Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

A Misspecification Test for Logit Based Route Choice Models

Anh Tien Mai
Emma Frejinger
Fabian Bastin

July 2014

CIRRELT-2014-32

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

A Misspecification Test for Logit Based Route Choice Models

Anh Tien Mai*, Emma Frejinger, Fabian Bastin

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Computer Science and Operations Research, Université de Montréal, P.O. Box 6128, Station Centre-Ville, Montréal, Canada H3C 3J7

Abstract. The multinomial logit (MNL) model is in general used for analyzing route choices in real networks in spite of the fact that path utilities are believed to be correlated. For this reason different attributes, such as path size, have been proposed to deterministically correct the utilities for correlation and they are often used in practice. Yet, statistical tests for model misspecification are rarely used. This paper shows how the information matrix test for model misspecification proposed by White (1982) can be applied for testing MNL route choice models. We consider the link-based recursive logit models and path-based logit models with sampled choice set. We prove the holding of the information matrix equality when the alternatives are sampled and the model is correctly specified, and the correction for sample bias is included. The numerical results are based on real and simulated observations in a real network with more than 3000 nodes and 7000 links. We use two different path-based models: MNL and path size logit. Similarly, we use two link-based models: recursive logit with and without the link size attribute. We conclude that, as expected, the models cannot be rejected for simulated observations while they are strongly rejected for real data. More interestingly, the models including the attributes correcting for correlation have significantly better model fit than those without, but they are still as strongly rejected by the information matrix test.

Keywords: Route choice, model misspecification testing, information matrix test, recursive logit, path size logit, link size.

Acknowledgements. This research was partially funded by Natural Sciences and Engineering Research Council of Canada (NSERC).

---

* Corresponding author: AnhTien.Mai@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2014

1 Introduction

The multinomial logit (MNL) model is in general used for analyzing route choices in real networks in spite of the fact that path utilities are believed to be correlated. For this reason different attributes, such as path size (Ben-Akiva and Bierlaire, 1999) and commonality factor (Cascetta et al., 1996), have been proposed to deterministically correct the utilities for correlation and they are often used in practice. Yet, statistical tests for model misspecification are rarely performed.

The information matrix (IM) test proposed by White (1982) is a general test for model misspecification. It exploits the well-known information matrix equality, which states that if a model is correctly specified, the expectation of the sum of the Hessian matrix and the outer product of scores is zero. The test statistic (Theorem 4.1 in White, 1982) has a quite complicated form and contains third derivatives which raises computational concerns. The first contribution of this paper is to prove that, when the models are correctly specified, the information matrix equality holds for the MNL models with sampled choice sets and the sampling corrections are added to the choice probabilities. The second contribution of this paper is to show how this test can be applied for path-based and link-based (recusive logit proposed by Fosgerau et al., 2013) MNL route choice models. We derive the analytical Hessian for the common case of linear-in-parameters utility functions so that we can compute third derivatives by finite difference. Moreover, the code for estimating route choice models and applying the test has been implemented in MATLAB and is freely available upon request.

There are different specification tests available for the MNL model. Hausman and McFadden (1984) present a test for the independence of irrelevant alternatives (IIA) property. They also show that IIA can be tested by comparing a nested logit and MNL models using likelihood ratio, Wald or Lagrange multiplier tests. The likelihood ratio test is easy to perform and is hence often used in practice. To give another example, Fosgerau (2008) describes how the Zheng test (Zheng, 1996) can be applied to discrete choice models. The IM test is complementary to these tests; it is a general test of model misspecification and unlike the other tests it does not require the estimation and comparison of multiple models. The IM test does however not give any guidance as to what is the source of misspecification. If the information matrix equality is rejected, the other tests can be used for further investigation, as also suggested by White (1982).

Several studies indicate that the finite sample distribution of the IM test when the likelihood function is correctly specified is quite different from its asymptotic χ^2 approximation. This makes the test prone to reject the null hypothesis when it is true (type I error). Davidson and MacKinnon (1992) and Horowitz (1994) provide a nice discussion on this issue. Horowitz (1994) use bootstrap to obtain finite sample critical values and compare those with the asymptotic critical values. The numerical results presented in this paper are based on both simulated data, for which the true model is known, and real data. The simulated data is used to numerically verify our proof of information matrix equality for sampled choice sets and investigate whether the incorrect size of the asymptotic critical value appears to be an issue for our application. It is important to note that there are several alternatives of White's test in the literature. Chesher (1983) and Lancaster (1984) show how to compute the IM test without the need of third derivatives. However, several later studies suggest that these methods do not perform well in many cases due to the finite sample size issue (see for instance Davidson and MacKinnon, 1992). Indeed, those methods are straightforward to apply with the analytical Hessians derived in our paper.

The paper is structured as follows. Section 2 reviews maximum likelihood estimation, the estimation of the variance-covariance matrix and the IM test. In Section 3 we present different approaches for route choice modeling and we derive the analytical Hessian for the MNL route choice models. Section 4 provides a proof for the information matrix equality for the MNL models with sampled choice sets. Estimation and test results are presented in Section 5 and finally Section 6 concludes.

2 Maximum likelihood estimation and information matrix test

In the context of maximum likelihood estimation (MLE), we aim to solve the following maximization problem

$$\max_{\beta} \widehat{LL}_N(\beta) = \frac{1}{N}\sum_{n=1}^{N} \ln f(y_n|\beta) \tag{1}$$

where $f(Y|\beta)$ is some probability density function (pdf), defined on Y, conditioned on a set of parameters β, and $y_1, \ldots, y_N$ are given observations. Using the terminology popular in stochastic programming (SP), (1) can be seen as the sample average approximation of the "true" problem

$$\max_{\beta} LL(\beta) = \mathbb{E}_Y[\ln f(y|\beta)]. \tag{2}$$

We note here that f does not necessarily correspond to the density of Y over the population, in which case the model is said to be misspecified. We can however still refer to the SP literature to establish that, under some regularity conditions, when N rises to infinity,

$$d(\hat{S}_N^*, S^*) \to 0 \text{ almost surely},$$

where d is a distance measure, $\hat{S}_N^*$ and S^* are the sets of first-order critical points of (1) and (2), respectively, assuming that $\hat{S}_N^*$ and S^* are not empty (see e.g. Shapiro, 2003 and Shapiro et al., 2009, Chapter 5). Moreover, if these sets are singletons, we denote by $\hat{\beta}_N^*$ the solution of (1) and by β^* the solution of (2). We then have that

$$\sqrt{N}(\hat{\beta}_N^* - \beta^*) \overset{A}{\backsim} \mathcal{N}(0, \Psi),$$

where $\overset{A}{\backsim}$ designs the convergence in distribution, and $\mathcal{N}$ refers to the normal distribution. Setting the gradient of (1) to zero, it can be shown that

