Information matrix test for logit based route choice models

Tien Mai * Emma Frejinger * Fabian Bastin *

March 15, 2014

Extended abstract for hEART 2014

*Department of Computer Science and Operational Research, Université de Montréal and CIRRELT, Canada

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), 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.

The main contribution of this paper is that we show how the information matrix (IM) test proposed by White (1982) can be applied for testing MNL route choice models. The IM test 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. For a finite population, this sum asymptotically follows a χ^2 distribution with η degrees of freedom. The test statistic (Theorem 4.1 in White, 1982) has a quite complicated form and contains third derivatives, but we establish that it can be easily computed for popular route choice models. We also review the consistency and asymptotic properties of the estimated parameters.

We now give a brief route choice modelling literature review in order to motivate the choice of models that we consider in the paper. Following the discussion in Fosgerau et al. (2013) we group the existing models into three approaches. First, the classic approach corresponds to path logit (PL) models where choice sets of paths are generated with some algorithm 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. 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. 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, PL and PSL with sampled choice sets and RL with and without the LS attribute. For each of these models, we derive analytical Hessian for linear-in-parameters utility functions. This allows us to compute third derivatives by finite difference and, consequently, the IM test is not only straightforward to apply, but also computationally efficient.

The numerical results are based on the Borlänge network in Sweden. The network is composed of 3077 nodes and 7459 links and the sample of real path observations consists of 1832 trips. In addition to the real data set we generate a set of simulated observations. We simulate as many observations as there are real ones and for the same OD pairs using RL, RL-LS and PL models. Furthermore, we sample choice sets using RL model so that we can know the path sampling probability and correct utilities for the sampling protocol. The estimation results for the simulated data show that the parameter estimates are not significantly different from the chosen values. The parameter estimates for the real data are plausible and similar to already published results on the same data (e.g. Frejinger and Bierlaire, 2007). We also note that the PSL model has a significantly better model fit than PL, and RL-LS has a significantly better model fit than RL.

The IM test results for the diagonal elements (corresponding to the variances) and for all of the elements (we only take the upper triangular elements because the matrix is symmetric) are reported in Table 1. For each sample and each model we present the test statistic value $\bar{\chi}^2_\eta$, the degrees of freedom η and the p-value. If the latter 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. Based on the results we do not reject the null hypothesis for simulated data and RL and PL. There are some numerical issues in the LS dimension (RL-LS model) when computing the Hessian and this is clearly shown in the results with a smaller p-value (0.06). This issue needs further investigation. All models are strongly rejected for the real data. It is interesting to note that the attributes designed to correct for correlation, path size and LS, significantly improve the model fit but they do not affect the result of this misspecification test.

Acknowledgement

This research was partially funded by NSERC.

Data	Model	Diagonal			Full matrix		
		$\bar{\chi}^2_\eta$	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 1: Information matrix test results

References

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

Flotterod, G. and Bierlaire, M. Metropolis-hastings sampling of paths. *Transportation Research Part B*, 48(1):53 – 66, 2013. ISSN 0191-2615.

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

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

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

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