A Multivariate Model for Multinomial Choices

Koen Bel[a*]   Richard Paap[a]

[a]*Econometric Institute*

Erasmus School of Economics

Erasmus University Rotterdam

13th October 2014

Abstract

Multinomial choices of individuals are likely to be correlated. Nonetheless, econometric models for this phenomenon are scarce. A problem of multivariate multinomial choice models is that the number of potential outcomes can become very large which makes parameter interpretation and inference difficult. We propose a novel Multivariate Multinomial Logit specification, where (i) the number of parameters stays limited; (ii) there is a clear interpretation of the parameters in terms of odds ratios; (iii) zero restrictions on parameters result in independence between the multinomial choices and; (iv) parameter inference is feasible using a composite likelihood approach even if the multivariate dimension is large. Finally, these nice properties are also valid in a fixed-effects panel version of the model.

Keywords: Discrete choices, Multivariate analysis, Multinomial Logit, Composite Likelihood
JEL-codes: C01, C35, C51

*Corresponding author: Econometric Institute (H11-2), Erasmus School of Economics, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. `email: bel@ese.eur.nl`, `phone: (+31) 10 4088902`

1   Introduction

It is common practice in applied research to use Multinomial Logit [MNL] models to describe multinomial choice data (McFadden, 1983, Chapter 24). These MNL models are suited to describe single multinomial choices. In practice we are often dealing with multiple correlated multinomial decisions. Answers to survey questions with two or more choice possibilities are likely to be correlated. The choice for job location may be correlated with residence choice. In corporate finance one may want to model simultaneously the strategy to takeover another company and the ways to finance this takeover. In marketing one may be interested in dependencies in brand choices for several product categories. Hence, simultaneous multinomial decisions occur in different areas of research.

In this paper we propose a relatively straightforward model to describe simultaneous multinomial decisions. As far as we know there are hardly any models available to model correlated multinomial decisions, see de Rooij & Kroonenberg (2003) for a similar conclusion. An obvious way to model simultaneous multinomial decisions is to use a correlated Multinomial Probit [MNP] approach, see Hausman & Wise (1978). Parameter estimation of such models implies solving high-dimensional integrals using numerical integration or simulation methods. Given the computational burden in univariate MNP models (Geweke et al., 1994, 1997), frequentist inference in a multivariate MNP model is unlikely to be feasible. Another option is to use mixed Logit models (Hensher & Greene, 2003) and let unobserved heterogeneity capture correlation among decisions. Again, computation of the choice probabilities implies solving integrals which becomes infeasible when the number of simultaneous decisions is already moderately large. A Nested Logit specification (Maddala, 1983, Chapter 3) is perhaps a more feasible approach. However, this model handles the data as if decisions are made sequentially, which is often not the case in practice. Finally, one may consider an MNL model for all possible combinations of the multinomial variables. The number of choice combinations however becomes easily large, see also Amemiya (1978) and Ben-Akiva & Lerman (1985, chapter 10). Clearly, the number of parameters and model interpretation get out of hand. Furthermore, parameter estimation becomes infeasible as the computation of choice probabilities requires summation over all potential outcomes.

As far as we know the multivariate MNL model of de Rooij & Kroonenberg (2003) is the only recent contribution to the simultaneous multinomial choice modelling literature. Their model is however specially designed for the problem and data set at hand and

cannot be applied in general. The work in Burda et al. (2008) is related although in their paper individuals make numerous choices on the same attribute. We focus on numerous separated multinomial choices. To fill the gap in the literature, we propose a general and novel Multivariate Multinomial Logit [MV-MNL] specification to describe simultaneous multinomial decisions. In essence, we extend the Multivariate (binary) Logit [MVL] model of Cox (1972) and Russell & Petersen (2000) to multivariate multinomial decisions. The advantages of this multivariate multinomial model specification are that (i) the number of parameters stays limited; (ii) there is a clear interpretation of the model parameters in terms of odds ratios and (iii) zero restrictions on a subset of parameters result in independence between the multinomial choices.

The model is related to the multivariate MNL specification of Amemiya (1978) and Ben-Akiva & Lerman (1985, chapter 10) but in contrast to these specifications we explicitly focus on the dependence structure in the multinomial choices. Furthermore, our proposed MV-MNL specification allows for an easy and computationally feasible parameter estimation method. Due to its special structure we can avoid the summation over all potential combinations of the multivariate multinomial choices by considering conditional probabilities in the estimation approach. Parameter estimates are obtained from a Composite Likelihood function (Lindsay, 1988) containing conditional probabilities, see Bel et al. (2014) for a similar approach in MVL models. Hence, the Composite Likelihood method avoids the computation of the joint probabilities over all possible combinations. Finally, the novel multivariate MNL specification can easily be extended to a fixed-effects specification for panel data. Parameter estimation stays feasible by using sufficient statistics in combination with the composite likelihood approach.

The remainder of this paper is organized as follows. In Section 2, we introduce the new MV-MNL specification. We also discuss parameter identification, interpretation and parameter inference. A small Monte Carlo study shows the accuracy of the parameter estimates and a small loss in efficiency due to the use of the composite instead of the true likelihood. An extension to panel data is discussed in Section 3. Section 4 provides two illustrations of the use of MV-MNL models. The first illustration concerns a cross-sectional survey on satisfaction about life and the second illustration deals with the choice for tuna using a household panel scanner data set. Finally, Section 5 concludes.

2 Model Specification

In this section we discuss the model specification for the Multivariate Multinomial Logit model. This model is an extension of the Multivariate Logit model introduced by Cox (1972) and Russell & Petersen (2000). To start the discussion, we first consider briefly this MVL specification in Section 2.1. The extension to a multinomial specification is proposed in Section 2.2. We discuss model specification, parameter identification and interpretation of the model parameters. Finally, Section 2.3 shows the model representation for the choice probabilities in a simple bivariate trinomial Logit model to clarify the structure of the model.

2.1 A Multivariate Binomial Logit Model

First, we consider the Multivariate Logit model to describe correlated binary decisions following the ideas in Russell & Petersen (2000). Let Y_i denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \ldots, N$, defined as

$$Y_i \quad = \quad \{Y_{i1}, \ldots, Y_{iK}\}, \tag{1}$$

where Y_{ik} describes the k-th binary choice for individual i for $k = 1, \ldots, K$. Note that there are 2^K possible realizations of the random variable Y_i. The set of possible realizations is called S.

The K choices in Y_i may be correlated. The starting point for modeling these dependencies is the conditional probabilities for each choice decision k given all choice decisions $l \neq k$, see Russell & Petersen (2000). These conditional probabilities are a Logit function of the individual characteristics X_i, the model parameters α, β and ψ and the other choices y_{il}, that is

$$\Pr[Y_{ik} = 1 | y_{il} \text{ for } l \neq k, X_i] \quad = \quad \frac{\exp(Z_{ik})}{1 + \exp(Z_{ik})} \tag{2}$$

with

$$Z_{ik} \quad = \quad \alpha_k + X_i \beta_k + \sum_{l \neq k} y_{il} \psi_{kl}, \tag{3}$$

where y_{il} are the actual realizations of Y_{il}, X_i is a vector of explanatory variables with corresponding parameter vector β_k, α_k are alternative-specific intercepts, and where ψ_{kl} are association parameters for $l \neq k$. Hence, the correlation between Y_{ik} and Y_{il} is captured

by the association parameter. Association means the relative change in the exponent Z_{ik} if choices k and l move together compared to being opposite. When $\psi_{kl} > 0$ this implies positive association and when $\psi_{kl} < 0$ we have negative association. For $\psi_{kl} = 0$ we have independence between Y_{ik} and Y_{il}. As we can only describe correlations, we have to impose $\psi_{kl} = \psi_{lk}$ for symmetry.

