Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Parameter estimation in multivariate logit models with many binary choices

Koen Bel, Dennis Fok & Richard Paap

ECONOMETRIC REVIEWS
2018, VOL. 37, NO. 5, 534–550
http://dx.doi.org/10.1080/07474938.2015.1093780

Parameter estimation in multivariate logit models with many binary choices

Koen Bel, Dennis Fok, and Richard Paap

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

ABSTRACT
Multivariate Logit models are convenient to describe multivariate correlated binary choices as they provide closed-form likelihood functions. However, the computation time required for calculating choice probabilities increases exponentially with the number of choices, which makes maximum likelihood-based estimation infeasible when many choices are considered. To solve this, we propose three novel estimation methods: (i) stratified importance sampling, (ii) composite conditional likelihood (CCL), and (iii) generalized method of moments, which yield consistent estimates and still have similar small-sample bias to maximum likelihood. Our simulation study shows that computation times for CCL are much smaller and that its efficiency loss is small.

KEYWORDS
Composite likelihood; generalized method of moments; multivariate logit model; stratified importance sampling

JEL CLASSIFICATION
C13; C31; C35; C51

1. Introduction

Multivariate choice models are widely used to describe correlated binary decision data in different fields of applied research. For example, grocery product choices by consumers are likely to be correlated across different brands or product categories (Chib et al., 2002). Choices for different types of insurances are correlated (Donkers et al., 2007), and effects of a medicine treatment on two or more physiological systems are also related (Ashford and Sowden, 1970). As a final example, Feddag (2013) investigates several "health-related quality of life" questions in a survey among cancer patients, and the answers to these questions are likely to be correlated. Hence, simultaneous binary decisions occur in many different fields of research.

The number of choices to be made in multivariate decision problems can be rather large. The number of products in a supermarket is large; individuals have to decide upon life, car, house insurances, and so forth; and the number of questions in a survey might also be large. There is therefore a need for a model that is applicable in these settings. In principle such models are available. However, current econometric estimation methods for multivariate choice models suffer from a computational burden if the number of choices grows large.

The standard econometric model to describe correlated multivariate binary choices is the Multivariate Probit model (Ashford and Sowden, 1970; Edwards and Allenby, 2003). The main disadvantage of this model is that the computation of the choice probabilities involves high-dimensional integrals which cannot be solved analytically. Numerical integration methods are not very accurate and slow in high dimensions, and simulation-based estimation methods are often used instead (Cappellari and Jenkins, 2006). However, the computational efforts to perform simulation-based estimation become excessive when a large number of correlated choices is considered. To avoid the evaluation of integrals, one may opt for multivariate binary decision models based on correlated logistic regressions. These models are

CONTACT Koen Bel bel@ese.eur.nl Erasmus Universiteit Rotterdam, Econometric Institute, Room H11-02, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

nonetheless difficult to generalize to higher dimensions (Carey et al., 1993; Glonek and McCullagh, 1995).

To avoid these difficulties we opt for the Multivariate Logit (MVL) model (Cox, 1972). Russell and Petersen (2000) show that this model can be written as a restricted Multinomial Logit (MNL) specification over all possible outcomes of the multivariate binary choices. The multivariate choice problem over K choices is reformulated as a *multinomial* choice model over 2^K alternatives.

The problem of this MVL specification is that the outcome space of the multivariate binary random variable, and thereby the computation time, increases exponentially with the number of choices. From a practical point of view, standard Maximum Likelihood (ML) parameter estimation becomes computationally infeasible even for a moderate number of choices. Further, numerical problems can occur as probabilities get too small for practical use. Russell and Petersen (2000) apply the model to four binary choices only and state that "as the number of categories becomes large, the approach taken in our research will clearly become infeasible" (p. 387). Guimarães et al. (2003) propose to use a more feasible approach based on Poisson regression. Unfortunately, this method only holds for the conditional logit specification where explanatory variables differ across choices. It therefore does not solve the infeasibility for all MVL specifications.

In this article, we propose three novel estimation methods for the MVL model which provide parameter estimates in an acceptable amount of time even if the number of binary choices is large. In the first proposed method, we use a sampling method to reduce the number of alternatives in the estimation routine. Using the method proposed by Ben-Akiva and Lerman (1985) we can still obtain consistent estimators for the model parameters. In the second method, we take advantage of the fact that the MVL model has simple conditional probabilities. We use these conditional probabilities in a Composite Conditional Likelihood (CCL) approach (Lindsay, 1988). In case of K binary choices, only K conditional probabilities have to be evaluated instead of 2^K joint probabilities, which reduces computing time. Furthermore, this method solves the problem of very small joint probabilities as these probabilities are not used within the estimation routine. Finally, we consider a Generalized Method of Moments (GMM) estimator based on the conditional probabilities, and hence this approach has the same advantages as the CCL approach. Monte Carlo results show that the three novel estimation methods are much faster, have similar small-sample biases as the standard ML approach of Russell and Petersen (2000), and that the loss in efficiency is very limited.

The remainder of this article is organized as follows. In Section 2 we describe the MVL model as discussed by Russell and Petersen (2000). Parameter inference is considered in Section 3. We first present standard ML parameter estimation followed by our three alternative methods. Section 4 describes the results of the Monte Carlo study which compares the estimation methods with respect to computation time, small-sample bias, and efficiency. Section 5 gives an illustration of the MVL model for a case with 10 binary choices for store choices of households in a shopping mall. Finally, Section 6 concludes.

2. Model specification

In this section, we discuss the model specification for the MVL model. We use the specification as introduced by Cox (1972) and further implemented by Russell and Petersen (2000).

Following Russell and Petersen (2000), we let Y_i denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \dots, N$, defined as $Y_i = \{Y_{i1}, \dots, Y_{iK}\}$, where Y_{ik} denotes the kth binary choice for individual i, for $k = 1, \dots, K$. The set of possible realizations of Y_i is called S which contains 2^K elements. It can immediately be seen that the number of possible realizations grows exponentially with the number of binary choices K.

The choices in Y_i may be correlated. To describe these dependencies, Russell and Petersen (2000) specify the conditional probabilities of the kth random variable Y_{ik} given all other choices, that is, y_{il} for $l \neq k$. These conditional probabilities are a Logit function of individual characteristics X_i, model parameters α, β, and ψ, and y_{il}, that is,

$$\Pr[Y_{ik} = 1 | y_{i1}, \dots, y_{ik-1}, y_{ik+1}, \dots, y_{iK}, X_i] = \frac{\exp(Z_{ik})}{1 + \exp(Z_{ik})} \quad (1)$$

with

$$Z_{ik} = \alpha_k + X_i\beta_k + \sum_{l\neq k} y_{il}\psi_{kl}, \tag{2}$$

where y_{il} is the realization of Y_{il}, α_k are alternative-specific intercepts, X_i is a $(1\times p)$-vector of explanatory variables with corresponding parameter vector β_k, and where ψ_{kl} are association parameters. The association parameters capture the correlation between Y_{ik} and Y_{il} for $l \neq k$. Positive association implies that options k and l tend to have similar values, and negative association implies that they tend to be different. Conditional independence between Y_{ik} and Y_{il} occurs when $\psi_{kl} = 0$. As we can only consider correlations and no causal impacts, we have to impose $\psi_{kl} = \psi_{lk}$ for symmetry, see also Russell and Petersen (2000). The model can be extended by including explanatory variables that differ across individuals and the different binary choices. Such an extension is straightforward, but to simplify notation we do not include such variables here.

