1973) is obtained by assuming that each $\varepsilon_{it1j'}$ is independently, identically distributed (iid) extreme value, giving

$$P_{it1s'} = \exp(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1s'}) / \sum_{j'=1}^J \exp(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1j'}) \quad \forall s', j' \quad (4)$$

where $\boldsymbol{\beta}_{11} = \mathbf{0}$ to normalize for the level. The MNL model has become very popular in business and applied economic research because expression (4) is in closed form. This, however, comes at the cost of the restriction that for each $it1$ case the $\varepsilon_{it1j'}$ are iid across all arrival states j' which is an assumption that, unfortunately, can easily be invalidated by unobserved components. Therefore, we may wish that the unobserved portion of utility is allowed to be correlated over the arrival states given our specification of representative utility. The MNP model does allow for these correlations and hence, we now turn to the specification of the MNP model.

The MNP model assumes for each case $it1$ that $\mathbf{e}_{it1} = [\varepsilon_{it11}, \dots, \varepsilon_{it1J}]'$ is distributed normal with a mean vector of zero and covariance matrix $\boldsymbol{\Omega}$. The MNP probabilities

$$P_{it1s'} = \int_{-\infty}^{\infty} d_{it1s'} \varphi(\mathbf{e}_{it1,s'}) d\mathbf{e}_{it1,s'} \quad (5)$$

with

$$\varphi(\mathbf{e}_{it1,s'}) = \left\{ 1 / \left((2\pi)^{(J-1)/2} \left| \boldsymbol{\Omega}_{1,s'}^* \right|^{1/2} \right) \right\} \exp \left(- (1/2) \mathbf{e}'_{it1,s'} \boldsymbol{\Omega}_{1,s'}^{*-1} \mathbf{e}_{it1,s'} \right) \quad (6)$$

do not have a closed-form expression and must be approximated numerically. Simulation has proven to be very general and useful for approximating probit probabilities (Train 2009, p. 115). The most accurate and widely used probit simulator is called GHK, after Geweke (1989, 1991), Hajivassiliou (as reported in Hajivassiliou and McFadden 1998) and Keane (1990, 1994), who developed the procedure (see also Train 2009, pp. 122–133, for a very clear presentation of the GHK simulator). In Sect. 3 we will discuss the derivation of the covariance matrix $\boldsymbol{\Omega}_{1,s'}^*$ in (6) for being used in the GHK simulator.

Thus far, we concentrated on the model to estimate the transition probabilities for the $it1$ cases, that is, those firms i at time t that face departure state 1. Two additional, but similar models will be considered for the two other departure states and hence, for the $it2$ and $it3$ cases. Notice again that a firm i at time t belongs to only one of the three departure states. The remainder of this section will be devoted to the discussion of the explanatory variables being included in the probability models. For purpose of comparison, we shall include the same variables as Van Nguyen et al. (2004) did in their “simple” MNL profit change models of which the estimation results are reported in their Tables 4 and 5. Therefore, our discussion below will be a short summary of their data description.

The empirical data concern firms from the German business-related services sectors in the second half of the 1990s. In contrast to the total German economy these firms were doing quite well at that time. To decompose their performance into firm and time characteristics, the following variables are considered: *Sector*—dummy variables corresponding to the 10 business-related service sectors: ACC (accounting); ADV (advertising); ARC (architecture); CAR (vehicle renting); CNS (management consulting); MCH (machine renting); PLN (technical planning = reference); SFT (software), TRN (transport); and WST (waste disposal); *Age*—examining the distribution of the year of foundation by sector in such a way that firms in the first quartile are considered old and in the fourth quartile young resulting in three dummy variables to represent firm age: OLD (before about 1981); YNG (after 1990); and MDL (1981–1990 = reference); *Size*—firm size is measured by the number of employees (mean = 86; maximum = 8,945; minimum = 1, and the 30 observations with a 0 are discarded) and is included in the model by its natural logarithm LSIZE; *Legal status*—two dummy variables: SINGPART (firms with single proprietorship or partnerships) and GmbH (limited liability companies = reference); *Banking relationship*—the dummy variable BANK takes the value 0 if the firm has only one creditor and 1 if it has at least two creditors; *Diversification*—the dummy variable DIVERS takes the value 1 if the firm is in at least two business fields and 0 otherwise; *Region*—EAST is a dummy variable that is coded 1 if a firm is located in East Germany and 0 if it is in West Germany; *Time*—to capture business cycle and seasonal effects, the dummy variables Y94–Y99 correspond to the years 1994–1999 (Y95 = reference, and the four quarters of 2000 are for out-of-sample forecast evaluation) and Q1–Q4 are quarter dummies (Q1 = reference). Cf. Appendix A in Van Nguyen et al. (2004).

The so-called “log-odds ratios” $\ln(P_{it1s'}/P_{it11}) = \mathbf{X}'_{it1} \boldsymbol{\beta}_{1s'}$ ($s' = 2, \dots, J$) can be derived from (4) to assess the effects of the firm and time characteristics on the transition probabilities. The MNL analysis in Van Nguyen et al. (2004) shows that reaching arrival state “up” is more likely for firms that are young, relatively larger, located in West Germany, have at least two creditors, belong to the CNS or SFT sectors, but not to the ARC or WST sectors, and are facing the economic situation in 1994 or 1998 and quarters 2–4. In contrast, the danger of reaching arrival state “down” is higher for firms that are not young, located in East Germany, active in at least two business fields, have single proprietorship or are partnerships, belong to the ADV, ARC, CNS or MCH sectors, but not to the ACC, CAR, SFT or TRN sectors, and are confronted with the economic situation of 1996, 1997 or 1999 and the first quarter. The effect of firm size depends on the starting state; getting into arrival state “down”

will be less (more) likely for larger firms if the starting state is “unchanged” (“up”). Consequently, given the earlier result that larger firms are also more likely to arrive in state “up” when the starting state is “up”, it can be concluded that larger firms are less likely to face unchanged profits after they experienced increasing profits in the previous quarter.

In this paper we first perform standard ML estimation on the MNL models to reproduce the estimates in [Van Nguyen et al. \(2004\)](#). Next, we estimate the MNL parameters and their 95% confidence levels by DE-MC to assess whether the DE-MC estimator is relatively efficient. Finally, based on the GHK simulated likelihood, the parameters of the MNP model and their 95% confidence levels are estimated by standard ML and DE-MC-within-Gibbs. The analysis has to show to what extent the MNP model can improve on the predictive validity of the traditional MNL model.

3 Parameter Estimation

3.1 Standard MLE for the MNL Model

Estimation of the parameters $\beta_{1j'}$ in the MNL model is straightforward (e.g. [Green 2008](#), pp. 843–847). The log-likelihood LL_1 is given by the sum of the natural logs of the probabilities in Eq. 4 for all $it1$ cases $n_1 = 1, \dots, N_1$ as

$$LL_1 = \sum_{n_1=1}^{N_1} \sum_{j'=1}^J d_{n_1j'} \ln(P_{n_1j'}) \tag{7}$$

from which, given the normalization $\beta_{11} = \mathbf{0}$, we obtain the first derivatives vector

$$\partial LL_1 / \partial \beta_{1l} = \sum_{n_1=1}^{N_1} (d_{n_1l} - P_{n_1l}) \mathbf{X}_{n_1} \quad \forall l = 2, \dots, J \tag{8}$$

and second derivatives matrix

$$\partial^2 LL_1 / \partial \beta_{1l} \partial \beta'_{1m} = - \sum_{n_1=1}^{N_1} P_{n_1l} (1(l = m) - P_{n_1m}) \mathbf{X}_{n_1} \mathbf{X}'_{n_1} \quad \forall l, m = 2, \dots, J \tag{9}$$

where $1(l = m)$ equals 1 if l equals m and 0 if not. The estimates of $\beta_1 = [\beta_{12'}, \dots, \beta_{1J'}]'$ can be calculated by the Newton-Raphson (NR) procedure (e.g. [Train 2009](#): ch. 8) as

$$\beta_{1\{g+1\}} = \beta_{1\{g\}} + \lambda \left[-\mathbf{H}_{1\{g\}}^{-1} \right] \mathbf{q}_{1\{g\}} \quad g = 0, \dots, G \tag{10}$$

where g indicates the iteration, G is the iteration at which convergence is achieved, \mathbf{H}_1 consists of the $(J - 1)^2 k \times k$ blocks as given in (9), \mathbf{q}_1 is $(J - 1)k \times 1$

consisting of the vectors in (8) and $\mathbf{H}_{1\{g\}}$ and $\mathbf{q}_{1\{g\}}$ are evaluated at $\beta_{1\{g\}}$. At each iteration a line search is performed along $\beta_{1\{g\}} + \lambda[-\mathbf{H}_{1\{g\}}^{-1}]\mathbf{q}_{1\{g\}}$, where the stepsize $\lambda \in \{2^{-40}, \dots, 2^{-2}, 2^{-1}, 1, 2, 4, \dots, 2^{40}\}$ starts with $\lambda = 1/2^{40}$. If $LL_1(\beta_{1\{g\}} + 2^{-40+i}[-\mathbf{H}_{1\{g\}}^{-1}]\mathbf{q}_{1\{g\}}) > LL_1(\beta_{1\{g\}} + 2^{-40+i-1}[-\mathbf{H}_{1\{g\}}^{-1}]\mathbf{q}_{1\{g\}})$ for $i = 1$, then i is increased by 1 as long as LL_1 continues to rise. As soon as $LL_1(\beta_{1\{g\}} + 2^{-40+i}[-\mathbf{H}_{1\{g\}}^{-1}]\mathbf{q}_{1\{g\}}) < LL_1(\beta_{1\{g\}} + 2^{-40+i-1}[-\mathbf{H}_{1\{g\}}^{-1}]\mathbf{q}_{1\{g\}})$, then λ is set equal to $2^{-40+i-1}$.

A refinement of the NR procedure is provided by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method according to which $-\mathbf{H}_1^{-1}$ is replaced by the updates of \mathbf{W}_1 obtained as (e.g. [Byatt et al. 2004](#)):

$$\mathbf{W}_{1\{g+1\}} = [\mathbf{W} + (1 + \mathbf{y}'\mathbf{W}\mathbf{y}/s'y) \mathbf{s}\mathbf{s}'/s'y - (\mathbf{s}'\mathbf{W} + \mathbf{W}\mathbf{s}')/s'y]_{1\{g\}} \tag{11}$$

where $\mathbf{s}_{1\{g\}} = \beta_{1\{g+1\}} - \beta_{1\{g\}}$, $\mathbf{y}_{1\{g\}} = \mathbf{q}_{1\{g+1\}} - \mathbf{q}_{1\{g\}}$ and the initial $\mathbf{W}_{1\{0\}}$ is set to $-\mathbf{I}$, where \mathbf{I} is the identity matrix. Convergence is achieved if $-\mathbf{q}'_{1\{G-1\}}\mathbf{W}_{1\{G-1\}}\mathbf{q}_{1\{G-1\}} < \delta$, where δ is some small value like 10^{-5} . The asymptotic covariance matrix of the maximum likelihood estimator (MLE) obtained at convergence is

$$\text{var}(\beta_{1\{G\}}) = -\mathbf{H}_{1\{G\}}^{-1} \tag{12}$$

so that division of the estimates of $\beta_{1\{G\}}$ by the asymptotic standard errors as given by the square root of the corresponding diagonal elements of $-\mathbf{H}_{1\{G\}}^{-1}$, yields the t values. The starting values $\beta_{1j'\{0\}}$ for each $j' = 2, \dots, J$, could be given by the means of the coefficients obtained by drawing N_1 -dimensional vectors of probabilities for running a 1000 log-odds ratio regressions $\ln(P_{it1j'}/P_{it11}) = \mathbf{X}'_{it1j'}\beta_{1j'}$. Except for cases of profound misspecification, the MNL likelihood will optimize at its global maximum and is not prone to optimization errors. Therefore, regarding all other issues than convergence to the global instead of a local optimum, we can nicely use the MNL model to compare the estimation performance of a global optimization algorithm with the gradient-based and quasi-Newton estimation algorithms discussed so far.

3.2 DE-MC for the MNL Model

The global maximization algorithm considered in this study is DE which, when combined with MCMC analysis, does not only find the optimal parameter estimates, but also their uncertainty distribution. In what follows we explain DE, its combination with MCMC and how the resulting DE-MC algorithm can be used to estimate the MNL parameters.