The theorem of Besag (1974) states that all properties of the joint distribution follow from the full set of conditional distributions. Russell & Petersen (2000) use this result to show that the conditional distributions in (2) imply the following Multinomial Logit model for the joint distribution of Y_i:

$$\Pr[Y_i = y_i | X_i] \quad = \quad \frac{\exp(\mu_{y_i})}{\sum_{s_i \in S} \exp(\mu_{s_i})}, \tag{4}$$

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

$$\mu_{y_i} \quad = \quad \sum_{k=1}^{K} y_{ik}(\alpha_k + X_i \beta_k) + \sum_{l>k} y_{ik} y_{il} \psi_{kl}. \tag{5}$$

Hence, the parameters α_k and β_k only occur if the corresponding choice equals 1. Furthermore, the association parameter ψ_{kl} only occurs if both $y_{ik} = 1$ and $y_{il} = 1$.

It can be shown that the association parameters ψ_{kl} equals the log odds ratio

$$\psi_{kl} = \ln \left(\frac{\Pr[Y_i = (0, \ldots, 0, y_k, 0, \ldots, 0, y_l, 0, \ldots, 0)' | X_i] \Pr[Y_i = (0, \ldots, 0)' | X_i]}{\Pr[Y_i = (0, \ldots, 0, y_k, 0, \ldots, 0)' | X_i] \Pr[Y_i = (0, \ldots, 0, y_l, 0, \ldots, 0)' | X_i]} \right) \tag{6}$$

which again illustrates that the parameter describes the simultaneity in the binary decisions. In the next subsection we will extend the idea of this section to the situation of simultaneous multinomial decisions and we will derive a Multivariate Multinomial Logit model.

2.2 A Multivariate Multinomial Logit Model

Assume now that we have K multinomial choices and that the k-th choice decision has J_k potential outcomes. Again we define a vector of random variables Y_i as in (1) but now $Y_{ik} = j$ if individual i chooses $j = 1, \ldots, J_k$ for the k-th choice. The number of potential outcomes of Y_i is $\prod_{k=1}^{K} J_k$. Let S again denote the set of possible realizations of Y_i. We consider the conditional probabilities for the k-th choice given all other choices y_{il}, that is

$$\Pr[Y_{ik} = j | y_{il} \text{ for } l \neq k, X_i] \quad = \quad \frac{\exp(Z_{ik,j})}{\sum_{l=1}^{J_k} \exp(Z_{ik,l})} \tag{7}$$

with

$$Z_{ik,j} \;=\; \alpha_{k,j} + X_i\beta_{k,j} + \sum_{l\neq k}\psi_{kl,jy_{il}}, \tag{8}$$

where $\alpha_{k,j}$ are alternative- and choice-specific intercepts, X_i a vector of explanatory variables with corresponding parameter vector $\beta_{k,j}$, y_{il} the choice decision of individual i for the l-th choice and where $\psi_{kl,jh}$ are association parameters between choosing j for the k-th choice and choosing h for the l-th choice. Not all parameters in (7) are identified. It is easy to see that when all $\psi_{kl,jh}$-parameters are 0, the conditional probabilities simplify to standard multinomial logit probabilities where the K choices are independent. Hence, to identify the parameters we have to impose the standard identification restrictions of the Multinomial (binary) Logit model, that is, $\alpha_{k,1} = 0$ and $\beta_{k,1} = 0$ for all k. Furthermore, using similar arguments as in the Multivariate Logit case we impose the symmetry restriction on the association parameters, that is $\psi_{kl,jh} = \psi_{lk,hj}$ for all j and h.

Finally, as utility differences determine choice, we cannot identify all association parameters. Without loss of generality we impose that $\psi_{kl,j1} = \psi_{kl,1h} = 0$ for all j and h. Note that it is possible to impose other identification restrictions. Our choice however (i) is a straightforward extension to the binomial example in the previous section; (ii) is universal, that is, can be applied for all possible values of K and J_k and (iii) yields direct interpretations of the association parameters via odds ratios.

The model in (7) is a straightforward extension of the MVL model discussed in Section 2.1. In Appendix A.1 we show that Besag (1974)'s Theorem can also be used in this multinomial setting leading to the joint probabilities given in (4) but now with

$$\mu_{y_i} \;=\; \sum_{k=1}^{K}\alpha_{k,y_{ik}} + X_i\beta_{k,y_{ik}} + \sum_{l>k}\psi_{kl,y_{ik}y_{il}}. \tag{9}$$

It is easy to see that the equation contains α_k and β_k corresponding to the specific choice for the k-th choice and $\psi_{kl,jh}$ corresponding to the observed choice pairs y_{ik} and y_{il}. The base alternative in this model is $y_i = (1, \ldots, 1)$ where under the identification restrictions the corresponding μ equals 0.

The discussion can easily be extended to a Multivariate Conditional Logit specification where the explanatory variables instead of parameters vary over alternative choices. Hence, the exponent in (7) then writes

$$Z_{ik,j} \;=\; \alpha_{k,j} + W_{ik,j}\gamma_k + \sum_{l\neq k}\psi_{kl,jy_{il}}, \tag{10}$$

6

where $W_{ik,j}$ denotes the value of the explanatory variables which now differs over i, k and j and γ_k denotes the corresponding parameter. The joint probabilities are then given by (4) with

$$\mu_{y_i} = \sum_{k=1}^{K} \alpha_{k,y_{ik}} + W_{ik,y_{ik}}\gamma_k + \sum_{l>k} \psi_{kl,y_{ik}y_{il}}. \tag{11}$$

The proof directly follows from the proof for the MV-MNL specification in Appendix A.1[1].

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

$$\ln\left(\frac{\Pr[Y_i = y_i | X_i]}{\Pr[Y_i = (1,\ldots,1)' | X_i]}\right) = \sum_{k=1}^{K} \alpha_{k,y_k} + X_i \beta_{k,y_k} + \sum_{l>k} \psi_{kl,y_k y_l}, \tag{12}$$

where we use that under the identification restrictions $\Pr[Y_i = (1,\ldots,1) | X_i] \propto 1$. Clearly, this odds ratio equals μ_{y_i} in (9) and provides the probability to observe y_i relative to the base set of choice decisions.

The parameters $\psi_{kl,jh}$ indicate the associations between choices k and l. $\psi_{kl,jh}$ is in theory an unbounded parameter and thus does not directly resemble correlation between choices j and h. To give a direct interpretation to these associations, we use log odds ratios. It is easy to show that

$$\psi_{kl,y_k y_l} = \ln\left(\frac{\Pr[Y_i = (1,\ldots,1,y_k,1,\ldots,1,y_l,1,\ldots,1)' | X_i] \Pr[Y_i = (1,\ldots,1)' | X_i]}{\Pr[Y_i = (1,\ldots,1,y_k,1,\ldots,1)' | X_i] \Pr[Y_i = (1,\ldots,1,y_l,1,\ldots,1)' | X_i]}\right). \tag{13}$$

Hence, a positive $\psi_{kl,jh}$ implies that the choices j and h more often move together than apart. Hence, this indeed implies positive $\psi_{kl,jh}$ for positive correlations and negative association parameters for negative correlations.

Finally, the model can easily be extended with individual-specific association parameters by replacing the expression for $\psi_{kl,jh}$ in (9) by

$$\psi_{i,kl,jh} = \xi_{kl,jh} + X_i \delta_{kl,jh}, \tag{14}$$

where $\xi_{kl,jh}$ and $\delta_{kl,jh}$ are additional parameters. The association between decisions j and h now depends on individual characteristics X_i. The resulting model comes closer to the specifications of Amemiya (1978) and Ben-Akiva & Lerman (1985, chapter 10).

---

[1]The proof requires that $Z_{ik,1} = 0$ which does not hold for this specification. We can however rewrite the model such that $Z_{ik,j} = \alpha_{k,j} + (W_{ik,j} - W_{ik,1})\gamma_k + \sum_{l\neq k} \psi_{kl,jy_{il}}$ with $Z_{ik,1} = 0$ such that the proof is similar as in Appendix A.1.