Using the results in Besag (1974), the joint distribution of Y_i follows directly from the full set of conditional distributions. Russell and Petersen (2000) show that the conditional distributions in (1) imply an MNL specification for the joint distribution of Y_i, that is,

$$\Pr[Y_i = y_i|X_i] = \frac{\exp\left(\mu_{y_i}\right)}{\sum_{s_i\in S}\exp\left(\mu_{s_i}\right)}, \tag{3}$$

where y_i is a possible realization from the outcome space S and where μ_{y_i} is defined as

$$\mu_{y_i} = \sum_{k=1}^{K} y_{ik}(\alpha_k + X_i\beta_k) + \sum_{l>k} y_{ik}y_{il}\psi_{kl}. \tag{4}$$

Hence, the parameters α_k and β_k only occur in the numerator of the probability function for $Y_{ik} = 1$. Further, the association parameter ψ_{kl} only occurs in the numerator when both $y_{ik} = 1$ and $y_{il} = 1$. Note that this implies that all pairs should occur in the available data to be able to estimate these association parameters.

The interpretation of the impact of the intercept parameters and X_i follows from the log odds ratio

$$\log\left(\frac{\Pr[Y_i = y_i|X_i]}{\Pr[Y_i = (0,\dots,0)|X_i]}\right) = \sum_{k=1}^{K} y_{ik}(\alpha_k + X_i\beta_k) + \sum_{l>k} y_{ik}y_{il}\psi_{kl}, \tag{5}$$

where we use that $\mu_{(0,\dots,0)} = 0$ for identification. Clearly, the odds ratio equals μ_{y_i} as defined in (4) and provides the probability to observe y_i relative to the base set of choices where all choices are 0.

The association parameter ψ_{kl} is in theory an unbounded parameter and thus does not directly represent a correlation. However, log odds ratios give a direct interpretation of these association parameters. That is, it is easy to show that

$$\log\left(\frac{\Pr[Y_i = (0,\dots,0,y_k = 1,0,\dots,0,y_l = 1,0,\dots,0)|X_i]\,\Pr[Y_i = (0,\dots,0)|X_i]}{\Pr[Y_i = (0,\dots,0,y_k = 1,0,\dots,0)|X_i]\,\Pr[Y_i = (0,\dots,0,y_l = 1,0,\dots,0)|X_i]}\right) = \psi_{kl}. \tag{6}$$

A positive ψ_{kl} thus implies that choices k and l more often move together than apart.

The MVL model can be used to find dependencies in multivariate choices. In the next section, we discuss several estimation methods to uncover these dependencies. We discuss why standard ML estimation using the joint probabilities in (3) is not computationally feasible in case K is large. New feasible methods are therefore introduced.

3. Parameter inference

This section proposes four estimation methods for the MVL model specification defined in Section 2. The first approach is a standard ML estimation procedure. This approach, however, is computationally infeasible when the number of choices K is large. We therefore propose three alternative novel estimation methods.

Standard ML

The first estimation method directly follows Russell and Petersen (2000). To estimate the model parameters, they suggest to use the joint probabilities in (3). That is, Russell and Petersen (2000) use the MNL specification on the full outcome space S which results in the log-likelihood function

$$\ell^r(\theta; y) = \sum_{i=1}^{N} \log \Pr[Y_i = y_i | X_i], \tag{7}$$

where the joint probabilities $\Pr[Y_i = y_i|X_i]$ are given in (3). Further, θ summarizes all model parameters. To distinguish between the several methods, we add the superscript r to the likelihood function. Standard errors of the estimator can be obtained in the same way as for standard MNL models, see, for example Amemiya (1985).

This estimation approach is very suitable when the number of choices K is small. However, the number of alternatives S increases exponentially with K. For example, 10 binary choices already lead to $2^{10} = 1024$ potential outcomes of Y_i. This leads to very small probabilities in (3) and a sum of many terms in the denominator, which may both lead to computational problems. Furthermore, the computation time of the probabilities and hence the log-likelihood function will increase rapidly with the number of choices. The dominating factor in the time spent computing the log likelihood function for a single observation in (7) is the sum over the exponents, which has order of complexity 2^K. We next propose three alternative novel estimation methods which avoid the computation of all joint probabilities.

Stratified Importance Sampling

The first alternative method reduces the number of elements in the denominator and thereby avoids large summations and the evaluation of small probabilities. To achieve this, we use a stratified subset of the full outcome space for parameter estimation, where the selection probabilities for outcomes differ. Straightforwardly, using such a selection may however result in an inconsistent ML estimator. We use the correction term of Ben-Akiva and Lerman (1985, Section 9.3) to correct for the stratification. This correction term is related to the sampling probability of the subset.

Formally, let D_i be a subset of the full outcome space S. We know from McFadden (1978) that maximization of the conditional log-likelihood

$$\ell^s(\theta; y) = \sum_{i=1}^{N} \log \Pr[Y_i = y_i | D_i, X_i] \tag{8}$$

yields consistent parameter estimates if $y_i \in D_i$. From Bayes' theorem, we can write

$$\begin{aligned}
\Pr[Y_i = y_i | D_i, X_i] &= \frac{\Pr[Y_i = y_i|X_i] \Pr[D_i|Y_i = y_i, X_i]}{\sum_{d_i \in D_i} \Pr[Y_i = d_i|X_i] \Pr[D_i|Y_i = d_i, X_i]} \\
&= \frac{\exp\left(\mu_{y_i} + \log\left(\Pr[D_i|Y_i = y_i, X_i]\right)\right)}{\sum_{d_i \in D_i} \exp\left(\mu_{d_i} + \log\left(\Pr[D_i|Y_i = d_i, X_i]\right)\right)},
\end{aligned} \tag{9}$$

where we use that $\Pr[Y_i = y_i|X_i]$ for all y_i in S follows from (3). Hence, the correction term in the MNL specification for using a subsample D_i instead of the full outcome space S is $\log\left(\Pr[D|Y_i = y_i, X_i]\right)$.

To select an appropriate subsample D_i, we follow Ben-Akiva and Lerman (1985). They propose to use Stratified Importance Sampling (SIS) for the creation of the subset D_i and to find the values for the correction term. This selection method creates disjoint strata containing comparable alternatives. One randomly selects (with equal probabilities) a fixed number of alternatives within each stratum. For stratum r, we select n_r alternatives. For the stratum that contains y_i, we make sure that y_i is contained in the selected set.