DE is a simple genetic algorithm for numerical global optimization in real parameter spaces. It was developed by [Storn and Price \(1997\)](#) to optimize real parameter, real valued functions for which the general problem formulation is: For an objective function $f : X \subseteq \mathbb{R}^D \rightarrow \mathbb{R}$ where the feasible region $X \neq \emptyset$, the maximization problem is to find $\mathbf{x}^* \in X$ such that $f(\mathbf{x}^*) \geq f(\mathbf{x}) \forall \mathbf{x} \in X$, where $f(\mathbf{x}^*) \neq \infty$. At each iteration (generation) g ($g = 0, 1, \dots, G$) DE applies a mutation, crossover and selection

operation on a number of NP parameter vectors of dimension D . Let these parameter vectors, which are called individuals, be denoted by $\mathbf{x}_{i\{g\}} = [x_{i1\{g\}}, \dots, x_{iD\{g\}}]'$ ($i = 1, \dots, NP$). Defining upper and lower bounds for each parameter of the initial population (that is, $g = 0$) : $x_j^L \leq x_{ij\{0\}} \leq x_j^U$ ($j = 1, \dots, D$), the initial parameter values are randomly selected uniformly on the intervals $[x_j^L, x_j^U]$ as follows:

$$x_{ij\{0\}} = rand_{ij\{0\}}[0, 1](x_j^U - x_j^L) + x_j^L \tag{13}$$

where $rand_{ij\{g\}}[0,1]$ denotes a uniformly distributed random value within range $[0, 1]$ that is chosen anew for each i, j and g .

For each vector $\mathbf{x}_{i\{g\}}$ the mutation operator determines a mutant vector $\mathbf{v}_{i\{g\}} = [v_{i1\{g\}}, \dots, v_{iD\{g\}}]'$ as follows

$$\mathbf{v}_{i\{g\}} = \mathbf{x}_{i'\{g\}} + F (\mathbf{x}_{i''\{g\}} - \mathbf{x}_{i'''\{g\}}) \tag{14}$$

where $i', i'', i''' \in \{1, 2, \dots, NP\}$ are mutually different randomly chosen indices that also need to differ from the current index i and hence, the number of individuals in a population must be at least four. New, random integer values for i', i'' and i''' are chosen for each value of the index i , that is, for each individual. Mutation scale factor F is a real constant $\in (0, 2]$ that controls the amplification of the difference vector $(\mathbf{x}_{i''\{g\}} - \mathbf{x}_{i'''\{g\}})$. If a parameter $v_{ij\{g\}}$ falls outside the interval $[x_j^L, x_j^U]$, a new value can be obtained as in (13). According to [Storn and Price \(1995\)](#) DE also works if the mutant vector $\mathbf{x}_{i'\{g\}}$ in (14) is replaced by the individual $\mathbf{x}_{i\{g\}}$ itself.

After mutation, a “binary” crossover operation forms the final trial vector out of the i th individual and its corresponding mutant vector according to

$$u_{ij\{g\}} = \begin{cases} v_{ij\{g\}} & \text{if } rand_{ij\{g\}}[0, 1] \leq CR \vee j = h \\ x_{ij\{g\}} & \text{otherwise} \end{cases} \tag{15}$$

where the constant CR is a real-valued crossover factor in the range $[0, 1]$ that controls the probability that the trial vector parameter $u_{ij\{g\}}$ will be equal to the mutant vector parameter $v_{ij\{g\}}$ instead of the population vector parameter $x_{ij\{g\}}$. Additionally, to ensure at least some crossover, one element of $\mathbf{u}_{i\{g\}}$, in (15) denoted as element h , is selected at random to be from $v_{i\{g\}}$, that is, $u_{ih\{g\}} = v_{ih\{g\}}$.

Finally, the selection operator decides whether the trial vector $\mathbf{u}_{i\{g\}}$ will be a member of the population of the next generation as follows

$$\mathbf{x}_{i\{g+1\}} = \begin{cases} \mathbf{u}_{i\{g\}} & \text{if } f(\mathbf{u}_{i\{g\}}) \geq f(\mathbf{x}_{i\{g\}}) \\ \mathbf{x}_{i\{g\}} & \text{otherwise} \end{cases} \tag{16}$$

showing that the vector with the higher objective function value obtains a place in the next generation’s population. Consequently, all the individuals of the next generation are as good or better than their counterparts in the current generation. The procedure terminates as soon as all individuals yield the same objective function value. Usually, however, a maximum number of generations is set. Notice that the control parameters

NP , F and CR remain constant during the search process. According to the literature reasonable values are: $NP = 10D$, $F = 0.8$ and $CR=0.5$. We refer to [Price et al. \(2005\)](#) for an extensive discussion of DE.

To turn DE into a population MCMC algorithm to obtain DE-MC, [Ter Braak \(2006\)](#) proposes the following adjustments: First, the proposal scheme in (14) is replaced by

$$\mathbf{v}_{1i\{g\}} = \beta_{1i\{g\}} + F (\beta_{1i''\{g\}} - \beta_{1i'''\{g\}}) + \mathbf{e}_{1i\{g\}} \tag{14'}$$

where the β_1 is the MNL model parameter vector and the first vector on the right-hand-side has now become the population vector $\beta_{1i\{g\}}$ itself instead of the mutant vector $\beta_{1i'\{g\}}$, $\mathbf{v}_{1i\{g\}}$ is the proposal vector, $\mathbf{e}_{1i\{g\}} \sim rand_{i\{g\}}[-10^{-4}, 10^4]^D$ and $F = 2.38/\sqrt{2D}$. Second, the selection operator is changed into the standard Metropolis acceptance rule

$$\beta_{1i\{g+1\}} = \begin{cases} \mathbf{v}_{1i\{g\}} & \text{if } \{LL_1(\mathbf{v}_{1i\{g\}}) - LL_1(\beta_{1i\{g\}})\} > \ln(rand_{i\{g\}} [0, 1]) \\ \beta_{1i\{g\}} & \text{otherwise} \end{cases} \tag{16'}$$

where $\mathbf{v}_{1i\{g\}}$ comes from (14') as the crossover operation is omitted. Third, at each iteration we randomly select one of the parameter vectors β_{1i} from the set of parameter vectors of which each parameter vector yields an LL_1 value that is higher than the highest LL_1 value of the five percent of the total population of parameter vectors with the lowest LL_1 values. As soon as each of the selected parameter vectors follow a Gaussian white noise process over the iterations, then the parameter values can be considered as independent drawings from a D -dimensional normal distribution. Consequently, for a sufficiently large number of iterations the sample mean of each parameter is the DE-MC parameter estimate and the 2.5 and 97.5 percentiles of the empirical distribution can be used to assess the significance of the parameter estimate. Finally, the LL_1 is computed for the DE-MC parameter estimates.

3.3 GHK Simulator for MNP Estimation by BFGS and DE-MC

Next, we turn to the estimation of the MNP model parameters. To outline the GHK simulator, we first discuss the derivation of the covariance matrix $\Omega_{1,s'}^*$ in (6), which is used for calculating the GHK simulations. Then, after describing the GHK simulator, we discuss MSL estimation by both BFGS and DE-MC-within-Gibbs.

To derive $\Omega_{1,s'}^*$, we consider our three-alternative model ($J = 3$), although generalization to more alternatives is straightforward (see, for example, [Train 2009](#): ch. 5). Recall that for each case $it1$ the vector of errors $\boldsymbol{\varepsilon}_{it1} = [\varepsilon_{it11}, \dots, \varepsilon_{it1J}]'$ is normally distributed with zero mean and covariance matrix

$$\Omega_1 = \begin{bmatrix} \sigma_{1,11} & \sigma_{1,12} & \sigma_{1,13} \\ \bullet & \sigma_{1,22} & \sigma_{1,23} \\ \bullet & \bullet & \sigma_{1,33} \end{bmatrix} \tag{18}$$

To take account of the fact that the level of utility is irrelevant, we take utility differences, cf. (3). Suppose that case $it1$ arrives in state 1, then the vector of error

differences to be considered is $\boldsymbol{\varepsilon}_{it1,1} = [\varepsilon_{it1,21}, \varepsilon_{it1,31}]'$. The covariance matrix of this vector takes the form

$$\boldsymbol{\Omega}_{1,1} = \mathbf{M}_1 \boldsymbol{\Omega}_1 \mathbf{M}'_1 = \begin{bmatrix} \theta_{1,122} & \theta_{1,123} \\ \bullet & \theta_{1,133} \end{bmatrix} \tag{18}$$

where \mathbf{M}_1 is a $J - 1 (= 3 - 1 = 2)$ identity matrix with an extra column of -1's added as the first column. The extra column makes the matrix have size $J - 1$ by J (which, in our case, is 2×3):

$$\mathbf{M}_1 = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{19}$$

We can normalize the scale of utility by setting $\theta_{1,122} = 1$, giving the following covariance matrix

$$\boldsymbol{\Omega}^*_{1,1} = \begin{bmatrix} 1 & \theta^*_{1,123} \\ \bullet & \theta^*_{1,133} \end{bmatrix} \tag{20}$$

There are two elements in $\boldsymbol{\Omega}^*_{1,1}$. Together with the $(J - 1)k = 2k$ parameters in the two vectors $\boldsymbol{\beta}_{1,21}$ and $\boldsymbol{\beta}_{1,31}$ these are the only identified parameters in the model since all the other vectors can be derived as follows: $\boldsymbol{\beta}_{1,12} = -\boldsymbol{\beta}_{1,21}$; $\boldsymbol{\beta}_{1,32} = \boldsymbol{\beta}_{1,31} - \boldsymbol{\beta}_{1,21}$; $\boldsymbol{\beta}_{1,13} = -\boldsymbol{\beta}_{1,31}$; and $\boldsymbol{\beta}_{1,23} = \boldsymbol{\beta}_{1,21} - \boldsymbol{\beta}_{1,31}$.

To make the model more convenient for simulation, we calculate the Cholesky factor of $\boldsymbol{\Omega}^*_{1,1}$, which is the lower-triangular matrix $\mathbf{L}^*_{1,1}$

$$\mathbf{L}^*_{1,1} = \begin{bmatrix} 1 & 0 \\ c_{1,121} & c_{1,122} \end{bmatrix} \tag{21}$$

such that $\mathbf{L}^*_{1,1} \mathbf{L}'_{1,1} = \boldsymbol{\Omega}^*_{1,1}$. Using this Choleski factor, the error differences can be rewritten as linear functions of *uncorrelated* standard normal deviates:

$$\varepsilon_{it1,21} = \eta_1 \tag{22a}$$

$$\varepsilon_{it1,31} = c_{1,121} \eta_1 + c_{1,122} \eta_2 \tag{22b}$$

where η_1 and η_2 are iid $N(0, 1)$ of which, for sake of simplicity, we omit the subscript "it1". Using (3), (22a) and (22b) we obtain

$$\begin{aligned} P_{it11} &= \text{Prob}(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} + \eta_1 < 0 \text{ and } \mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_1 + c_{1,122} \eta_2 < 0) \\ &= \text{Prob}(\eta_1 < -\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} \text{ and } \eta_2 < -(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_1) / c_{1,122}) \\ &= \text{Prob}(\eta_1 < -\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21}) \\ &\quad \times \text{Prob}(\eta_2 < -(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_1) / c_{1,122} | \eta_1 < -\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21}) \end{aligned} \tag{23}$$

Now the GHK simulator is computed as follows (see also Train 2009, pp. 122–133):

1. Calculate $\text{Prob}(\eta_1 < -\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}) = \Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})$, where $\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})$ is the standard normal cumulative distribution evaluated at $-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}$.
2. Draw a value of η_1 , labeled $\eta_{1\{r\}}$, from a truncated standard normal truncated at $-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}$. This draw is obtained by drawing a standard uniform $\mu_{1\{r\}}$ and calculating $\eta_{1\{r\}} = \Phi^{-1}(\mu_{1\{r\}}\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}))$.
3. Calculate $\text{Prob}(\eta_2 < -(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_1)/c_{1,122} | \eta_1 = \eta_{1\{r\}}) = \Phi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122})$.
4. Calculate $\tilde{P}_{it11\{r\}} = \Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\Phi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122})$.
5. Repeat steps 1–4 many times, for $r = 1, \dots, R$.
6. The simulated probability is $\tilde{P}_{it11} = \sum_{r=1}^R \tilde{P}_{it11\{r\}}/R$.