$$\Psi = H(\beta^*)^{-1} I(\beta^*) H(\beta^*)^{-1},$$

where $H(\beta^*) = \mathbb{E}_Y[\nabla_{\beta\beta} f(Y|\beta^*)]$ and $I(\beta^*) = \mathbb{E}_Y[\nabla_\beta f(Y|\beta^*) \nabla_\beta f(Y|\beta^*)^T]$ is the outer product of scores, also called the Fisher information matrix (see e.g. Newey and McFadden, 1986). The asymptotic variance-covariance can therefore be estimated using

$$\text{Cov}(\hat{\beta}_N^*) = \frac{[H_N(\hat{\beta}_N^*)]^{-1} I_N(\hat{\beta}_N^*) [H_N(\hat{\beta}_N^*)]^{-1}}{N} \tag{3}$$

where

$$H_N(\hat{\beta}_N^*) = \frac{1}{N} \sum_{n=1}^{N} \nabla^2_{\beta\beta} \ln f(y_n|\hat{\beta}_N^*)$$

and

$$I_N(\hat{\beta}_N^*) = \frac{1}{N} \sum_{n=1}^{N} [\nabla_\beta \ln f(y_n|\hat{\beta}_N^*)][\nabla_\beta \ln f(y_n|\hat{\beta}_N^*)]^T$$

are the samples average estimates of the Hessian and the information matrix, respectively. We refer to the variance-covariance matrix given by (3) as the robust variance-covariance matrix.

The well-known information matrix equality implies that if the model is well specified, i.e. $f(Y|\beta^*)$ is the density of Y over the population, the Fisher information matrix is equal to the opposite of the Hessian matrix,

$$I(\beta^*) = -H(\beta^*). \tag{4}$$

The robust variance-covariance of the maximum likelihood estimator then becomes $-[H_N(\hat{\beta}^*_N)]^{-1}/N$.

The information matrix equality test proposed by White (1982) is based on the idea that the information matrix equality (4) can be reformulated as

$$H(\beta^*) + I(\beta^*) = 0$$

and the sum can be consistently estimated by

$$I_N(\hat{\beta}^*_N) + H_N(\hat{\beta}^*_N).$$

The test statistic (Theorem 4.1 in White, 1982) is designed based on the jointly normally asymptotically distributed property of $D_N(\hat{\beta}^*_N) = I_N(\hat{\beta}^*_N) + H_N(\hat{\beta}^*_N)$

$$\sqrt{N} D^\eta_N(\hat{\beta}^*_N) \overset{A}{\backsim} \mathcal{N}(0, \mathcal{V}_N(\hat{\beta}^*_N)).$$

Here we note that for a matrix A, vector A^η is defined by taking η indicators of interest in A. An asymptotic χ^2 statistic test is

$$\wp_N = N D^\eta_N(\hat{\beta}^*_N)^T \mathcal{V}_N(\hat{\beta}^*_N)^{-1} D^\eta_N(\hat{\beta}^*_N) \overset{A}{\backsim} \chi^2_\eta \tag{5}$$

where χ^2_η is chi-square distribution with η degrees of freedom. The value of $D^\eta_N(\hat{\beta}^*_N)$ and $\mathcal{V}_N(\hat{\beta}^*_N)$ are defined by

$$\begin{aligned} D^\eta_N(\hat{\beta}^*_N) &= \frac{1}{N}\sum_{n=1}^{N} d^\eta_n(y_n|\hat{\beta}^*_N) \\ \mathcal{V}_N(\hat{\beta}^*_N) &= \frac{1}{N}\sum_{n=1}^{N}\left[\psi_n(\hat{\beta}^*_N)\psi_n(\hat{\beta}^*_N)^T\right] \end{aligned} \tag{6}$$

where

$$d_n(y_n|\hat{\beta}^*_N) = [\nabla_\beta \ln f(y_n|\hat{\beta}^*_N)][\nabla_\beta \ln f(y_n|\hat{\beta}^*_N)]^T + \nabla^2_{\beta\beta}(\ln f(y_n|\hat{\beta}^*_N))$$

and $\psi_n(\hat{\beta}^*_N) = d^\eta_n(y_n|\hat{\beta}^*_N) - \nabla_\beta D^\eta_N(\hat{\beta}^*_N) H_N(\hat{\beta}^*_N)^{-1} \nabla_\beta \ln f(y_n|\hat{\beta}^*_N)$. Note that (5) contains third derivatives through $\nabla_\beta D^\eta_N(\hat{\beta}^*_N)$ which raises computational concerns. In the following section we present MNL route choice models and derive the corresponding analytical Hessians so that we can evaluate (5) by computing third derivatives by finite difference on the Hessian.

3 Route choice models

Following the discussion in Fosgerau et al. (2013) we group the numerous route choice models proposed in the literature into three approaches. First, the classic approach that corresponds to path based models where choice sets of paths are generated with some algorithm (typically a repeated shortest path search) and treated as the actual choice sets. Frejinger et al. (2009) argue that this does not yield consistent parameter estimates since they have empirically observed that parameter estimates vary significantly for a same dataset depending on the definition of the choice sets. The second approach, proposed by Frejinger et al. (2009), is based on importance sampling of alternatives. The idea is to correct the path utilities for the sampling protocol that is used so that the parameter estimates do not significantly change when the definition of the choice sets change. This approach has so far been used for estimating the MNL and path size logit (PSL) models. Related to this approach is the work in Flotterod and Bierlaire (2013). They propose an approach for sampling paths according to a pre-defined probability distribution. It is important to note that the derivation of the sampling correction is based on the MNL model. Guevara and Ben-Akiva (2013) present a correction for sampling of alternatives for the more general multivariate extreme value models but this approach has not yet been used for route choice analysis.

The third approach is the link-based recursive logit (RL) model proposed by Fosgerau et al. (2013). This model is based on the same underlying assumption as the sampling approach, namely, that any path in the network is feasible and belongs to the universal choice set. However, is does not require any choice sets of paths. The RL model is theoretically superior to the sampling approach because it can be consistently estimated and efficiently used for prediction. Fosgerau et al. (2013) also proposed an heuristic correction for correlation, similar to path size, but that is link additive and they call it link size (LS). The models based on sampling of alternatives can also be used for prediction using the approach proposed by Flotterod and Bierlaire (2013) as long as the path costs can be computed independently of other paths which excludes PSL.

In this paper we use the two comparable approaches, namely, MNL and PSL with sampled choice sets and RL with and without LS.

