A maximum likelihood approach to estimate a bounded route choice model

Thomas Kjær Rasmussen*, David Paul Watling, Otto Anker Nielsen
*corresponding author: tkra@dtu.dk.

1. Introduction

It is crucial to have available a well-functioning transport model which gives a realistic representation of the route choices of travellers when evaluating potential infrastructure investments or policies. Traditional discrete choice models applied for route choice, such as the Multinomial Probit or Multinomial Logit and variants thereof, are based on an unbounded distribution of the error-term and thereby assign a non-zero probability to all alternatives, no matter how irrelevant they may be. Thus, when the parameters of such models are estimated through maximum likelihood, we in principle need to enumerate all possible non-cyclic alternatives, in order to calculate the choice probabilities. Full enumeration is unfeasible even for medium-sized networks, and consequently in practice a subset is typically identified and used to calculate the probability of the observed route (e.g., Bekhor et al., 2006; Prato et al., 2014; Rasmussen et al., 2016). This potentially induces inconsistencies and bias in the estimated parameters. Fosgerau et al. (2013) proposed a link-choice based approach that avoids full path enumeration, but this is at the price of assuming that infinite cycles are permitted, which seems not very realistic.

Recently, a stream of research has addressed an issue of existing, commonly-used choice models (including those described above), that all paths are assigned a non-zero probability, no matter how irrelevant they are (Watling et al., 2015, 2018; Rasmussen et al., 2015). To address this, Watling et al. (2018) proposed a novel choice model, the Bounded Choice Model (BCM), alongside a corresponding set of conditions for a Bounded SUE network equilibrium. Derived from a Random Utility Theory framework and with a continuous mapping of the probability, the model introduces a *bound* on the random utility of the set of alternatives used, allocating zero probability to alternatives violating the bound. Thereby, the model consistently identifies the choice sets by predicting which paths/alternatives are used and unused, and allocates flow to ensure accordance to the assumed Discrete Choice Model.

Procedures as to how the BCM may be estimated have yet to be proposed. This paper accordingly present a procedure for consistent estimation of the bounded choice model by maximum likelihood, using observed route choices tracked passively through GPS or mobile phone devices. A likelihood-function consistent with the BCM is formulated and a corresponding estimation procedure is subsequently applied to two simulation experiments, allowing us to investigate whether parameters of a 'true model' from which observations are simulated can be reproduced.

2. Bounded Choice Model and Likelihood formulation

2.1. Bounded Choice Model

In this subsection, the BCM is briefly introduced in a route choice context[1]. Consider a network with *M* OD-relations indexed by *m* (*m*=1,2,…,*M*). Define the set R_m as all simple paths (without cycles) for each OD-pair *m*. Each path $r \in R_m$ is associated with a cost c_{mr}, which is a weighted sum (by parameter vector $\boldsymbol{\beta}$) of variables **x,** i.e. $c_{mr}(\mathbf{x}, \boldsymbol{\beta})$**.** Applying a bound δ_m to the differences in random utility for routes to receive positive flow induces the following continuous choice probability expression P_{mr} of path *r* for OD-relation *m*:

$$P_{mr}(\mathbf{c}(\mathbf{x}, \boldsymbol{\beta}), \theta, \boldsymbol{\delta}) = \frac{\left(\exp\left(-\theta(c_{mr} - \min(c_{ms}: s \in R_m) - \delta_m)\right) - 1\right)_+}{\sum_{t \in R_m}\left(\exp\left(-\theta(c_{mt} - \min(c_{ms}: s \in R_m) - \delta_m)\right) - 1\right)_+} \qquad (r \in R_m;\ m = 1,2,\dots,M)$$

where θ is a scale parameter of the BCM.

The model consistently predicts which paths are used and unused, i.e. the choice sets are equilibrated, by allocating zero probability/flow to paths violating the cost bound, while ensuring that flow are allocated according to a discrete choice model among paths not violating the cost bound.