While, for sake of simplicity, omitting the last subscript “ j ” ($j = 1, 2, 3$) of \tilde{P}_{it1j} , which is the subscript that indicates the observed arrival state of case $it1$, and noticing that “ n_1 ” and “ $it1$ ” are two different notations for the same case, the simulated log-likelihood (SLL_1) function for departure state 1 takes the form

$$SLL_1 = \sum_{n_1=1}^{N_1} \ln(\tilde{P}_{n_1}(\boldsymbol{\gamma}_1)) \tag{24}$$

where $\tilde{P}_{n_1}(\boldsymbol{\gamma}_1)$ is the simulated probability of the observed arrival state of case n_1 and $\boldsymbol{\gamma}_1 = (\boldsymbol{\beta}'_{1,21}, \boldsymbol{\beta}'_{1,31}, c_{1,121}, c_{1,122})'$ is a $k(J - 1) + (J - 1)J/2 - 1 = 2k + 2$ dimensional vector of unknown coefficients. If a case arrives at another state than the first one, then the parameters to be used in the simulation procedure are obtained as follows

$$\boldsymbol{\beta}_{1,j'} = \mathbf{A}_{1j'}\boldsymbol{\beta}_{1,1} \tag{25a}$$

$$\boldsymbol{\Omega}^*_{1,j'} = \mathbf{M}_{j'}\mathbf{L}_{1,1}\mathbf{L}'_{1,1}\mathbf{M}'_{j'} \tag{25b}$$

$\forall j' = 2, 3$, where $\boldsymbol{\beta}_{1,1} = (\boldsymbol{\beta}'_{1,21}, \boldsymbol{\beta}'_{1,31})'$, $\mathbf{A}_{12} = \begin{bmatrix} -\mathbf{I}_k & \mathbf{0}_k \\ -\mathbf{I}_k & \mathbf{I}_k \end{bmatrix}$, $\mathbf{A}_{13} = \begin{bmatrix} \mathbf{0}_k & -\mathbf{I}_k \\ \mathbf{I}_k & -\mathbf{I}_k \end{bmatrix}$ and $\mathbf{L}_{1,1}$ is a $J \times J$ matrix containing $\mathbf{L}^*_{1,1}$ (see (21)) as follows

$$\mathbf{L}_{1,1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c_{1,121} & c_{1,122} \end{bmatrix} \tag{26}$$

To reduce the computation time of the GHK simulator, we refer to Schervish (1984) and Genz (1992) for reordering the $J - 1$ variables so that the variables associated with the largest integration intervals are the innermost variables.

Of course, the accuracy of the GHK simulator can always be established by a large number of repetitions R . Hess et al. (2006) advocates the use of a modified Latin hypercube sampling method to reduce the R required. In this paper we will use weighted

systematic sampling to establish coverage at a limited R . Accordingly, $\mu_{1\{r\}}$ takes the values $0.5/R, 1.5/R, \dots, (R-0.5)/R$ and the simulated probability is computed as

$$\begin{aligned} \tilde{P}_{it11} &= \sum_{r=1}^R \Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}) \Phi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122}) \varphi(\eta_{1\{r\}}) / \\ &\times \sum_{r'=1}^R \varphi(\eta_{1\{r'\}}) \end{aligned} \tag{27}$$

where the weights $\varphi(\eta_{1\{r\}})/\sum_{r'=1}^R \varphi(\eta_{1\{r'\}})$ correct for the increasing dispersion among the $\eta_{1\{r\}}$ outcomes in the outer percentiles of φ (in our empirical analysis a grid size of $R = 10$ delivers estimates that are comparable to those of DE-MC-within-Gibbs and we find no higher likelihoods when using larger grid sizes).

We are now going to use the GHK simulator first to describe MSL estimation by the BFGS method and then to outline how to perform MSL estimation by the DE-MC-within-Gibbs procedure. For the BFGS method we could use numerical first derivatives of the SLL , but clearly, computation time will be considerably reduced when analytical expressions are used instead (Bolduc 1999). Considering the weights in (27), that is, $\varphi(\eta_{1\{w\}})/\sum_{r=1}^R \varphi(\eta_{1\{r\}})$ with $w = 1, \dots, R$, as fixed constants and taking into account that $\eta_{1\{r\}} = \Phi^{-1}(\mu_{1\{r\}}\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}))$ can be rewritten as $\Phi(\eta_{1\{r\}}) = \mu_{1\{r\}}\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})$ so that $\partial\eta_{1\{r\}}/\partial\boldsymbol{\beta}_{1,21} = -\mu_{1\{r\}}\varphi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\mathbf{X}_{it1}/\varphi(\eta_{1\{r\}})$, we obtain the following first derivative vector of \tilde{P}_{it11} :

$$\begin{aligned} \partial\tilde{P}_{it11}/\partial\boldsymbol{\beta}_{1,21} &= \sum_{r=1}^R [-\varphi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\mathbf{X}_{it1} \\ &\times \Phi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122})\varphi(\eta_{1\{r\}}) \\ &+ \Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\varphi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122}) \\ &\times (c_{1,121}/c_{1,122})\mu_{1\{r\}}\varphi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\mathbf{X}_{it1}] / \sum_{r'=1}^R \varphi(\eta_{1\{r'\}}) \end{aligned} \tag{28a}$$

$$\begin{aligned} \partial\tilde{P}_{it11}/\partial\boldsymbol{\beta}_{1,31} &= \sum_{r=1}^R [-\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21}) \\ &\times \varphi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122})(\mathbf{X}_{it1}/c_{1,122}) \\ &\times \varphi(\eta_{1\{r\}})] / \sum_{r'=1}^R \varphi(\eta_{1\{r'\}}) \end{aligned} \tag{28b}$$

$$\begin{aligned} \partial\tilde{P}_{it11}/\partial c_{1,121} &= \sum_{r=1}^R \{-\Phi(-\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,21})\varphi(-(\mathbf{X}'_{it1}\boldsymbol{\beta}_{1,31} + c_{1,121}\eta_{1\{r\}})/c_{1,122}) \\ &\times (\eta_{1\{r\}}/c_{1,122})\varphi(\eta_{1\{r\}})\} / \sum_{r'=1}^R \varphi(\eta_{1\{r'\}}) \end{aligned} \tag{28c}$$

$$\begin{aligned} \partial \tilde{P}_{it11} / \partial c_{1,122} &= \sum_{r=1}^R \{ -\Phi(-\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21}) \varphi(-(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_{1\{r\}}) / c_{1,122}) \\ &\quad \times ((\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_{1\{r\}}) / c_{1,122}^2) \varphi(\eta_{1\{r\}}) \} / \sum_{r'=1}^R \varphi(\eta_{1\{r'\}}) \end{aligned} \tag{28d}$$

Instead of using (25b) if a case arrives in state 2, we can also simply pre-multiply the equation system

$$U_{it1,21} = \mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} + \eta_1 \tag{29a}$$

$$U_{it1,31} = \mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} + c_{1,121} \eta_1 + c_{1,122} \eta_2 \tag{29b}$$

by $\mathbf{A}_{12} = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix}$ (cf. (25a)) to obtain

$$U_{it1,12} = -\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} - \eta_1 \tag{30a}$$

$$U_{it1,32} = \mathbf{X}'_{it1} \boldsymbol{\beta}_{1,31} - \boldsymbol{\beta}_{1,21} + (c_{1,121} - 1) \eta_1 + c_{1,122} \eta_2 \tag{30b}$$

from which it follows that

$$\begin{aligned} P_{it12} &= \text{Prob}\{-\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} - \eta_1 < 0 \text{ and } \mathbf{X}'_{it1} (\boldsymbol{\beta}_{1,31} - \boldsymbol{\beta}_{1,21}) \\ &\quad + (c_{1,121} - 1) \eta_1 + c_{1,122} \eta_2 < 0\} \\ &= \text{Prob}(-\eta_1 < \mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21} \text{ and } \eta_2 < \{\mathbf{X}'_{it1} (\boldsymbol{\beta}_{1,21} - \boldsymbol{\beta}_{1,31}) + (1 - c_{1,121}) \eta_1\} / c_{1,122}) \end{aligned} \tag{31}$$

so that

$$\tilde{P}_{it12\{r\}} = \Phi(\mathbf{X}'_{it1} \boldsymbol{\beta}_{1,21}) \Phi(\{\mathbf{X}'_{it1} (\boldsymbol{\beta}_{1,21} - \boldsymbol{\beta}_{1,31}) + (c_{1,121} - 1) \kappa_{1\{r\}}\} / c_{1,122}) \tag{32}$$

where $\kappa_1 = -\eta_1$. Given the illustration for \tilde{P}_{it11} above, it is now straightforward to compute the first derivatives vector of \tilde{P}_{it12} as well. And for arrival state 3 we have that $\tilde{P}_{it13} = 1 - \tilde{P}_{it11} - \tilde{P}_{it12}$. Now MSL estimation by the BFGS method proceeds as described above for the MNL model. At convergence, however, the variance-covariance matrix of the $\boldsymbol{\beta}_{1,1} = (\boldsymbol{\beta}'_{1,21}, \boldsymbol{\beta}'_{1,31})'$ parameters can be computed by the GLS formula

$$\text{var}(\beta_{1,1}) = \left[\begin{array}{ccc} \left[\begin{array}{cc} \mathbf{X}_1 & \mathbf{0}_{k \times 1} \\ \mathbf{0}_{k \times 1} & \mathbf{X}_1 \end{array} \right] & \cdots & \left[\begin{array}{cc} \mathbf{X}_{N_1} & \mathbf{0}_{k \times 1} \\ \mathbf{0}_{k \times 1} & \mathbf{X}_{N_1} \end{array} \right] \\ \vdots & & \vdots \\ \left[\begin{array}{cc} \mathbf{X}'_{N_1} & \mathbf{0}_{1 \times k} \\ \mathbf{0}_{1 \times k} & \mathbf{X}'_{N_1} \end{array} \right] & & \left[\begin{array}{cc} \mathbf{X}'_{N_1} & \mathbf{0}_{1 \times k} \\ \mathbf{0}_{1 \times k} & \mathbf{X}'_{N_1} \end{array} \right] \end{array} \right]^{-1} \quad (33)$$

as follows from the linear specification of the utility model. In our model, however, the SUR does not require this formula as each equation contains the same regressors so that $\text{var}(\beta_{1,21}) = (\mathbf{X}'_1 \mathbf{X}_1)^{-1}$ and $\text{var}(\beta_{1,31}) = (c_{1,121}^2 + c_{1,122}^2)(\mathbf{X}'_1 \mathbf{X}_1)^{-1}$, where $\mathbf{X}_1 = [\mathbf{X}_1, \dots, \mathbf{X}_{N_1}]'$ is an $(N_1 \times k)$ matrix.

For the DE-MC method first derivatives do not have to be derived. In fact, we could compute the probabilities in (27), (32) and $\tilde{P}_{it13} = 1 - \tilde{P}_{it11} - \tilde{P}_{it12}$ to obtain the SLI_1 in (24) to serve as the LI_1 in (16'). Unfortunately, this procedure appears to be already rather time-consuming at a grid size of $R = 10$ in the GHK approximation of the integrals. Using a Gibbs sampling scheme, however, we can circumvent the approximation of the density function integrals by drawing from conditional density functions and doing so with the help of DE-MC within Gibbs when no standard distribution is available to draw from. The idea behind the Gibbs sampler is that it is far easier to sequentially draw from conditional densities—which, in a bivariate random variable (x, y) example, proceeds as follows: $x_{\{r\}} \sim \text{Prob}(x|y = y_{\{r-1\}})$ followed by $y_{\{r\}} \sim \text{Prob}(y|x = x_{\{r\}})$ followed by $x_{\{r+1\}} \sim \text{Prob}(x|y = y_{\{r\}})$, etc., converging, for a sufficiently large r , to be draws from the marginal distributions $\text{Prob}(x)$ and $\text{Prob}(y)$ —than it is to obtain the marginals by integration of the joint density $\text{Prob}(x, y)$, that is, $\text{Prob}(x) = \int \text{Prob}(x, y)dy$ and $\text{Prob}(y) = \int \text{Prob}(x, y)dx$. In the MNP model this opportunity of using the Gibbs sampling scheme is made possible by data augmentation, that is, just as drawing values for the unknown coefficients through their conditional density functions, draws for the latent variables (the utility differences) can be obtained by conditional density functions as well. In fact, the GHK simulator nicely allows for data augmentation by drawing the utility differences $U_{it1,21}$ and $U_{it1,31}$ conditional on the coefficients $(\beta'_{1,21}, \beta'_{1,31}, c_{1,121}, c_{1,122})'$ as follows:

1. Calculate $\text{Prob}(\eta_1 < -\mathbf{X}'_{it1} \beta_{1,21\{r-1\}}) = \Phi(-\mathbf{X}'_{it1} \beta_{1,21\{r-1\}})$.
2. Draw $\eta_{1\{r\}} = \Phi^{-1}(\mu_{1\{r\}} \Phi(-\mathbf{X}'_{it1} \beta_{1,21\{r-1\}}))$.
3. Calculate $\text{Prob}(\eta_2 < -(\mathbf{X}'_{it1} \beta_{1,31\{r-1\}} + c_{1,121\{r-1\}} \eta_{1\{r\}}) / c_{1,122\{r-1\}}) = \Phi(-(\mathbf{X}'_{it1} \beta_{1,31\{r-1\}} + c_{1,121\{r-1\}} \eta_{1\{r\}}) / c_{1,122\{r-1\}})$.
4. Draw $\eta_{2\{r\}} = \Phi^{-1}(\mu_{2\{r\}} \Phi(-(\mathbf{X}'_{it1} \beta_{1,31\{r-1\}} + c_{1,121\{r-1\}} \eta_{1\{r\}}) / c_{1,122\{r-1\}}))$.
5. Compute the draws for $U_{it1,21}$ and $U_{it1,31}$ by the utility models $U_{it1,21\{r\}} = \mathbf{X}'_{it1} \beta_{1,21\{r-1\}} + \eta_{1\{r\}}$ and $U_{it1,31\{r\}} = \mathbf{X}'_{it1} \beta_{1,31\{r-1\}} + c_{1,121\{r-1\}} \eta_{1\{r\}} + c_{1,122\{r-1\}} \eta_{2\{r\}}$.

