

# New Misspecification Tests for Multinomial Logit Models

Dennis Fok\*      Richard Paap

<sup>a</sup>*Econometric Institute  
Erasmus School of Economics  
Erasmus University Rotterdam*

Econometric Institute Report EI2019-24

## Abstract

Misspecification tests for Multinomial Logit [MNL] models are known to have low power or large size distortion. We propose two new misspecification tests. Both use that preferences across binary pairs of alternatives can be described by independent binary logit models when MNL is true. The first test compares Composite Likelihood parameter estimates based on choice pairs with standard Maximum Likelihood estimates using a Hausman (1978) test. The second tests for overidentification in a GMM framework using more pairs than necessary. A Monte Carlo study shows that the GMM test is in general superior with respect to power and has correct size.

**Keywords:** Discrete choices, Multinomial Logit, IIA, Hausman test, Composite Likelihood

**JEL-codes:** C25, C12, C52

---

\*Corresponding author: Econometric Institute (ET-35), Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. email: [dfok@ese.eur.nl](mailto:dfok@ese.eur.nl), phone: (+31)104081333.

# 1 Introduction

Multinomial logit [MNL] models (McFadden, 1973) are the most popular models to describe multinomial choices in empirical research. The main reason for this is that the MNL model yields closed-form expressions for choice probabilities and marginal effects. Moreover, its log-likelihood function is well-behaved. A serious drawback of the multinomial logit model is, however, the implied *Independence of Irrelevant Alternatives* [IIA] assumption (Ray, 1973). IIA implies that the odds ratio of two alternatives is not influenced by the characteristics of any other alternative. In case IIA does not hold, the MNL model is misspecified and its estimation results should not be used. As an alternative one may resort to more complex models like the Multinomial Probit [MNP] (Thurstone, 1927; Hausman and Wise, 1978), Nested Logit (McFadden, 1978; Ben-Akiva and Lerman, 1985) or the Mixed Logit model (McFadden and Train, 2000). Testing whether the MNL is well specified is important for applied research.

There are not so many tests available for misspecification in MNL models. The existing tests are based on the idea that deleting an alternative from the choice set should not affect the parameters. McFadden et al. (1977) were the first to suggest such a test. However, their Likelihood Ratio [LR] based test for the difference between parameters estimated on a full set of alternatives versus estimates based on a reduced set, suffers from severe size distortion. This size distortion is corrected in the LR test proposed by Small and Hsiao (1985), by estimating the parameters of the model with fewer alternatives on only a subset of the data. The most often applied test is the Hausman and McFadden (1984) test. This test amounts to deleting an alternative and testing whether the parameters stay the same using a Hausman (1978) type test.

For all existing tests one has to decide which categories to remove. In practice, researchers often perform several tests each time deleting different alternatives but this leads to difficulties controlling the size of the test procedure. For the Small and Hsiao (1985) test one additionally has to split the sample randomly in 2 parts, which implies that the test is not reproducible unless one uses a random seed and the same software. It is our aim to solve these drawbacks.

We propose two alternative tests for misspecification in MNL models. Both tests use the fact that if the MNL specification holds, the preferences across each binary pair of alternatives can be described by a binary logit model, see *e.g.* Wooldridge (2002, p. 498). This result is a direct consequence of the IIA property of the model.

For our first test we estimate the parameters of the MNL model using a composite likelihood [CL] function (Lindsay, 1988) based on a set of choice pairs. Under the null

hypothesis of a correctly specified model, the corresponding CL estimator [CLE] is consistent but not efficient (Varin et al., 2011). As the Maximum Likelihood estimator [MLE] is both consistent and efficient under the null hypothesis, we can construct a Hausman (1978) type test for misspecification in the MNL model. For our second approach we construct moments conditions based on choice pairs and estimate the parameters of the MNL using General Methods of Moments [GMM]. As in general there are more pairs than categories, we can use a GMM test for overidentification (Hansen, 1982) to test for misspecification in the specification of the choice probabilities.

Both new tests have practical advantages over the existing tests for misspecification: one does not have to split the sample and the testing procedure does not involve the researcher's decision on which alternative to remove.

To investigate the size and power properties of our new tests in comparison to the Hausman and McFadden and Small and Hsiao tests, we conduct a Monte Carlo experiment. The results show that the Hausman (1978) type tests have size distortions even in relatively large samples, while the Small and Hsiao (1985) and especially our GMM test for overidentification have very accurate empirical size. With respect to power, the GMM test using all choice pairs has in general best power against a wide range of alternatives (multinomial probit, nested logit, and mixed logit). We would recommend using this test in applied research.

As a byproduct we show that the CLE and the GMM pairwise estimators perform quite well in small samples. Their empirical bias is small and there is little loss in efficiency compared with standard Maximum Likelihood estimation. This result is very useful, especially if one wants to model the choice among many alternatives. The maximization of the full likelihood function can be quite demanding in those cases. The computation of the choice probabilities may give numerical problems as the smallest choice probabilities may get very small and the denominator of the MNL probability specification contains a large sum of exponential terms. In contrast, the computation of the GMM objective function and the composite likelihood function remains feasible when the dimension of the choice set increases.

The outline of this paper is as follows. In Section 2 we discuss parameter estimation of MNL models using maximum likelihood, composite likelihood using choice pairs, and choice-paired based GMM. Furthermore, existing and new test for misspecification in MNL models are discussed. Section 3 analyses the size and power properties of the misspecification tests and the small sample properties of the estimators. Finally, Section 4 concludes.

## 2 Theory

Section 2.1 introduces the MNL model and highlights the IIA property and the fact that preferences across binary pairs follow binary logit models in case the model is correctly specified. Section 2.2 deals with parameter estimation where we focus on using choice pairs in a CL approach and a GMM approach. Finally, Section 2.3 discusses existing misspecification test for the MNL model as well as our newly proposed tests.