2.2. Likelihood formulation

Imagine we have available a set of observed routes, e.g. collected through GPS units or smart phones, and consider a situation where it is not needed to distinguish individuals in their preferences (the approach is, of course, readily generalised to permit multiple user classes differing in their parameters). For each observation *k*=1,2,…,*N*, let R_k define the set of all non-cyclic paths between the origin and destination of observation *k*. Suppose the data are contained in a vector of size N with elements:

$$x_k = \{i \quad \text{if alternative } i \in \{1,2,\dots,R_k\} \text{ is chosen}\} \quad (k = 1,2,\dots,N).$$

By assuming a generic cost bound $\delta_m = \delta$, the likelihood for sample of size N can be formulated as:

$$L(\boldsymbol{\beta}, \theta, \delta) = \prod_{k=1}^{N} P_{x_k}(\mathbf{c}(\mathbf{x}, \boldsymbol{\beta}), \theta, \delta)$$

$$= \begin{cases} 0 & if\ c_{x_k}(\mathbf{x}, \boldsymbol{\beta}) - \min(c_s(\mathbf{x}, \boldsymbol{\beta}): s \in R_k) \geq \delta\ for\ any\ k \\ \displaystyle\prod_{k=1}^{N} \frac{\exp\left(-\theta\left(c_{x_k}(\mathbf{x}, \boldsymbol{\beta}) - \min(c_s(\mathbf{x}, \boldsymbol{\beta}): s \in R_k) - \delta\right)\right) - 1}{\sum_{t \in R_k}\left(\exp\left(-\theta(c_t - \min(c_s: s \in R_k) - \delta)\right) - 1\right)_+} & else \end{cases}$$

The lower part of the last equality will always be positive, and thus a maximum likelihood estimate will always be on this part. This means that the maximum likelihood estimate will also lie on the log-likelihood of this part, and hence the log-likelihood to be maximized becomes:

$$LL(\boldsymbol{\beta}, \theta, \delta) = \ln\left(\prod_{k=1}^{N} P_{x_k}(\mathbf{c}(\mathbf{x}, \boldsymbol{\beta}), \theta, \delta)\right) = \sum_{k=1}^{N} ln\left(\frac{\exp\left(-\theta\left(c_{x_k}(\mathbf{x}, \boldsymbol{\beta}) - \min(c_s(\mathbf{x}, \boldsymbol{\beta}): s \in R_k) - \delta\right)\right) - 1}{\sum_{t \in R_k}\left(\exp\left(-\theta(c_t - \min(c_s: s \in R_k) - \delta)\right) - 1\right)_+}\right).$$

[1] See Watling et al. (2018) for more details on the derivation and theoretical properties of the model.

3. Simulation study

The formulated likelihood function is investigated in a simulation study, evaluating the likelihood-surface and the possibility to estimate reasonable parameters that reproduces observed behaviour.

3.1. Experiment setup

In general, the approach samples observations according to an assumed 'true' model, and then use these in combination with the log-likelihood function to evaluate the ability to reproduce the assumed 'true' parameters. The simulation study consists of three steps, namely:

(i) Postulate true BCM choice model including specification and parameter values. For each relevant OD-relation, identify corresponding set of choice sets[2] to be used in the estimation.

(ii) Sample a set of observed routes according to the true model using the equilibrated network costs and choice sets identified in *(i)*.

(iii) Apply maximum likelihood estimation approach to obtain parameter estimates based on observed speeds, sampled observed routes and generated (universal) choice sets.

The generic estimation procedure is detailed in the following pseudo-code:

Algorithm 1: Pseudo-code for simulation experiments

Step 1: Initialisation

1.1 Assume true parameters $(\boldsymbol{\beta}, \theta, \delta)$ of the BCM and solve for corresponding BCM SUE using these. Based on equilibrated network costs[3], generate the universal choice set and store the associated route attributes and choice probabilities (in true model).

1.2 Based on the choice probabilities of routes in the universal choice set (obtained in 1.1), sample *N* observed routes.

1.3 Define initial set of parameter values $(\tilde{\boldsymbol{\beta}}, \tilde{\theta}, \tilde{\delta})_1$ $\boldsymbol{\beta}_1$ for MLE (starting point), set *i*=1.