To obtain draws for the utility differences when differenced against arrival state 2, that is, $U_{it1,12}$ and $U_{it1,32}$, recall that $\beta_{1,12}$ and $\beta_{1,32}$ can be derived from $\beta_{1,21}$ and $\beta_{1,31}$ through (25a), giving $\beta_{1,12} = -\beta_{1,21}$ and $\beta_{1,32} = \beta_{1,31} - \beta_{1,21}$. Furthermore, we use (21), (25b) and (26) to derive the coefficients $c_{1,212}$ and $c_{1,222}$ from $c_{1,112}$ and $c_{1,122}$. Then, steps 1–5 as above can be followed to obtain the draws $U_{it1,12\{r\}}$ and $U_{it1,32\{r\}}$ from which the against the arrival state 1 utility differenced utilities of arrival states 2 and 3 can be derived by the definitions $U_{it1,21\{r\}} = -U_{it1,12\{r\}}$ and $U_{it1,31\{r\}} = U_{it1,32\{r\}} - U_{it1,12\{r\}}$. However, to obtain an MCMC or DE-MC sequence of coefficient draws converging towards a stationary distribution our experience is that the utility differences when differenced against the last arrival state, in our model this is arrival state 3 and hence, the utility differences are $U_{it1,13}$ and $U_{it1,23}$, must be drawn as follows:

1. Pre-multiply equation system (29a)–(29b) by $\mathbf{A}_{13} = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$ (cf. (25a)) to obtain: $U_{it1,13} = \mathbf{X}'_{it1}\beta_{1,13} - c_{1,121}\eta_1 - c_{1,122}\eta_2$ and $U_{it1,23} = \mathbf{X}'_{it1}\beta_{1,23} + (1 - c_{1,121})\eta_1 - c_{1,122}\eta_2$ which allow $P_{it13} = \text{Prob}(U_{it1,13} < 0 \text{ and } U_{it1,23} < 0)$ to be expressed as $\text{Prob}(-\eta_2 < (\mathbf{X}'_{it1}\beta_{1,31} + c_{1,121}\eta_1)/c_{1,122} \text{ and } -\eta_2 < (\mathbf{X}'_{it1}\beta_{1,32} + (c_{1,121} - 1)\eta_1)/c_{1,122})$.
2. Since $\eta_1 < -\mathbf{X}'_{it1}\beta_{1,21}$ in case of arrival state 1, see (23), while $\eta_1 > -\mathbf{X}'_{it1}\beta_{1,21}$ when 2 is the arrival state, see (30a), a draw for η_1 can simply be obtained by a draw $\eta_{1\{r\}}$ from the standard normal distribution.
3. Given that $P_{it13} = \text{Prob}(-\eta_2 < (\mathbf{X}'_{it1}\beta_{1,31} + c_{1,121}\eta_1)/c_{1,122} \text{ and } -\eta_2 < (\mathbf{X}'_{it1}\beta_{1,32} + (c_{1,121} - 1)\eta_1)/c_{1,122})$ a draw $\eta_{2\{r\}} = -\Phi^{-1}(\mu_{2\{r\}}\Phi((\mathbf{X}'_{it1}\beta_{1,31\{r-1\}} + c_{1,121\{r-1\}}\eta_{1\{r\}})/c_{1,122\{r-1\}}))$ is taken if $\mathbf{X}'_{it1}\beta_{1,31\{r-1\}} > \mathbf{X}'_{it1}\beta_{1,32\{r-1\}} - \eta_{1\{r\}}$ and $\eta_{2\{r\}} = \Phi^{-1}(\mu_{2\{r\}}\Phi((\mathbf{X}'_{it1}\beta_{1,32\{r-1\}} + (c_{1,121\{r-1\}} - 1)\eta_{1\{r\}})/c_{1,122\{r-1\}}))$ otherwise.
4. Derive $U_{it1,13\{r\}} = \mathbf{X}'_{it1}\beta_{1,13\{r-1\}} - c_{1,121\{r-1\}}\eta_{1\{r\}} - c_{1,122\{r-1\}}\eta_{2\{r\}}$ and $U_{it1,23\{r\}} = \mathbf{X}'_{it1}\beta_{1,23\{r-1\}} + (1 - c_{1,121\{r-1\}})\eta_{1\{r\}} - c_{1,122\{r-1\}}\eta_{2\{r\}}$.

Also these negative utility differences $U_{it1,13\{r\}}$ and $U_{it1,23\{r\}}$ can be transformed to values of $U_{it1,21\{r\}}$ and $U_{it1,31\{r\}}$, since $U_{it1,21\{r\}} = U_{it1,23\{r\}} - U_{it1,13\{r\}}$ and $U_{it1,31\{r\}} = -U_{it1,13\{r\}}$.

Now that the draws $U_{it1,21\{r\}}$ and $U_{it1,31\{r\}}$ have been obtained for each case n_1 , we can estimate the coefficients $(\beta'_{1,21}, \beta'_{1,31})'$ by performing a regression of $\mathbf{U}_{1,21\{r\}}$ on $\mathbf{X}_{1,1}$, where $\mathbf{U}_{1,21\{r\}} = [U_{1,21\{r\}}, \dots, U_{N_1,21\{r\}}]'$ is an N_1 -dimensional vector and $\mathbf{X}_{1,1}$ is the $(N_1 \times k)$ matrix defined earlier, to obtain the estimate of $\beta_{1,21}$, denoted as $E_{\{r\}}(\beta_{1,21})$, and by performing a regression of $\mathbf{U}_{1,31\{r\}}$ on $\mathbf{X}_{1,1}$ to obtain the estimate of $\beta_{1,31}$, denoted as $E_{\{r\}}(\beta_{1,31})$. As a consequence, we now know that $\beta_{1,21}$ is normally distributed with mean vector $E_{\{r\}}(\beta_{1,21})$ and covariance matrix $\text{var}(\beta_{1,21}) = (\mathbf{X}'_{1,1}\mathbf{X}_{1,1})^{-1}$, see below formula (33). Using the Cholesky decomposition $\mathbf{C}_{1,21}\mathbf{C}'_{1,21} = \text{var}(\beta_{1,21})$, where $\mathbf{C}_{1,21}$ is a lower-triangular matrix, a draw of $\beta_{1,21}$ is given by $\beta_{1,21\{r\}} = E_{\{r\}}(\beta_{1,21}) + \mathbf{C}_{1,21}\mathbf{z}_{1,21\{r\}}$, where $\mathbf{z}_{1,21\{r\}}$ is a k -dimensional vector of standard normal random values. Similarly, a draw of $\beta_{1,31}$ is obtained by $\beta_{1,31\{r\}} = E_{\{r\}}(\beta_{1,31}) + \mathbf{C}_{1,31\{r-1\}} \times \mathbf{z}_{1,31\{r\}}$, where $\mathbf{C}_{1,31\{r-1\}}$ is the Cholesky factor of the covariance matrix $\text{var}(\beta_{1,31})_{\{r-1\}} = (c^2_{1,121\{r-1\}} + c^2_{1,122\{r-1\}})(\mathbf{X}'_{1,1}\mathbf{X}_{1,1})^{-1}$ and $\mathbf{z}_{1,31\{r\}}$ is a vector of k standard normal random numbers.

Last to be drawn are the coefficients $c_{1,121}$ and $c_{1,122}$ of the covariance matrix. As a consequence of the identification restriction $\theta_{1,122} = 1$ the covariance matrix $\Omega_{1,1}^*$, conditional on $\mathbf{U}_{\cdot,1,21\{r\}}$, $\mathbf{U}_{\cdot,1,31\{r\}}$, $\boldsymbol{\beta}_{1,21\{r\}}$ and $\boldsymbol{\beta}_{1,31\{r\}}$, is no longer Wishart distributed (cf. Koop 2003, p. 224). To solve this problem we may employ the DE-MC-within-Gibbs method in which the objective function LL_1 in (16') is simply given by the conditional density function (cf. (6)):

$$\begin{aligned} & \ln\{\text{Prob}(\boldsymbol{\varepsilon}_{\cdot,1,1}|\mathbf{U}_{\cdot,1,21\{r\}}, \mathbf{U}_{\cdot,1,31\{r\}}, \boldsymbol{\beta}_{1,21\{r\}}, \boldsymbol{\beta}_{1,31\{r\}})\} \\ &= -N_1((J - 1)/2) \ln(2\pi) - (N_1/2) \ln |\Omega_{1,1}^*| \\ & \quad - 0.5 \sum_{n_1=1}^{N_1} (\boldsymbol{\varepsilon}'_{n_1,1\{r\}} \Omega_{1,1}^{*-1} \boldsymbol{\varepsilon}_{n_1,1\{r\}}) \end{aligned} \tag{34}$$

where $\boldsymbol{\varepsilon}_{n_1,1\{r\}} = [U_{n_1,21\{r\}} - \mathbf{X}'_{n_1} \boldsymbol{\beta}_{1,21\{r\}}, U_{n_1,31\{r\}} - \mathbf{X}'_{n_1} \boldsymbol{\beta}_{1,31\{r\}}]'$, such that the $v_{1i\{g\}}$ and $\boldsymbol{\beta}_{1i\{g\}}$ individuals in (16') are now 2-dimensional vectors of estimates for the $\Omega_{1,1}^*$ parameters $c_{1,121}$ and $c_{1,122}$ (so $D = 2$) at iteration (generation) $r(=g)$ to be selected (for the next iteration of the Gibbs sampling scheme) according to the selection rule in (16') and as described in the remainder of the paragraph below (16'). Another solution to draw values of $c_{1,121}$ and $c_{1,122}$ is called the ‘‘full Gibbs sampler’’ (see McCulloch et al. 2000). This method extends the regression of $\mathbf{U}_{\cdot,1,31\{r\}}$ on $\mathbf{X}_{\cdot,1}$ by including the residuals of the regression of $\mathbf{U}_{\cdot,1,21\{r\}}$ on $\mathbf{X}_{\cdot,1}$ as a regressor so that $c_{1,121}$ can be drawn along with the coefficients $\boldsymbol{\beta}_{1,31} : (\boldsymbol{\beta}'_{1,31\{r\}}, c_{1,121\{r\}})' = E_{\{r\}}([\boldsymbol{\beta}'_{1,31}, c_{1,121}]') + \mathbf{C}^*_{1,31\{r-1\}} \mathbf{z}^*_{1,31\{r\}}$, where $\mathbf{C}^*_{1,31\{r-1\}}$ is the Cholesky factor of the covariance matrix $\text{var}([\boldsymbol{\beta}'_{1,31}, c_{1,121}]')_{\{r-1\}} = c_{1,122\{r-1\}}^2 \{[\mathbf{X}_{\cdot,1}, \mathbf{U}_{\cdot,1,21\{r\}} - \mathbf{X}_{\cdot,1} E_{\{r\}}(\boldsymbol{\beta}_{1,21})]' [\mathbf{X}_{\cdot,1}, \mathbf{U}_{\cdot,1,21\{r\}} - \mathbf{X}_{\cdot,1} E_{\{r\}}(\boldsymbol{\beta}_{1,21})]\}^{-1}$ and $\mathbf{z}^*_{1,31\{r\}}$ is a vector of $k + 1$ standard normal random numbers, while a draw of $c_{1,122}$ can be obtained by using the following conditional distribution

$$\begin{aligned} & \left(\sum_{n_1=1}^{N_1} \{U_{n_1,31\{r\}} - [\mathbf{X}'_{n_1}, U_{n_1,21\{r\}} - \mathbf{X}'_{n_1} E_{\{r\}}(\boldsymbol{\beta}_{1,21})]'\} \right. \\ & \left. [\boldsymbol{\beta}'_{1,31\{r\}}, c_{1,121\{r\}}]'\right)^2 / c_{1,122}^2 |\mathbf{U}_{\cdot,1,21\{r\}}, \mathbf{U}_{\cdot,1,31\{r\}}, \mathbf{X}_{\cdot,1}, \boldsymbol{\beta}_{1,31\{r\}}, c_{1,121\{r\}}| \sim \chi_{N_1-1}^2 \end{aligned} \tag{35}$$

from which it follows that

$$\begin{aligned} c_{1,122\{r\}} = & \left\{ \sum_{n_1=1}^{N_1} (\mathbf{U}_{n_1,31\{r\}} - [\mathbf{X}'_{n_1}, U_{n_1,21\{r\}} - \mathbf{X}'_{n_1} E_{\{r\}}(\boldsymbol{\beta}_{1,21})]')' [\boldsymbol{\beta}_{1,31\{r\}}, c_{1,121\{r\}}]'\right)^2 \\ & / \sum_{i=1}^{N_1-1} (z_{i\{r\}})^2 \Big\}^{1/2} \end{aligned} \tag{36}$$

where the $z_{i\{r\}}$ are standard normal random numbers so that $\sum_{i=1}^{N_1-1} z_{i\{r\}}^2$ is a draw from $\chi_{N_1-1}^2$. As both the DE-MC within Gibbs sampler and the full Gibbs sampler appear

to generate identical results in the empirical analysis, we shall refer to these samplers as the DE-MC-within-Gibbs sampler.