Specifically, we create strata of singles, pairs, triplets, etc., in the multivariate binary choice data. Even though there may be many triplets, SIS allows us to limit the number of triplets we actually need to consider.

Formally, let R be the number of disjoint strata, and let q_r be the stratum-specific probability to be in subset D_i based on the fixed amount of alternatives to be drawn. This probability equals n_r divided by the number of alternatives in stratum r. Then, referring to Ben-Akiva and Lerman (1985), $\Pr[D_i|Y_i = y_i, X_i] \propto 1/q_{r(y_i)}$, where $r(y_i)$ is the stratum containing the joint set of binary choices under consideration.

Hence, the correction term equals the negative logarithm of the stratum-specific selection probabilities. The joint probabilities in (9) are then given by

$$\Pr[Y_i = y_i|D_i, X_i] = \frac{\exp\left(\mu_{y_i} - \log\left(q_{r(y_i)}\right)\right)}{\sum_{d_i \in D_i} \exp\left(\mu_{d_i} - \log\left(q_{r(d_i)}\right)\right)}. \tag{10}$$

Replacing the joint probabilities in (7) by (10) provides a stratified log-likelihood. The stratified importance sampling ML estimator is consistent, but there is loss in efficiency compared to full ML due to the sampling.

It is easy to see the advantages of this approach over the standard ML approach of Russell and Petersen (2000). Using only a subset D_i in SIS reduces the dimension in the MVL model and thereby avoids the large summation in the denominator of (3). The order of complexity of a likelihood contribution calculation reduces from 2^K to the size of D_i, which can be chosen considerably smaller than 2^K. Furthermore, an optimal choice of strata R and sampling probabilities q_r will not imply large efficiency losses. Nonetheless, small sampling probabilities q_r decreases computation time but increases efficiency loss. A Monte Carlo study has to shed light on the effect of the size of D_i on efficiency losses. In the remainder of this section, we introduce two alternative novel estimation methods.

Composite Conditional Likelihood

Given the structure of the MVL model, it is possible to use CCL (Lindsay, 1988) for parameter estimation. Where both the method by Russell and Petersen (2000) and the method proposed in the previous paragraph write the MVL model as a Multinomial Logit specification on a large outcome space, the CCL representation uses the conditional probabilities in (1) as separate, nonetheless correlated, choices. Hence, CCL avoids summation over the complete outcome space. It can be shown that the CCL approach provides consistent estimators at the cost of a loss in efficiency (Varin et al., 2011).

Following Molenberghs and Verbeke (2005, Chapter 12), the conditional probabilities in (1) lead to the composite log-likelihood function for the MVL model, that is,

$$\begin{aligned} \ell^c(\theta; y) &= \sum_{i=1}^{N} \ell^c(\theta; y_i) = \sum_{i=1}^{N} \sum_{k=1}^{K} \ell^c(\theta; y_{ik}) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \log \Pr[Y_{ik} = y_{ik}|y_{il} \qquad \text{for } l \neq k, X_i], \end{aligned} \tag{11}$$

where the superscript c stands for CCL. The estimator $\hat{\theta}$ which follows from maximizing (11) is consistent as $N \to \infty$ (Varin et al., 2011).

Varin et al. (2011) furthermore show that standard errors in CCL can be computed using the Godambe (1960) information matrix, which has a sandwich form and equals

$$G^c_{\hat{\theta}} = H^c_{\hat{\theta}} \left(J^c_{\hat{\theta}}\right)^{-1} H^c_{\hat{\theta}} \tag{12}$$

with

$$H_{\hat{\theta}}^{c} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\nabla\ell^{c}(\hat{\theta};y_{ik})\nabla\ell^{c\prime}(\hat{\theta};y_{ik}) \qquad \text{and} \qquad J_{\hat{\theta}}^{c} = \frac{1}{N}\sum_{i=1}^{N}\nabla\ell^{c}(\hat{\theta};y_{i})\nabla\ell^{c\prime}(\hat{\theta};y_{i}), \tag{13}$$

where $\nabla\ell^{c}(\hat{\theta};y_{ik})$ and $\nabla\ell^{c}(\hat{\theta};y_{i})$ denote the first derivatives of the corresponding log-likelihood contributions in (11). The covariance matrix of the parameter estimates then follows from $(-G_{\hat{\theta}}^{c})^{-1}$.

Although the CCL does not correspond to the correct likelihood function, it still takes dependencies in the MVL model into account. The advantage over the full multinomial representation in (3) is that CCL avoids the large summation in the denominator. The order of complexity for a likelihood contribution is further reduced to K because of the separation of conditional choices. It is therefore possible to compute CCL even when there is a large number of choices. Nonetheless, since the composite instead of the true likelihood function is used, the estimator is not efficient. A Monte Carlo study in Section 4 will however show a rather small and acceptable efficiency loss.

Generalized Method of Moments

The final estimation method, we consider for the MVL model is GMM (Hansen, 1982). To reduce the computation time, we base the moment conditions only on the conditional probabilities. Assuming exogeneity of the explanatory variables, the moment conditions

$$\begin{aligned}
&\mathbb{E}(Y_{ik} - \Pr[Y_{ik}=1|y_{il} \quad \text{for } l\neq k, X_i]) = 0 \qquad \forall\, k=1,\ldots,K,\\
&\mathbb{E}((Y_{ik} - \Pr[Y_{ik}=1|y_{il} \quad \text{for } l\neq k, X_i])X_i) = 0 \qquad \forall\, k=1,\ldots,K,\\
&\mathbb{E}((Y_{ik} - \Pr[Y_{ik}=1|y_{il} \quad \text{for } l\neq k, X_i])Y_{il}) = 0 \qquad \forall\, l\neq k
\end{aligned} \tag{14}$$

are valid to estimate the parameters in θ. We denote the sample analogue of these moment conditions for observation i by $m_i(\theta)$, which is a $(p+K)\times K$-dimensional vector.

The number of moment conditions equals $(p+K)\times K$. When $K>1$, the number of moment conditions exceeds the number of parameters in the model, and we use a two-step GMM approach (Cameron and Trivedi, 2005, Chapter 6). First, we estimate the parameters assigning equal weight to all moment conditions. In the second step, we optimally weigh the moment conditions according to the covariance matrix of the moment conditions to obtain the final parameter estimates. That is, in the second step we solve

$$\min_{\theta} M(\theta)'WM(\theta), \tag{15}$$

where $M(\theta)=\frac{1}{N}\sum_{i=1}^{N}m_i(\theta)$. The weighting matrix W is estimated as the matrix $\left(\frac{1}{N}\sum_{i=1}^{N}m_i(\theta)m_i(\theta)'\right)^{-1}$ evaluated at the first round estimate of θ, see, for example, Cameron and Trivedi (2005, Chapter 6.3).