### 2.1 MNL model

Consider a multinomial logit [MNL] model for the random variable  $Y_i \in \{1, \dots, J\}$  with  $J > 2$

$$\Pr[Y_i = j | x_i, w_i] = \frac{\exp(\beta_{0,j} + x_i' \beta_{1,j} + w_{ij}' \gamma)}{\sum_{l=1}^J \exp(\beta_{0,l} + x_i' \beta_{1,l} + w_{il}' \gamma)} \quad (1)$$

for  $j = 1, \dots, J$  and  $i = 1, \dots, N$ , where  $x_i$  is a  $k_x$ -dimensional vector of individual-specific explanatory variables and  $w_{ij}$  a  $k_w$ -dimensional vector of alternative- and individual-specific variables with  $w_i = (w_{i1}, \dots, w_{iJ})$ . The  $J$   $\beta_0$  parameters represent the intercept parameters and the  $k_x$  dimensional  $\beta_{1,j}$  parameters and  $k_w$  dimensional  $\gamma$  parameter vector describe the effect of the  $x$  and  $w$  on the choices, respectively. Hence, the model is a mix of a conditional logit model (McFadden, 1973) and the strict multinomial logit model (Theil, 1969). For parameter identification we impose  $\beta_{0,J} = 0$  and  $\beta_{1,J} = 0$ .

It is easy to derive that the odds ratio between choosing  $j$  and  $m$  is given by

$$\frac{\Pr[Y_i = j | x_i, w_i]}{\Pr[Y_i = m | x_i, w_i]} = \frac{\exp(\beta_{0,j} + x_i' \beta_{1,j} + w_{ij}' \gamma)}{\exp(\beta_{0,m} + x_i' \beta_{1,m} + w_{im}' \gamma)}, \quad (2)$$

which clearly does not depend on characteristics of the other alternatives (unequal to  $j$  and  $m$ ). Hence, the MNL specification implies IIA, see Ray (1973) for a discussion. Another property of the MNL model which follows directly from the odds ratio (2) is that the probability that  $Y_i = j$  conditional on  $Y_i \in \{j, m\}$  follows a binary logit specification, that is,

$$\begin{aligned} \Pr[Y_i = j | Y_i \in \{j, m\}, x_i, w_i] \\ = \frac{\exp((\beta_{0,j} - \beta_{0,m}) + x_i'(\beta_{1,j} - \beta_{1,m}) + (w_{ij} - w_{im})' \gamma)}{1 + \exp((\beta_{0,j} - \beta_{0,m}) + x_i'(\beta_{1,j} - \beta_{1,m}) + (w_{ij} - w_{im})' \gamma)} \end{aligned} \quad (3)$$

see, for example, Wooldridge (2002, p. 498). This property holds for all pairs  $j$  and  $m$  given that the MNL is the true model. The probabilities in (3) can be used for parameter estimation as well as for model misspecification testing.

## 2.2 Parameter estimation

Before we turn to our misspecification test, we first consider 3 ways to estimate the parameters of the MNL model (1).

### Maximum Likelihood

The standard way to estimate the parameters of the MNL model is to use maximum likelihood. Given observed choices  $y_1, \dots, y_N$ , the log-likelihood function equals

$$\ell_{MNL}(\theta) = \sum_{i=1}^N \sum_{j=1}^J I[y_i = j] \ln \Pr[Y_i = j | x_i, w_i], \quad (4)$$

where  $\theta = (\beta_{0,1}, \dots, \beta_{0,J-1}, \beta_{1,1}, \dots, \beta_{1,J-1}, \gamma)$ ,  $\Pr[Y_i = j | x_i, w_i]$  is given in (1) and where  $I[\cdot]$  is an indicator which equals 1 when the argument is true and 0 otherwise. When the model is correctly specified, the ML estimator  $\hat{\theta}_{ML}$  is consistent and asymptotically efficient. It is of course also possible to use a method of moments estimator based on the difference  $I[y_i = j] - \Pr[Y_i = j | x_i, w_i]$  for all  $i$  and  $j$ , although this may lead to efficiency loss, see, for example, Lee (1996, Section 5.3).

### Composite Likelihood

Another way to estimate the parameters is to make use of the binary logit specifications in (3). When we, for example, consider the pairs of alternatives  $j$  and  $J$  for  $j = 1, \dots, J-1$  we can identify all model parameters given a data set. A possible way to estimate the parameters is to use a composite likelihood approach (Lindsay, 1988). The CL function can be composed by conditional probabilities (3) for all pairs  $\{j, J\}$  (see, for example, Molenberghs and Verbeke, 2005) resulting in

$$\begin{aligned} \ell_{CL}(\theta) &= \sum_{i=1}^N \ell_{i,CL}(\theta) = \sum_{i=1}^N \sum_{j=1}^{J-1} \ell_{ij,CL}(\theta) \\ &= \sum_{i=1}^N \sum_{j=1}^{J-1} I[y_i \in \{j, J\}] \left( I[y_i = j] \ln \Pr[Y_i = j | Y_i \in \{j, J\}, x_i, w_i] \right. \\ &\quad \left. + I[y_i = J] \ln(1 - \Pr[Y_i = j | Y_i \in \{j, J\}, x_i, w_i]) \right), \end{aligned} \quad (5)$$

where  $\Pr[Y_i = j | Y_i \in \{j, m\}, x_i, w_i]$  is given in (3). The CLE,  $\hat{\theta}_{CL}$ , maximizes (5) and is consistent but less efficient than the MLE, see Varin et al. (2011). As shown by Varin

et al. (2011) the asymptotic covariance matrix can be estimated using the Godambe (1960) information matrix and equals

$$V_{CL}(\hat{\theta}_{CL}) = \left( -H_{CL}(\hat{\theta}_{CL}) \left( J_{CL}(\hat{\theta}_{CL}) \right)^{-1} H_{CL}(\hat{\theta}_{CL}) \right)^{-1} \quad (6)$$

with

$$H_{CL}(\hat{\theta}_{CL}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{J-1} \nabla \ell_{ij,CL}(\hat{\theta}_{CL}) \nabla \ell_{ij,CL}(\hat{\theta}_{CL})' \quad (7)$$

and

$$J_{CL}(\hat{\theta}_{CL}) = \frac{1}{N} \sum_{i=1}^N \nabla \ell_{i,CL}(\hat{\theta}_{CL}) \nabla \ell_{i,CL}(\hat{\theta}_{CL})', \quad (8)$$

where  $\nabla \ell_{ij,CL}(\hat{\theta}_{CL})$  and  $\nabla \ell_{i,CL}(\hat{\theta}_{CL})$  denote the first-order derivatives of the corresponding composite log-likelihood contributions in (5) with respect to  $\theta$  evaluated at the composite likelihood estimator. A CL estimation approach was successfully used by Bel et al. (2018) to estimate the parameters of large-dimensional multivariate binary logit models instead of MNL models.