Step 2: Recalculate choice probabilities and LL

Calculate route costs $c_{x_k}(\mathbf{x}, \tilde{\boldsymbol{\beta}}_i)$ using $\tilde{\boldsymbol{\beta}}_i$ and recalculate choice probabilities of the observed routes $P_{x_k}(\mathbf{c}(\mathbf{x}, \tilde{\boldsymbol{\beta}}_i), \tilde{\theta}_i, \tilde{\delta}_i)$. Based on these, calculate LL_i[4].

Step 3: Compute new set of parameters

Based on LL_k and the associated parameters $(\tilde{\boldsymbol{\beta}}, \tilde{\theta}, \tilde{\delta})_k$ for all $k \leq i$, compute a new set of parameters $(\tilde{\boldsymbol{\beta}}, \tilde{\theta}, \tilde{\delta})_{i+1}$ to test in the following iteration.

---

[2] Note that the estimation procedure in principle needs to enumerate and store the universal choice sets in order allow to evaluate the log-likelihood for very large values of the bound δ (e.g. in cases when BCM approaches MNL). For larger networks this is not feasible and a subset consisting of paths with costs below some value being considerably larger than the bounding cost of the true model can be used.

[3] In the simulation experiments, the network costs are defined by the costs in an equilibrated network, based on some assumed parameter values. In general, it is not required to use such an approach, the network costs might as well be defined otherwise (e.g. using observed travel times or some other assumed values).

[4] In cases where an observed route is assigned zero probability, i.e. violates δ_n of the current iteration, the log-likelihood contribution of this is set at a large negative value (-999) as ln(0) is not defined.

Step 4: Stopping criterion If | $LL_i - LL_{i-1}$ | < ε, Stop. Else set i=i+1 and return to *Step 2*.

The number of observed routes to sample (N) is exogenously defined. In general, step 3 could apply procedures from standard numerical maximization methods to identify the parameter-adjustment to evaluate in the next iteration, e.g. gradient approaches such as Newton-Raphson or BHHH.

Note that *Algorithm 1* computes one set of parameter estimates, estimated from one set of sampled routes. Unlike for some other choice models such as the MNL, it is not possible to calculate the standard error of the estimates analytically for the BCM. This is because the model violates the regularity conditions that establish asymptotic standard errors of the MLEs as the inverse of the Fisher information. Instead, the robustness of the parameters estimated can be investigated numerically by applying *Algorithm 1* multiple times and then analysing the variation of the estimated parameters.

3.2. Case Studies: Sioux Falls and Anaheim

The estimation experiments use two case study networks that are often used to study transportation network problems, namely the Sioux Falls network and the Anaheim network[5]. The Sioux Falls network consists of 76 links and 528 OD-relations between which there is non-zero demand. The corresponding numbers are 914 and 1406 for the Anaheim network.

N is specified as N=500 and N=1400 observations for the Sioux Falls and Anaheim networks, respectively, and the origin and destination for each observed route is simulated randomly among the OD-relations with non-zero demand. For the Anaheim network, the effect on estimates of varying the number of observations N is also investigated (see section 4.2.2).

The cost c_{nt} of alternative t for observation n is specified as a weighed sum of the length l_{nt} and the (congested) travel time t_{nt}, i.e.

$$c_{nt} = \beta_l \cdot l_{nt} + \beta_t \cdot t_{nt}$$

Rather than applying some approach to adjust the parameters in Step 3, the following applications apply a fine-grained tracing of the log-likelihood surface and then identify the parameter-configuration inducing the maximum log-likelihood as solution.

[5] The transportation network test problems have been downloaded from https://github.com/bstabler/TransportationNetworks, where it is also possible to find detailed descriptions of the networks.

4. Results

4.1. Sioux Falls application

The model requires the specification of four parameters, namely δ, θ, β_t and β_l. To ensure identification, β_t is fixed at β_t=0.7 throughout.