3.1 Path based logit with sampled choice sets

The MNL model can be consistently estimated on a sample of alternatives (McFadden, 1978). The probability that an individual n choses a path σ

given a sampled choice set of paths D_n is given by

$$P(\sigma|D_n) = \frac{e^{\frac{1}{\mu}v_{\sigma n}+\ln\pi(D_n|\sigma)}}{\sum_{j\in D_n} e^{\frac{1}{\mu}v_{jn}+\ln\pi(D_n|j)}} \tag{7}$$

where $v_{\sigma n} = v(x_{\sigma n}, \beta)$ is the deterministic utility component associated with each path σ, $x_{\sigma n}$ is a vector of observed attributes and β is a vector of unknown parameters to be estimated. $\ln\pi(D_n|\sigma)$ is the correction for sampling bias where $\pi(D_n|\sigma)$ is the probability of sampling choice set D_n given that σ is the chosen alternative. When $\pi(D_n|\sigma) = \pi(D_n|j)$ for all $j \in D_n$, this correction term can be safely ignored, otherwise it is required to ensure that the information matrix equality still holds, as we will see in Section 4. Frejinger et al. (2009) show that in a path sampling context $\pi(D_n|\sigma) = \frac{r_{\sigma n}}{q(\sigma)}$ where $r_{\sigma n}$ is the number of times path σ was drawn when sampling set D_n and $q(\sigma)$ the path sampling probability, and report improvement in numerical results when incorporating the correction in the model estimation. In this paper we refer to (7) as the path logit (PL) model.

PSL is often used for route choice analysis by researchers and practitioners and it is choice set dependent. Frejinger et al. (2009) propose a heuristic sampling correction of the path size attribute called expanded path size (EPS). A path σ is a sequence of links $(k_0, \ldots, k_I)$ and the probability that n chooses σ according to PSL is

$$P(\sigma|D_n) = \frac{e^{\frac{1}{\mu}(v_{\sigma n}+\beta_{PS}\ln EPS_{\sigma n})+\ln\pi(D_n|\sigma)}}{\sum_{j\in D_n} e^{\frac{1}{\mu}(v_{jn}+\beta_{PS}EPS_{jn})+\ln\pi(D_n|j)}} \tag{8}$$

where

$$EPS_{\sigma n} = \sum_{i=0}^{I_\sigma} \frac{L(k_i)}{L(\sigma)\sum_{j\in D_n}\delta(j,k_i)\phi^{\sigma}_{jn}}.$$

$L(\cdot)$ is the length link k_i or path σ and $\delta(j, k_i)$ equals one if k_i is on path j and zero otherwise so that $\sum_{j\in D_n}\delta(j, k_i)$ is the number of paths in D_n that use link k_i. $\phi^{\sigma}_{jn} = \max\{1, \frac{1-\mathbb{I}[j=\sigma]}{q(j)|D_n|}\}$ is the expansion factor where $\mathbb{I}[\cdot]$ is an indicator function and $|D_n|$ is cardinality of D_n.

The model can be estimated by maximum likelihood and given a set of observations $n = 1, \ldots, N$ the log-likelihood function (1) becomes

$$\widehat{LL}_N(\beta) = \frac{1}{N}\sum_{n=1}^{N} \ln P(\sigma_n|D_n) = \frac{1}{N}\sum_{n=1}^{N} \ln \left(\frac{e^{\frac{1}{\mu}(v_{\sigma n}+\beta_{PS}\ln EPS_{\sigma n})+\ln \pi(D_n|\sigma)}}{\sum_{j\in D_n} e^{\frac{1}{\mu}(v_{jn}+\beta_{PS}EPS_{jn})+\ln \pi(D_n|j)}} \right). \tag{9}$$

We denote

$$\upsilon_{jn}^{\sigma}(\beta) = \exp\left(\frac{1}{\mu}(v_{jn} - v_{\sigma n} + \beta_{PS}(EPS_{jn} - EPS_{\sigma n})) + \ln \frac{\pi(D_n|j)}{\pi(D_n|\sigma)}\right)$$

so that (9) can be more concisely written as

$$\widehat{LL}_N(\beta) = -\frac{1}{N}\sum_{n=1}^{N} \ln \left(\sum_{j\in D_n} e^{\upsilon_{jn}^{\sigma}(\beta)} \right) \tag{10}$$

and its gradient is

$$\nabla_{\beta}\widehat{LL}_N(\beta) = -\frac{1}{N}\sum_{n=1}^{N} \frac{\sum_{j\in D_n} \nabla_{\beta}\upsilon_{jn}^{\sigma}(\beta) e^{\upsilon_{jn}^{\sigma}(\beta)}}{\sum_{j\in D_n} e^{\upsilon_{jn}^{\sigma}(\beta)}}. \tag{11}$$

In case of a linear-in-parameters utility functions $v_{jn}(\beta)$ and $v_{\sigma n}(\beta)$, we can write $\upsilon_{jn}^{\sigma}(\beta) = (x_{jn}^{\sigma})^T\beta + \kappa$ (where κ is some constant) and the first order derivative is trivial. Taking the derivative of (11) we obtain an analytical expression of the Hessian

$$\nabla_{\beta\beta}^2\widehat{LL}_N(\beta) = \frac{1}{N}\sum_{n=1}^{N} \frac{\psi_n(\beta)\psi_n(\beta)^T - \Omega_n(\beta)\eta_n(\beta)}{(\eta_n(\beta))^2} \tag{12}$$

where $\psi_n(\beta) = \sum_{j\in D_n} \nabla_{\beta}\upsilon_{jn}^{\sigma}(\beta)e^{\upsilon_{jn}^{\sigma}(\beta)}$, $\Omega_n(\beta) = \sum_{j\in D_n}(\nabla_{\beta}\upsilon_{jn}^{\sigma})(\nabla_{\beta}\upsilon_{jn}^{\sigma})^T e^{\upsilon_{jn}^{\sigma}(\beta)}$ and $\eta_n(\beta) = \sum_{j\in D_n} e^{\upsilon_{jn}^{\sigma}(\beta)}$. We note that this expression is fast and straightforward to compute.

3.2 Recursive logit

Fosgerau et al. (2013) recently proposed the RL model where the path choice problem is formulated as a sequence of link choices and modeled in a dynamic discrete choice framework. In the following we describe the model briefly but in enough detail to allow us to derive the analytical Hessian.

The network corresponds to a directed connected graph (not assumed acyclic) $G = (A, \mathbb{V})$ where A is the set of link and $\mathbb{V}$ if the set of nodes. We denote links $k, a \in A$ and the set outgoing links from the sink node of k $A(k)$. An instantaneous utility $u(a|k;\beta) = v(a|k;\beta) + \varepsilon(a)$ is associated with action $a \in A(k)$. The deterministic utility $v(a|k;\beta)$ is assumed negative for all links except the dummy link d (absorbing state) that has been added to the destination and equals zero. The random terms are independently and identically distributed extreme value type I so that the choice model at each decision stage is MNL.