## Generalized Method of Moments

The conditional probabilities in (3) can also be used to construct moment conditions and estimate the parameters using GMM (Hansen, 1982). If we again take all pairs  $\{j, J\}$  for  $j = 1, \dots, J-1$  we can construct exactly enough moment conditions to identify all parameters using the moment conditions

$$E \left[ (I[y_i = j] - \Pr[Y_i = j | Y_i \in \{j, J\}, x_i, w_i]) Z_{ij} \right] = 0, \text{ for all } i \text{ where } y_i \in \{j, J\}, \quad (9)$$

for  $j = 1, \dots, J-1$ , where  $Z_{ij} = (1, x_i', (w_{ij} - w_{iJ})')'$ . Note that we have to condition on  $y_i \in \{j, J\}$ . We can rewrite the moment condition such that it holds for all observations, that is,

$$E \left[ (I[y_i = j] - \Pr[Y_i = j | Y_i \in \{j, J\}, x_i, w_i]) Z_{ij} I[y_i \in \{j, J\}] \right] = 0. \quad (10)$$

In total we have  $p = (J-1)(k_x + 1 + k_w)$  moment conditions. If  $k_w \neq 0$  and  $J > 2$  these conditions overidentify the  $\gamma$  parameters. The other parameters are exactly identified by these conditions. The GMM estimator follows from minimizing

$$\mathcal{G}(\theta) = \left( \frac{1}{N} \sum_{i=1}^N m_i(\theta) \right)' S \left( \frac{1}{N} \sum_{i=1}^N m_i(\theta) \right), \quad (11)$$

with respect to  $\theta$ , where  $m_i(\theta)$  is a  $p$ -dimensional vector containing the sample analogue of the argument of the expectation in (10) for observation  $i$  and  $S$  is a  $(p \times p)$  positive definite weighting matrix. The optimal weighting matrix  $S_{opt}$  is given by the inverse covariance matrix of the moment conditions. In practice one can replace this matrix by an estimate. The resulting GMM estimator  $\hat{\theta}_{GMM}$  is consistent when the MNL model is the true model but may suffer from some efficiency loss with respect to the ML estimator. The asymptotic covariance matrix of the GMM estimator with optimal weight matrix can be estimated by

$$V_{GMM}(\hat{\theta}_{GMM}) = \frac{1}{N} \left( \left( \nabla \sum_{i=1}^N m_i(\hat{\theta}_{GMM}) \right) \hat{S}_{opt} \left( \nabla \sum_{i=1}^N m_i(\hat{\theta}_{GMM}) \right)' \right)^{-1}, \quad (12)$$

where  $\nabla \sum_{i=1}^N m_i(\hat{\theta}_{GMM})$  denotes the first-order derivative of the moment conditions with respect to  $\theta$  evaluated in the GMM estimator and

$$\hat{S}_{opt} = \left( \frac{1}{N} \sum_{i=1}^N m_i(\hat{\theta}_{GMM}) m_i(\hat{\theta}_{GMM})' \right)^{-1}, \quad (13)$$

see Hansen (1982) for details.

For the discussed CL and GMM estimators we took  $J - 1$  pairs, all including the final choice  $J$ , to estimate the parameters. In general you can take any  $J - 1$  pairs as long as all relevant parameters are identified. The covariance of the estimator of course depends on the chosen pairs. As a rule of thumb, for large  $N$  it is probably best to choose choice pairs which have equal sample frequencies as the asymptotic covariance matrix of the binary logit model is roughly proportional to the inverse of the logit probability times 1 minus the logit probability which is smallest for probabilities equal to  $\frac{1}{2}$ . As a practical rule we therefore recommend to relabel the  $Y_i$  variables according to their sample frequencies (small to large) and the take  $J - 1$  consecutive pairs, that is, the pairs  $\{j, j + 1\}$  for  $j = 1, \dots, J - 1$  based on the relabeled  $Y$ .

To increase efficiency it is possible to take more than  $J - 1$  pairs or even *all* possible pairs. This choice yields over-identifying restrictions for the  $\beta_0$  and  $\beta_1$  parameters. Again, note that we in general already have overidentifying restrictions for the  $\gamma$  parameters even in case we use  $J - 1$  pairs.

## 2.3 Testing for misspecification

The structure of the MNL model leads to the IIA property. A famous test for misspecification of the MNL model is the Hausman (1978) type test for IIA proposed in Hausman

and McFadden (1984). For this test, we first delete one (or more choice alternatives) and estimate the model parameters for the remaining alternatives. Similar to the results in (3) one can show that the resulting model is also an MNL model parameterized by a subset of the parameters of the full MNL when the MNL is the true model. The estimator for the limited data set is again consistent but less efficient compared with the full model which uses more information. Let  $\hat{\theta}_r$  denote the ML parameter estimate where one or more alternatives are deleted and  $\hat{\theta}_f$  the ML estimates of the *same* set of parameters based on the all choice alternatives (hence  $\hat{\theta}_f$  is a subset of  $\hat{\theta}_{ML}$ ). Let  $V_{ML,r}(\hat{\theta}_r)$  and  $V_{ML}(\hat{\theta}_f)$  be the corresponding estimates of the covariance matrix of the estimator. The Hausman test is given by

$$H = (\hat{\theta}_f - \hat{\theta}_r)'(V_{ML,r}(\hat{\theta}_r) - V_{ML}(\hat{\theta}_f))^{-1}(\hat{\theta}_f - \hat{\theta}_r). \quad (14)$$

Under the null hypothesis, the test statistic asymptotically has a  $\chi^2$  distribution where the degrees of freedom is equal to the number of parameters in  $\hat{\theta}_r$ . Large values of the test statistic imply misspecification and hence rejection of IIA and the MNL model.