2.3   A Bivariate Trinomial Logit Model

To illustrate the properties of the proposed Multivariate Multinomial Logit model and the need for identification restrictions we consider a bivariate trinomial Logit specification. Hence, we assume that $K = 2$ and $J_1 = J_2 = 3$. The conditional probabilities with the proper identification restrictions imposed are defined as

$$
\begin{aligned}
\Pr[Y_{i1} = 1 | y_{i2}, X_i] &\propto 1 \\
\Pr[Y_{i1} = 2 | y_{i2}, X_i] &\propto \exp(\alpha_{1,2} + X_i \beta_{1,2} + \psi_{12,2y_{i2}}) \\
\Pr[Y_{i1} = 3 | y_{i2}, X_i] &\propto \exp(\alpha_{1,3} + X_i \beta_{1,3} + \psi_{12,3y_{i2}}) \\
\Pr[Y_{i2} = 1 | y_{i1}, X_i] &\propto 1 \\
\Pr[Y_{i2} = 2 | y_{i1}, X_i] &\propto \exp(\alpha_{2,2} + X_i \beta_{2,2} + \psi_{12,y_{i1}2}) \\
\Pr[Y_{i2} = 3 | y_{i1}, X_i] &\propto \exp(\alpha_{2,3} + X_i \beta_{2,3} + \psi_{12,y_{il}3}).
\end{aligned}
\tag{15}
$$

These conditional probabilities imply the following 9 choice probabilities:

$$
\begin{aligned}
\Pr[Y_i = (1,1)' | X_i] &\propto 1 \\
\Pr[Y_i = (1,2)' | X_i] &\propto \exp(\alpha_{2,2} + X_i \beta_{2,2}) \\
\Pr[Y_i = (1,3)' | X_i] &\propto \exp(\alpha_{2,3} + X_i \beta_{2,3}) \\
\Pr[Y_i = (2,1)' | X_i] &\propto \exp(\alpha_{1,2} + X_i \beta_{1,2}) \\
\Pr[Y_i = (2,2)' | X_i] &\propto \exp(\alpha_{1,2} + \alpha_{2,2} + X_i(\beta_{1,2} + \beta_{2,2}) + \psi_{12,22}) \\
\Pr[Y_i = (2,3)' | X_i] &\propto \exp(\alpha_{1,2} + \alpha_{2,3} + X_i(\beta_{1,2} + \beta_{2,3}) + \psi_{12,23}) \\
\Pr[Y_i = (3,1)' | X_i] &\propto \exp(\alpha_{1,3} + X_i \beta_{1,3}) \\
\Pr[Y_i = (3,2)' | X_i] &\propto \exp(\alpha_{1,3} + \alpha_{2,2} + X_i(\beta_{1,3} + \beta_{2,2}) + \psi_{12,32}) \\
\Pr[Y_i = (3,3)' | X_i] &\propto \exp(\alpha_{1,3} + \alpha_{2,3} + X_i(\beta_{1,3} + \beta_{2,3}) + \psi_{12,33}).
\end{aligned}
\tag{16}
$$

As we have 9 probabilities we can only identify 8 different intercept parameters. The imposed identification restrictions result in exactly 4 α-parameters and 4 ψ-parameters and thus cause identifiability. It is easy to see that imposing $\psi_{12,22} = \psi_{12,23} = \psi_{12,32} = \psi_{12,33} = 0$ implies that the joint probabilities can be written as the product of two independent Multinomial Logit probabilities. Furthermore we see that

$$
\psi_{12,jh} = \ln\left(\frac{\Pr[Y_i = (j,h)' | X_i] \Pr[Y_i = (1,1)' | X_i]}{\Pr[Y_i = (j,1)' | X_i] \Pr[Y_i = (1,h)' | X_i]} \right).
\tag{17}
$$

Hence, a positive value of $\psi_{12,jh}$ implies positive association between choosing j for choice 1 and h for choice 2.

2.4 Parameter Inference

To estimate the parameters of the Multivariate binary Logit model Russell & Petersen (2000) suggest to use Maximum Likelihood using a log-likelihood function based on the joint probabilities, that is

$$\ell(\theta; y) = \sum_{i=1}^{N} I[Y_i = y_i] \Pr[Y_i = y_i | X_i], \tag{18}$$

where $I[A] = 1$ if A holds true and 0 otherwise, $\Pr[Y_i = y_i | X_i]$ is given in (4) and where θ summarizes the model parameters.

The same approach is of course possible for our MV-MNL specification. The disadvantage is however that the computation of these joint probabilities may be a burden if the dimensions of the Logit specification are large. For example, for $K = 10$ and $J_k = 5$ for all k we have to take the sum of 5^{10} different terms in the denominator of the joint probabilities. The outcome space of the multivariate multinomial random variable rapidly grows large and the computation time thereby increases exponentially with the number of choices.

To avoid this large computation time, we propose another estimation approach based on the ideas in Bel et al. (2014) for the MVL specification. Bel et al. (2014) propose to use a Composite Likelihood approach (Lindsay, 1988) using all conditional probabilities (2) in the likelihood specification (Molenberghs & Verbeke, 2005, chapter 12) instead of the joint probabilities (4). The resulting Composite Conditional Likelihood [CCL] representation only uses conditional probabilities and hence it avoids summation over the complete outcome space. It can be shown that the CCL approach provides consistent estimators (Varin et al., 2011) but at the cost of loss in efficiency.

The conditional probabilities in (7) lead to the composite log-likelihood function of the MV-MNL specification, that is

$$\begin{aligned} \ell^c(\theta; y) &= \sum_{i=1}^{N} \ell^c(\theta; y_i) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \ell^c(\theta; y_{ik}) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{J_k} I[Y_{ik} = j] \log P[Y_{ik} = j | y_{il} \text{ for } l \neq k, X_i]. \end{aligned} \tag{19}$$

The estimator $\hat{\theta}$ which follows from maximizing (19) is consistent. Varin et al. (2011) show that standard errors in CCL can be computed using the Godambe (1960) information

matrix, which has a sandwich form and writes

$$G_{\hat{\theta}} \;=\; H_{\hat{\theta}} J_{\hat{\theta}}^{-1} H_{\hat{\theta}} \tag{20}$$

with

$$H_{\hat{\theta}} \;=\; \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \nabla \ell^c(\hat{\theta}; y_{ik}) \nabla \ell^{c'}(\hat{\theta}; y_{ik}) \tag{21}$$

and

$$J_{\hat{\theta}} \;=\; \frac{1}{N} \sum_{i=1}^{N} \nabla \ell^c(\hat{\theta}; y_i) \nabla \ell^{c'}(\hat{\theta}; y_i). \tag{22}$$

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 (19). The covariance matrix of the parameter estimates is then given by

$$(-G_{\hat{\theta}})^{-1}. \tag{23}$$

To test for independence in the multinomial decisions one can use a Likelihood Ratio [LR] statistic for the restriction that the association parameters ψ equal 0. This LR-statistic does not have a standard distribution when the CCL estimation approach is used. Based on results by Satterthwaite (1946) and Kent (1982), Varin et al. (2011) propose to use an adjusted LR-statistic which for our test for independence boils down to

$$LR = \frac{\nu}{Q\bar{\lambda}} 2 \left(\ell^c(\hat{\theta}; y) - \ell^c(\hat{\alpha}, \hat{\beta}; y) \right), \tag{24}$$

where $\ell^c(\hat{\theta}; y)$ is the value of the CCL evaluated in the estimate under the alternative hypothesis and $\ell^c(\hat{\alpha}, \hat{\beta}; y)$ the value of the CCL evaluated in the estimate under the null and where Q is the number of ψ parameters. This LR-statistics is asymptotically $\chi^2(\nu)$ distributed with

$$\nu = \frac{\left(\sum_{q=1}^{Q} \lambda_q \right)^2}{\sum_{q=1}^{Q} \lambda_q^2}, \tag{25}$$

where $\lambda_1, \ldots, \lambda_Q$ are eigenvalues of $(G_{\psi}(H^{-1})_{\psi})^{-1}$ with G_{ψ} the $Q \times Q$ submatrix of the Godambe information matrix corresponding to ψ. Moreover, $\bar{\lambda}$ denotes the average of the eigenvalues.

Although the Composite Conditional Likelihood does not correspond to the true likelihood function, it still takes the correlation between choice decisions in the Multivariate

Multinomial Logit model into account. The advantage over the full multinomial representation in (4) is that CCL avoids the large summation in the denominator. Therefore, CCL will be more robust in computation time in case of a large number of choices and alternatives. Nonetheless, since the composite instead of the true likelihood function is used, the estimator is not efficient. Bel et al. (2014) show that the loss in efficiency is quite small for MVL models. In the next subsection we conduct a small Monte Carlo study to analyze the efficiency loss for the MV-MNL specification.

