Estimating flexible route choice models using sparse data

Anders Karlström *

Dept. of Transport Science,

Royal Institute of Technology, Stockholm, Sweden

Marcus Sundberg

Dept. of Transport Science,

Royal Institute of Technology, Stockholm, Sweden

Masoud Fadaei Oshyani

Dept. of Transport Science,

Royal Institute of Technology, Stockholm, Sweden

*Email: anders.karlstrom@abe.kth.se

1 Extended Abstract

A growing range of data gathering techniques is increasingly available to be used in transportation planning. For example, GPS and other nomad devices are used for the collection of spatial-temporal data regarding individuals. Individuals' whereabouts are not registered at every instant; rather they are sampled in space and/or time. For many different applications, such as route guidance, traffic management or collaborative filtering applications, sparse data provides a challenge since the path between the recorded observations cannot easily be detected. In addition, measurement errors provide yet another challenge.

The purpose of this paper is to develop a consistent estimator for a flexible route choice model for the case when we have sparse GPS data with measurement errors. That is, we do not have access to the paths chosen by the individuals (or vehicles), and the sampled points are associated with a spatial uncertainty.

Different approaches have been suggested to deal with the issues related to choice sets. McFadden [11] suggested estimating multinomial logit (MNL) models on subsets of

alternatives. Frejinger et al. [7] explained that estimations based on subsets could return biased results and how to correct for the bias in MNL models. Considering the issue of correlation between overlapping paths, path-based approaches using discrete choice methods are elaborated in the literature. (See, e.g., [1], [2], and [3]).

Alternatively, one may view a route as a sequence of links, see Dial [4]. Link costs are random and the summation of link costs gives the cost of the paths. In practice, Fosgerau et al. [5] and Karlström et al. [9] use such link-based route cost models and assign the random terms to the links rather than the paths. These approaches incurs a natural correlation structure between paths, but also allow for the consideration of the full choice set. The link-based approach is used in this paper as well, and the route choice is provided by choosing the shortest path. The estimator proposed in this paper enables the estimation of flexible route choice models provided only sparse GPS data. Most closely related to our work are Frejinger and Bierlaire [6] and Newman et al. [12], who also consider GPS data for route choice estimation. Yet, our approach allows for the estimation of models with a flexible correlation structure, full choice set consideration, and bias corrects for data manipulations.

GPS data has many virtues. To use such a data set, some technical obstacles need to be addressed. In a route choice context, the data need to be associated with a transportation network, usually by some map matching technique. For the purpose of this paper, the map-matching technique of Krumm et al. [10] is used. The path with the minimum calculated distance, from the observed GPS points, is stated the matched path.

The estimation procedure that is proposed in this paper is based on indirect inference. Indirect inference is a simulation based estimation method, see Gourieroux et al. [8] and Smith [13]. The methodology is applicable even to, analytically intractable, or too difficult to evaluate likelihood functions, which is the case for our flexible route choice model. Like other simulation-based methods, a major prerequisite of the indirect inference approach is that it should be possible to simulate data from the model of interest for different values of the parameters involved in the model.

The main characteristic of the indirect inference method is that it uses an approximate or auxiliary model in order to form a criterion function. The aim of the indirect inference is to find parameters for the model of interest such that the simulated and observed data look the same from the auxiliary model's point of view. Karlström et al. [9] proposed an indirect inference estimator, for the estimation of flexible route choice models, in this paper the methodology is extended to incorporate the use of sparse GPS data.

Monte-Carlo evidence is provided, showing that the estimation procedure regenerates consistent parameter estimates. We use a network of Borlänge city as described by Frejinger et al. [7].

Given the network and the model, we may assign a value to the underlying behavioral parameters β, (say $\beta = 1$) and simulate paths. Given paths we simulate GPS observations along the paths. Then, the proposed estimator based on indirect inference including map-matching is applied. We investigate if the estimated parameter is consistent with the assigned parameter of the data generating process (i.e. 1).

The observed paths contains $N = 3000$ routes which are made based on $\beta = 1$.