An alternative likelihood ratio-based test is proposed by Small and Hsiao (1985) who improve upon the incorrectly sized LR test of McFadden et al. (1977). The idea of this test is to additionally divide the sample of  $N$  observations randomly in (asymptotically equal) parts with  $N_1$  and  $N_2$  observations, such that  $N_1 + N_2 = N$ . First one estimates the parameters of the MNL model for all alternatives on both subsamples separately resulting in  $\hat{\theta}_{ML}^{(1)}$  and  $\hat{\theta}_{ML}^{(2)}$ . Next, one deletes one or more of the alternatives and uses only the second subsample ( $N_2$  observations) to compute the corresponding maximum likelihood estimator  $\hat{\theta}_r^{(2)}$ . The loglikelihood function corresponding to the  $N_2$  sample is denoted by  $\ell_{ML}^{(2)}(\cdot)$ . The Likelihood Ratio statistic is now given by

$$SH = -2 \times (\ell_{ML}^{(2)}(\hat{\theta}_{ML}^{(12)}) - \ell_{ML}^{(2)}(\hat{\theta}_r^{(2)})), \quad (15)$$

where  $\hat{\theta}_{ML}^{(12)} = \sqrt{1/2}\hat{\theta}_{ML}^{(1)} + (1 - \sqrt{1/2})\hat{\theta}_{ML}^{(2)}$ . The Small and Hsiao (1985) test is asymptotically  $\chi^2$  distributed with the dimension of  $\theta_r$  as degrees of freedom. Again large values of the test statistic means rejection of IIA.

A disadvantage of the above mentioned test procedures is that one has to decide which alternative(s) to delete from the choice set. For the SH test the additional disadvantage is that the conclusion may depend on the chosen split of the data. In this paper we propose two new tests for misspecification in the MNL model using the choice pairs approach. These tests solve both mentioned disadvantages of the existing tests. A drawback of the new tests is perhaps that one has to decide which pairs to use, but one can always take all pairs in case one does not want to make this decision.



Our first test is based on the composite likelihood approach. Let  $\hat{\theta}_{ML}$  denote the maximum likelihood estimator and  $\hat{\theta}_{CL}$  the composite likelihood estimator with estimated covariance matrix  $V_{ML}(\hat{\theta}_{ML})$  and  $V_{CL}(\hat{\theta}_{CL})$ , respectively.  $V_{CL}(\hat{\theta}_{CL})$  is given in (6) and  $V_{ML}(\hat{\theta}_{ML})$  is the regular covariance matrix estimate of an ML estimator for the MNL model. Under the null hypothesis that the MNL is the true model, both estimators are consistent but only the ML estimator is efficient. We can therefore again use a Hausman type test. We propose

$$Q_1 = (\hat{\theta}_{ML} - \hat{\theta}_{CL})'(V_{CL}(\hat{\theta}_{CL}) - V_{ML}(\hat{\theta}_{ML}))^{-1}(\hat{\theta}_{ML} - \hat{\theta}_{CL}). \quad (16)$$

The test statistic is asymptotically  $\chi^2$  distributed with  $(J - 1)(k_x + 1) + k_w$  (dimension of  $\theta$ ) degrees of freedom. Note that this test can already be performed in case one only takes  $J - 1$  pairs in the composite likelihood function but can also be applied for more pairs. In the Monte Carlo experiments below we will analyze whether taking more pairs leads to a higher power of the test.

The second test is a variant of the standard GMM overidentification test. In case one takes  $J - 1$  pairs to estimate the model parameters of the MNL model one has  $(J - 1)(k_x + 1) + (J - 1)k_w$  moment restrictions for  $(J - 1)(k_x + 1) + k_w$  parameters. In case  $(J - 2)k_w > 0$  and the MNL is the true model, we have that

$$Q_2 = \mathcal{G}(\hat{\theta}_{GMM}) \stackrel{asy}{\sim} \chi^2((J - 2)k_w), \quad (17)$$

where  $\mathcal{G}(\theta)$  is given in (11) with  $S$  replaced by (13). When  $J \leq 2$  and/or  $k_w = 0$  there is no overidentification ( $Q_2 = 0$ ) and one has to use more than  $J - 1$  pairs to apply the test. It is of course always possible to take more than  $J - 1$  pairs when  $J > 2$ . The overidentification test can then be used even if  $k_w = 0$ , the correct degrees of freedom can easily be obtained as the number of overidentifying moment conditions. Again we will use a Monte Carlo experiment to analyze whether taking more pairs results in better power properties of the test.

The above mentioned tests are designed to test for general types of misspecification of the MNL model and are often referred to as test for IIA. Multinomial discrete choice models in which IIA does not automatically hold are, for example, multinomial/conditional probit models (Hausman and Wise, 1978), nested logit models (McFadden, 1978) and mixed logit models (McFadden and Train, 2000). It is of course also possible to compare the MNL model directly with these alternatives although in practice one may not want to estimate the parameters of the more complex models such as MNP due to the computational burden, (see, for example, Geweke et al., 1994). Furthermore, it is our explicit

goal to construct general tests which have power against several alternative discrete choice model specifications.

In the next section we conduct a Monte Carlo experiment to analyze the size and power of our newly proposed tests in comparison to the two often applied existing tests.

### 3 Monte Carlo Study

This section discusses a Monte Carlo study to determine the empirical size and power of the proposed test (16) and (17). In Section 3.1 we discuss the data generation processes we use. Section 3.2 deals with the (small) sample properties of the proposed CL (5) and GMM estimators (11) using choice pairs. In Section 3.3 we analyze the empirical size of the misspecification tests. Finally, in Section 3.4 we perform a power study of our tests.

#### 3.1 Data Generating Process

We consider a Data Generating Process [DGP] for the choice among 4 alternatives. The general setup is given by specifying the utility for the  $J = 4$  choice options

$$u_{ij} = \beta_{0,j} + x_i' \beta_{1,j} + w_{ij}' \gamma + \eta_{ij} + \varepsilon_{ij}, \quad (18)$$

for  $j = 1, \dots, 4$  with  $\beta_{0,4} = 0$  and  $\beta_{1,4} = 0$  where  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{i4})'$  and  $\eta_i = (\eta_{i1}, \dots, \eta_{i4})'$  are random terms. Discrete choices are generated according to

$$y_i = j \text{ if } u_{ij} > u_{ik} \text{ for all } k \neq j \quad (19)$$

We again collect all parameters in  $\theta = (\beta_{0,1}, \dots, \beta_{0,3}, \beta_{1,1}, \dots, \beta_{1,3}, \gamma)'$ .