2.5 Monte Carlo Study

In this section we conduct a Monte Carlo study to investigate the properties of the Composite Likelihood estimator for the parameters of a Multivariate Multinomial Logit specification. We focus on potential small sample bias and loss in efficiency caused by using the composite instead of the exact log-likelihood specification in the estimation procedure. Finally, we check whether the normal distribution can be used to approximate the small sample distribution of the CCL estimator.

For our Monte Carlo study we consider the MV-MNL specification (4) with (9). The number of choices K is fixed to 3 and the number of choice alternatives per choice are $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$. We consider a relatively small sample size $N = 250$ and a large sample $N = 5000$. 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. The second variable is made discrete based on a zero threshold. The parameters of our DGP are chosen such that there is an unequal distribution over the choice alternatives but still substantial choice probabilities for every choice combination, see Tables 1 and 2 for the values of our DGP parameters.

Tables 1 to 4 display the mean and root mean squared error of the CCL estimator. The final two tables show that for $N = 5000$ the bias in the estimator is quite small. For a smaller sample size $N = 250$, the deviation from the DGP parameters is larger. Unreported results[2] show that the bias is almost the same as the bias in a regular Maximum Likelihood approach. The RMSE shows that there is a large variance of the estimator for small sample sizes. This is not a surprise as we in fact try to estimate the parameters of an MNL model with $3 \times 4 \times 5 = 60$ choice alternatives using only 250 observations.

---

[2]Detailed results are available upon request.

To analyze the loss in efficiency between CCL and the regular likelihood approach, we consider the ratio of the RMSEs of both approaches. Table 5 shows that the ratios are close to 1 and hence the loss in efficiency is rather limited even in small samples. For example, for the largest difference, CCL is only 1.3 percent worse in RMSE than regular ML. Hence, CCL seems to be a valid alternative for Maximum Likelihood to estimate the parameters of an MV-MNL model. The small sample bias is similar and the loss in efficiency is very small.

Apart from bias and efficiency, we also consider the validity of using a normal distribution for testing for significance of the parameters. Table 6 displays the empirical size of the t-tests for $N = 250$ for both tails of t-statistics. The table shows that even for $N = 250$ size distortions are rather small. For example, a theoretical 90 percent confidence interval for $\psi_{13,33}$ turns out to have coverage of 88.8 percent. This size distortion is still acceptable.

In sum, the simulation study shows that the Composite Likelihood estimator has similar small sample biases as the Maximum Likelihood estimator and that efficiency losses are limited. Inference based on t-statistics seems to be valid even in relatively small samples. Because of the advantages of CCL over ML when dimensions increase, CCL is a good alternative for the estimation of parameters in a Multivariate Multinomial Logit specification. In Section 4.1, we will use the CCL approach in a small application.

3   A Panel Specification

The MV-MNL model can easily be extended to a fixed-effects panel data specification. Let Y_{it} denote the K-dimensional random variable describing the joint set of choices for individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$ and let $Y_{itk} = j$ if individual i chooses $j = 1, \ldots, J_k$ for the k-th choice at time t. The choice probabilities are given by

$$\Pr[Y_{it} = y_{it}|X_{it}] \;\; = \;\; \frac{\exp(\mu_{y_{it}})}{\sum_{s_{it} \in S} \exp(\mu_{s_{it}})}, \tag{26}$$

where y_{it} is a possible realization from the outcome space S and where $\mu_{y_{it}}$ is defined as

$$\mu_{y_{it}} \;\; = \;\; \sum_{k=1}^{K} \alpha_{ik,y_{itk}} + X_{it}\beta_{k,y_{itk}} + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}}. \tag{27}$$

Hence, both the intercepts and the association parameters are individual specific. A special case of the model is where the association parameters are pooled across the individuals in which case we replace $\psi_{ikl,y_{itk}y_{itl}}$ in (27) by $\psi_{kl,y_{itk}y_{itl}}$.

3.1 Parameter Estimation

In practice the number of cross sections is usually limited and hence parameter estimation suffers from the incidental parameter problem. To solve this, we follow Chamberlain (1980) and Lee (2002, Chapter 6) who condition on a sufficient statistic which eliminates the fixed effects from the model specification. We extend the solution of Chamberlain (1980) for a univariate panel MNL model to our multivariate multinomial setting in (26). The appropriate sufficient statistics are given by

$$
v_{i,s}^{(1)} \;=\; \sum_{t=1}^{T} I[Y_{it} = s] = c_{i,s} \qquad \forall s \in S, \tag{28}
$$

where $c_{i,s}$ is the number of times the combination of choices s occurs for individual i. Thus, only the alternatives containing the same choice sets over time as observed for individual i are used in the logit specification. That is, only the permutations of choices of individual i over time are taken into account. Since no permutations can be made for individuals where no change takes place over time, these observations are not of interest and discarded. Appendix A.2 shows that the choice probabilities conditionally on these sufficient statistics are given by

$$
\Pr[Y_i = y_i | v_i^{(1)}, X_i] \;=\; \frac{\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{K} X_{it}\beta_{k,y_{itk}}\right)}{\sum_{d_i \in B} \exp\left(\sum_{t=1}^{T_i}\sum_{k=1}^{K} X_{it}\beta_{k,d_{itk}}\right)}, \tag{29}
$$

where B is the set of alternatives for which $v_i^{(1)}$ holds. Hence, the individual-specific parameters (intercepts and association parameters) are removed from the probabilities and the β-parameters can be estimated consistently using a log-likelihood function where we condition on the sufficient statistics. Note that this approach only works if X_{it} does not depend on lagged dependent variables.

In case the association parameters are of core interest, these should not be discarded from the specification. Therefore, we make $\psi_{kl,jh}$ not individual-specific and we have to consider other sufficient statistics

$$
v_{i,k,j}^{(2)} \;=\; \sum_{t=1}^{T} I[Y_{itk} = j] = c_{i,k,j} \qquad \forall k, j, \tag{30}
$$

where $c_{i,k,j}$ now is the number of times that individual i chooses option j for the k-th choice. Appendix A.2 shows that when we condition on these sufficient statistics the

choice probabilities are given by

$$\Pr[Y_i = y_i | v_i^{(2)}, X_i] \quad = \quad \frac{\exp\left(\sum_{t=1}^T \sum_{k=1}^K X_{it} \beta_{k,y_{itk}} + \sum_{l>k} \psi_{kl,y_{itk}y_{itl}}\right)}{\sum_{d_i \in B} \exp\left(\sum_{t=1}^T \sum_{k=1}^K X_{it} \beta_{k,d_{itk}} + \sum_{l>k} \psi_{kl,d_{itk}d_{itl}}\right)}, \quad (31)$$

where $\psi_{kl,jh}$ does not drop out since the combination of choices may differ over the alternatives in set B where $v_i^{(2)}$ holds. Hence, we now can find estimates of both $\beta_{k,j}$ and the association parameters $\psi_{kl,jh}$ describing the relation of the choices in the Multivariate Multinomial Logit specification. Again this approach is only valid if X_{it} does not contain lagged dependent variables.

The disadvantage of using the log-likelihood function conditional on the sufficient statistics for parameter estimation is again the sum over the alternatives in the denominator of the choice probabilities. In Appendix A.3 we however show that the Composite Likelihood method can also be applied in a panel data setting thereby avoiding the extensive sum and making parameter estimation of MV-MNL models feasible in a panel context.

In the next section we illustrate the possibilities of the MV-MNL model by applications of its panel version discussed in this section and its cross-sectional counterpart from Section 2 to household panel scanner data and a survey on life satisfaction, respectively.

4 Illustration

This section considers two illustrations of our newly proposed MV-MNL model. First, we apply the model on cross sectional survey data on satisfaction. Satisfaction is measured at an ordinal scale and satisfaction on different items are likely to be correlated. Hence, the MV-MNL model specification from Section 2 and the CCL estimation procedure from Section 2.4 can be used. Second, we investigate the product choice of canned tuna fish in a household panel scanner data set. Various multinomial choices on the characteristics of canned tuna fish are made. As these decisions are made simultaneously the model presented in Section 3 is highly applicable.