GPS error(m)	0	10	25	50
Mean	0.99	0.98	1.01	1.03
Std	0.03	0.03	0.04	0.03
RMSE	-0.33	-0.48	0.17	1.10
ZETA	0.03	0.03	0.04	0.04

Table 1: Sampling interval of 30 seconds.

GPS error(m)	0	10	25	50
Mean	1.02	1.05	1.01	1.24
Std	0.03	0.04	0.04	0.48
RMSE	0.82	1.09	0.15	0.49
ZETA	0.04	0.06	0.04	0.52

Table 2: Sampling interval of 120 seconds.

In Table 1 and 2 we report the Monte Carlo evidence with a sampling interval of 30 and 120 seconds. The true parameter is assumed to equal 1. GPS errors are assumed to have a Gaussian distribution with standard deviation 0, 10,25, or 50m. As is evident, the estimator is quite precise also with rather large measurement errors.

We also apply the method to real route choice data. The data consists of actual reported route choices. Given these actual route choices we may simulate GPS observations and apply the estimator. Furthermore, the estimator may be applied without GPS-errors, and in that case should replicate estimates performed directly on actual route choice data without GPS related errors and sparseness. In Table 3 we see that the estimator performs well.

Note: We are currently working on calculations of robust standard errors, this work is under completion and will be included in the paper.

Data collection	Estimation	
period(sec)	0 meters error	10 meters error
0	1.27	-
30	1.34	1.23
60	1.23	1.25
90	1.19	1.09
120	1.16	1.14

Table 3: Monte Carlo evidence with GPS points sampled with an interval of 90 seconds. GPS errors is assumed to have a Gaussian distribution with standard deviation 0, 10m.

The main conclusion is that indirect inference is a useful option in the tool box for flexible route choice estimation which can be used for estimating path choice using low frequency GPS sampling data with measurement errors.

References

[1] Ben-Akiva, M. and Bierlaire, M., (1999). "Discrete choice methods and their applications to short term travel decisions" *Chapter for the Transportation Science Handbook.*

[2] Ben-Akiva, M. and Ramming, S., (1998). "Lecture notes: Discrete choice models of traveler behavior in networks" *Prepared for Advanced Methods for Planning and Management of Transportation Networks*, Capri, Italy.

[3] Cascetta, E., Nuzzolo, A., Russo, F. and Vitetta, A., (1996). A modified "logit route choice model overcoming path overlapping problems. Spec- ification and some calibration results for interurban networks, in J. B. Lesort (ed.)" *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France.

[4] Dial, R.B., (1971). "A probabilistic multipath traffic assignment model which obviates path enumeration" *Transportation Research*, 5(2):83–111.

[5] Fosgerau, M., Frejinger, E. and Karlström, A., (2011). "A logit model for the choice among infinitely many routes in a network" *Technical report*, Center for Transport Studies, Royal Institute of Technology KTH, Stockholm, Sweden.

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

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

[8] Gourieroux, C., Monfort, A. and Renault, E., (1993). "Indirect inference" *Journal of Applied Econometrics*, 8(S1):S85–S118.

[9] Karlström. A., Sundberg, M., Wang, Q., (2011). "Consistently estimating flexible route choice models using an MNL lens" *International Choice Modelling Conference*, Leeds.

[10] Krumm J., Letchner J., Horvitz E., (2007). "Map matching with travel time constraints" (Paper 2007-01-1102) *Society of automotive engineers (SAE) 2007 world congress*, Detroit, MI, USA.

[11] McFadden, D., (1978). "Modeling the choice of residential location" *Cowles foundation discussion paper no. 477.*

[12] Newman, J., Chen, J., Bierlaire, M., (2009). "Generating probabilistic path observation from GPS data from route choice modeling" *Proceedings of the European Transport Conference ETC.*

[13] Smith, Anthony A., Jr., (2008). "Indirect inference" *The New Palgrave Dictionary of Economics Online, Palgrave Macmillan* DOI:10.1057/9780230226203.0778.