For the exogenous variables we specify  $w_{ij} \sim NID(0, I_{k_w})$  and  $x_i \sim N(0, I_{k_x})$ , where  $k_w = k_x = 1$ . We consider six different DGPs. The first one is a regular MNL specification which allows us to validate the empirical size of the misspecification tests. This DGP is also used to study the small sample properties of the composite likelihood and GMM estimator for a correctly specified model. The remaining five DGPs are alternatives to the MNL specification where the error terms have normal or generalized extreme value distributions. We use *iEV* to denote the (independent) extreme value distribution and *GEV* to denote the Generalized Extreme Value distribution which induces dependence across alternatives. Together these alternatives represent a wide variety of deviations from IIA. In each DGP we make sure that the ordering of the alternatives is not informative, if necessary we make use of random permutations to ensure this. This also makes our

results insensitive to our choice of deleting the first choice category in the Hausman and McFadden (1984) and Small and Hsiao (1985) tests. The DGPs are given by

1. MNL:  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i = 0$  and  $\varepsilon_{ij} \sim iEV$
2. MNP:  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i = 0$  and  $\varepsilon_i \sim N(0, \pi^2/6 \times I_4)$
3. MNP:  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i = 0$  and  $\varepsilon_i \sim N(0, \pi^2/6 \times \Sigma)$  with  $\Sigma$  a diagonal matrix containing the values  $(2, 2/3, 3/2, 1/2)$  in random order.
4. MNP:  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i = 0$  and  $\varepsilon_i \sim N(0, \pi^2/6 \times \Sigma)$  where we take for  $\Sigma$  random permutations of the matrix

$$\begin{pmatrix} 1 & 0.5 & 0.5 & -0.5 \\ 0.5 & 1 & 0.5 & -0.5 \\ 0.5 & 0.5 & 1 & -0.5 \\ -0.5 & -0.5 & -0.5 & 1 \end{pmatrix}.$$

5. Mixed Logit (intercepts only):  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i \sim N(0, \pi^2/6 \times \Sigma)$  with  $\Sigma$  as in DGP 4 and  $\varepsilon_{ij} \sim iEV$ .
6. Nested Logit:  $\theta \sim N(0, 0.25I_7)$ ,  $\eta_i = 0$  and  $\varepsilon_i \sim GEV$  with 2 clusters of size 2 with as  $\tau$  parameters a random permutation of  $(\sqrt{0.2}, \sqrt{0.8})$ .

We consider sample sizes  $N = 400, 1,000, 4,000$  and  $10,000$ . In each replication of our Monte Carlo study we draw a new value for the  $\theta$  parameters and hence we cover a broad range of different distributions of choices over the 4 categories. To prevent numerical problems in computing the covariance matrix for the GMM estimator in the case where we take all pairs, we impose that there are at least 25 observations in every choice category.

### 3.2 Empirical Bias and Variance

We start with an analysis of the small sample performance of the proposed estimators. We take DGP 1 from the previous section and compute the small sample bias and variance of the regular ML estimator of the MNL model and compare it with the alternative estimators which were proposed in Section 2.2. The alternative estimators include the composite likelihood estimator and the GMM estimator, where we take different choices for the (number of) pairs.

Table 1: Bias and RMSE, multiplied by 100, of the ML, pairwise CL & pairwise GMM estimators of the parameters of an MNL model (10,000 replications) for DGP 1.<sup>a</sup>

Estimator	Sample size Specification <sup>b</sup>	$N = 400$		$N = 1,000$		$N = 4,000$		$N = 10,000$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ML	MNL	0.25	5.38	0.20	2.13	0.05	0.52	0.03	0.21
CL	3 pairs	0.31	6.33	0.15	2.46	0.06	0.60	0.04	0.24
CL	3 sorted pairs	0.31	6.86	0.19	2.65	0.05	0.65	0.03	0.25
CL	all pairs	0.26	5.65	0.21	2.22	0.05	0.54	0.04	0.22
GMM	3 pairs	0.67	6.87	0.10	2.53	0.08	0.60	0.04	0.24
GMM	3 sorted pairs	0.32	7.41	0.20	2.71	0.05	0.65	0.03	0.25
GMM	all pairs	0.31	7.64	0.37	2.54	0.05	0.55	0.05	0.21

<sup>a</sup> The table displays 100 times the square root of the average bias squared (20) and 100 times the average RMSE (21).

<sup>b</sup> In case of 3 pairs the pairs (1, 4), (2, 4) and (3, 4) are considered. Sorted pairs corresponds to the pairs (1, 2), (2, 3) and (3, 4), after sorting the alternatives on choice frequency.

Table 1 displays ( $100\times$ ) the small sample bias, which is defined as

$$\text{Bias} = \sqrt{\frac{1}{K} \sum_{k=1}^K \left( \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_k^{(r)} - \theta_k^{(r)}) \right)^2}, \quad (20)$$

where  $K = 7$  denotes the number of parameters,  $R = 10,000$  the number of replications, and  $\hat{\theta}_k^{(r)}$  and  $\theta_k^{(r)}$  denote the estimated and true value of the  $k$ th parameter in the  $r$ th replication, respectively. In words, the bias is 100 times the square root of the average bias squared. The table also shows ( $100\times$ ) the root mean squared error [RMSE], defined as

$$\text{RMSE} = \sqrt{\frac{1}{KR} \sum_{k=1}^K \sum_{r=1}^R \left( (\hat{\theta}_k^{(r)} - \theta_k^{(r)}) - \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_k^{(r)} - \theta_k^{(r)}) \right)^2}. \quad (21)$$

Note that the formulas are different than usual due to the fact that we have different true parameters across the replications.

For the choice of the pairs in the CL and GMM method we consider three options. The first option, named “3 pairs”, uses the pairs (1,4), (2,4), and (3,4). Note that, due to the setup of the DGP, there is no information in the ordering of alternatives. The second option first sorts the alternatives according to their overall choice proportion and next creates consecutive pairs. We label this option as “3 sorted pairs”. In the final option we use all pairs, with 4 alternatives this implies the use of 6 pairs.

Due to the scaling in the table, all values have to be divided by 100 to make them comparable with the values of the parameters. We first focus on the sample size of 10,000. The best estimator is the regular Maximum Likelihood estimator with the smallest average bias and the smallest RMSE. The bias in the other estimators is however also quite small. For the CL and GMM estimators, the sorted pairs approach has a slightly smaller average bias than the unsorted pairs or all pairs approach. The RMSE is however slightly higher. Taking all pairs leads to smaller RMSE but the average bias increases a bit.