4.1 Survey Data on Satisfaction

To illustrate the MV-MNL model discussed in Section 2, we consider modeling satisfaction of 2012 Dutch respondents to an extensive survey from 2004[3]. Satisfaction is represented by 5 ordinal dependent variables: Satisfaction about *Life*, *Income*, the *Social* security system, *Democracy* and the *Government*. For *Life*, *Income* and *Democracy* respondents can be *Satisfied*, *Unsatisfied* or *In between*. *Social* and *Goverment* have two options: either the respondent is *Satisfied* or (s)he is *Unsatisfied*. The base category is *Satisfied* such that a positive β-parameter indicates less satisfaction if x_i is large and positive. To describe relations in satisfaction level we consider the MV-MNL model of Section 2.2 with $K = 5$, $J_1 = J_2 = J_4 = 3$ and $J_3 = J_5 = 2$. As explanatory variables we have *Gender*, *Age*, *Unemployment*, (self-reported) *Health* status, *Religion*, *Political interest* and *Income*.

Since our dependent variables are ordered multinomial variables we opt for a Stereo-type Logit specification (Anderson, 1984). That is, we adjust our model specification in (9) such that the parameter estimates are restricted to be monotically increasing or decreasing over the choice options. Formally, we change (9) into

$$
\mu_{y_i} = \sum_{k=1}^{K} \alpha_{k,y_{ik}} + \phi_{k,y_{ik}}(X_i\beta_k) + \sum_{l>k} \psi_{kl,y_{ik}y_{il}}, \tag{32}
$$

where $0 = \phi_{k,1} < \cdots < \phi_{k,J_k} = 1$ for ordering and identification purposes. This addition to the model specification does not change the general setup of our proposed estimation procedures.

We use the Composite Likelihood method to estimate the model parameters in (32). First, we test for independence among the five satisfaction levels. The *LR*-statistic for the restriction that all $\psi_{kl,jh}$ are 0 equals 1808.94. Since the degrees of freedom of the approximate χ^2-distribution is 50.44, independence is clearly rejected. Hence, we find positive support for association between the levels of satisfaction under consideration.

Tables 7 and 8 display the parameter estimates and estimated standard errors from the CCL method. The majority of respondents is satisfied about life, income, social security and the government, which results in negative estimates of the choice-specific intercepts although the effect for *Government* is modest. The positive estimate of the α_2 intercept shows less baseline satisfaction on democracy.

---

[3]This data is freely available at the website of the The Netherlands Institute for Social Research: http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Culturele_veranderingen_in_Nederland_CV

Several relations between the explanatory variables and satisfaction are found. Note that since *Satisfied* is the base category, a negative β-parameter indicates that the probability to be satisfied gets larger when x_i increases. For example, individuals with low (high) self-reported *Health* status are ceteris paribus more likely to report low (high) satisfaction about life. Furthermore, both women and respondents of higher age are more satisfied about their income than respectively men and respondents of average age. Unemployed respondents are more likely to report low satisfaction on the social security system. Respondents with low political interest tend to have ceteris paribus less satisfaction on democracy. Finally, religious respondents report to be more satisfied about the (at that time Christian-Liberal) government than nonreligious respondents.

The estimates of the association parameters ψ in Table 8 indicate the relation between reported satisfaction levels for the five dependent variables. Clear interpretations can be given. All parameter values that are significantly different from 0 are positive. That is, there is a positive relation between the reported satisfaction levels of respondents. For example, $\phi_{Life\ Income,33}$ indicates that respondents who report *Unsatisfied* on *Income* are likely also to be unsatisfied about life. Respondents unsatisfied about the social security system are more likely also to be unsatisfied about both *Democracy* and *Government*. This can be explained by the Labor party ending second in the previous elections with 27% of the votes but not being in charge.

4.2   Household Panel Scanner Data

To illustrate the MV-MNL model in a panel data setting we consider product choices of canned tuna in 21 supermarkets belonging to 4 chains for 1092 individuals during the period 1986(week 25)–1987(week 23) in Springfield, Missouri[4]. For each household we take the first 5 purchases in the sample and hence $T = 5$. The product choice of canned tuna concerns choosing from four characteristics: *Brand* (Chicken of the Sea, Star-Kist, CTL), whether it is *Oil*-based or not, whether it is a *Light*-product or not and *Volume* of the can. There are three choice options for *Brand* and two for the remaining characteristics. We assume that individuals make choices for these characteristics simultaneously and hence the Multivariate Multinomial choice model of Section 3 is applicable. That is, we consider a panel data MV-MNL model with $K = 4$, $J_1 = 3$ and $J_2 = \cdots = J_4 = 2$ with $N = 1092$

---

[4]This data set is from the ERIM Database and publicly available at http://research.chicagobooth.edu/kilts/marketing-databases/erim/erim-dataset

and $T = 5$. The base category for each of the 4 choices is taken to be the characteristic of the market leader.

As explanatory variables for product choice, we take the product-specific marketing-mix variables *Price* of the product, *Display* and *Feature*. Hence, (27) becomes

$$
\mu_{y_{it}} = \sum_{k=1}^{K} \alpha_{ik,y_{itk}} + W_{ity_{it}}\gamma + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}}, \tag{33}
$$

where $W_{ity_{it}}$ are now choice-specific variables. We consider two model specifications. In the first specification the ψ-parameters are individual-specific. The second specification contains ψ-parameters for all households. Hence, we respectively use $v_{i,s}^{(1)}$ and $v_{i,s}^{(2)}$.

Table 9 displays the parameter estimates and estimated standard errors from the model specification with individual-specific association parameters. Parameter estimates are obtained using a likelihood approach using (28) as sufficient statistic. Hence, the individual-specific association parameters ψ are not estimated.

To interpret the parameter estimates, we opt for the conditional marginal effects

$$
\frac{\partial \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}]}{\partial w_{ity_{it}}} = \gamma \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}] \times
$$
$$
(1 - \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, X_{it}, W_{ity_{it}}]) . \tag{34}
$$

By averaging these over y_{itl} ($l \neq k$) and the explanatory variables, that is,

$$
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \Pr[Y_{itk} = j | y_{itl} \text{ for } l \neq k, W]}{\partial w} \tag{35}
$$

we obtain an estimate for the average marginal effects. Table 10 reports these effects. An increase in *Price* leads to a decrease in the probability for each product characteristic. Equation (34) shows that the maximum marginal effect takes place when $\Pr[Y_{itk} = j | X_{it}, W_{ity_{it}}] = 0.5$ and equals $1/4$ of the parameter estimate in Table 9. The effect is on average larger for the probability to buy large *Volume* products and relative small for water-based canned tuna. Both increases in *Display* and *Feature* have a positive effect on the probability for each product characteristic, where the effect of *Feature* is larger. A product with characteristics *Brand* Star-Kist, *Oil*-based, *Light* and large volume would especially gain from advertisements, given the relatively large marginal effects.

Table 11 displays the parameter estimates and standard errors from the model specification with fixed association parameters. The parameter estimates of the marketing-mix variables are very similar to the previous specification. The advantage of this specification is that we also can interpret the association between characteristics of tuna sales.

For example, given that $\hat{\psi}_{12,22} = 1.548$, it is likely that if individuals buy *Brand* Star-Kist they also choose for the *Oil*-based tuna. The opposite conclusion holds for *Brand* CTL ($\psi_{12,32} = -1.491$). Obviously, the choice for *Oil*-based tuna is negatively associated with the *Light* product. Given the large association parameter estimate for $\psi_{13,22}$ *Brand* Star-Kist apparently is market leader in low fat tuna.

To conclude, the two examples in this section show that the MV-MNL model can be used to model simultaneous multinomial decisions in a cross-sectional and in a panel context.

5   Conclusion

In this paper we have introduced a novel Multivariate Multinomial Logit specification to describe simultaneous multinomial decisions. The advantages of the new model specification over other potential model specifications are that (i) the number of parameter stays limited; (ii) there is a clear interpretation of model parameters and; (iii) parameter estimation is feasible even if the multivariate dimension is large.