4 Empirical Results

In the empirical analysis we compare two models, MNL and MNP, and two estimation methods, M(S)L by the BFGS algorithm and M(S)L estimation by the DE-MC or DE-MC-within-Gibbs sampler. We compare the coefficient estimates and their significance and we evaluate the within-sample and out-of-sample model predictions of the average probabilities of transition from state j to state j' . Recall that for each firm we look at the change in profit between quarter $t - 1$ and quarter t . We consider three categories of changes: profit does not change (state 1), profit goes up (state 2) and profit goes down (state 3). Then, for each firm, we consider each pair of subsequent states, the first state called the departure (or starting) state and the next state called the arrival state. Purpose of estimation is the probability of transition from departure state j to arrival state j' ($j, j' = 1, 2, 3$). This requires three models, one model for each departure state to estimate the probabilities of reaching arrival states 1, 2 and 3, where one of the probabilities is implied by the other two given the adding-up restriction according to which the sum of the three probabilities is one.

In Fig. 1a–f and Tables 1–7 we compare the MNL ML estimation results obtained by the BFGS algorithm with those computed by the DE-MC algorithm. First of all, it should be noticed that the BFGS estimates reproduce the results in Tables 4 and 5 of Van Nguyen et al. (2004). Comparing the BFGS estimates with those obtained by DE-MC, we observe that the point estimates of the coefficients appear to be the same. This is also true for the upper and lower levels of the 95% confidence intervals, except for the departure state 1 estimates, where the DE-MC interval is somewhat wider, in particular for the constant term (which is the last coefficient, coefficient no. 25), see Fig. 1a and b and Tables 1 and 2. Nevertheless, the confidence level differences hardly lead to any other conclusions regarding the significance of the estimates.

The performance of the DE-MC algorithm is further investigated in Fig. 2a–c. The DE-MC algorithm was applied with $NP = 2D + 5 = 2 \times 50 + 5 = 105$ individuals, where D equals the number of elements in the parameter vector $\beta_i = [\beta'_{i2}, \beta'_{i3}]'$, where β_{ij} is a (25×1) vector for each arrival state j ($j = 2, 3$) and each departure state i ($i = 1, 2, 3$) so that $D = 2 \times 25 = 50$. This population size is clearly smaller than the advice $NP = 10D$, but given that the MNL likelihood function is globally concave, a smaller NP , which increases computational speed, may suffice to let DE-MC generate the marginal distributions of the coefficients. In Fig. 2a three log-likelihoods (LLs) are displayed for iterations 1–5000. At each iteration the LLs of 105 parameter vectors β_1 are computed. The parameter vectors with the five smallest LLs are discarded. Next, one out of the 100 parameter vectors left is randomly selected. The LL of this selected parameter vector is indicated as LL1_DRAW_COEFF_i in Fig. 2a. The LL of the parameter vector that is obtained by averaging the selected parameter values over iterations $1 - i$ ($i = 1, \dots, 5000$), is denoted as LL1_MEAN_COEFF_1_i, while the LL resulting from the parameter vector that is derived by averaging the values of each parameter over the best 100 individuals of

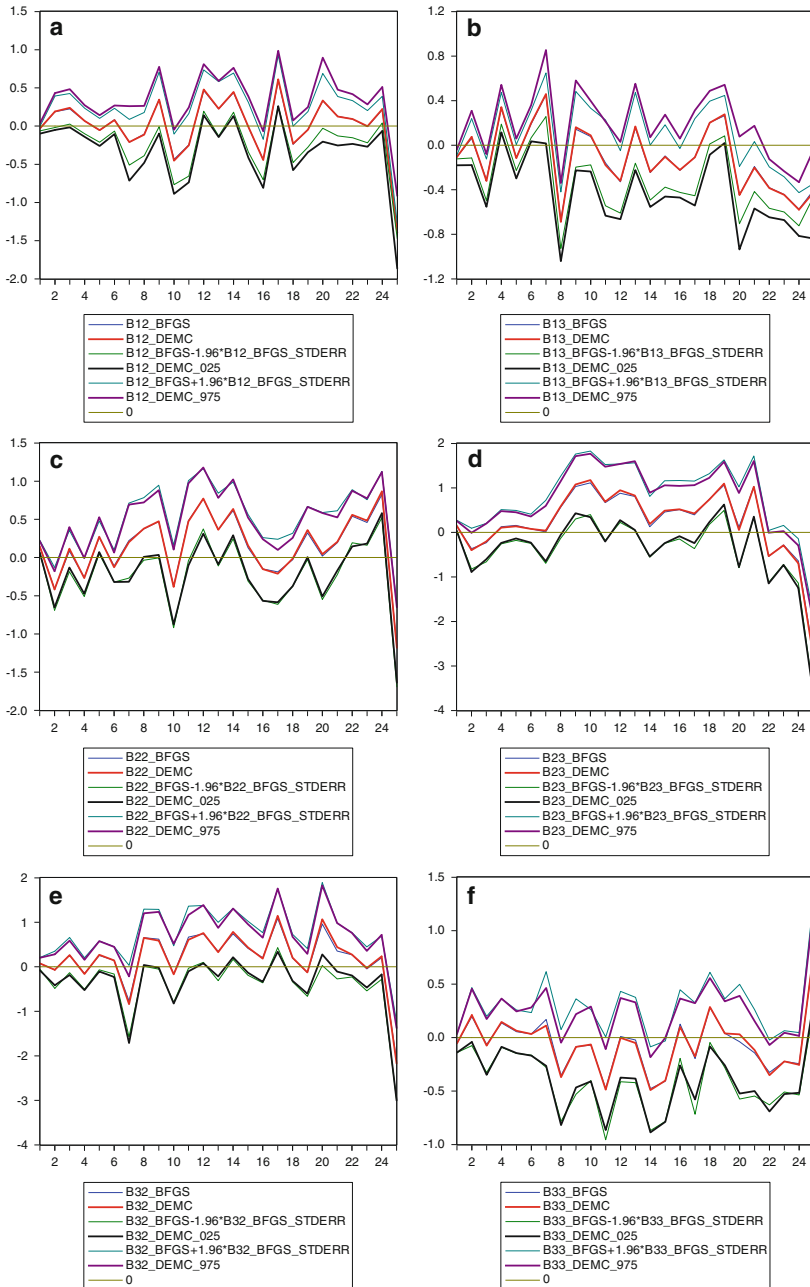


Fig. 1 **a** Comparison MNL coefficients β_{12} estimated by BFGS and DE-MC (see also Table 1). **b** Comparison MNL coefficients β_{13} estimated by BFGS and DE-MC (see also Table 2). **c** Comparison MNL coefficients β_{22} estimated by BFGS and DE-MC (see also Table 3). **d** Comparison MNL coefficients β_{23} estimated by BFGS and DE-MC (see also Table 4). **e** Comparison MNL coefficients β_{32} estimated by BFGS and DE-MC (see also Table 5). **f** Comparison MNL coefficients β_{33} estimated by BFGS and DE-MC (see also Table 6)

Table 1 Comparison MNL coefficients β_{12} estimated by BFGS and DE-MC (see also Fig. 1a)

Coeff	B12_BFGS	B12_DEMC	B12_BFGS -1.96 *B12_BFGS_ STDERR	B12_DEMC_ 025	B12_BFGS +1.96 *B12_BFGS_ STDERR	B12_DEMC_ 975
1	-0.030575	-0.028592	-0.062299	-0.097431	0.001148	0.040533
2	0.188491	0.190758	-0.018170	-0.050215	0.395151	0.431921
3	0.225821	0.239607	0.025390	-0.015740	0.426252	0.483317
4	0.067205	0.067220	-0.099503	-0.141109	0.233912	0.273329
5	-0.054853	-0.054202	-0.210567	-0.259764	0.100860	0.141991
6	0.083471	0.079710	-0.065405	-0.104875	0.232347	0.269505
7	-0.211484	-0.208623	-0.511224	-0.712943	0.088257	0.261793
8	-0.104919	-0.111695	-0.387713	-0.479797	0.177874	0.265301
9	0.348338	0.346556	-0.007241	-0.096566	0.703918	0.776058
10	-0.435776	-0.452571	-0.766273	-0.886859	-0.105279	-0.047635
11	-0.244782	-0.251215	-0.654617	-0.735277	0.165052	0.244142
12	0.465373	0.480230	0.195747	0.142612	0.734998	0.811037
13	0.225922	0.227497	-0.129838	-0.144066	0.581681	0.590369
14	0.438245	0.447934	0.181568	0.131110	0.694921	0.762086
15	-0.015849	-0.011974	-0.341447	-0.415542	0.309748	0.386817
16	-0.437137	-0.443743	-0.701248	-0.809684	-0.173027	-0.066798
17	0.596428	0.613056	0.262958	0.259879	0.929899	0.986160
18	-0.238376	-0.231126	-0.476359	-0.577092	-0.000394	0.074759
19	-0.043205	-0.047863	-0.276100	-0.342234	0.189691	0.249948
20	0.330737	0.334490	-0.027553	-0.204025	0.689028	0.895011
21	0.129303	0.125388	-0.129102	-0.252717	0.387707	0.475989
22	0.090274	0.092410	-0.152029	-0.232207	0.332577	0.418801
23	-0.007856	-0.000292	-0.218983	-0.271170	0.203270	0.283300
24	0.215664	0.222629	0.041828	-0.063492	0.389500	0.510836
25	-1.372298	-1.397091	-1.474124	-1.865073	-1.270473	-0.913140

Note: The coefficients numbers are attached to the independent variables as follows: 1=LSIZE; 2=OLD; 3=YNG; 4=EAST; 5=BANK; 6=DIVERS; 7=SINGPART; 8=ACC; 9=ADV; 10=ARC; 11=CAR; 12=CNS; 13=MCH; 14=SFT; 15=TRN; 16=WST; 17=1994; 18=1996; 19=1997; 20=1998; 21=1999; 22=Q2; 23=Q3; 24=Q4; 25=Intercept

generation i , is named $LL1_MEAN_COEFF_i_i$. Figure 2a shows that after 1000 iterations the graphs of the three LLs follow a stationary pattern ($LL1_DRAW_COEFF_i$) or almost horizontal line ($LL1_MEAN_COEFF_i_i$ and $LL1_MEAN_COEFF_1_i$). In addition, Fig. 2b shows the simulated values of coefficient β_{121} (that is, coefficient no. 1 in the β_{12} vector, which is the coefficient of the explanatory variable LSIZE) over iterations 1–5000. Their average over iterations $1 - i$ does not really change anymore after about 3500 iterations. This result is representative for the other coefficients as well. Hence, we discard iterations 1–4000 as burn-in and use the last 1000 simulated values for estimation and statistical inference purposes. For these 1000 values no

Table 2 Comparison MNL coefficients β_{13} estimated by BFGS and DE-MC (see also Fig. 1b)

Coeff	B13_BFGS	B13_DEMC	B13_BFGS -1.96 *B13_BFGS_ STDERR	B13_DEMC_ 025	B13_BFGS +1.96 *B13_BFGS_ STDERR	B13_DEMC_ 975
1	-0.106802	-0.107232	-0.122859	-0.180529	-0.090745	-0.042571
2	0.062232	0.074770	-0.114019	-0.179319	0.238482	0.308286
3	-0.309977	-0.321078	-0.496595	-0.552660	-0.123360	-0.079984
4	0.331582	0.342562	0.189420	0.115155	0.473744	0.542718
5	-0.110529	-0.116941	-0.227362	-0.298502	0.006304	0.059859
6	0.190320	0.199796	0.066946	0.034541	0.313693	0.360309
7	0.452829	0.459032	0.256938	0.016803	0.648721	0.854968
8	-0.673254	-0.688264	-0.927355	-1.040339	-0.419153	-0.340654
9	0.143231	0.161757	-0.196315	-0.225092	0.482776	0.580786
10	0.077567	0.089384	-0.176295	-0.235673	0.331430	0.397528
11	-0.158696	-0.178761	-0.543799	-0.632386	0.226407	0.213185
12	-0.328688	-0.320216	-0.608484	-0.663508	-0.048893	0.028872
13	0.155814	0.168911	-0.161626	-0.223848	0.473253	0.551778
14	-0.245706	-0.240183	-0.492318	-0.553775	0.000906	0.071060
15	-0.096225	-0.104067	-0.376518	-0.460821	0.184068	0.274643
16	-0.225818	-0.222627	-0.422809	-0.469603	-0.028827	0.059596
17	-0.107190	-0.108540	-0.453404	-0.541476	0.239025	0.311637
18	0.202947	0.202415	0.010593	-0.085429	0.395302	0.487867
19	0.265857	0.277183	0.084656	0.019846	0.447057	0.542499
20	-0.447987	-0.447690	-0.705497	-0.935730	-0.190477	0.077585
21	-0.191713	-0.201551	-0.416298	-0.567656	0.032873	0.173809
22	-0.380736	-0.384774	-0.565982	-0.646730	-0.195491	-0.125183
23	-0.441895	-0.444840	-0.600098	-0.672419	-0.283692	-0.234038
24	-0.574778	-0.578139	-0.723908	-0.815012	-0.425647	-0.332333
25	-0.398542	-0.418935	-0.454747	-0.838134	-0.342338	-0.032951

Note: See note in Table 1

structural serial correlation is being diagnosed (see Fig. 2c), nor are there any serious deviations from the normal distribution (see Fig. 2d). And also these results are found for the other coefficients as well. Nevertheless, Table 7 still reveals that, although not significantly so, the within-sample fit and out-of-sample predictions are slightly better for the BFGS algorithm as compared to the DE-MC sampler. Therefore, there is little to gain by using DE-MC instead of BFGS for MNL ML estimation. This, however, may not be true for the MNP model as the DE-MC-within-Gibbs algorithm with data augmentation avoids the curse of dimensionality in numerical integration.