For smaller sample sizes the ML estimator still has, as expected, smaller average bias and RMSE than the other estimators. For all methods, the bias and RMSE increase when the sample size gets smaller. In general the average bias and RMSE seem to be a bit smaller in the CL approach than in the GMM approach. For the CL estimator the RMSE is in general smallest when all pairs are used. For the GMM approach the 3 unsorted pairs seem to provide the smallest RMSE. It is somewhat surprising that the use of additional (valid) moment conditions does not pay off. This result is probably due to the fact that estimating the optimal weighting matrix is less accurate in small samples. Furthermore, taking sorted pairs is not always better than just taking unsorted pairs. This may be due to the fact that in small samples some of the sorted pairs may have little observations (both categories may have a low choice frequency).

We analyze the small sample distribution of the several estimators using the results in Table 2. The table reports the empirical size of a joint Wald test for the parameters being equal to their true value. In general the empirical size is very close to the asymptotically theoretical size. There is some size distortion for  $N \leq 1,000$  when applying the GMM estimator with all pairs. Again this is mainly due to the fact that the estimation of the covariance matrix of the moment conditions is inaccurate in small samples. This leads to suboptimal weighting of the moment conditions.

In sum, the small sample average bias and RMSE of the CL estimators and GMM estimators are larger than for the Maximum Likelihood approach but the differences are in general quite small. This especially holds for CL when all pairs of alternatives are used. Furthermore, the simulation experiments suggest that the small sample distribution of the estimators is close to normal. Using all pairs in a GMM context in small sample sizes may lead to size distortions. This is probably due to a poor estimation of the optimal weighting matrix when there is only a limited number of observations for some pairs. In general, the results suggest that both the CL estimator and the GMM estimator can actually be used as an alternative to the ML estimator. In practice this may be very useful. GMM and CL provide an easy way to estimate the parameters of a MNL model in case the

Table 2: Empirical size of a Wald test statistic for the parameters of an MNL model (10,000 replications) for DGP 1.<sup>a</sup>

Estimator	Sample size Specification <sup>b</sup>	$N = 400$		$N = 1,000$		$N = 4,000$		$N = 10,000$	
		10%	5%	10%	5%	10%	5%	10%	5%
ML	MNL	0.09	0.05	0.10	0.05	0.10	0.05	0.10	0.05
CL	3 pairs	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
CL	3 sorted pairs	0.09	0.05	0.10	0.05	0.10	0.05	0.10	0.05
CL	all pairs	0.09	0.05	0.10	0.05	0.10	0.05	0.09	0.05
GMM	3 pairs	0.12	0.06	0.10	0.05	0.10	0.05	0.10	0.05
GMM	3 sorted pairs	0.12	0.06	0.11	0.06	0.10	0.05	0.10	0.05
GMM	all pairs	0.30	0.22	0.20	0.13	0.13	0.07	0.11	0.06

<sup>a</sup> The null hypothesis is that all parameters are equal to their true value. The table displays the empirical size when using asymptotically 10% and 5% valid critical values.

<sup>b</sup> In case of 3 pairs the pairs (1, 4), (2, 4) and (3, 4) are considered. Sorted pairs corresponds to the pairs (1, 2), (2, 3) and (3, 4), after sorting the alternatives on choice frequency.

number of choice alternatives is large. For large number of alternatives the evaluation of the MNL probabilities may give numerical problems due to the high number of exponent terms in the denominator and/or due to the fact that some choice probabilities get close to zero. Such probabilities do not have to be calculated for our proposed GMM and CL estimators.

### 3.3 Empirical Size of Misspecification Tests

In this section we use DGP 1 to analyze the empirical size of our misspecification tests. We apply the Small and Hsiao (1985) and Hausman and McFadden (1984) test with first category deleted next to our two new tests. As the McFadden et al. (1977) test is well known to be biased and is hardly applied in practice, we do not consider this test in our simulation study.

Cheng and Long (2007) performed an extensive study on the size of the Small and Hsiao (1985) and Hausman and McFadden (1984) tests and notice several size distortions. The same was observed by Fry and Harris (1996). It is not our goal to replicate their results but to compare the performance of our newly proposed tests with the existing tests.

Table 3 reports the empirical size of the misspecification tests under consideration

Table 3: Empirical size of several tests for misspecification in MNL specifications based on DGP 1 (10,000 replications)<sup>a</sup>.

Estimator	Type Test <sup>b</sup>	Sample size Specification <sup>c</sup>	$N = 400$		$N = 1,000$		$N = 4,000$		$N = 10,000$	
			10%	5%	10%	5%	10%	5%	10%	5%
MNL	Hausman	without cat. 1	0.19	0.15	0.19	0.15	0.15	0.11	0.13	0.08
ML	Small/Hsiao	without cat. 1	0.12	0.07	0.10	0.06	0.10	0.05	0.10	0.05
CL-ML	Hausman	3 pairs	0.11	0.09	0.14	0.11	0.16	0.12	0.15	0.11
CL-ML	Hausman	3 sorted pairs	0.16	0.13	0.17	0.14	0.16	0.12	0.15	0.10
CL-ML	Hausman	all pairs	0.04	0.03	0.09	0.08	0.18	0.15	0.18	0.14
GMM	Overidentification	3 pairs	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
GMM	Overidentification	3 sorted pairs	0.10	0.05	0.10	0.05	0.10	0.05	0.11	0.05
GMM	Overidentification	all pairs	0.09	0.04	0.10	0.05	0.10	0.05	0.10	0.05

<sup>a</sup> The table displays the empirical size when using asymptotically valid 10% and 5% critical values.

<sup>b</sup> The Hausman test in ML, the Small/Hsiao test in ML, the Hausman test for CL/ML and the GMM test are given in (14), (15), (16) and (17), respectively.

<sup>c</sup> In case of 3 pairs the pairs (1, 4), (2, 4) and (3, 4) are considered. Sorted pairs corresponds to the pairs (1, 2), (2, 3) and (3, 4), after sorting the alternatives on choice frequency.

for different sample sizes. The first thing we notice from the table is that only the misspecification tests using overidentification restrictions in the GMM approach (17) are correctly sized for all sample sizes. The Small and Hsiao test (15) only has little size distortion for the smallest sample size. All Hausman based tests show large size distortions for samples size of 4,000 and smaller. This holds for the regular test in the Maximum Likelihood approach as well as the test for the difference in estimates between the ML and CL estimator (16).