To estimate the parameters of the MV-MNL model we have proposed to use a Composite Likelihood function. This method limits the computational burden of a regular likelihood approach and is computationally feasible even if the multivariate dimension is large. The resulting maximum Composite Likelihood estimator is consistent. A small Monte Carlo study shows that the small sample bias of this estimator is comparable with a regular Maximum Likelihood estimator and that the loss in efficiency is small.

The applicability of the novel MV-MNL specification is illustrated in an application to self-reported satisfaction about life, income, social security, democracy and government. The proposed extension to panel data is illustrated using a household panel scanner data set, where we describe the purchase choice of canned tuna which we disentangle in several characteristics like brand, oil/water based and can size.

Finally, the present model specification can be extended in several directions. A possible extension is to include dynamics to the panel data model. Parameter estimation will be straightforward unless one opts for dynamics together with individual-specific effects (Honore & Kyriazidou, 2000; Carro, 2007). Other potential extensions are to adjust the model for multivariate ordered and rank ordered data or to take into account that not all choice options have to be in the consideration set of each individual.

References

Amemiya, T. (1978). On a two-step estimation of a multivariate logit model. *Journal of Econometrics*, *8*, 13 – 21.

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *46*, 1–30.

Bel, K., Fok, D., & Paap, R. (2014). *Parameter Estimation in Multivariate Logit models with Many Binary Choices*. Econometric Institute Report 2014-25 Erasmus University Rotterdam.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. 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.

Burda, M., Harding, M., & Hausman, J. (2008). A bayesian mixed logitprobit model for multinomial choice. *Journal of Econometrics*, *147*, 232 – 246.

Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, *140*, 503–528.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, *47*, 225–238.

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

Geweke, J., Keane, M., & Runkle, D. (1994). Alternative computational approaches to inference in the multinomial probit model. *The Review of Economics and Statistics*, *76*, 609–632.

Geweke, J. F., Keane, M. P., & Runkle, D. E. (1997). Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, *80*, 125–165.

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

Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica (pre-1986)*, *46*, 403.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: The state of practice. *Transportation*, *30*, 133.

Honore, B. E., & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, *68*, 839–874.

Kent, J. T. (1982). Robust properties of likelihood ratio test. *Biometrika*, *69*, 19–27.

Lee, M.-J. (2002). *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*. San Diego: Academic Press.

Lindsay, B. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes Contemporary Mathematics* (pp. 221–239). volume 80.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics* volume 3. Cambridge [etc.]: Cambridge University Press.

McFadden, D. L. (1983). *Handbook of Econometrics*. North-Holland Publishing Company, Amsterdam.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York NY Springer.

de Rooij, M., & Kroonenberg, P. M. (2003). Multivariate multinomial logit models for dyadic sequential interaction data. *Multivariate Behavioral Research*, *38*, 463–504.

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

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110–114.

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

A   Derivations

A.1   Joint probabilities in MV-MNL

In this section we derive the joint probability $\Pr[Y = y]$ in the MV-MNL model taking as starting point the conditional probabilities. To derive the joint probability $\Pr[Y = y]$ (from now on abbreviated as $\Pr[y]$) in the MV-MNL model, we use the identity

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} \quad = \quad \prod_{k=1}^{K} \frac{\Pr[y_k | y_1, \ldots, y_{k-1}, 1, \ldots, 1]}{\Pr[1 | y_1, \ldots, y_{k-1}, 1, \ldots, 1]}. \tag{36}$$

which follows from the theorem of Besag (1974). The denominator in the conditional probabilities (7) is the same in both the numerator and denominator of (36) and hence drops out of the ratio. Second, the numerator of $\Pr[1 | y_1, \ldots, y_{k-1}, 1, \ldots, 1]$ is simply proportional to 1 due to our identification restrictions. Therefore (36) simplifies to

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} \quad = \quad \prod_{k=1}^{K} \exp\left(\alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l<k} \psi_{kl,y_k y_l} + \sum_{l>k} \psi_{kl,y_k 1} \right). \tag{37}$$

Due to the restriction $\psi_{kl,y_k 1} = 0$ we obtain after rewriting

$$\frac{\Pr[y]}{\Pr[\mathbf{1}]} \quad = \quad \exp\left(\sum_{k=1}^{K} \alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l>k} \psi_{kl,y_k y_l} \right). \tag{38}$$

To obtain $\Pr[y]$ we use the identity

$$\Pr[y] \quad = \quad \frac{\Pr[y]/\Pr[\mathbf{1}]}{\sum_{s \in S} \Pr[s]/\Pr[\mathbf{1}]}, \tag{39}$$

where S is the set of all possible choice combinations. Substituting (38) in (39) results in

$$\Pr[y] \quad = \quad \frac{\exp(\mu_y)}{\sum_{s \in S} \exp(\mu_s)}, \tag{40}$$

where

$$\mu_y \quad = \quad \sum_{k=1}^{K} \alpha_{k,y_k} + X\beta_{k,y_k} + \sum_{l>k} \psi_{kl,y_k y_l}. \tag{41}$$

A.2 Choice probability conditional on sufficient statistic

In this section we derive the panel joint choice probabilities conditional on the proposed sufficient statistics in a fixed-effects MV-MNL model of Section 3. If we condition on the sufficient statistic in (28) or (30), only the choice alternatives where the sufficient statistic holds are relevant, that is

$$\Pr[y_i|v_i^{(r)}] \;=\; \frac{\Pr[y_i]}{\sum_{d_i \in B} \Pr[d_i]}, \tag{42}$$

where $r = \{1, 2\}$, and where B is the subset of alternatives which corresponds to $v_i^{(r)}$. Since we assume no dynamics we can write

$$\Pr[y_i|v_i^{(r)}] \;=\; \frac{\prod_{t=1}^{T} \Pr[y_{it}]}{\sum_{d_i \in B} \prod_{t=1}^{T} \Pr[d_{it}]} \tag{43}$$

and as the denominator of the probabilities in both the numerator and denominator are the same, this simplifies to

$$\Pr[y_i|v_i^{(r)}] \;=\; \frac{\exp(\sum_{t=1}^{T} \mu_{y_{it}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \mu_{d_{it}})}. \tag{44}$$

If we opt for the sufficient statistics in (28), we can substitute (27) for $\mu_{y_{it}}$ and rewrite this as

$$\Pr[y_i|v_i^{(1)}] \;=\; \frac{\exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{ik,y_{itk}} + \sum_{l>k} \psi_{ikl,y_{itk}y_{itl}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{ik,d_{itk}} + \sum_{l>k} \psi_{ikl,d_{itk}d_{itl}})} \times$$
$$\frac{\exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,y_{itk}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,d_{itk}})}. \tag{45}$$

As the combination of $\alpha_{ik,j}$ and $\psi_{ikl,jh}$ is by assumption constant over time, it drops out of the equation and hence we obtain

$$\Pr[y_i|v_i^{(1)}] \;=\; \frac{\exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,y_{itk}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,d_{itk}})}. \tag{46}$$

For the sufficient statistics in (30), we follow the same approach and substituting (27) for $\mu_{y_{it}}$ results in

$$\Pr[y_i|v_i^{(2)}] \;=\; \frac{\exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{ik,y_{itk}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \alpha_{ik,d_{itk}})} \times$$
$$\frac{\exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,y_{itk}} + \sum_{l>k} \psi_{kl,y_{itk}y_{itl}})}{\sum_{d_i \in B} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} X_{it}\beta_{k,d_{itk}} + \sum_{l>k} \psi_{ikl,d_{itk}d_{itl}})}. \tag{47}$$

As now only $\alpha_{ik,j}$ is constant over time, only the intercepts drop out of the equation and we obtain

$$\Pr[y_i|v_i^{(2)}] \;\;=\;\; \frac{\exp(\sum_{t=1}^{T}\sum_{k=1}^{K}X_{it}\beta_{k,y_{itk}} + \sum_{l>k}\psi_{kl,y_{itk}y_{itl}})}{\sum_{d_i\in B}\exp(\sum_{t=1}^{T}\sum_{k=1}^{K}X_{it}\beta_{k,d_{itk}} + \sum_{l>k}\psi_{ikl,d_{itk}d_{itl}})}. \tag{48}$$

A.3   Composite Conditional Likelihood in panel data setting

In this section we show that the composite likelihood approach is also applicable in a fixed-effects panel MV-MNL model. This section presents a panel data analog, where composite likelihood and the use of sufficient statistics is combined.