At each choice stage a traveler observes the deterministic utilities of the outgoing links, the random terms and the expected maximum utility until the destination, value function $V(a)$, and chooses the link with the maximum utility. The key is how to compute these value functions and Fosgerau et al. (2013) show that they are conveniently the solution to a system of linear equations. Note that we use a simplified notation here, but the value functions are destination specific, and possibly, origin-destination (OD) specific depending on the definition of the instantaneous utilities.

We define a matrix M with entries

$$M_{ka}(\beta) = \begin{cases} \delta(a|k)e^{\frac{1}{\mu}v(a|k)} & \forall k \in A \\ 0 & k = d \end{cases}$$

where $\delta(a|k) = 1$ if $a \in A(k)$ and equal 0 otherwise. $z(\beta)$ is a vector with elements $z_k = e^{\frac{1}{\mu}V(k)}$ and b is a vector of size $[(|A| + 1) \times 1]$ with zero values for all states except for the destination that equals one, $b_d = 1$. $z(\beta)$ is the solution to the system of linear equations

$$[I - M(\beta)]z(\beta) = b \tag{13}$$

where I is the identity matrix. Moreover, similar to the idea of a PS attribute, the LS attribute can be added to the instantaneous utilities. This attribute is OD specific and simply corresponds to the link flow when the demand at the origin is one

$$LS^{od} = F^{od}(\widetilde{\beta}) \tag{14}$$

where F^{od} is the vector of link flows computed by solving a system of linear equations for some chosen parameter values $\underset{\sim}{\beta}$. We refer the reader to Fosgerau et al. (2013) for more details.

A path σ is a sequence links $(k_0, \ldots, k_I)$ with $k_{i+1} \in A(k_i)$ for all $i < I$ and at each link k the choice of next link is given by MNL

$$P(a|k) = \frac{e^{\frac{1}{\mu}(v(a|k)+V(a))}}{\sum_{a \in A(k)} e^{\frac{1}{\mu}(v(a|k)+V(a))}} = e^{\frac{1}{\mu}(v(a|k)+V(a)-V(k))}.$$

The RL model can be estimated based on path observations $n = 1, \ldots, N$ by maximum likelihood. The likelihood of a path is

$$\begin{aligned} P(\sigma, \beta) &= \prod_{i=0}^{I-1} P(k_{i+1}|k_i) \\ &= \prod_{i=0}^{I-1} e^{\frac{1}{\mu}[v(k_{i+1}|k_i)+V(k_{i+1})-V(k_i)]} \\ &= e^{-\frac{1}{\mu}V(k_0)} \prod_{i=0}^{I-1} e^{\frac{1}{\mu}v(k_{i+1}|k_i)} \end{aligned}$$

and the log-likelihood function (1) is

$$\widehat{LL}_N(\beta) = \frac{1}{N}\sum_{n=1}^{N} \ln P(\sigma_n, \beta) = \frac{1}{\mu N}\sum_{n=1}^{N} (v(\sigma_n, \beta) - V(k_0^n, \beta)) \tag{15}$$

where $v(\sigma_n, \beta) = \sum_{i=0}^{I_n-1} v(k_{i+1}^n|k_i^n)$ is the sum of the deterministic link utilities of observed path σ_n.

As we did for the PSL model, we now derive analytical expressions for the gradient and Hessian assuming a linear-in-parameters formulation of the deterministic utilities.[1] In this case the log-likelihood function (15) can be written as

$$\widehat{LL}_N(\beta) = \frac{1}{\mu N}\sum_{n=1}^{N} \left(x_{\sigma_n}^T \beta - V(k_0^n, \beta)\right) \tag{16}$$

where $x_{\sigma_n} = \sum_{i=0}^{I_n-1} x(k_{i+1}^n|k_i^n)$. We note that $V(k_0^n, \beta) = \mu \ln z^n(k_0^n, \beta)$, thus the gradient of $V(k_0^n, \beta)$ can be written as

$$\nabla_\beta V(k_0^n, \beta) = \mu \frac{\nabla_\beta z^n(k_0^n, \beta)}{z^n(k_0^n, \beta)}$$

So the gradient of (16) is

$$\nabla_\beta \widehat{LL}_N(\beta) = \frac{1}{N}\sum_{n=1}^{N} \left(x_{\sigma_n} - \frac{\nabla_\beta z^n(k_0^n, \beta)}{z^n(k_0^n, \beta)} \right) \tag{17}$$

and in the following we derive $\nabla_\beta z^n(k_0^n, \beta)$. As mentioned above, for each n the vector $z^n(\beta)$ is computed by solving the system of linear equations (13) either for each destination or for each OD pair depending on if a LS attribute

[1] The gradient is also derived in Fosgerau et al. (2013).

is included or not. The first derivative of this equation with respect to one parameter β_q is

$$\frac{\partial z^n(\beta)}{\partial \beta_p} = [I - M^n(\beta)]^{-1} \left[\frac{\partial M^n(\beta)}{\partial \beta_p}\right] z^n(\beta) \tag{18}$$

We denote the Jacobian as $\nabla_\beta z^n(\beta)$ and $\nabla_\beta V(k_0^n, \beta)$ of (17) is simply the row corresponding to state k_0^n, $\nabla_\beta z^n(k_0^n, \beta)$.

The analytical Hessian has a more complicated form but can be derived based on (17)

$$\begin{aligned}\nabla^2_{\beta\beta}\widehat{LL}_N(\beta) &= -\frac{1}{\mu N}\sum_{n=1}^{N} \nabla_{\beta\beta} V(k_0^n, \beta) \\ &= \frac{1}{N}\sum_{n=1}^{N} - \left(\frac{\nabla^2_{\beta\beta} z^n(k_0^n, \beta)}{z^n(k_0^n, \beta)}\right) + \left(\frac{\nabla_\beta z^n(k_0^n, \beta) \nabla_\beta z^n(k_0^n, \beta)^T}{z^n(k_0^n, \beta)^2}\right).\end{aligned} \tag{19}$$

$\nabla_\beta z^n(k_0^n, \beta)$ can be calculated by (17) and $\nabla^2_{\beta\beta} z^n(k_0^n, \beta)$ obtained by taking derivative of (18). The second derivative of this equation with respect to two parameters β_p, β_q is

$$\begin{aligned}\frac{\partial^2 z^n(\beta)}{\partial \beta_p \partial \beta_q} &= \frac{\partial [I - M^n(\beta)]^{-1}}{\partial \beta_q} \frac{\partial M^n(\beta)}{\partial \beta_p} z^n(\beta) \\ &+ [I - M^n(\beta)]^{-1} \frac{\partial M^n(\beta)}{\partial \beta_p} \frac{\partial z^n(\beta)}{\partial \beta_q} \\ &+ [I - M^n(\beta)]^{-1} \frac{\partial^2 M^n(\beta)}{\partial \beta_p \partial \beta_q} z^n(\beta).\end{aligned} \tag{20}$$