The covariance matrix of the parameter estimates from GMM follows from

$$\left(H_{\hat{\theta}}^{g\prime}\left(J_{\hat{\theta}}^{g}\right)^{-1}H_{\hat{\theta}}^{g}\right)^{-1} \tag{16}$$

with $H_{\hat{\theta}}^{g}=\sum_{i=1}^{N}\nabla m_i(\hat{\theta})$ and $J_{\hat{\theta}}^{g}=\sum_{i=1}^{N}m_i(\hat{\theta})m_i'(\hat{\theta})$ where the superscript g stands for GMM.

The GMM approach uses conditional probabilities (1) instead of joint probabilities (3), and hence the large summation in the denominator of (3) is avoided. GMM therefore has the same computational advantages as the CCL approach. The order of complexity for a single observation equals the number of moment conditions. Hence, this is lower than 2^K if $K>4$ and p reasonably small. As the suggested GMM approach has more moment conditions than parameters, it is possible to use a standard test for overidentifying restrictions to test for the validity of the MVL model specification.

In sum, in this section we have proposed four parameter estimation methods for the MVL model. Since the standard ML method is computationally infeasible when the number of choices is large, we have proposed three novel estimation methods. In the next section, we compare these new estimation methods with the standard ML approach in a Monte Carlo study. We focus on small-sample bias, loss in efficiency and computation time for several numbers of correlated binary choices K and sample sizes N.

4. Monte Carlo study

In this section, we conduct a Monte Carlo study to investigate the properties of the four estimation methods described in the previous sections. First, we compare computation times of the four methods. Second, we examine small-sample bias and efficiency losses by looking at the average parameter estimates and the root mean squared error (RMSE) over the replications. Since the standard ML method uses the full information likelihood function, this method is expected to be most efficient. We compare the three alternative novel estimation methods to this method to analyze loss in efficiency. Finally, we check whether standard errors provided by the methods allow for valid inference in small samples.

For our Monte Carlo study, we consider the MVL specification in (3) and (4). The number of choices is either small ($K = 4$), medium ($K = 8$), or large ($K = 12$). We consider a relatively small sample size ($N = 500$) and a large sample ($N = 5{,}000$). As explanatory variables X_i we take two positively correlated random variables, one continuous and one discrete. Both variables are drawn from a bivariate normal distribution with variances 0.25 and correlation 0.75, and the second variable is made discrete based on a zero threshold. To avoid the need to consider many different Data Generating Processes (DGPs), the DGP parameters are chosen in such a way that different types of correlation structures occur within our set of K binary variables, see Tables 1–3 for the values of the DGP-parameters. For all K, positive and negative as well as large and small association parameters are used. Note that the size of the association parameters depends on K and thus differs over K. The GMM approach uses the discussed two-step estimator. For the stratified sampling approach we have to choose R and q_r. Since the sets of binary choices within a stratum should be comparable, we create strata of singles, pairs, triplets, etc. An intuitive choice for q_r is the relative fraction of stratum r in the data. We consider two alternatives: one where the size of subset D_i is $2^{K/2}$ and one where it is $2^{K/3}$.

All estimation methods are implemented in Matlab R2013a on a quad-core Intel Xeon 2.67Ghz processor with 8GB RAM. Before we discuss the results of the Monte Carlo study, we first focus on computation time. Table 4 displays the average computation time over 100 replications and $N = 1{,}000$ observations for different values of K, where we use the DGP from Tables 1–3. Since large summations in the denominator of (3) and small joint probabilities do not occur for small K, standard ML estimation is still computationally feasible. However, for larger K, differences in computation time grow rapidly. For instance, the computation time for standard ML when $K = 12$ is on average 25.6 minutes and

Table 1. Average parameter estimates and RMSE in a simulation study with 4 binary choices (5,000 replications)[a].

	DGP	ML		$SIS_{2^{K/2}}$		$SIS_{2^{K/3}}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−0.35	−0.358	0.230	−0.354	0.257	−0.365	0.298	−0.358	0.230	−0.381	0.239
β_2	−1	−1.018	0.277	−1.027	0.320	−1.037	0.364	−1.018	0.277	−0.990	0.274
	−0.5	−0.503	0.251	−0.508	0.286	−0.508	0.315	−0.504	0.252	−0.498	0.252
$\psi_{1,4}$	0.35	0.354	0.220	0.357	0.259	0.361	0.277	0.354	0.220	0.355	0.236
$\psi_{2,4}$	−0.9	−0.912	0.231	−0.926	0.260	−0.930	0.277	−0.913	0.231	−0.851	0.239
$\psi_{3,4}$	0.55	0.559	0.212	0.562	0.248	0.567	0.279	0.559	0.212	0.562	0.230
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−0.35	−0.350	0.071	−0.349	0.079	−0.351	0.091	−0.350	0.071	−0.353	0.071
β_2	−1	−1.003	0.085	−1.003	0.098	−1.003	0.108	−1.003	0.086	−0.998	0.085
	−0.5	−0.499	0.077	−0.500	0.088	−0.501	0.095	−0.499	0.077	−0.499	0.076
$\psi_{1,4}$	0.35	0.351	0.068	0.352	0.079	0.353	0.085	0.351	0.068	0.352	0.069
$\psi_{2,4}$	−0.9	−0.902	0.071	−0.904	0.081	−0.903	0.084	−0.902	0.071	−0.894	0.070
$\psi_{3,4}$	0.55	0.551	0.067	0.552	0.078	0.553	0.086	0.551	0.067	0.551	0.069

[a]To save space, we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 2. Average parameter estimates and RMSE in a simulation study with 8 binary choices (5,000 replications)[a].

	DGP	ML		$SIS_{2K/2}$		$SIS_{2K/3}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−0.95	−0.972	0.269	−0.974	0.286	−0.973	0.316	−0.972	0.270	−1.014	0.287
β_3	−1	−1.024	0.330	−1.032	0.352	−1.050	0.393	−1.026	0.333	−0.986	0.331
	−0.5	−0.511	0.295	−0.517	0.310	−0.521	0.345	−0.512	0.296	−0.504	0.299
$\psi_{1,8}$	0	−0.009	0.262	−0.008	0.275	−0.011	0.299	−0.009	0.263	0.003	0.271
$\psi_{2,7}$	0.15	0.146	0.257	0.148	0.269	0.151	0.294	0.146	0.257	0.152	0.266
$\psi_{3,5}$	−0.9	−0.928	0.296	−0.936	0.309	−0.959	0.331	−0.931	0.297	−0.824	0.302
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−0.95	−0.949	0.082	−0.949	0.087	−0.949	0.096	−0.949	0.082	−0.954	0.084
β_3	−1	−1.003	0.099	−1.004	0.105	−1.005	0.115	−1.003	0.099	−0.994	0.100
	−0.5	−0.501	0.090	−0.502	0.093	−0.503	0.103	−0.501	0.090	−0.499	0.090
$\psi_{1,8}$	0	−0.001	0.080	−0.001	0.084	−0.001	0.090	−0.001	0.080	0.002	0.082
$\psi_{2,7}$	0.15	0.149	0.079	0.149	0.082	0.148	0.087	0.149	0.079	0.150	0.080
$\psi_{3,5}$	−0.9	−0.905	0.092	−0.906	0.094	−0.908	0.101	−0.905	0.092	−0.875	0.097

[a]To save space, we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 3. Average parameter estimates and RMSE in a simulation study with 12 binary choices (5,000 replications)[a].