The MNP model lifts the IIA restriction of the MNL model. Nevertheless, to find out whether or not the IIA assumption is violated by insufficient dissimilarity of the unexplained utility components ε_{ij} in

Table 3 Comparison MNL coefficients β_{22} estimated by BFGS and DE-MC (see also Fig. 1c)

Coeff	B22_BFGS	B22_DEMC	B22_BFGS -1.96 *B22_BFGS_ STDERR	B22_DEMC_ 025	B22_BFGS +1.96 *B22_BFGS_ STDERR	B22_DEMC_ 975
1	0.149031	0.148696	0.069398	0.068848	0.228664	0.216936
2	-0.409623	-0.417105	-0.690221	-0.650435	-0.129025	-0.178090
3	0.086502	0.116910	-0.184541	-0.130347	0.357545	0.397403
4	-0.259684	-0.269246	-0.509187	-0.473388	-0.010181	-0.006205
5	0.263843	0.273514	0.048710	0.072244	0.478977	0.528064
6	-0.110611	-0.127870	-0.321980	-0.319482	0.100757	0.064920
7	0.222183	0.201901	-0.268791	-0.316689	0.713157	0.693129
8	0.374514	0.379607	-0.034804	0.009548	0.783831	0.722053
9	0.470773	0.474722	-0.005243	0.034423	0.946789	0.881962
10	-0.376507	-0.384270	-0.917851	-0.874780	0.164837	0.104500
11	0.482489	0.473323	-0.042199	-0.103056	1.007177	0.972042
12	0.769694	0.773031	0.374539	0.312979	1.164850	1.179328
13	0.366350	0.363092	-0.110115	-0.089408	0.842814	0.781535
14	0.616443	0.636481	0.243710	0.292320	0.989176	1.024336
15	0.128774	0.156935	-0.307166	-0.280494	0.564713	0.524565
16	-0.148734	-0.151534	-0.558493	-0.567400	0.261026	0.241712
17	-0.186902	-0.210454	-0.613559	-0.585759	0.239755	0.099743
18	-0.028835	-0.013378	-0.375197	-0.374091	0.317528	0.258763
19	0.322355	0.359884	-0.024513	0.005560	0.669224	0.663754
20	0.020183	0.044222	-0.548865	-0.509608	0.589232	0.584167
21	0.194510	0.205570	-0.223467	-0.170008	0.612487	0.526405
22	0.541557	0.560253	0.193499	0.146134	0.889616	0.873606
23	0.458908	0.483373	0.160400	0.183186	0.757415	0.775130
24	0.825344	0.864870	0.522607	0.578546	1.128082	1.118766
25	-1.156336	-1.180662	-1.692288	-1.637143	-0.620385	-0.649994

Note: See note in Table 1

$$U_{ij} = \mathbf{X}'_i \boldsymbol{\beta}_j + \varepsilon_{ij} \tag{2'}$$

for, say, individual i ($i = 1, \dots, N$) and alternative j ($j = 1, \dots, J$), we may test for the joint absence of the non-diagonal elements in the covariance matrix of $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \dots, \varepsilon_{iJ}]'$. Consider $J = 3$. Then, $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}]'$ is assumed to be distributed as a trivariate normal, identically and independently across the N individuals, with zero mean and covariance matrix

$$Cov(\boldsymbol{\varepsilon}_i) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \tag{17'}$$

Table 4 Comparison MNL coefficients β_{23} estimated by BFGS and DE-MC (see also Fig. 1d)

Coeff	B23_BFGS	B23_DEMC	B23_BFGS -1.96 *B23_BFGS_ STDERR	B23_DEMC_ 025	B23_BFGS +1.96 *B23_BFGS_ STDERR	B23_DEMC_ 975
1	0.149943	0.156534	0.023603	0.043370	0.276283	0.260362
2	-0.363760	-0.398394	-0.820361	-0.886670	0.092840	-0.004089
3	-0.230006	-0.200750	-0.661233	-0.601720	0.201221	0.199890
4	0.128510	0.109282	-0.256589	-0.235605	0.513609	0.481288
5	0.157295	0.137554	-0.178941	-0.134081	0.493531	0.455299
6	0.082068	0.079233	-0.244846	-0.226921	0.408981	0.356051
7	0.012939	0.040730	-0.693073	-0.647629	0.718952	0.598064
8	0.560829	0.581483	-0.135540	-0.060685	1.257199	1.144722
9	1.031449	1.076132	0.302166	0.429915	1.760733	1.716094
10	1.115353	1.177605	0.401246	0.343282	1.829460	1.767834
11	0.669271	0.691270	-0.184896	-0.202921	1.523439	1.478710
12	0.881817	0.947164	0.224616	0.275031	1.539017	1.539959
13	0.808336	0.827969	0.050006	0.057062	1.566666	1.601469
14	0.128193	0.187908	-0.554578	-0.541399	0.810963	0.892714
15	0.458648	0.489778	-0.244356	-0.242484	1.161653	1.058154
16	0.511140	0.522511	-0.146076	-0.083844	1.168356	1.043936
17	0.395354	0.424802	-0.362167	-0.238237	1.152875	1.063264
18	0.749541	0.739112	0.179444	0.229613	1.319639	1.224830
19	1.061360	1.097922	0.494207	0.626457	1.628512	1.589323
20	0.117011	0.056153	-0.791827	-0.776984	1.025849	0.885706
21	1.030048	1.026730	0.342390	0.355520	1.717706	1.604685
22	-0.538779	-0.532588	-1.122197	-1.145820	0.044639	-0.004042
23	-0.281165	-0.289587	-0.720338	-0.729644	0.158009	0.030312
24	-0.639255	-0.693923	-1.137649	-1.244901	-0.140860	-0.290921
25	-2.721050	-2.776905	-3.611728	-3.629466	-1.830372	-1.982308

Note: See note in Table 1

Normalizing to alternative 1 gives

$$Cov \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i1} \end{pmatrix} = \begin{bmatrix} \sigma_{22} - 2\sigma_{12} + \sigma_{11} & \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} \\ \sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_{11} & \sigma_{33} - 2\sigma_{13} + \sigma_{11} \end{bmatrix} \quad (37)$$

The IIA restriction implies $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$, reducing (37) to

$$Cov \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i1} \end{pmatrix} = \begin{bmatrix} \sigma_{22} + \sigma_{11} & \sigma_{11} \\ \sigma_{11} & \sigma_{33} + \sigma_{11} \end{bmatrix} \quad (38)$$

Table 5 Comparison MNL coefficients $\beta_{3,21}$ estimated by BFGS and DE-MC (see also Fig. 1e)

Coeff	B32_BFGS	B32_DEMC	B32_BFGS -1.96 *B32_BFGS_ STDERR	B32_DEMC_ 025	B32_BFGS +1.96 *B32_BFGS_ STDERR	B32_DEMC_ 975
1	0.076603	0.075248	-0.054195	-0.082357	0.207400	0.203213
2	-0.066102	-0.067015	-0.484319	-0.415297	0.352115	0.280986
3	0.263120	0.264342	-0.132257	-0.184798	0.658497	0.588926
4	-0.151908	-0.159066	-0.505859	-0.523969	0.202044	0.158662
5	0.255508	0.269381	-0.067355	-0.104227	0.578371	0.576476
6	0.143330	0.144737	-0.164661	-0.226900	0.451321	0.446812
7	-0.768745	-0.840465	-1.570686	-1.710490	0.033196	-0.216860
8	0.654458	0.650056	0.012660	0.041191	1.296255	1.201436
9	0.619651	0.587709	-0.050892	-0.018616	1.290194	1.235587
10	-0.163862	-0.168586	-0.798061	-0.825559	0.470337	0.517587
11	0.666294	0.609438	-0.031663	-0.102396	1.364251	1.165500
12	0.740013	0.758228	0.100693	0.080501	1.379334	1.387844
13	0.345409	0.329311	-0.308483	-0.215272	0.999301	0.874840
14	0.739716	0.784406	0.167659	0.214015	1.311773	1.307073
15	0.417347	0.437140	-0.188736	-0.131483	1.023429	0.956042
16	0.204735	0.184915	-0.358860	-0.334628	0.768330	0.655548
17	1.089279	1.147111	0.431596	0.338643	1.746962	1.758981
18	0.195797	0.206024	-0.336521	-0.315222	0.728116	0.673011
19	-0.126777	-0.126124	-0.663149	-0.602720	0.409595	0.289080
20	0.962102	1.069152	0.029564	0.279601	1.894641	1.829789
21	0.353779	0.441375	-0.272897	-0.108515	0.980454	0.982830
22	0.270545	0.275930	-0.227546	-0.196552	0.768636	0.759800
23	-0.046259	-0.026053	-0.540067	-0.465178	0.447550	0.357856
24	0.206010	0.233235	-0.279290	-0.171292	0.691310	0.718476
25	-2.106812	-2.172419	-2.984981	-3.003585	-1.228643	-1.376334

Note: See note in Table 1

In addition to the IIA restriction the MNL model also imposes homoskedasticity, that is, $\sigma_{11} = \sigma_{22} = \sigma_{33}$. Moreover, normalizing for the scale of utility by setting $\sigma_{22} + \sigma_{11} = 1$ we finally obtain

$$Cov \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i1} \end{pmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \tag{39}$$

with Cholesky factor (cf. (21))

$$L_1 = \begin{bmatrix} 1 & 0 \\ c_{121} & c_{122} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0.5 & \sqrt{0.75} \end{bmatrix} \tag{40}$$

Table 6 Comparison MNL coefficients β_{33} estimated by BFGS and DE-MC (see also Fig. 1f)

Coeff	B33_BFGS	B33_DEMC	B33_BFGS −1.96 *B33_BFGS_ STDERR	B33_DEMC_ 025	B33_BFGS +1.96 *B33_BFGS_ STDERR	B33_DEMC_ 975
1	−0.053790	−0.053921	−0.139214	−0.139836	0.031634	0.033004
2	0.195873	0.210530	−0.075156	−0.040890	0.466903	0.455749
3	−0.064685	−0.074865	−0.327664	−0.348130	0.198294	0.173284
4	0.136749	0.146402	−0.090002	−0.086080	0.363499	0.364431
5	0.054904	0.063524	−0.148553	−0.145137	0.258361	0.244764
6	0.034888	0.032743	−0.164202	−0.168311	0.233978	0.281801
7	0.169141	0.112641	−0.278714	−0.263660	0.616995	0.464650
8	−0.353582	−0.370084	−0.782509	−0.817688	0.075344	−0.047556
9	−0.083813	−0.087188	−0.529172	−0.465822	0.361546	0.219306
10	−0.069788	−0.064200	−0.401478	−0.408382	0.261902	0.290106
11	−0.474901	−0.486821	−0.955955	−0.864165	0.006153	−0.108485
12	0.009547	−0.001831	−0.412984	−0.372943	0.432078	0.370336
13	−0.022221	−0.050777	−0.420768	−0.383799	0.376325	0.329190
14	−0.477400	−0.489777	−0.869202	−0.886688	−0.085598	−0.183896
15	−0.407900	−0.401452	−0.784329	−0.787765	−0.031471	−0.001335
16	0.126155	0.099966	−0.194628	−0.260763	0.446938	0.365819
17	−0.195259	−0.175989	−0.716576	−0.578008	0.326058	0.321285
18	0.282560	0.286756	−0.045804	−0.082580	0.610924	0.556647
19	0.040763	0.039905	−0.281155	−0.258709	0.362680	0.339398
20	−0.038058	0.030714	−0.574433	−0.523868	0.498316	0.390155
21	−0.142955	−0.116422	−0.546473	−0.500543	0.260564	0.154176
22	−0.325923	−0.351013	−0.628644	−0.690177	−0.023203	−0.068849
23	−0.222418	−0.222390	−0.509195	−0.528715	0.064359	0.045660
24	−0.244852	−0.254189	−0.536244	−0.515566	0.046540	0.015923
25	0.824404	0.833865	0.299280	0.362420	1.349528	1.260988

Note: See note in Table 1

so that we may estimate the MNP model subjected to the restrictions $c_{121} = 0.5$ and $c_{122} = \sqrt{0.75}$. Notice, however, that if the restrictions of non-correlation $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ in (38) are replaced by imposing the restrictions of equicorrelation $\sigma_{12} = \sigma_{13} = \sigma_{23}$ on (37), then we also obtain (39) and (40) as was nicely pointed out by [Monfardini and Santos Silva \(2008\)](#). Consequently, if the restrictions in (40) are not rejected, then the IIA assumption can still be invalidated by equicorrelation. Nevertheless, rejection of (40) invalidates the MNL model anyway and favors the general covariance matrix allowed for by MNP.