Note that $\frac{\partial [I - M^n(\beta)]^{-1}}{\partial \beta_q} = [I - M^n(\beta)]^{-1} \frac{\partial M^n(\beta)}{\partial \beta_q} [I - M^n(\beta)]^{-1}$, substitution into (18) gives a concise analytical expression of $\frac{\partial^2 z^n(\beta)}{\partial \beta_p \partial \beta_q}$

$$\frac{\partial^2 z^n(\beta)}{\partial \beta_p \partial \beta_q} = [I - M^n(\beta)]^{-1} \left(\frac{\partial^2 M^n(\beta)}{\partial \beta_p \partial \beta_q} z^n(\beta) + \frac{\partial M^n(\beta)}{\partial \beta_p} \frac{\partial z^n(\beta)}{\partial \beta_q} + \frac{\partial M^n(\beta)}{\partial \beta_q} \frac{\partial z^n(\beta)}{\partial \beta_p}\right) \tag{21}$$

which may seem complicated but is fairly fast to compute.

4 Information matrix equality and sampled choice sets

We have introduced the information equality test which is based on the idea that for a correctly specified model at the true parameters, $H(\beta^*) + I(\beta^*) = 0$.

We aim to apply the test for link-based and path-based MNL route choice models. For the linked-based approach (Fosgerau et al., 2013), the choice probabilities are based on the universal choice set and it does not require sampling alternatives. The sampling correction does not appear in the formula of choice probability, meaning that the information matrix equality can be established similarly to the general maximum likelihood estimation problem (see for instance Amemiya (1985), Chapter 1). However, to our best knowledge, there is currently no proof when the choice set is sampled. In the following, we show that the information matrix equality still holds for the MNL models when the alternatives are sampled and the model is correctly specified.

As in Section 3.1, we denote the probability of sampling a set of alternatives D_n, given observed choice σ for individual n, by $\pi(D_n|\sigma)$. We assume that the positive conditioning property (McFadden, 1978) holds:

$$\text{if } \pi(D_n|\sigma) > 0, \text{ then for all } j \in D,\ \pi(D_n|j) > 0.$$

We also assume that the chosen path is always included in D_n, meaning that $\pi(D_n|j) = 0$ if $j \notin D_n$. The conditional probability $P(\sigma, \beta|D_n)$ for an observed choice σ, given a choice set D_n, is expressed by (7):

$$P(\sigma, \beta|D_n) = \frac{e^{\frac{1}{\mu}v(x_{\sigma n},\beta)+\ln\pi(D_n|\sigma)}}{\sum_{j\in D_n} e^{\frac{1}{\mu}v(x_{jn},\beta)+\ln\pi(D_n|j)}}$$

or, equivalently,

$$P(\sigma, \beta|D_n) = \frac{\pi(D_n|\sigma)e^{\frac{1}{\mu}v(x_{\sigma n},\beta)}}{\sum_{j\in D_n} \pi(D_n|j)e^{\frac{1}{\mu}v(x_{jn},\beta)}}$$

The expectation over alternative samples of $\ln P(\sigma, \beta|D_n)$ is

$$\sum_{D\subseteq\Omega} \pi(D|\sigma)\ln P(\sigma, \beta|D),$$

where Ω is the universal choice set, and taking the expectation over the possible choices, we have

$$\sum_{\sigma\in\Omega}\sum_{D\subseteq\Omega} P(\sigma)\pi(D|\sigma)\ln P(\sigma, \beta|D), \tag{22}$$

where $P(\sigma)$ is the true choice probablity of σ: $P(\sigma) = P(\sigma, \beta^*) = P(\sigma, \beta^*|\Omega)$. The expectation of (22) over the population Y is then

$$\mathbb{E}_Y \left[\sum_{\sigma \in \Omega} \sum_{D \subseteq \Omega} P(\sigma) \pi(D|\sigma) \ln P(\sigma, \beta|D) \right]. \tag{23}$$

Replacing $P(\sigma)$ using its logit expression in (23), we can write

$$\mathbb{E}_Y \left[\sum_{\sigma \in \Omega} \sum_{D \subseteq \Omega} \frac{\pi(D|\sigma) e^{\frac{1}{\mu} v(x_{\sigma n}, \beta^*)}}{\sum_{j \in \Omega} e^{\frac{1}{\mu} v(x_{jn}, \beta^*)}} \frac{\sum_{j \in D} \pi(D|j) e^{\frac{1}{\mu} v(x_{jn}, \beta^*)}}{\sum_{j \in D} \pi(D|j) e^{\frac{1}{\mu} v(x_{jn}, \beta^*)}} \ln P(\sigma, \beta|D) \right].$$

or, by definition of $P(\sigma|D)$ $(= P(\sigma, \beta^*|D))$,

$$\mathbb{E}_Y \left[\sum_{\sigma \in \Omega} \sum_{D \subseteq \Omega} P(\sigma|D) \frac{\sum_{j \in D} \pi(D|j) e^{\frac{1}{\mu} v(x_{jn}, \beta^*)}}{\sum_{j \in \Omega} e^{\frac{1}{\mu} v(x_{jn}, \beta^*)}} \ln P(\sigma, \beta|D) \right]. \tag{24}$$

The unconditional probability to select the sample D is

$$\pi(D) = \sum_{j \in \Omega} P(j) \pi(D|j),$$

so (24) is equivalent to

$$\mathbb{E}_Y \left[\sum_{\sigma \in \Omega} \sum_{D \subseteq \Omega} P(\sigma|D) \pi(D) \ln P(\sigma, \beta|D) \right].$$

The expectation over the population of the Hessian of (22) is therefore

$$H(\beta) = \mathbb{E}_Y \left[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} \pi(D) P(\sigma|D) \nabla^2_{\beta\beta} \ln P(\sigma, \beta|D) \right],$$

and the expected outer product of the gradient is

$$I(\beta) = \mathbb{E}_Y \left[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} \pi(D) P(\sigma|D) \nabla_\beta \ln P(\sigma, \beta|D) \nabla_\beta \ln P(\sigma, \beta|D)^T \right].$$

We now show that $H(\beta^*) + I(\beta^*) = 0$. For any β, we have

$$\sum_{\sigma \in D} P(\sigma, \beta|D) = 1 \tag{25}$$

We assume that for any β under consideration, $P(\sigma, \beta|D) > 0$ for all $\sigma \in D$. Taking the derivative of (25) with respect to β gives

$$\sum_{\sigma \in D} \nabla_\beta P(\sigma, \beta|D) = 0,$$

and since

$$\nabla_\beta \ln P(\sigma, \beta|D) = \frac{1}{P(\sigma, \beta|D)} \nabla_\beta P(\sigma, \beta|D), \tag{26}$$

we can write

$$\sum_{\sigma \in D} P(\sigma, \beta|D) \nabla_\beta \ln P(\sigma, \beta|D) = 0.$$