	DGP	ML[b]		$SIS_{2K/2}$[b]		$SIS_{2K/3}$		CCL		GMM	
$N = 500$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−1.55	–	–	–	–	−1.602	0.368	−1.591	0.314	−1.645	0.347
β_4	−1	–	–	–	–	−1.074	0.451	−1.040	0.386	−0.995	0.390
	−0.5	–	–	–	–	−0.525	0.401	−0.508	0.340	−0.518	0.352
$\psi_{3,12}$	−0.35	–	–	–	–	−0.405	0.432	−0.390	0.397	−0.346	0.395
$\psi_{5,10}$	0.15	–	–	–	–	0.136	0.398	0.133	0.368	0.114	0.371
$\psi_{7,8}$	0.55	–	–	–	–	0.570	0.390	0.554	0.349	0.486	0.374
$N = 5000$	θ	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse	$\hat{\theta}$	rmse
α_1	−1.55	–	–	–	–	−1.558	0.106	−1.555	0.094	−1.561	0.097
β_4	−1	–	–	–	–	−1.007	0.128	−1.005	0.116	−0.993	0.115
	−0.5	–	–	–	–	−0.503	0.117	−0.502	0.103	−0.505	0.103
$\psi_{1,4}$	−0.35	–	–	–	–	−0.355	0.121	−0.352	0.116	−0.341	0.116
$\psi_{2,4}$	0.15	–	–	–	–	0.151	0.113	0.150	0.107	0.139	0.109
$\psi_{3,4}$	0.55	–	–	–	–	0.548	0.111	0.547	0.103	0.519	0.110

[a]To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.
[b]As estimation for *ML* and $SIS_{2K/2}$ take too long (see Table 4), we do not include them in the 5,000 replications simulation.

Table 4. Average computation time over 100 replications (1,000 observations)[a].

	Estimation method				
Number of choices K	ML	$SIS_{2K/2}$	$SIS_{2K/3}$	CCL	GMM
4	0.79	1.02	0.89	0.25	1.22
8	37.33	15.89	8.17	1.66	7.25
12	1538.94	200.76	70.94	5.57	33.73

[a]In seconds in Matlab R2013a on a Quad-Core Intel Xeon 2.67Ghz processor (8GB RAM) running Windows 7 64 bits.

the other three methods have a clear advantage. The computation time of CCL is more than 275 times faster (only 5.6 seconds). These computation times are in line with the (objective) order of complexity presented in Section 3. If the small-sample bias and losses in efficiency are both small, the alternative estimation methods are sound alternatives for parameter estimation in the large MNL specification with large K. Note that the difference in computation time will further increase if we include more explanatory variables in the model or consider even larger K.

Table 5. Empirical size of the distribution of the four estimators of the MVL model with 4 binary choices (5,000 observations, 5,000 replications)[a].

		Percentiles					
Model	Theoretical	0.025	0.05	0.1	0.9	0.95	0.975
ML	α_1	0.026	0.052	0.099	0.896	0.949	0.977
	β_2	0.025	0.048	0.098	0.894	0.947	0.972
		0.024	0.048	0.097	0.902	0.950	0.975
	$\psi_{1,4}$	0.024	0.050	0.099	0.901	0.949	0.976
	$\psi_{2,4}$	0.023	0.047	0.097	0.896	0.946	0.972
	$\psi_{3,4}$	0.026	0.052	0.099	0.898	0.949	0.977
$SIS_{2^{K/2}}$	α_1	0.028	0.051	0.100	0.897	0.949	0.975
	β_2	0.024	0.049	0.096	0.898	0.947	0.972
		0.024	0.049	0.098	0.898	0.949	0.975
	$\psi_{1,4}$	0.027	0.051	0.103	0.900	0.953	0.975
	$\psi_{2,4}$	0.023	0.046	0.096	0.892	0.944	0.972
	$\psi_{3,4}$	0.025	0.050	0.100	0.900	0.949	0.976
$SIS_{2^{K/3}}$	α_1	0.026	0.051	0.098	0.896	0.948	0.974
	β_2	0.022	0.049	0.099	0.899	0.948	0.975
		0.025	0.047	0.096	0.906	0.952	0.977
	$\psi_{1,4}$	0.024	0.049	0.097	0.899	0.949	0.975
	$\psi_{2,4}$	0.025	0.050	0.101	0.898	0.948	0.973
	$\psi_{3,4}$	0.027	0.049	0.101	0.895	0.946	0.975
CCL	α_1	0.027	0.052	0.099	0.896	0.948	0.977
	β_2	0.025	0.049	0.098	0.893	0.946	0.972
		0.025	0.048	0.098	0.903	0.950	0.975
	$\psi_{1,4}$	0.025	0.050	0.099	0.900	0.949	0.974
	$\psi_{2,4}$	0.023	0.048	0.099	0.895	0.945	0.972
	$\psi_{3,4}$	0.025	0.053	0.099	0.898	0.949	0.977
GMM	α_1	0.029	0.057	0.106	0.888	0.943	0.972
	β_2	0.027	0.053	0.105	0.889	0.942	0.970
		0.027	0.050	0.100	0.903	0.950	0.973
	$\psi_{1,4}$	0.032	0.062	0.111	0.888	0.940	0.969
	$\psi_{2,4}$	0.033	0.063	0.116	0.881	0.933	0.965
	$\psi_{3,4}$	0.032	0.061	0.111	0.885	0.940	0.970

[a]To save space, we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Tables 1–3 display the average and RMSE of the estimators over 5,000 replications. Since results are highly comparable and to save space, a diverse selection of parameters from the DGPs is displayed.[1] The DGP with $N = 5{,}000$ shows that the bias is quite small for all estimation methods. For small sample sizes, the deviation of the parameter estimates from the DGP values is larger. Nonetheless, all methods find comparably accurate estimates. Our newly introduced estimation methods thus are capable of finding estimates comparable to the regular likelihood approach.

To further analyze the loss in efficiency between the three novel estimation methods and standard ML, we consider best and worst cases of the RMSEs across all parameters. As expected, standard ML is most efficient. The subset approach used in SIS causes a loss of information and thereby an increase in RMSE. Obviously, the smaller the subset, the larger the loss in efficiency. In the best and worst case, the RMSE of ML and SIS with a subset D of size $2^{K/2}$ differ 3.7% and 7.0%, respectively. The smaller subset of size $2^{K/3}$ yields efficiency losses between 12.0% and 20.4%. For CCL and GMM, only small efficiency losses occur. The differences of GMM with ML in terms of RMSE are between 0.02% and 7.3%. These differences are smallest for the parameters of the covariates. For CCL, the minimum and maximum differences are only 0.1% and 0.9%, respectively.