4.1.1. Investigating two parameters, δ and θ

In this subsection the focus is on the two parameters δ, θ, assuming $\beta_l = 0.5$. Figure 1A illustrates the distribution of cost differences to the corresponding minimum cost path for one set of simulated observed paths (*Step 1* in *Algorithm 1*) when assuming δ=*15* and θ=*0.2*. As can be seen, most sampled paths correspond to the minimum cost path (68%), while there is also observed paths that have costs that are considerably different from the cost on the corresponding cheapest path. Figure 1B illustrates the distribution of the largest of such cost differences across 100 simulation experiments (applications of *Algorithm 1)* when δ=*15* and θ=*0.2.* For that parameter configuration it is clear that almost all sets of sampled paths contains at least one path with a cost close to the cost bound. Figure 1C and 1D illustrate, for δ=*25*, the influence of varying the scale parameter of the BCM. Comparing the two it can be seen that the distribution of the maximum cost difference to the corresponding minimum cost path is highly dependent on the choice of the scale parameter θ. This is expected, as increasing θ corresponds to increasing the perception error. Intuitively, estimating δ for large θ should be more difficult, as there are fewer observations with costs close to the cost bound.

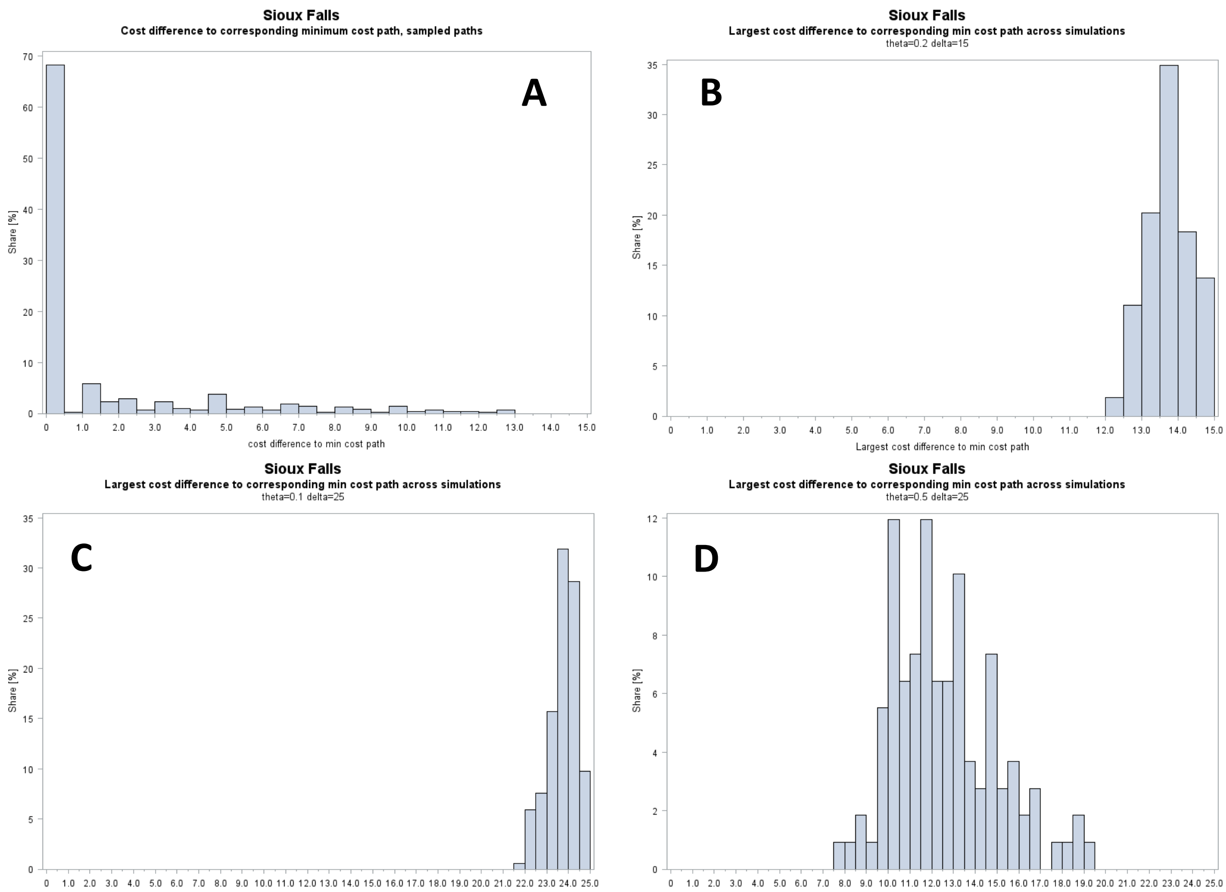