The empirical results, however, do not lead towards rejection of (40). The simulated loglikelihoods (SLLs) of the unrestricted (restricted) MNP models are obtained by the BFGS GHK estimation algorithm and found to be -3132.80 (-3132.78), -1670.35

Table 7 MNL estimation performance BFGS and DE-MC

Performance indicators	Starting state					
	1 (unchanged)		2 (up)		3 (down)	
	BFGS	DE-MC	BFGS	DE-MC	BFGS	DE-MC
<i>LL</i>	-3132.46	-3132.61	-1670.71	-1671.13	-1969.85	-1970.20
Arrival state 1						
Within-sample						
Observations	2077	2077	726	726	705	705
Transition P_{actual}	0.605	0.605	0.393	0.393	0.326	0.326
Transition $P_{pred.}$	0.605	0.608	0.394	0.393	0.326	0.326
Out-of-sample						
Observations	306	306	79	79	81	81
Transition P_{actual}	0.664	0.664	0.348	0.348	0.312	0.312
Transition $P_{pred.}$	0.625	0.627	0.435	0.434	0.371	0.375
Arrival state 2						
Within-sample						
Observations	619	619	911	911	249	249
Transition P_{actual}	0.180	0.180	0.494	0.494	0.115	0.115
Transition $P_{pred.}$	0.180	0.180	0.494	0.497	0.115	0.112
Out-of-sample						
Observations	70	70	122	122	30	30
Transition P_{actual}	0.152	0.152	0.537	0.537	0.115	0.115
Transition $P_{pred.}$	0.191	0.191	0.510	0.512	0.117	0.112
Arrival state 3						
Within-sample						
Observations	736	736	208	208	1211	1211
Transition P_{actual}	0.214	0.214	0.113	0.113	0.559	0.559
Transition $P_{pred.}$	0.214	0.213	0.113	0.110	0.559	0.562
Out-of-sample						
Observations	85	85	26	26	260	260
Transition P_{actual}	0.184	0.184	0.115	0.115	0.573	0.573
Transition $P_{pred.}$	0.184	0.182	0.055	0.054	0.512	0.513

Note: Within-sample (out-of-sample) concerns the period 94:2-99:4 (00:1-00:4)

(-1671.19) and -1669.17 (-1669.66) for departure states 1, 2 and 3, respectively. It is easy to see that these outcomes do not lead to significant likelihood ratio test statistics ($= 2(SLL_{unrestricted} - SLL_{restricted}) \sim \chi_r^2$ distributed with $r = 2$ restrictions) (in fact, for departure state 1 the restricted SLL is even a little bit higher than the unrestricted one), a result that complies with the nonrejection of the Hausman tests of the IIA hypothesis as performed by Van Nguyen et al. (2004, pp. 281, 282) for the MNL model. The similarity between the MNL and MNP estimates is also revealed by comparing the DE-MC MNL results with the DE-MC-within-Gibbs MNP results, see

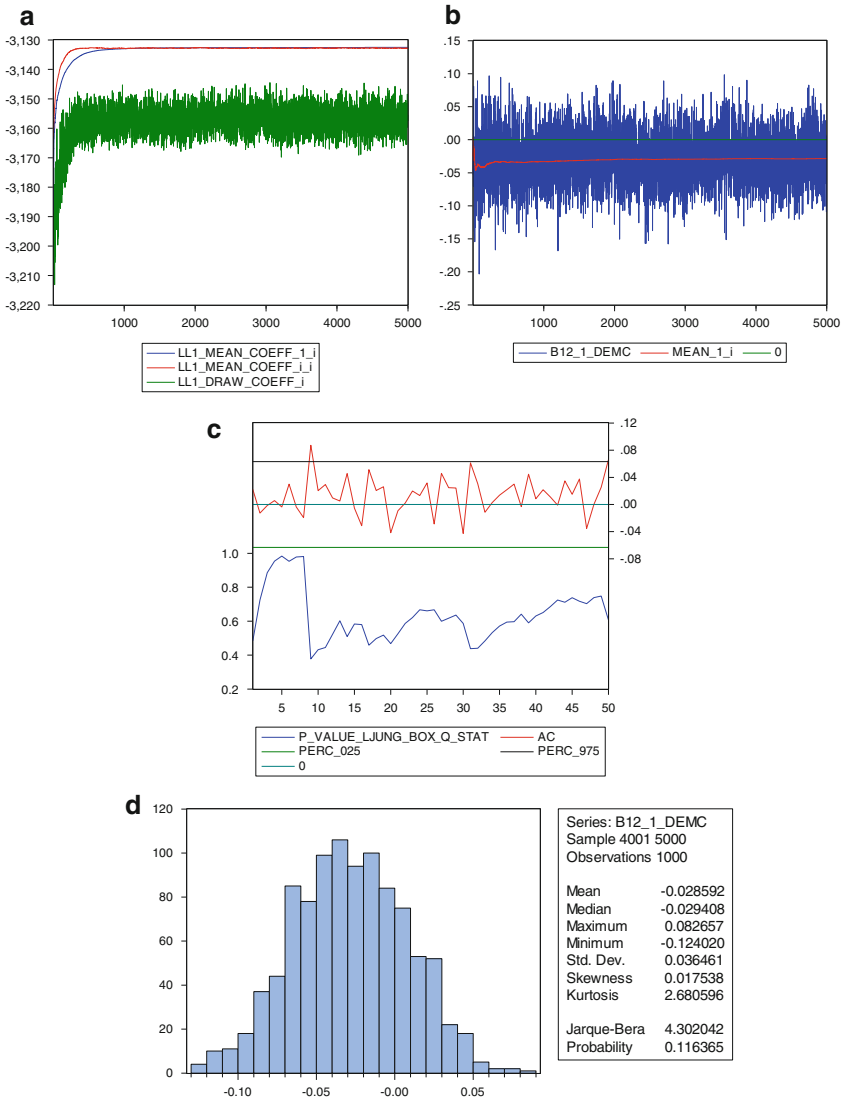


Fig. 2 **a** DE-MC ML estimation of MNL model for departure state 1: Loglikelihood when coefficients are averaged over iteration 1 up to and including iteration i ($i = 1, \dots, 5000$) ($LL1_MEAN_COEFF_1_i$), loglikelihood when coefficients are averaged over the individuals of generation (= iteration) i ($LL1_MEAN_COEFF_i_i$), and loglikelihood of coefficient vector (=individual) drawn from generation's i population ($LL1_DRAW_COEFF_i$). **b** DE-MC simulated values of first element of MNL coefficient vector β_{12} and its mean values averaged over iteration 1 up to and including iteration i ($i = 1, \dots, 5000$). **c** For DE-MC simulated values of first element of MNL coefficient vector β_{12} w.r.t. iterations 4001–5000 (see Fig. 2b): P -value of the Ljung-Box Q-statistic testing for the absence of autocorrelation up to and including order k ($k = 1, \dots, 50$), autocorrelation (AC) of order 1 up to and including the AC of order 50, and their 95% confidence interval. **d** For DE-MC simulated values of first element of MNL coefficient vector β_{12} w.r.t. iterations 4001-5000 (see Fig. 2b): Histogram and result for and p -value (Probability) of the Jarque-Bera statistic testing for normality.

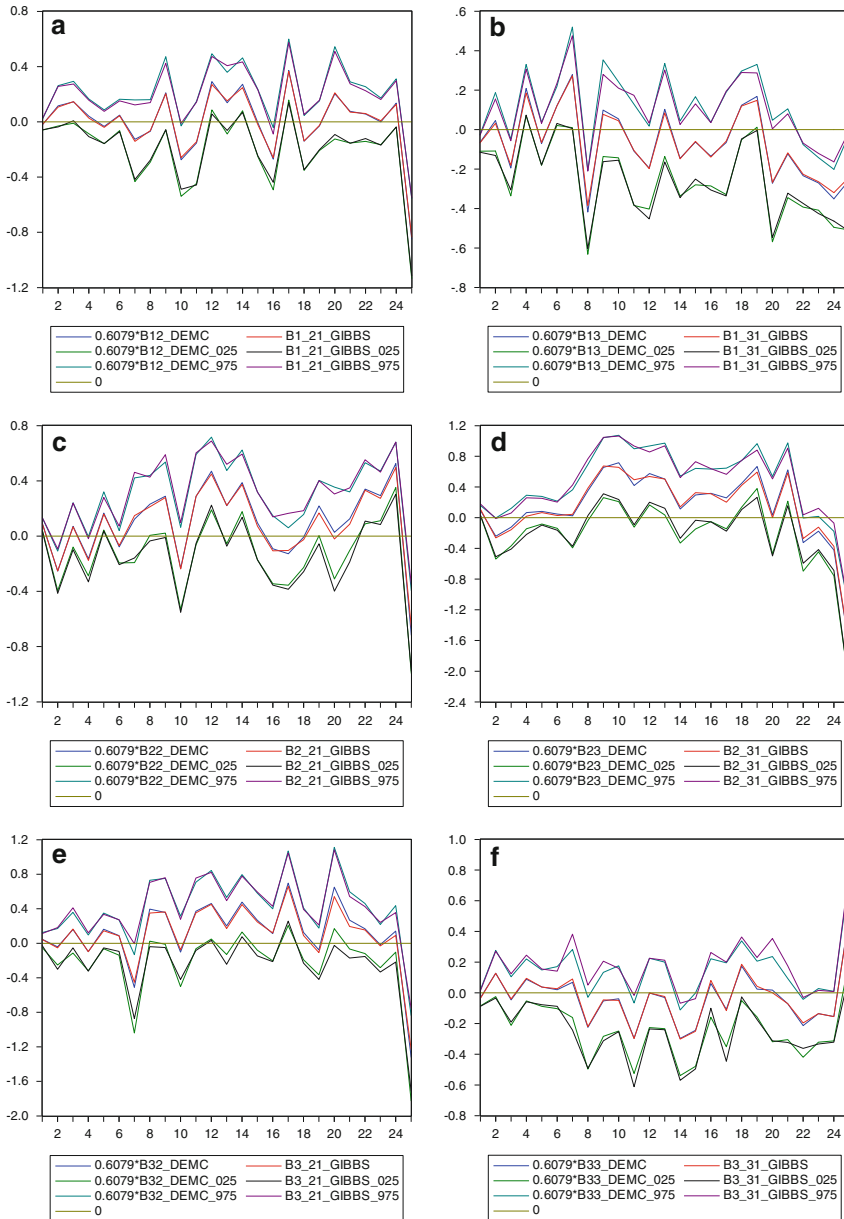


Fig. 3 **a** Comparison MNL coefficients β_{12} estimated by DE-MC and MNP coefficients $\beta_{1,21}$ estimated by DE-MC-within-Gibbs. **b** Comparison MNL coefficients β_{13} estimated by DE-MC and MNP coefficients $\beta_{1,31}$ estimated by DE-MC-within-Gibbs. **c** Comparison MNL coefficients β_{22} estimated by DE-MC and MNP coefficients $\beta_{2,21}$ estimated by DE-MC-within-Gibbs. **d** Comparison MNL coefficients β_{23} estimated by DE-MC and MNP coefficients $\beta_{2,31}$ estimated by DE-MC-within-Gibbs. **e** Comparison MNL coefficients β_{32} estimated by DE-MC and MNP coefficients $\beta_{3,21}$ estimated by DE-MC-within-Gibbs. **f** Comparison MNL coefficients β_{33} estimated by DE-MC and MNP coefficients $\beta_{3,31}$ estimated by DE-MC-within-Gibbs