In practice one usually opts for the most efficient approach. However, the estimation method should also be computationally feasible such that parameter estimates can be obtained in a reasonable amount of time. The large summation over all possible alternatives in the standard ML method may lead

[1]The results for the other parameters are similar and available upon request.

to numerical problems and long computation times for large K. CCL and GMM seem to be useful alternatives for standard ML and produce useful parameter estimates in little time. The small-sample bias is similar, and the loss in efficiency is rather small. For SIS, there is a clear tradeoff between the size of the subset and the loss in efficiency.

Apart from bias and efficiency, we also consider the validity of the standard errors with respect to significance testing of the model parameters. Tables 5–7 display the empirical size of the t-test for $N = 5{,}000$ for both tails of the t-statistic. The table shows that size distortions are rather small. The largest size distortions are found for the GMM approach. For example, a theoretical 90% confidence interval for $\psi_{3,12}$ in GMM turns out to have a coverage of 84.2%. This size distortion is still acceptable. For the other approaches, the size distortions are smaller. The same coverage probability is 89.9% for the CCL approach. Unreported results show that even for small N size distortions of ML, SIS and CCL are still negligible. Hence, hypothesis tests can be carried out in the usual manner for these estimation methods. In accordance with existing literature (Altonji and Segal, 1996), size distortion for the GMM approach are larger in small samples.

In sum, the Monte Carlo study shows that the novel estimation methods are sound alternatives for the regular likelihood approach. Where computation times in standard ML increase exponentially over the number of choices, the computation time stays limited using CCL, GMM, or SIS. Further, small-sample biases are comparable to full ML, and efficiency losses are rather small and acceptable. Given the win in computation time, the avoidance of numerical problems, small small-sample biases, and negligible losses in efficiency, CCL is the most promising alternative estimation method.

Table 6. Empirical size of the distribution of the four estimators of the MVL model with 8 binary choices (5,000 observations, 5,000 replications)[a].

		Percentiles					
Model	Theoretical	0.025	0.05	0.1	0.9	0.95	0.975
ML	α_1	0.022	0.048	0.098	0.900	0.948	0.972
	β_3	0.021	0.044	0.099	0.899	0.949	0.978
		0.026	0.051	0.101	0.899	0.954	0.977
	$\psi_{1,8}$	0.025	0.048	0.096	0.901	0.952	0.976
	$\psi_{2,7}$	0.025	0.052	0.104	0.891	0.947	0.975
	$\psi_{3,5}$	0.022	0.048	0.100	0.898	0.944	0.974
$SIS_{2K/2}$	α_1	0.027	0.052	0.102	0.900	0.949	0.975
	β_3	0.023	0.047	0.099	0.900	0.948	0.976
		0.026	0.050	0.102	0.892	0.950	0.976
	$\psi_{1,8}$	0.027	0.053	0.096	0.899	0.952	0.978
	$\psi_{2,7}$	0.025	0.056	0.103	0.894	0.948	0.975
	$\psi_{3,5}$	0.025	0.047	0.093	0.893	0.945	0.974
$SIS_{2K/3}$	α_1	0.023	0.050	0.105	0.902	0.948	0.976
	β_3	0.022	0.045	0.098	0.897	0.948	0.973
		0.027	0.050	0.100	0.900	0.954	0.979
	$\psi_{1,8}$	0.026	0.047	0.098	0.899	0.951	0.977
	$\psi_{2,7}$	0.023	0.049	0.099	0.898	0.947	0.975
	$\psi_{3,5}$	0.025	0.049	0.098	0.890	0.946	0.974
CCL	α_1	0.023	0.048	0.100	0.900	0.948	0.972
	β_3	0.022	0.044	0.100	0.898	0.949	0.976
		0.026	0.051	0.103	0.899	0.952	0.977
	$\psi_{1,8}$	0.026	0.049	0.099	0.896	0.951	0.974
	$\psi_{2,7}$	0.027	0.054	0.105	0.888	0.945	0.975
	$\psi_{3,5}$	0.024	0.049	0.100	0.897	0.942	0.970
GMM	α_1	0.029	0.057	0.109	0.887	0.941	0.967
	β_3	0.028	0.054	0.107	0.886	0.941	0.970
		0.029	0.055	0.105	0.892	0.949	0.976
	$\psi_{1,8}$	0.035	0.060	0.117	0.874	0.931	0.961
	$\psi_{2,7}$	0.039	0.069	0.119	0.868	0.931	0.963
	$\psi_{3,5}$	0.034	0.064	0.121	0.873	0.930	0.958

[a]To save space we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

Table 7. Empirical size of the distribution of the four estimators of the MVL model with 12 binary choices (5,000 observations, 5,000 replications)[a].

Model	Theoretical	Percentiles 0.025	0.05	0.1	0.9	0.95	0.975
$SIS_{2K/3}$	α_1	0.025	0.048	0.093	0.900	0.950	0.975
	β_4	0.023	0.051	0.098	0.903	0.957	0.977
		0.023	0.044	0.095	0.898	0.949	0.975
	$\psi_{3,12}$	0.024	0.046	0.093	0.902	0.949	0.975
	$\psi_{5,10}$	0.021	0.046	0.094	0.901	0.953	0.977
	$\psi_{7,8}$	0.024	0.042	0.094	0.904	0.947	0.974
CCL	α_1	0.025	0.050	0.095	0.894	0.948	0.974
	β_4	0.024	0.051	0.106	0.894	0.946	0.975
		0.024	0.048	0.097	0.902	0.949	0.971
	$\psi_{3,12}$	0.024	0.048	0.098	0.891	0.947	0.974
	$\psi_{5,10}$	0.023	0.049	0.101	0.895	0.948	0.974
	$\psi_{7,8}$	0.025	0.050	0.098	0.898	0.950	0.972
GMM	α_1	0.036	0.066	0.119	0.876	0.935	0.965
	β_4	0.030	0.065	0.120	0.882	0.938	0.967
		0.028	0.055	0.102	0.892	0.943	0.968
	$\psi_{3,12}$	0.044	0.076	0.127	0.862	0.918	0.953
	$\psi_{5,10}$	0.043	0.069	0.129	0.862	0.920	0.954
	$\psi_{7,8}$	0.045	0.072	0.125	0.870	0.925	0.954

[a]To save space, we only report results of six parameters selected from a wide range of the parameter space. The results for the other parameters are similar and available upon request.

5. Application

In this section we illustrate the use of an MVL model with many choices. We consider survey data of 2,046 individuals on store visits in a particular Dutch specialized shopping mall. Visits to different stores are likely to be correlated, and hence, it is convenient to model these simultaneous decisions using a MVL specification. In this application, we consider simultaneous choices for ten different stores. All stores fall under the general theme of home decoration and do-it-yourself. Table 8 details the types of stores. Our dependent variable can take $2^{10} = 1{,}024$ different values. As explanatory variables, we have *Family size*, *Age*, *Gender*, *Income*, *Number of visits*, and *Appreciation* of the shopping mall.