### 3.4 Power of Misspecification Tests

Fry and Harris (1996) conclude that the power of the regular IIA tests are relatively poor in small samples. In this section we check whether the power for our newly proposed test are better. Table 4 displays the power of the test for DGPs 2-6 defined in Section 3.1. As some of our tests have size distortions we report size corrected power. In other words, we obtain the critical values as percentiles of the test statistics generated under DGP 1 as discussed above.

DGPs 2 to 4 correspond to MNP models, where the only difference is the covariance matrix of the random part of the utilities. As expected the power of the misspecification tests is in general higher in case the covariance matrix is different from a scaled identity

Table 4: Empirical power (size corrected) of several tests for misspecification in MNL specifications (10,000 replications; best performing test in bold)<sup>a</sup>.

Estimator	Type Test <sup>b</sup>	Sample size	$N = 400$		$N = 1,000$		$N = 4,000$		$N = 10,000$	
		Sign. level	10%	5%	10%	5%	10%	5%	10%	5%
<i>DGP 2 (Homoskedastic MNP without correlation)</i>										
ML	Hausman	without cat. 1	0.09	0.05	0.10	0.05	0.15	0.08	0.28	0.18
ML	Small/Hsiao	without cat. 1	0.10	0.05	0.11	0.05	0.12	0.06	0.14	0.08
CL-ML	Hausman	3 pairs	0.10	0.05	0.11	0.05	0.11	0.05	0.27	0.14
CL-ML	Hausman	3 sorted pairs	0.09	0.04	0.09	0.05	0.13	0.06	0.28	0.15
CL-ML	Hausman	all pairs	0.09	0.05	0.14	<b>0.09</b>	<b>0.46</b>	<b>0.36</b>	<b>0.71</b>	<b>0.64</b>
GMM	Overidentification	3 pairs	0.12	0.06	0.14	<b>0.09</b>	0.23	0.16	0.36	0.27
GMM	Overidentification	3 sorted pairs	<b>0.13</b>	<b>0.07</b>	<b>0.16</b>	<b>0.09</b>	0.28	0.19	0.43	0.35
GMM	Overidentification	all pairs	<b>0.13</b>	<b>0.07</b>	0.15	0.08	0.28	0.19	0.49	0.39
<i>DGP 3 (Heteroskedastic MNP without correlation)</i>										
ML	Hausman	without cat. 1	0.13	0.06	0.23	0.11	0.61	0.51	0.81	0.76
ML	Small/Hsiao	without cat. 1	0.11	0.06	0.14	0.07	0.24	0.16	0.38	0.30
CL-ML	Hausman	3 pairs	0.15	0.07	0.20	0.09	0.49	0.35	0.69	0.63
CL-ML	Hausman	3 sorted pairs	0.12	0.06	0.19	0.10	0.54	0.41	0.73	0.69
CL-ML	Hausman	all pairs	0.07	0.04	0.07	0.04	0.07	0.06	0.09	0.07
GMM	Overidentification	3 pairs	0.25	0.16	0.42	0.33	0.69	0.63	0.81	0.77
GMM	Overidentification	3 sorted pairs	0.29	<b>0.21</b>	0.43	0.35	0.64	0.59	0.77	0.73
GMM	Overidentification	all pairs	<b>0.30</b>	<b>0.21</b>	<b>0.53</b>	<b>0.43</b>	<b>0.81</b>	<b>0.77</b>	<b>0.91</b>	<b>0.89</b>
<i>DGP 4 (Heteroskedastic MNP with correlation)</i>										
ML	Hausman	without cat. 1	0.21	0.10	0.45	0.27	0.79	0.74	0.89	0.87
ML	Small/Hsiao	without cat. 1	0.14	0.08	0.20	0.13	0.32	0.25	0.45	0.37
CL-ML	Hausman	3 pairs	0.19	0.10	0.33	0.20	0.60	0.54	0.72	0.70
CL-ML	Hausman	3 sorted pairs	0.19	0.09	0.34	0.21	0.71	0.65	0.83	0.81
CL-ML	Hausman	all pairs	0.08	0.05	0.05	0.03	0.05	0.03	0.07	0.05
GMM	Overidentification	3 pairs	0.38	0.29	0.54	0.47	0.73	0.68	0.82	0.79
GMM	Overidentification	3 sorted pairs	0.48	0.39	0.66	0.59	0.83	0.80	0.90	0.88
GMM	Overidentification	all pairs	<b>0.51</b>	<b>0.41</b>	<b>0.72</b>	<b>0.66</b>	<b>0.90</b>	<b>0.88</b>	<b>0.96</b>	<b>0.95</b>
<i>DGP 5 (Mixed Logit)</i>										
ML	Hausman	without cat. 1	0.10	0.05	0.15	0.07	0.36	0.24	0.58	0.49
ML	Small/Hsiao	without cat. 1	0.10	0.05	0.12	0.05	0.15	0.09	0.21	0.15
CL-ML	Hausman	3 pairs	0.10	0.05	0.12	0.06	0.29	0.17	0.53	0.43
CL-ML	Hausman	3 sorted pairs	0.10	0.05	0.13	0.07	0.35	0.20	0.59	0.50
CL-ML	Hausman	all pairs	0.06	0.03	0.08	0.05	0.20	0.17	0.33	0.31
GMM	Overidentification	3 pairs	0.15	0.08	0.22	0.15	0.42	0.34	0.57	0.50
GMM	Overidentification	3 sorted pairs	<b>0.17</b>	<b>0.10</b>	<b>0.25</b>	<b>0.16</b>	0.51	0.43	0.67	0.62
GMM	Overidentification	all pairs	0.15	0.08	0.24	<b>0.16</b>	<b>0.53</b>	<b>0.45</b>	<b>0.72</b>	<b>0.66</b>
<i>DGP 6 (Nested logit)</i>										
ML	Hausman	without cat. 1	0.19	0.10	0.40	0.25	0.74	0.69	0.82	0.81
ML	Small/Hsiao	without cat. 1	0.12	0.06	0.18	0.10	0.34	0.26	0.49	0.42
CL-ML	Hausman	3 pairs	0.20	0.11	0.30	0.19	0.52	0.45	0.62	0.59
CL-ML	Hausman	3 sorted pairs	0.22	0.11	0.36	0.25	0.62	0.57	0.72	0.69
CL-ML	Hausman	all pairs	0.12	0.08	0.11	0.07	0.15	0.13	0.17	0.16
GMM	Overidentification	3 pairs	0.42	0.35	0.54	0.47	0.72	0.67	0.81	0.78
GMM	Overidentification	3 sorted pairs	0.44	0.37	0.53	0.47	0.70	0.65	0.79	0.76
GMM	Overidentification	all pairs	<b>0.71</b>	<b>0.64</b>	<b>0.84</b>	<b>0.80</b>	<b>0.95</b>	<b>0.94</b>	<b>0.98</b>	<b>0.97</b>