The same holds for the expectation over the sampled choice sets:

$$\sum_{D \subseteq \Omega} \pi(D) \sum_{\sigma \in D} P(\sigma, \beta|D) \nabla_\beta \ln P(\sigma, \beta|D) = 0.$$

Now, taking the expectation over the population, we obtain

$$\mathbb{E}_Y \left[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} \pi(D) P(\sigma, \beta|D) \nabla_\beta \ln P(\sigma, \beta|D) \right] = 0.$$

Let's assume that some regularity conditions allowing to permute the expectation and the derivative operators are satisfied. We then have

$$\begin{aligned} 0 = \mathbb{E}_Y \Bigg[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} & \pi(D) P(\sigma, \beta|D) \nabla^2_{\beta\beta} \ln P(\sigma, \beta|D) \\ & + \pi(D) \nabla_\beta P(\sigma, \beta|D) \nabla_\beta \ln P(\sigma, \beta|D)^T \Bigg] \end{aligned} \tag{27}$$

At β^*, using (26), (27) becomes

$$\begin{aligned} & \mathbb{E}_Y \left[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} \pi(D) P(\sigma|D) \nabla^2_{\beta\beta} \ln P(\sigma|D) \right] \\ & = -\mathbb{E}_Y \left[\sum_{D \subseteq \Omega} \sum_{\sigma \in D} \pi(D) P(\sigma|D) \nabla_\beta \ln P(\sigma|D) \nabla_\beta \ln P(\sigma|D)^T \right], \end{aligned}$$

i.e. $H(\beta^*) = -I(\beta^*)$, as announced. In other terms, the information matrix equality holds for sampled choice sets.

5 Numerical results

The numerical results presented in this section are based on simulated observations and real data and aims at analyzing estimation results and providing the corresponding information matrix equality test results using the four different models described in the previous section: PL, PSL, RL with LS and RL without LS. The purpose of using simulated data is to validate the estimation and the information matrix test since the true model is known.

5.1 Network

For both simulated and real observations we use the Borlänge network in Sweden. It is composed of 3077 nodes and 7459 links and it is uncongested so travel times are assumed static and deterministic. There are 21452 link pairs and therefore as many non-zero elements in the M matrix of (13). The sample consists of 1832 trips corresponding to simple paths with a minimum of five links. There are 466 destinations, 1420 different OD pairs and more than 37,000 link choices in this sample. The same data has been used in Frejinger and Bierlaire (2007) and Fosgerau et al. (2013).

The route choice data was collected by GPS monitoring only so there is no socio-economic information on the drivers. We use network specific attributes link travel time $TT(a)$ and turn angle from link k to a. Based on the latter attribute with define left turn $LT(a|k)$ (angle larger than 40 degrees and less than 177 degrees) and u-turn $UT(a|k)$ (angle larger than 177 degrees) dummies.

5.2 Model specifications

In order to have comparable estimation results we use the same specification of the deterministic utilities for the four different models and hence only include link additive path attributes with the exception of path size. The deterministic utilities for link a given state k and path $\sigma_n = (k_0^n, \ldots, k_{I_n}^n)$ are

$$\begin{aligned}
v^{\mathrm{RL}}(a|k) &= \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) \\
v^{\mathrm{RL\text{-}LS}}(a|k) &= \beta_{TT}TT(a) + \beta_{LT}LT(a|k) + \beta_{LC}LC(a) + \beta_{UT}UT(a|k) \\
&\quad + \beta_{LS}LS(a) \\
v^{\mathrm{PL}}(\sigma_n) &= \mu[\beta_{LL}TT(\sigma_i) + \beta_{LT}LT(\sigma_i) + \beta_{LC}LC(\sigma_i) + \beta_{UT}UT(\sigma_i)] \\
&\quad + \ln\left(\frac{r_{\sigma_n}}{q(\sigma_n)}\right) \\
v^{\mathrm{PSL}}(\sigma_n) &= \mu[\beta_{LL}TT(\sigma_n) + \beta_{LT}LT(\sigma_n) + \beta_{LC}LC(\sigma_n) + \beta_{UT}UT(\sigma_n) \\
&\quad + \beta_{PS}PS(\sigma_n)] + \ln\left(\frac{r_{\sigma_n}}{q(\sigma_n)}\right)
\end{aligned}$$

where $TT(\sigma_n) = \sum_{i=0}^{I_n} TT(k_i^n)$, $LT(\sigma_n) = \sum_{j=0}^{I_n-1} LT(k_{i+1}^n|k_i^n)$, $LC(\sigma_n) = \sum_{i=0}^{I_n} LC(k_i^n)$ and $UT(\sigma_n) = \sum_{i=0}^{I_n-1} UT(k_{i+1}^n|k_i^n)$. The term $\ln(r_{\sigma_n}/q(\sigma_n))$ is the sampling correction as presented in Section 3.1.

5.3 Generation of simulated observations

We generate two sets of observations using two different models: RL and RL-LS. In order to be consistent with the underlying choice set assumption, any feasible path in the network can be chosen. Since the RL and PL models are equivalent we can estimate them on the same data but we cannot simulate observations in a real network for PSL since this would require the universal choice set to be known. Note that simulated data for PSL was used in Frejinger et al. (2009) but for a small cycle-free network so that all paths could be enumerated.

We simulate as many observations as there are real observations (1832) using the same OD pairs. We use the deterministic utilities v^{RL} and $v^{\mathrm{RL\text{-}LS}}$ as described in the previous section with the following chosen parameter values $\widetilde{\beta}_{TT}^{obs} = -2.0$, $\widetilde{\beta}_{LT}^{obs} = -1.0$, $\widetilde{\beta}_{LC}^{obs} = -1.0$, $\widetilde{\beta}_{UT}^{obs} = -20.0$ and $\widetilde{\beta}_{LS}^{obs} = -0.2$. These values are the same as the ones used in Fosgerau et al. (2013) and we exclude u-turns by fixing the parameter to a large negative value. The differences with their study are that we simulate data for a model with LS and we use the same OD pairs as in the real data as opposed to only one OD pair. The main focus of this paper is the real data and the simulated observations are mainly used as validation. We therefore use one sample but note that more samples could be used for an in-depth analysis.