We use sufficient statistics to remove the individual-specific effects from the conditional probabilities. The sufficient statistics imply that we have to consider permutations of the choices over time. Given the panel equivalence of the specification in (7) any permutation over time of the choices Y_{itk}, $k=1,\ldots,K$, yields the same set of intercepts but a different set of association parameters. Hence, we can only deal with the situation of individual-specific intercepts $\alpha_{ik,j}$ but the $\psi_{kl,jh}$ parameters have to pooled. Using sufficient statistic (30) we get

$$\Pr[y_{ik}|y_{il} \text{ for } l\neq k, X_i, v_{ik}^{(2)}] \;\;=\;\; \frac{\exp(\sum_{t=1}^{T}\alpha_{ik,y_{ik}})}{\sum_{d_{ik}\in B}\exp(\sum_{t=1}^{T}\alpha_{ik,d_{ik}})} \times$$
$$\frac{\exp(\sum_{t=1}^{T}X_{it}\beta_{k,y_{ik}} + \sum_{l\neq k}\psi_{kl,y_{ik}y_{il}})}{\sum_{d_{ik}\in B}\exp(\sum_{t=1}^{T}X_{it}\beta_{k,d_{ik}} + \sum_{l\neq k}\psi_{kl,d_{ik}y_{il}})}. \tag{49}$$

As the set of intercepts $\alpha_{ik,j}$ is constant over time, they drop out of the equation resulting in

$$\Pr[y_{ik}|y_{il} \text{ for } l\neq k, X_i, v_{ik}^{(2)}] \;\;=\;\; \frac{\exp(\sum_{t=1}^{T}X_{it}\beta_{k,y_{ik}} + \sum_{l\neq k}\psi_{kl,y_{ik}y_{il}})}{\sum_{d_{ik}\in B}\exp(\sum_{t=1}^{T}X_{it}\beta_{k,d_{ik}} + \sum_{l\neq k}\psi_{kl,d_{ik}y_{il}})}. \tag{50}$$

Hence, using the full set of conditional probabilities $\Pr[y_{ik}|y_{il} \text{ for } l\neq k, X_i, v_{ik}^{(2)}]$ in Composite Likelihood estimation yields an approximation of the full likelihood conditional on the sufficient statistics. As shown by the simulation study in Section 2.5 Composite Likelihood estimation in cross-sectional data finds accurate parameter estimates with only small loss of efficiency. Unreported results show that the same holds in panel data setting.

B   Tables

Table 1: Mean and RMSE of the estimator for the MV-MNL model parameters based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

		$k=1$				$k=2$				$k=3$		
		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE
	$\alpha_{1,2}$	0.15	0.165	0.684	$\alpha_{2,2}$	0.15	0.172	0.758	$\alpha_{3,2}$	0.150	0.164	0.827
	$\alpha_{1,3}$	0.25	0.278	0.690	$\alpha_{2,3}$	0.25	0.290	0.753	$\alpha_{3,3}$	0.250	0.262	0.830
					$\alpha_{2,4}$	0.375	0.426	0.757	$\alpha_{3,4}$	0.375	0.424	0.781
									$\alpha_{3,5}$	0.475	0.540	0.777
X_1	$\beta_{1,2}$	1.05	1.110	0.525	$\beta_{2,2}$	1.05	1.122	0.661	$\beta_{3,2}$	1.05	1.138	0.723
	$\beta_{1,3}$	1.45	1.533	0.526	$\beta_{2,3}$	1.45	1.545	0.650	$\beta_{3,3}$	1.45	1.556	0.710
					$\beta_{2,4}$	1.75	1.864	0.650	$\beta_{3,4}$	1.75	1.878	0.677
									$\beta_{3,5}$	1.95	2.083	0.674
X_2	$\beta_{1,2}$	0.25	0.258	0.464	$\beta_{2,2}$	0.25	0.272	0.635	$\beta_{3,2}$	0.25	0.274	0.702
	$\beta_{1,3}$	0.45	0.479	0.452	$\beta_{2,3}$	0.45	0.490	0.602	$\beta_{3,3}$	0.45	0.486	0.671
					$\beta_{2,4}$	0.65	0.696	0.594	$\beta_{3,4}$	0.65	0.697	0.637
									$\beta_{3,5}$	0.80	0.850	0.628

[a] The DGP is given in Section 2.5 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 2: Mean and RMSE of the estimator for the association parameters based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

$k=1$		2			3			4	
	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	0.475	0.513	0.612	0.250	0.267	0.598	0	-0.001	0.601
3	0.250	0.271	0.617	0.475	0.492	0.597	0.25	0.249	0.588

$k = 2$

$k=1$		2			3			4			5	
	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	-0.375	-0.409	0.704	-0.15	-0.155	0.663	0	-0.002	0.636	0.15	0.152	0.646
3	-0.150	-0.168	0.672	-0.375	-0.389	0.668	-0.15	-0.166	0.636	0	-0.010	0.642

$k = 3$

$k=2$		2			3			4			5	
	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE	ψ	$\hat{\psi}$	RMSE
2	0.475	0.512	0.814	0.250	0.289	0.827	0	0.008	0.799	-0.25	-0.262	0.780
3	0.250	0.278	0.833	0.475	0.529	0.808	0.25	0.268	0.772	0	0.004	0.756
4	0	0.011	0.862	0.250	0.288	0.834	0.475	0.512	0.781	0.25	0.265	0.761

$k = 3$

[a] The DGP is given in Section 2.5 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 3: Mean and RMSE of the estimator for the MV-MNL model parameters based on a Monte Carlo study with $N = 5000$ (10000 replications)[a]

		$k = 1$				$k = 2$				$k = 3$		
		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE		θ	$\hat{\theta}$	RMSE
	$\alpha_{1,2}$	0.15	0.141	0.131	$\alpha_{2,2}$	0.150	0.143	0.154	$\alpha_{3,2}$	0.150	0.138	0.166
	$\alpha_{1,3}$	0.25	0.243	0.137	$\alpha_{2,3}$	0.250	0.242	0.149	$\alpha_{3,3}$	0.250	0.241	0.167
					$\alpha_{2,4}$	0.375	0.375	0.145	$\alpha_{3,4}$	0.375	0.367	0.149
									$\alpha_{3,5}$	0.475	0.467	0.158
X_1	$\beta_{1,2}$	1.05	1.049	0.107	$\beta_{2,2}$	1.05	1.062	0.129	$\beta_{3,2}$	1.05	1.046	0.144
	$\beta_{1,3}$	1.45	1.447	0.101	$\beta_{2,3}$	1.45	1.462	0.128	$\beta_{3,3}$	1.45	1.454	0.141
					$\beta_{2,4}$	1.75	1.762	0.130	$\beta_{3,4}$	1.75	1.752	0.138
									$\beta_{3,5}$	1.95	1.953	0.136
X_2	$\beta_{1,2}$	0.25	0.254	0.098	$\beta_{2,2}$	0.25	0.243	0.122	$\beta_{3,2}$	0.25	0.258	0.147
	$\beta_{1,3}$	0.45	0.455	0.099	$\beta_{2,3}$	0.45	0.443	0.120	$\beta_{3,3}$	0.45	0.455	0.137
					$\beta_{2,4}$	0.65	0.644	0.121	$\beta_{3,4}$	0.65	0.658	0.132
									$\beta_{3,5}$	0.80	0.806	0.133

[a] The DGP is given in Section 2.5 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 4: Mean and RMSE of the estimator for the association parameters based on a Monte Carlo study with $N = 5000$ (10000 replications)[a]

$k = 2$

$k = 1$	2			3			4		
	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE
2	0.475	0.483	0.122	0.250	0.258	0.120	0	0.001	0.122
3	0.250	0.256	0.131	0.475	0.485	0.122	0.25	0.252	0.119

$k = 3$

$k = 1$	2			3			4			5		
	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE
2	-0.375	-0.374	0.147	-0.150	-0.151	0.136	0	0.003	0.129	0.15	0.153	0.134
3	-0.150	-0.143	0.137	-0.375	-0.371	0.138	-0.15	-0.146	0.126	0	0.001	0.130

$k = 3$

$k = 2$	2			3			4			5		
	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE	ψ	$\hat\psi$	RMSE
2	0.475	0.482	0.164	0.25	0.259	0.163	0	0.004	0.160	-0.25	-0.237	0.155
3	0.250	0.256	0.166	0.475	0.482	0.152	0.25	0.256	0.153	0	0.008	0.150
4	0	0.001	0.161	0.25	0.255	0.162	0.475	0.477	0.157	0.25	0.257	0.149

[a] The DGP is given in Section 2.5 with $K = 3$ and $J_1 = 3$, $J_2 = 4$ and $J_3 = 5$.