The simulation study in Section 4 showed that for this size of the outcome space, large differences in computation time occur. Hence, one may not be willing to use standard ML estimation. Based on the simulation results, we consider the CCL approach (fast and accurate) to estimate the model parameters[2]. As benchmark, we will also consider the standard ML approach. The standard ML approach takes about 1.6 hours on a dual-core Intel 3.4Ghz processor with 4GB RAM, which shows that this method is not very convenient if you want to investigate several model specifications. The CCL approach on the other hand only takes 2.3 minutes.

First, we test for independence among the choices for store visits. The *LR*-statistic in the ML approach for the restriction that all $\psi = 0$ is 1,373.4 (45 degrees of freedom). This statistic clearly shows that independence is rejected. Hence, we find evidence for correlations between visiting the different store types and the MVL model from Section 2 thus is applicable to the data. An adjusted *LR*-test for CCL (Varin et al., 2011) yields the same conclusion.

Tables 8–11 display the parameter estimates and standard errors for the two estimation methods. The parameter estimates are very similar and both methods find the same parameter estimates to be significantly different from 0. The standard errors in the CCL approach are slightly smaller than in the standard ML estimation approach, but this may be due to the relatively small sample size. Unreported results show that the GMM and SIS approaches also provide similar results. The results of SIS indicate that subset D_i should be large to get results close to standard ML.

[2]The results of the other two approaches are available upon request.

Table 8. Parameter estimates and standard errors of the MVL model using the standard ML method for shopping mall data.

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen		Other	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
intercept	−1.373	0.241	−1.649	0.245	−0.699	0.253	−1.339	0.256	−1.854	0.266	−2.196	0.248	−2.547	0.288	−1.459	0.243	−3.042	0.350	−2.199	0.437
family size	0.163	0.043	−0.060	0.043	0.059	0.047	−0.063	0.047	0.028	0.045	0.040	0.042	−0.019	0.048	−0.012	0.043	−0.160	0.061	0.078	0.075
young (<35)	−0.064	0.118	0.101	0.121	−0.035	0.130	−0.165	0.129	0.213	0.125	0.237	0.118	0.143	0.130	−0.099	0.121	−0.061	0.165	0.442	0.214
old (>54)	−0.362	0.166	−0.095	0.156	−0.222	0.170	−0.039	0.175	−0.323	0.186	−0.185	0.163	0.280	0.173	0.041	0.162	−0.190	0.234	−0.007	0.332
female	0.086	0.106	0.128	0.105	−0.086	0.117	−0.136	0.114	−0.193	0.116	−0.126	0.106	0.049	0.118	0.165	0.107	−0.001	0.150	0.121	0.214
low income ($<$ *modal*)	0.074	0.125	0.094	0.123	−0.100	0.133	−0.097	0.129	−0.025	0.135	0.143	0.124	0.034	0.136	0.098	0.127	−0.228	0.181	0.167	0.240
high income ($>$ *modal*)	−0.219	0.170	−0.144	0.173	0.021	0.188	0.598	0.188	0.029	0.183	−0.095	0.171	−0.225	0.194	−0.448	0.175	−0.293	0.250	0.401	0.295
# visits	0.003	0.003	0.023	0.003	0.018	0.004	0.030	0.004	−0.001	0.003	0.007	0.003	0.005	0.003	0.002	0.003	0.004	0.003	0.009	0.006
positive apprec.	−0.113	0.153	0.098	0.152	0.239	0.165	0.066	0.162	0.495	0.177	0.808	0.163	0.359	0.190	0.363	0.154	−0.102	0.234	0.292	0.306
negative apprec.	−0.142	0.216	−0.123	0.214	0.031	0.228	−0.065	0.225	−0.109	0.263	0.231	0.233	0.152	0.261	−0.106	0.223	0.000	0.330	0.281	0.386

Table 9. Estimates of the association parameters and standard errors of the MVL model using the standard ML method for shopping mall data.

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
Paint/Wallpaper	0.101	0.111																
Materials	−0.367	0.121	0.584	0.117														
Hardware	−0.162	0.121	0.484	0.115	1.739	0.116												
Furniture	0.656	0.115	−0.268	0.126	−0.413	0.130	−0.109	0.131										
Lamps	0.505	0.108	0.635	0.108	−0.357	0.119	0.366	0.118	0.435	0.120								
Garden	−0.070	0.127	0.274	0.123	0.341	0.141	0.529	0.134	0.075	0.137	0.708	0.123						
Curtain/Carpet	0.783	0.106	0.794	0.111	−0.146	0.120	−0.377	0.120	0.782	0.118	0.468	0.108	−0.183	0.130				
Kitchen	0.451	0.163	0.201	0.169	0.873	0.201	0.406	0.184	0.876	0.169	0.139	0.165	0.538	0.154	0.171	0.173		
Other	−0.200	0.235	−0.704	0.278	−0.890	0.264	−0.611	0.289	−0.466	0.271	−0.333	0.251	−0.128	0.277	−0.449	0.249	0.423	0.345

Table 10. Parameter estimates and standard errors of the MVL model using the CCL method for shopping mall data.

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen		Other	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
intercept	−1.373	0.234	−1.647	0.237	−0.697	0.244	−1.337	0.248	−1.854	0.258	−2.195	0.241	−2.547	0.281	−1.459	0.237	−3.042	0.351	−2.199	0.434
family size	0.162	0.041	−0.055	0.042	0.066	0.045	−0.055	0.045	0.027	0.043	0.041	0.041	−0.018	0.046	−0.013	0.041	−0.161	0.060	0.078	0.074
young ($<$ 35)	−0.065	0.114	0.102	0.117	−0.034	0.128	−0.164	0.125	0.213	0.121	0.237	0.114	0.143	0.126	−0.099	0.117	−0.061	0.162	0.442	0.209
old ($>$ 54)	−0.362	0.160	−0.094	0.150	−0.222	0.164	−0.039	0.166	−0.323	0.180	−0.185	0.158	0.280	0.166	0.041	0.155	−0.190	0.228	−0.007	0.308
female	0.086	0.103	0.129	0.102	−0.085	0.113	−0.135	0.110	−0.193	0.112	−0.125	0.103	0.049	0.114	0.165	0.104	−0.001	0.145	0.121	0.208
low income ($<$ *modal*)	0.074	0.120	0.094	0.119	−0.100	0.129	−0.097	0.126	−0.025	0.130	0.143	0.119	0.034	0.131	0.098	0.121	−0.228	0.176	0.167	0.229
high income ($>$ *modal*)	−0.219	0.163	−0.144	0.167	0.021	0.183	0.599	0.180	0.029	0.177	−0.095	0.165	−0.225	0.188	−0.448	0.168	−0.293	0.242	0.401	0.288
# visits	0.003	0.003	0.021	0.003	0.015	0.003	0.026	0.004	−0.001	0.003	0.007	0.003	0.005	0.002	0.002	0.002	0.004	0.003	0.008	0.006
positive apprec.	−0.113	0.150	0.099	0.147	0.241	0.159	0.068	0.158	0.495	0.169	0.808	0.158	0.359	0.183	0.363	0.150	−0.103	0.224	0.292	0.303
negative apprec.	−0.142	0.210	−0.122	0.208	0.031	0.220	−0.064	0.218	−0.110	0.254	0.231	0.226	0.152	0.253	−0.106	0.218	0.000	0.310	0.281	0.381

Table 11. Estimates of the association parameters and standard errors of the MVL model using the CCL method for shopping mall data.