5.4 Sampling of path alternatives

In order to estimate PL and PSL, sets of path alternatives D_n need to be sampled. Since the RL model is very efficient for this purpose we use it with a new set of parameters ($\widetilde{\beta}_{TT}^{sampl}$=-1.8, $\widetilde{\beta}_{LT}^{sampl}$=-0.9, $\widetilde{\beta}_{LC}^{sampl} = -0.8$ and $\widetilde{\beta}_{UT}^{sampl}$=-4) and make 50 draws for each observation. The magnitude of the parameter values is smaller than the model used to simulate observations. This is to ensure that a diversity of paths are sampled. Figure 1 shows the frequency of $|D_n|$ over all 1832 choice sets. Some choice sets have very few paths but these correspond to origins and destinations that are close to each other.

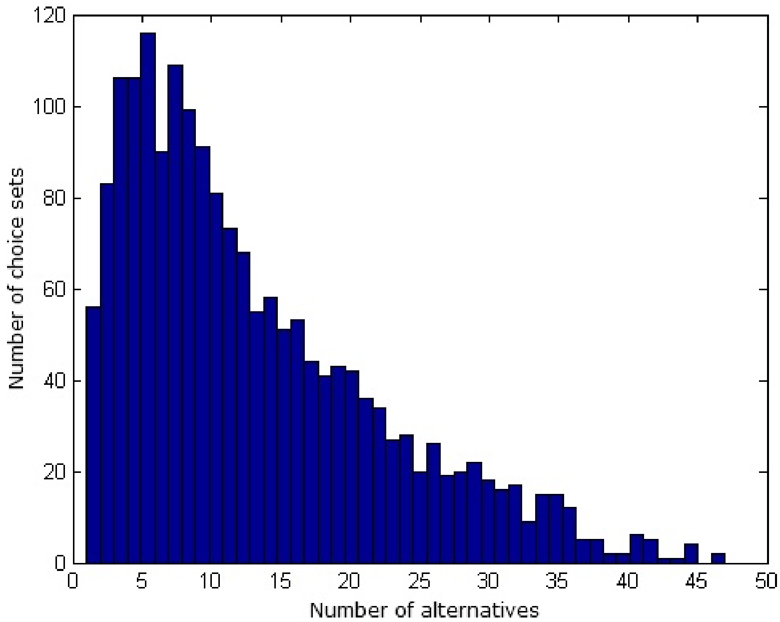

Figure 1: *Histogram of* $|D_n|$ *(50 draws)*

5.5 Estimation results

In this section we present the estimation results for both data sets. Table 1 reports the results for simulated data: parameter estimates, robust standard errors and robust t-tests with respect to zero and the chosen (true) values. As expected the parameter estimates are not significantly different from their true values. The PL model estimated on the data simulated with the RL model is comparable to the estimates of RL which also validates the sampling correction.

Parameters	True value	RL-LS	RL	PL
$\widehat{\beta}_{TT}$	-2.0	-2.033	-1.971	-1.960
Rob. Std. Err.		0.088	0.070	0.070
Rob. t-test (0)		-23.102	-28.157	-28.0
Rob. t-test (-2)		-0.375	0.414	0.571
$\widehat{\beta}_{LT}$	-1.0	-1.000	-1.018	-1.029
Rob. Std. Err.		0.037	0.039	0.039
Rob. t-test (0)		-27.027	-26.103	-26.385
Rob. t-test (-1)		0.000	-0.462	-0.744
$\widehat{\beta}_{LC}$	-1.0	-1.004	-0.995	-0.989
Rob. Std. Err.		0.020	0.020	0.020
Rob. t-test (0)		-50.2	-49.750	-49.45
Rob. t-test (-1)		-0.200	0.250	0.550
$\widehat{\beta}_{LS}$	-0.2	-0.223		
Rob. Std. Err.		0.014		
Rob. t-test (0)		-15.929		
Rob. t-test (-0.2)		-1.643		

Table 1: *Estimation results for simulated observations (β_{UT} fixed to -20)*

We now turn our attention to the estimation results based on real data reported in Table 2. Here we estimate the parameter associated with u-turns β_{UT} because there are some observations with u-turns. Estimating this parameter leads to a significant increase in model fit compared to a fixed parameter of -20 (see results in Fosgerau et al., 2013) while the other parameter estimates are similar for both RL and RL-LS. The parameter estimates are stable for the four models (excluding LS and PS). Furthermore, the model fit for RL-LS is significantly better than RL and the same for PSL compared to PL.

In summary the estimation results for simulated and real observations are as expected and comparable to earlier published results. In the next section we analyze the results of the IM test.

5.6 Information matrix equality test

It is often assumed that paths share unobserved attributes due to the spatial overlap in the network. If true, the utilities are correlated which in turn implies that a MNL route choice model is misspecified. In this section we

Parameters	RL-LS	RL	PSL	PL
$\widehat{\beta}_{TT}$	-3.060	-2.494	-2.738	-2.431
Rob. Std. Err.	0.103	0.098	0.086	0.083
Rob. t-test (0)	-27.709	-25.449	-31.837	-29.289
$\widehat{\beta}_{LT}$	-1.057	-0.933	-1.000	-0.920
Rob. Std. Err.	0.029	0.030	0.027	0.029
Rob. t-test (0)	-36.448	-31.100	-37.037	-31.724
$\widehat{\beta}_{LC}$	-0.353	-0.411	-0.545	-0.429
Rob. Std. Err.	0.011	0.013	0.012	0.013
Rob. t-test (0)	-32.091	-31.615	-45.417	-33.000
$\widehat{\beta}_{UT}$	-4.531	-4.459	-4.366	-4.375
Rob. Std. Err.	0.126	0.114	0.118	0.119
Rob. t-test (0)	-35.960	-39.114	-37.000	-36.765
$\widehat{\beta}_{LS}$	-0.227			
Rob. Std. Err.	0.013			
Rob. t-test (0)	-17.462			
$\widehat{\beta}_{PS}$			1.461	
Rob. Std. Err.			0.082	
Rob. t-test (0)			17.817	
$\widehat{LL}_N(\widehat{\beta})$	-3.300	-3.441	-1.601	-1.688

Table 2: *Estimation results for real data*

test the information matrix equality using (5) and the results show that, indeed, the information matrix equality holds for the simulated data set and it is rejected for the real data set.

In order to apply the information matrix equality test we compute the value $\wp_N$ (5) and the corresponding p-value based on the χ^2_η distribution where η is the degree of freedom. If the p-value does not exceed a given critical value (for instance a significance level of 0.05), we reject the null hypothesis that the information matrix equality holds. The vector $\nabla_\beta D^\eta_N(\hat{\beta})$ requires third order derivatives which we compute by finite difference on the analytical Hessian given by (19) for the RL and RL-LS models and by (12) for the PL and PSL models.