Table 5: Relative RMSE of the maximum CCL and the regular ML estimator[a]

	Sample size	
Parameter	250	5000
$\alpha_{1,2}$	1.007	1.003
$\alpha_{2,3}$	1.007	1.003
$\alpha_{3,4}$	1.007	1.000
$\beta_{1,3}$	1.008	1.000
	1.002	1.000
$\beta_{2,4}$	1.013	1.000
	1.004	1.001
$\beta_{3,5}$	1.013	1.003
	1.002	1.001
$\psi_{12,22}$	1.005	1.001
$\psi_{13,33}$	1.005	1.000
$\psi_{23,44}$	1.006	1.002

[a] We only report results for a subset of parameters. The results for the other parameters are similar and available upon request.

Table 6: Empirical size of the distribution of the estimators based on a Monte Carlo study with $N = 250$ (10000 replications)[a]

Theoretical	0.025	0.05	0.95	0.975
$\alpha_{1,2}$	0.028	0.055	0.951	0.976
$\alpha_{2,3}$	0.027	0.053	0.954	0.977
$\alpha_{3,4}$	0.022	0.048	0.957	0.980
$\beta_{1,3}$	0.031	0.058	0.954	0.979
	0.027	0.053	0.953	0.978
$\beta_{2,4}$	0.031	0.056	0.910	0.957
	0.029	0.055	0.908	0.958
$\beta_{3,5}$	0.029	0.054	0.962	0.982
	0.031	0.055	0.961	0.982
$\psi_{12,22}$	0.034	0.059	0.952	0.977
$\psi_{13,33}$	0.026	0.054	0.943	0.971
$\psi_{23,44}$	0.029	0.056	0.954	0.979

[a] We only report results for a subset of parameters. The results for the other parameters are similar and available upon request.

Table 7: Parameter estimates of the MV-MNL model for a cross-sectional survey on satisfaction[a][b] (standard errors in parentheses)

	Life		Income		Social Sec.		Democracy		Government	
α_2[c]	-0.526	(0.238)	-1.096	(0.226)	-2.070	(0.270)	1.246	(0.228)	-0.667	(0.233)
α_3	-2.125	(0.351)	-2.307	(0.351)	–		-0.811	(0.341)	–	
ϕ_2	0.737	(0.073)	0.519	(0.062)	–		0.625	(0.058)	–	
gender	-0.266	(0.128)	-0.520	(0.143)	-0.023	(0.107)	0.870	(0.187)	0.189	(0.103)
younger than 35	-0.231	(0.168)	0.342	(0.175)	-0.243	(0.133)	0.010	(0.241)	-0.291	(0.130)
older than 54	-0.348	(0.175)	-1.023	(0.189)	-0.578	(0.146)	0.065	(0.257)	-0.127	(0.143)
unemployment	0.257	(0.165)	0.049	(0.171)	0.368	(0.134)	-0.304	(0.247)	-0.226	(0.135)
low health status	0.888	(0.159)	0.531	(0.171)	0.286	(0.132)	-0.052	(0.233)	-0.129	(0.129)
high health status	-1.521	(0.196)	0.051	(0.200)	-0.063	(0.144)	-0.916	(0.232)	-0.175	(0.133)
religion	-0.041	(0.138)	-0.209	(0.148)	-0.174	(0.112)	0.016	(0.194)	-0.335	(0.110)
low political interest	0.159	(0.137)	-0.105	(0.150)	0.023	(0.115)	0.982	(0.206)	-0.117	(0.111)
high political interest	-0.015	(0.191)	-0.557	(0.218)	0.083	(0.160)	-0.289	(0.251)	0.526	(0.161)
low income	0.104	(0.151)	0.970	(0.162)	-0.029	(0.127)	0.273	(0.223)	-0.158	(0.125)
high income	-0.677	(0.166)	-0.924	(0.202)	-0.002	(0.139)	-0.204	(0.235)	0.044	(0.129)

[a] The data is from an extensive survey on cultural development from 2004 by the The Netherlands Institute for Social Research.

[b] The model specification is described in (32).

[c] Higher subscripts indicate lower satisfaction (the baseline is 'satisfied').

Table 8: Estimates of the association parameters of the MV-MNL model for a cross-section survey on satisfaction[a] (standard errors in parentheses)

	Life		Income		Social Sec.	Democracy	
	2	3	2	3	2	2	3
Income							
2	0.327 (0.116)	0.454 (0.184)					
3	0.740 (0.184)	1.948 (0.217)					
Social							
2	0.179 (0.120)	0.342 (0.166)	0.430 (0.120)	0.887 (0.155)			
Democracy							
2	0.464 (0.196)	0.280 (0.292)	0.348 (0.203)	-0.119 (0.270)	0.361 (0.214)		
3	0.556 (0.217)	0.408 (0.313)	0.444 (0.224)	0.340 (0.288)	1.058 (0.225)		
Government							
2	0.125 (0.114)	0.599 (0.171)	0.377 (0.114)	0.642 (0.163)	0.951 (0.113)	0.261 (0.178)	1.445 (0.200)

[a] As the association parameters are symmetric only the lower triangular matrix is given.

Table 9: Parameter estimates of the MV-MNL model for a household panel scanner data set on canned tuna product choice (standard errors in parentheses)[a]

	$\hat{\gamma}$	s.e.
Price	-0.366	(0.017)
Display	0.888	(0.117)
Feature	1.416	(0.087)

[a] Results are obtained using sufficient statistics (28).

Table 10: Average marginal effects of 1 unit change in explanatory variables to marginal probabilities for the tuna product characteristic choices (standard deviations in parentheses)

	Brand			Oil		Light		Volume	
	Ckn of Sea	Star-Kist	CTL	No	Yes	No	Yes	Small	Large
Price	-0.025 (0.025)	-0.027 (0.022)	-0.022 (0.027)	-0.011 (0.014)	-0.027 (0.026)	-0.012 (0.012)	-0.030 (0.026)	-0.009 (0.009)	-0.036 (0.028)
Display	0.061 (0.061)	0.064 (0.053)	0.054 (0.065)	0.026 (0.035)	0.065 (0.063)	0.029 (0.028)	0.073 (0.063)	0.021 (0.022)	0.087 (0.067)
Feature	0.097 (0.097)	0.103 (0.085)	0.087 (0.104)	0.041 (0.056)	0.104 (0.100)	0.047 (0.045)	0.116 (0.100)	0.033 (0.036)	0.139 (0.107)

Table 11: Parameter estimates of the MV-MNL model for a panel data set on tuna sales (standard errors in parentheses)[a]

	Brand		Oil		Light		Volume	
	Star-Kist	CTL	Yes		Yes		Large	
Association parameters[b]								
Brand								
Star-Kist			1.548	(0.150)	2.862	(0.436)	1.937	(0.253)
CTL			-1.491	(0.203)	0.259	(0.526)	0.275	(0.327)
Oil								
Yes					-1.653	(0.755)	-1.190	(0.221)
Light								
Yes							$-\infty$[c]	–
Product-specific characteristics								
Price	-0.297	(0.014)						
Display	0.882	(0.106)						
Feature	1.508	(0.086)						

[a] Results are obtained using sufficient statistics (30).

[b] As the association parameters are symmetric only the upper triangular matrix is given.

[c] This combination of choices does not occur in the dataset.