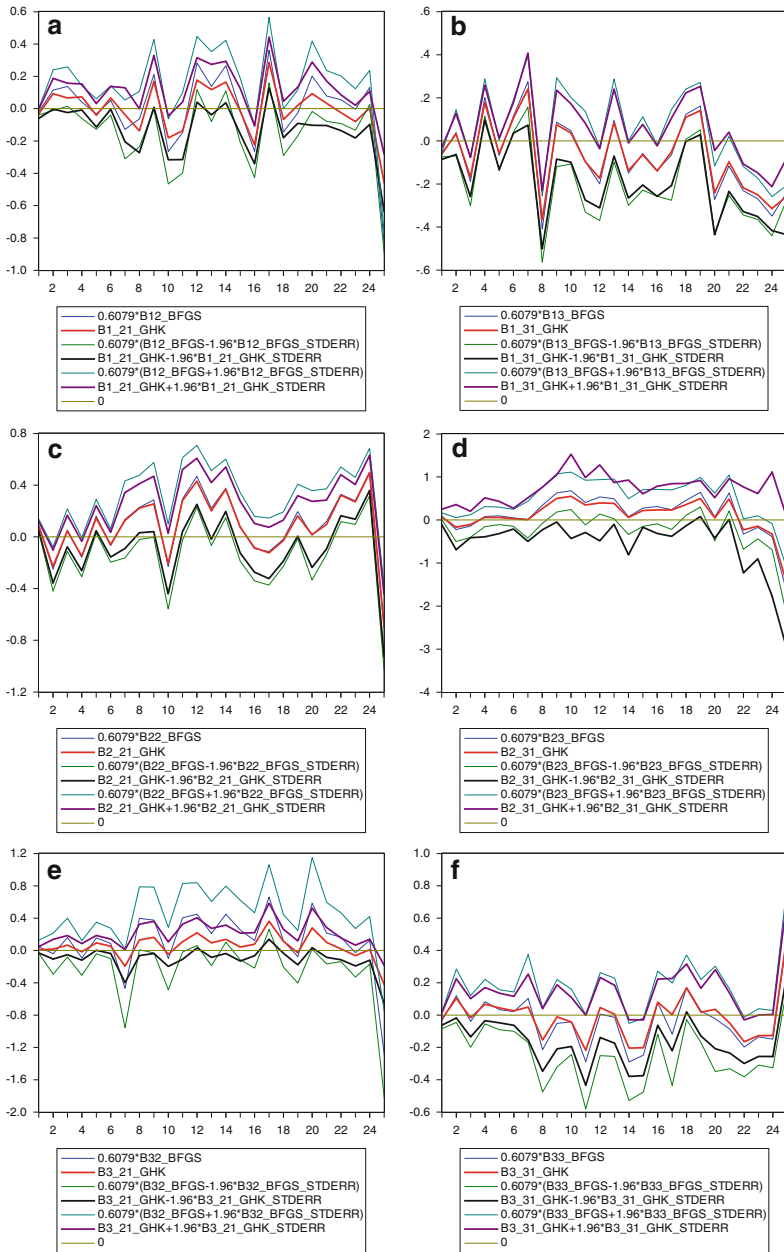


Fig. 4 **a** Comparison MNL coefficients β_{12} estimated by BFGS and MNP coefficients $\beta_{1,21}$ estimated by BFGS GHK. **b** Comparison MNL coefficients β_{13} estimated by BFGS and MNP coefficients $\beta_{1,31}$ estimated by BFGS GHK. **c** Comparison MNL coefficients β_{22} estimated by BFGS and MNP coefficients $\beta_{2,21}$ estimated by BFGS GHK. **d** Comparison MNL coefficients β_{23} estimated by BFGS and MNP coefficients $\beta_{2,31}$ estimated by BFGS GHK. **e** Comparison MNL coefficients β_{32} estimated by BFGS and MNP coefficients $\beta_{3,21}$ estimated by BFGS GHK. **f** Comparison MNL coefficients β_{33} estimated by BFGS and MNP coefficients $\beta_{3,31}$ estimated by BFGS GHK

Table 8 MNP estimation performance BFGS GHK and DE-MC within Gibbs

Performance indicators	Starting state					
	1 (unchanged)		2 (up)		3 (down)	
	BFGS GHK	DE-MC within Gibbs	BFGS GHK	DE-MC within Gibbs	BFGS GHK	DE-MC within Gibbs
<i>LL</i>	-3132.80	-3132.92	-1670.35	-1672.54	-1969.17	-1970.70
Arrival state 1						
Within-sample						
Observations	2077	2077	726	726	705	705
Transition P_{actual}	0.605	0.605	0.393	0.393	0.326	0.321
Transition $P_{pred.}$	0.606	0.601	0.393	0.391	0.326	0.326
Out-of-sample						
Observations	306	306	79	79	81	81
Transition P_{actual}	0.664	0.664	0.348	0.348	0.312	0.312
Transition $P_{pred.}$	0.626	0.619	0.436	0.427	0.370	0.362
Arrival state 2						
Within-sample						
Observations	619	619	911	911	249	249
Transition P_{actual}	0.180	0.180	0.494	0.494	0.115	0.115
Transition $P_{pred.}$	0.180	0.179	0.494	0.491	0.115	0.113
Out-of-sample						
Observations	70	70	122	122	30	30
Transition P_{actual}	0.152	0.152	0.537	0.537	0.115	0.115
Transition $P_{pred.}$	0.192	0.193	0.509	0.516	0.115	0.118
Arrival state 3						
Within-sample						
Observations	736	736	208	208	1211	1211
Transition P_{actual}	0.214	0.214	0.113	0.113	0.559	0.559
Transition $P_{pred.}$	0.214	0.220	0.113	0.118	0.560	0.566
Out-of-sample						
Observations	85	85	26	26	260	260
Transition P_{actual}	0.184	0.184	0.115	0.115	0.573	0.573
Transition $P_{pred.}$	0.182	0.188	0.055	0.058	0.515	0.521

Notes: Within-sample (out-of-sample) concerns the period 94:2–99:4 (00:1–00:4). In this table the estimates are obtained for the following categorization: Alternative 1 = Arrival state 1 (unchanged profit); Alternative 2 = Arrival state 2 (profit goes up); Alternative 3 = Arrival state 3 (profit goes down). Normalization for level is achieved by taking utility differences against the utility of Alternative 1

Fig. 3a–f and compare the DE-MC results in Table 7 with the DE-MC-within-Gibbs ones in Table 8. The MNP estimates in Fig. 3a–f and, to be discussed below, Figures 4a–f, are always obtained by estimating the model such that its first equation for the utility differences always concerns the arrival state under study (differences are

always taken with respect to arrival state 1) so that the residuals are standard normal (as a consequence of the normalization for scale). In the MNL model the residuals are distributed extreme value with variance $\pi^2/6$ so the MNL point and confidence interval estimates are multiplied by $6/\pi^2 \approx 0.6079$ for comparison with the MNP results. The similarity between the estimation results is striking and hence, clearly provides further evidence in favor of the MNL model.

Although the DE-MC-within-Gibbs algorithm generates the β estimates by the first equation for the differences in utility (we always difference against the arrival state 1 utility), the unrestricted residual variance of the second utility differences equation is exploding by the exponentially increasing draws of c_{122} . Fortunately, restricting the draws of $\ln(c_{122})$ (the natural log transformation lifts the lower level 0 restriction) by an upper level of, say, $\ln(4)$ (if a draw of $\ln(c_{122})$ has exceeded $\ln(4)$, then the draws of the utility differences at iteration i are obtained by conditioning on the average value of the iteration $1 - (i - 1)$ draws of $\ln(c_{122})$ instead of the single $\ln(c_{122})$ value drawn at iteration $i - 1$), does not affect the β estimates of the first utility equation as we observed by looking at higher levels than $\ln(4)$ (e.g. $\ln(25)$, $\ln(100)$ or even $\ln(1000)$). At levels lower than $\ln(4)$ the SLL (computed by the estimates obtained by averaging the draws of a sample of size 250 composed by taking the drawn values at every 50th iteration starting at iteration 12,550 and ending with iteration 25,000) decreases, while for higher levels the SLL shows ignorable increments. Consequently, the exact upper level of c_{122} does not really matter as long as it is not too low. In turn, the identification problem of the covariance parameters also arises in the BFGS GHK algorithm. There the c_{121} and c_{122} estimates appear to be rather sensitive for different starting values, also at grid sizes larger than $R = 10$ (see (27)), whereas the SLL hardly changes. Moreover, comparing the BFGS MNL estimates with those of BFGS GHK MNP, see Figures 4a–f and compare the BFGS results in Table 7 with the BFGS GHK estimates in Table 8, does not always reveal striking similarities. In stead, Figure 4a as well as Figure 4e display considerable differences. Therefore, we conclude that the DE-MC and DE-MC-within-Gibbs algorithms as outlined in this article has been quite useful in showing the validity of the MNL model regarding the Van Nguyen et al. (2004) dataset.

5 Conclusions

In this article we outline the estimation of the MNL model by DE-MC and the MNP model by the DE-MC-within-Gibbs sampler. We also describe how the MNL model can be estimated by the BFGS algorithm and how this algorithm can estimate the MNP model with the help of the GHK simulator.

The DE-MC estimator of the MNL model only requires the log-likelihood function to be specified and returns the full marginal distributions of the identified coefficients. In the empirical analysis the mean values and 95% confidence levels by these distributions are shown to be quite similar to those obtained by the BFGS MNL ML estimator.

The covariance matrix of the unexplained components of the utility differences in the MNP model has a nonstandard distribution when its first diagonal element is

set equal to one for scale identification. Sampling the identified covariance matrix coefficients by DE-MC-within-Gibbs, however, only requires the specification of the normal density of the unexplained utility differences, while the other coefficients can easily be drawn with the help of standard distributions, data augmentation and the GHK simulator. In the empirical analysis we show how this algorithm can be used in view of remaining identification problems with covariance matrix coefficients. The resulting MNP estimates appear to be rather similar to the DE-MC estimates of the MNL model so that we come to the conclusion that the MNP model does not improve the MNL model estimates in Van Nguyen et al. (2004).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Ben-Akiva, M., McFadden, D., Abe, M., Böckenholt, U., Bolduc, D., Gopinath, D., et al. (1997). Modeling methods for discrete choice analysis. *Marketing Letters*, 8, 273–286.
- Bolduc, D. (1999). A practical technique to estimate multinomial probit models in transportation. *Transportation Research Part B*, 33, 63–79.
- Börsch-Supan, A., & Hajivassiliou, V. (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics*, 58, 347–368.
- Byatt, D., Coope, I. D., & Price, C. J. (2004). Effect of limited precision on the BFGS quasi-Newton algorithm. *ANZIAM Journal*, 45, C283–C295.
- Cramer, J. S. (1991). *The Logit Model: An Introduction for Economists*. London: Edward Arnold.
- Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2), 141–149.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57, 1317–1339.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and student-*t* distributions subject to linear constraints. In E. M. Deramidas (Ed.), *Computer science and statistics: Proceedings of the twenty-third symposium on the interface* (pp. 571–578). Fairfax: Interface Foundation of North America, Inc.
- Geweke, J., Keane, M., & Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *Review of Economics and Statistics*, 76, 609–632.
- Green, W. H. (2008). *Econometric Analysis, Sixth Edition*. Upper Saddle River (NJ): Prentice Hall.
- Hajivassiliou, V., & McFadden, D. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 66, 863–896.
- Hess, S., Train, K. W., & Polak, J. W. (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 Part B*, 40, 147–163.
- Keane, M. (1990). *Four essays in empirical macro and labor economics*. Ph.D. Thesis, Brown University.
- Keane, M. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62, 95–116.
- Koop, G. (2003). *Bayesian Econometrics*. Chichester: Wiley.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. (pp. 105–142). New York: Academic Press.
- McCulloch, R. E., Polson, N. G., & Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99, 173–193.
- McFadden, D. (1984). Econometric analysis of qualitative response models. In Z. Griliches & M. Intriligator (Eds.), *Handbook of Econometrics, Vol. 2, Chap. 18*. (pp. 1395–1457). Amsterdam: North Holland.

- Monfardini, C., & Santos Silva, J. M. C. (2008). What can we learn about correlations from multinomial probit estimates. *Economics Bulletin*, 3(28), 1–9.
- Nguyen, P. Van, Laisney, F., & Kaiser, U. (2004). The performance of German firms in the business-related service sector: a dynamic analysis. *Journal of Business & Economic Statistics*, 22(3), 274–295.
- Nerlove, M. & Press, S. (1973). Univariate and multivariate log-linear and logistic models. RAND-R1306-EDA/NIH, Santa Monica.
- Price, K. V., Storn, R. M., & Lampinen, J. A. (2005). *Differential Evolution—A Practical Approach to Global Optimization*. Berlin: Springer.
- Schervish, M. (1984). Multivariate normal probabilities with error bound. *Applied Statistics*, 33, 81–87.
- Storn, R., & Price, K. (1995). Differential Evolution—A simple and efficient adaptive scheme for global optimization over continuous spaces. International Computer Science Institute, Berkeley, TR-95-012. <http://www.icsi.berkeley.edu~storn/litera.html>
- Storn, R., & Price, K. (1997). Differential Evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11, 341–359.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540.
- Ter Braak, C. J. F. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16, 239–249.
- Train, K. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge: Cambridge University Press.