	Bed		Paint/Wallpaper		Materials		Hardware		Furniture		Lamps		Garden		Curtain/Carpet		Kitchen	
	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
Paint/Wallpaper	0.102	0.107																
Materials	−0.366	0.114	0.586	0.114														
Hardware	−0.162	0.114	0.486	0.111	1.741	0.110												
Furniture	0.656	0.110	−0.268	0.121	−0.413	0.125	−0.109	0.126										
Lamps	0.505	0.105	0.636	0.105	−0.356	0.115	0.367	0.114	0.435	0.115								
Garden	−0.070	0.121	0.275	0.119	0.342	0.134	0.529	0.129	0.075	0.131	0.708	0.118						
Curtain/Carpet	0.782	0.103	0.795	0.107	−0.145	0.116	−0.376	0.116	0.781	0.113	0.468	0.105	−0.184	0.123				
Kitchen	0.450	0.152	0.200	0.158	0.872	0.189	0.405	0.171	0.875	0.155	0.138	0.157	0.537	0.149	0.170	0.158		
Other	−0.201	0.209	−0.704	0.244	−0.890	0.230	−0.611	0.252	−0.466	0.243	−0.333	0.230	−0.128	0.257	−0.450	0.227	0.423	0.305

The negative estimates of the choice-specific intercepts in Tables 8 and 10 show that most stores are visited only by a minority of the individuals. The order of the intercepts shows that stores selling kitchens are visited least, where stores selling building materials are visited by the most individuals.

Several relations between the explanatory variables and store visits are found. For example, the more frequent visitors of the mall visit more stores selling paint/wallpaper, building materials, and hardware. These can be seen as the fanatic handymen. Furthermore, visitors who very much appreciate the mall are more likely to also buy their furniture, lamps, and floor and wall decorations at this shopping mall.

The association parameters in Table 11 show the relations between the visits to different stores. Clear interpretations can be given. For example, individuals who visit a store selling an odd jobs article (paint/wallpaper, building materials, or hardware) are likely also to visit other odd jobs stores. The same holds for stores selling lamps, curtains/carpets, and furniture since the corresponding association parameters are positive. Negative and significant association parameters are for instance found for the combination hardware and curtains/carpets. Apparently, individuals seem to be unlikely to visit both these store types in this shopping mall.

In sum, the MVL model gives understandable and interpretable parameter estimates for the data of store visits in a Dutch shopping mall. Furthermore, the standard ML and CCL approaches yield very similar estimation results and conclusions. The clear advantage of the CCL approach is the time it takes to obtain consistent parameter estimates with small loss in efficiency. The reduction in computation time is large, and with the CCL method it becomes feasible to easily consider several model specifications. In case the number of stores had been larger, ML estimation would have broken down, while CCL could still be used.

6. Conclusion

The MVL model is used to model correlated simultaneous binary choices. In this article, we propose three novel estimation methods for this model: estimation by (i) SIS; (ii) CCL; and by (iii) GMM. The new estimation methods are especially of interest when the dimension of the choice problem is large. Methods available in the literature go together with a large computational burden. The new methods in this article circumvent this problem.

Results from a Monte Carlo study show that the new estimation methods yield comparable small-sample biases as the standard (full information) ML approach as proposed by Russell and Petersen (2000). Furthermore, efficiency losses compared to the full likelihood approach are rather small. Because of these findings, the decrease in computation time and avoidance of numerical problems are clear advantages of our proposed estimation methods. The CCL approach turns out to have the largest decrease in computation time, leads to a very small loss in efficiency, and provides accurate standard errors.

In an application, we applied the methods to store visits in a shopping mall. Multivariate binary choice data occur widely in practice. Hence, other applications in different fields of research can be given. Since the dimension of the choice problem will often be large, our methods are highly useful in applied research.

Several extensions to the current research are possible. For instance, a Conditional Logit specification can easily be derived. Furthermore, the association parameters can also depend on exogenous variables or be individual-specific (in panel data models). Finally, instead of binary choices, this model can be extended to a multivariate multinomial specification. The feasible estimation methods proposed in this article can be used in all these cases. Especially CCL is applicable to extensions of the current model specification if a clear composition of the conditional probabilities can be given.

Acknowledgment

We would like to thank Bert de Bruijn and Andreas Pick for their useful suggestions to earlier versions of this manuscript. We further thank the anonymous reviewers and (associate) editor of *Econometric Reviews* for their cooperative and valuable comments.

References

Altonji, J. G., Segal, L. M. (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business and Economic Statistics* 14:353–366.

Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Basil Blackwell.

Ashford, J. R., Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics* 26:535–546.

Ben-Akiva, M., Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: The MIT Press.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 36:192–236.

Cameron, A., Trivedi, P. (2005). *Microeconometrics Methods and Applications*. New York: Cambridge University Press.

Cappellari, L., Jenkins, S. P. (2006). Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal* 6:156–189.

Carey, V., Zeger, S. L., Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* 80:517–526.

Chib, S., Seetharaman, P., Strijnev, A. (2002). Analysis of multi-category purchase incidence decisions using IRI market basket data. *Econometric Models in Marketing* 16:57–92.

Cox, D. R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 21:113–120.

Donkers, B., Verhoef, P., Jong, M. (2007). Modeling CLV: A test of competing models in the insurance industry. *Quantitative Marketing and Economics* 5:163–190.

Edwards, Y. D., Allenby, G. M. (2003). Multivariate analysis of multiple response data. *Journal of Marketing Research* 40:321–334.

Feddag, M.-L. (2013). Composite likelihood estimation for multivariate probit latent traits models. *Communications in Statistics — Theory and Methods* 42:2551–2566.

Glonek, G. F. V., McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society. Series B (Methodological)* 57:533–546.

Godambe, V. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31:1208–1211.

Guimarães, P., Figueirdo, O., Woodward, D. (2003). A tractable approach to the firm location decision problem. *The Review of Economics and Statistics* 85:201–204.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1054.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics* 80:221–239.

McFadden, D. (1978). *Modelling the Choice of Residential Location*. Berkeley, California. Institute of Trartation Studies, University of California.

Molenberghs, G., Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Russell, G. J., Petersen, A. (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing* 76:367–392.

Varin, C., Reid, N., Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* 21:5–42.