<sup>a</sup> Results are based on DGPs 2-6 given in Section 3.1.

<sup>b</sup> The Hausman test in ML, the Small/Hsiao test in ML, the Hausman test for CL/ML and the GMM test are given in (14), (15), (16) and (17), respectively. In case of 3 pairs the pairs (1, 4), (2, 4) and (3, 4) are considered. Sorted pairs corresponds to the pairs (1, 2), (2, 3) and (3, 4), after sorting the alternatives on choice frequency.



matrix. There are however some exceptions.

For the homoskedastic MNP without correlations the CL-ML test with all pairs seems to have the best power in large samples. In smaller samples the GMM test based on all pairs has the highest power. Considering the relatively small difference between this form of the MNP and the MNL it is rather remarkable that a relatively high power can be obtained.

In case we make the variances of the random utility different across the choice categories or allow for correlation among the choice options (DGP 3 and 4) we see that the GMM test with all pairs performs best. In large samples the regular Hausman test has quite some power for DGPs 3 and 4 but in small samples not. The Small and Hsiao test has in general lower power than the other tests.

For the mixed logit specification again the GMM approach with all pairs has in general highest power although the power is quite small for small sample sizes. Note that GMM with 3 pairs only has overidentifying restrictions on the  $\gamma$  parameter and not the  $\beta$  parameters in (1). The Small and Hsiao test has again low power. The (size corrected) power of the regular Hausman test is fine in large samples but weak in small samples.

The power against a nested logit specification (DGP 6) is in general higher than for the mixed logit specification. The GMM is again best and even has high power in small samples. The regular Hausman and McFadden test performs well in large samples but has much lower power in small samples. Again the Small and Hsiao test lags behind in power.

In sum, apart from DGP 2 it seems that an overidentification test in a GMM framework using all pairs seems to have best power in large and small samples. The Hausman test usually has lack of power in small samples. Given the fact that the GMM tests are correctly sized, this suggests that the new GMM based test for misspecification using several extra moments is to be preferred in empirical research. For this test an empirical researcher can safely use the standard  $\chi^2$  critical values. For all other tests this would lead to an overrejection under the null hypothesis.

## 4 Conclusion

In this paper we have proposed two new tests for misspecification in MNL models. Both tests are based on the fact that preferences across binary pairs of alternatives can be described by a binary logit model when the MNL is the true model. The first test is a Hausman-type test where we compare the parameter estimates of the efficient ML

estimator with the estimates of a CL approach based on the choice pairs. The second test is a GMM test for overidentification using moment conditions based on several choice pairs in a GMM framework.

We compare the size and power of the new tests with the regular Hausman and McFadden (1984) and Small and Hsiao (1985) in a Monte Carlo study. The results show that the GMM overidentification test is in general superior with respect to power, while having no size distortions. As a byproduct we show that the CL and GMM estimators using choice pairs have quite good small sample properties. Hence, these methods can be used to estimate the parameters of MNL models with many choice categories as for large number of choice categories the evaluation of the MNL probabilities provide numerical problems due to the large summation of exponents in the denominator and/or the potential small value of these probabilities. The calculation of such probabilities is not required for the CL or GMM estimators.

For applied researchers we advice the routine use of the GMM-based test for misspecification. Standard critical value can be used and this test has the highest power across a variety of alternatives.

## References

- Bel, K., D. Fok, and R. Paap (2018). Parameter estimation in multivariate logit models with many binary choices. *Econometric Reviews* 37, 534–550.
- Ben-Akiva, M. and S. R. Lerman (1985). *Discrete choice analysis: Theory and application to travel demand*. The MIT Press.
- Cheng, S. and J. S. Long (2007). Testing for IIA in the multinomial logit model. *Sociological Methods & Research* 35(4), 583–599.
- Fry, T. R. L. and M. N. Harris (1996). A Monte Carlo study of tests for the independence of irrelevant alternatives property. *Transportation Research B* 30(1), 19–30.
- Geweke, J., M. Keane, and D. Runkle (1994). Alternative approaches to inference in the multinomial probit model. *The Review of Economics and Statistics* 76(4), 609–632.
- Godambe, V. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4), 1208–1211.

- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50(4), 1029–1054.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* 46, 1251–1272.
- Hausman, J. A. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica* 52, 1219–1240.
- Hausman, J. A. and D. Wise (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica* 45, 319–339.
- Lee, M.-J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. New York: Springer.
- Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics* 80, 220–239.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, Chapter 4, pp. 105–142. New York: Academic Press.
- McFadden, D. (1978). Modelling the choice of residential location. In A. Karlquist, L. Lundquist, F. Snickars, and J. W. Weibull (Eds.), *Spatial Interaction Theory and Residential Location*, pp. 75–96. North Holland, Amsterdam.
- McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5), 447–470.
- McFadden, D., K. Train, and W. Tye (1977). An application for diagnostic tests for independent from irrelevant alternatives property of the multinomial logit model. *Transportation Research Record* 637, 39–46.
- Molenberghs, G. and G. Verbeke (2005). *Models for discrete longitudinal data*. New York NY Springer.
- Ray, P. (1973). Independence of irrelevant alternatives. *Econometrica* 41(5), 987–991.
- Small, K. A. and C. Hsiao (1985). Multinomial logit specification tests. *International Economic Review* 26(3), 619–627.

- Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review* 10(3), 251–59.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review* 34(3), 273–286.
- Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica* 21, 5–42.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.