The test results for the diagonal elements (corresponding to the variances) and for the upper triangular elements (matrix is symmetric) are reported in Table 3. The first three rows report the results for the simulated observations. The p-value is close to one for both RL and PL models and the information matrix equality cannot be rejected. Based on these results we also conclude that the sample size seems large enough for these models and that we do not seem to have finite sample issues.

The first row (RL-LS model and simulated data) requires more careful analysis. Indeed, the p-value for RL-LS based on synthetic data is close to the criteria value (0.05) to reject the null hypothesis. This raises a concern about the finite sample issue, which is mentioned in the introduction section. In order to validate if it is the case we increase the number of samples up to 10 times larger (18320 observations, we generate 10 observations per each OD pair taken from the real data). We obtain the p-value for the diagonal test is 0.71 and for the full matrix test is 0.41. They are quite far from the criteria value, meaning that the finite sample issue seems to affect a bit to the validity of the information matrix equality for the RL-LS model.

The fours last rows of Table 3 report the results for the real data. The information matrix equality is strongly rejected for all models. Similar to the case of the simulated observations there is an important difference in magnitude of the test value for the RL-LS model compared to the others. It is interesting to note that the LS and PS attributes significantly improve the model fit but they do not influence the test results.

		Diagonal			Full matrix		
Data	Model	$\bar{\chi}^2_\eta$	η	p-value	$\bar{\chi}^2_\eta$	η	p-value
Simulated data	RL-LS	5.541	4	0.24	17.7	10	0.06
	RL	0.19	3	0.98	0.28	6	0.99
	PL	0.14	3	0.99	0.18	6	0.99
Real data	RL-LS	69.17	5	1.53e-13	159.3	15	0
	RL	21.09	4	3.04e-04	89.16	10	7.8e-15
	PSL	42.63	5	4.39e-08	148.3	15	0
	PL	15.93	4	3.10e-03	63.89	10	6.60e-10

Table 3: *Information matrix test statistic*

Given that the main challenge of applying this test is the computational complexity, we report the computational time in Table 4. The code is in MATLAB and we have used an Intel(R) machine, Core^{TM} i5-3210M CPU 2.50GHz, running Window 8. The machine is multi-core processor but we only use one processor for the computation. We note that the LS attribute is OD dependence and we store them in a sequence of 1832 matrices. Reading data from this sequence increases the cost for log-likelihood computation. It leads that the computational time for the RL-LS model is approximately 5 times longer than for the RL model. The computational time for RL and RL-LS models is higher than for PL and PSL models due to the presence of system of linear equations e.g.(13).

Data	Model	η	Computational time
Simulated data	RL-LS	4	4 hours 21 mins
	RL	3	50 mins
	PL	3	20 seconds
Real data	RL-LS	5	7 hours 21 mins
	RL	4	1 hours 30 mins
	PSL	5	44 seconds
	PL	4	30 seconds

Table 4: *Computational time for IM test*

6 Conclusions

This paper shows the information matrix equality for sampled choice sets and how the IM test can be applied to path- and link-based MNL route choice models. The test statistic contains third derivatives and we address this issue by deriving the analytical Hessian for linear-in-parameters utility functions of MNL and RL so that we can compute third derivatives by finite difference. This test and the estimation algorithm are implemented in MATLAB (this code is available upon request) and can easily be used to test for model misspecification.

We present results for simulated and real observations. As expected, the information matrix equality is not rejected for simulated data but it is strongly rejected for real data. We find that including a path size or LS attribute increases model fit but it does change the outcome of the test. Moreover, the results for simulated data numerically validate our proof for the information matrix equality when the choice sets are sampled.

It is important to note that the IM test is general and does not provide any guidance on the source of misspecification. It can be due to correlated utilities, which is usually assumed to be case in route choice models. However, it could also be a misspecification due to, for example, omitted variables. As mentioned in the introduction other tests could be used for further investigation.

Future research will be dedicated to modeling the correlation and test whether this affect the validity of the information matrix equality. The first step in this direction is the investigation of a nested logit version of the RL model.

Acknowledgement

This research was partially funded by Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Amemiya, T. *Advanced Econometrics*. Harvard University Press, Cambridge, MA, United States of America, 1985.

Ben-Akiva, M. and Bierlaire, M. Discrete choice methods and their applications to short-term travel decisions. In Hall, R. W., editor, *Handbook of Transportation Science*, pages 5–34, Norwell, MA, USA, 1999. Kluwer.

Cascetta, E., Nuzzolo, A., Russo, F., and Vitetta, A. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In Lesort, J.-B., editor, *Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France*, pages 697–711, New York, NY, United States of America, 1996. Pergamon.

Chesher, A. The information matrix test: Simplified calculation via a score test interpretation. *Economics Letters*, 13(1):45–48, 1983.

Davidson, R. and MacKinnon, J. G. A new form of the information matrix test. *Econometrica: Journal of the Econometric Society*, pages 145–157, 1992.

Flotterod, G. and Bierlaire, M. Metropolis-hastings sampling of paths. *Transportation Research Part B*, 48(1):53–66, 2013.

Fosgerau, M. Specification testing of discrete choice models: a note on the use of a nonparametric test. *Journal of Choice Modelling*, 1(1):26–39, 2008.

Fosgerau, M., Frejinger, E., and Karlstrom, A. A link based network route choice model with unrestricted choice set. *Transportation Research Part B*, 56:70–80, 2013.

Frejinger, E. and Bierlaire, M. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 41(3):363–378, 2007.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. Sampling of alternatives for route choice modeling. *Transportation Research Part B*, 43(10):984–994, 2009.

Guevara, C. A. and Ben-Akiva, M. E. Sampling of alternatives in multivariate extreme value (MEV) models. *Transportation Research Part B*, 48(1):31–52, 2013.

Hausman, J. and McFadden, D. Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, pages 1219–1240, 1984.

Horowitz, J. L. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, 61(2):395–411, 1994.

Lancaster, T. The covariance matrix of the information matrix test. *Econometrica: Journal of the Econometric Society*, pages 1051–1053, 1984.

McFadden, D. L. Modelling the choice of residential location. In A. Karlquist et al., editor, *Spatial Interaction Theory and Residential Location*, pages 75–96. North Holland, Amsterdam, The Netherlands, 1978.

Newey, W. K. and McFadden, D. Large sample estimation and hypothesis testing. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 36, pages 2111–2245. Elsevier, Amsterdam, The Netherlands, 1986.

Shapiro, A. Monte carlo sampling methods. In Shapiro, A. and Ruszczyński, A., editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, chapter 6, pages 353–425. Elsevier, 2003.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. *Lectures on Stochastic Programming*. SIAM, Philadelphia, PA, USA, 2009.

White, H. Maximum likelihood of misspecified models. *Econometrica*, 50(1): 1–25, 1982.

Zheng, J. X. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2):263–289, 1996.