Figure 1 – Sioux Falls application. A: Distribution of cost differences to corresponding minimum cost across observations, $\boldsymbol{\delta}$=*15* and $\boldsymbol{\theta}$=*0.2*. B: Distribution of largest cost difference to corresponding minimum 100 simulation experiments, $\boldsymbol{\delta}$=*15* and $\boldsymbol{\theta}$=*0.2.* C: Distribution of largest cost difference to corresponding minimum 100 simulation experiments, $\boldsymbol{\delta}$=*25* and $\boldsymbol{\theta}$=*0.1.* D: Distribution of largest cost difference to corresponding minimum 100 simulation experiments, $\boldsymbol{\delta}$=*25* and $\boldsymbol{\theta}$=*0.5.*

Evaluating the likelihood function at various parameter-configurations enables to visualise the surface of the log-likelihood function for one set of simulated routes. Assuming true parameters δ=15 and θ=0.2 in the simulation of a set of observed routes, Figure 2 visualise the surface of the objective function.

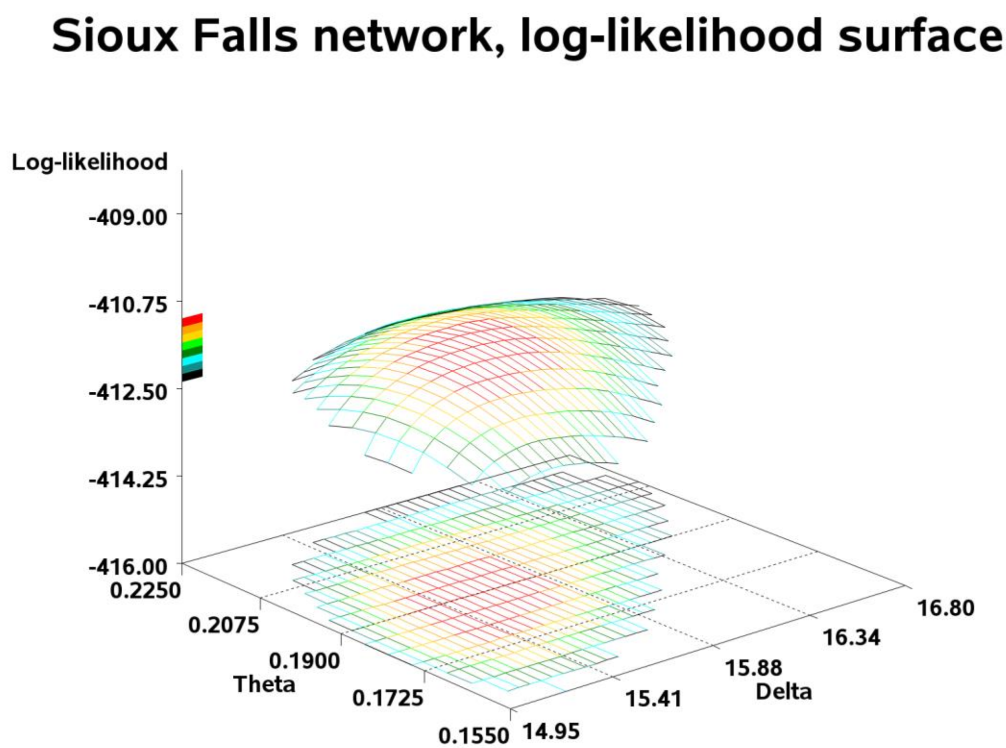

Figure 2 - Log-Likelihood surface, Sioux Falls network simulation experiment with δ=15 and θ=0.2.

As can be seen, the surface is smooth and increasing around the true parameters. The estimated parameters are reasonably close to the true parameters, namely $\tilde{\delta} = 15.52$ and $\tilde{\theta} = 0.205$.

Applying the procedure outlined in *Algorithm 1* numerous times allows analysing the stability of the estimated parameters. Table 1 reports, for various settings of the true parameters, the mean value, standard error and the Mean Squared Error (MSE) of the estimates across *m* applications of *Algorithm 1*.

Table 1 - Estimation results for two parameters, multiple experiments Sioux Falls network.

δ_{true}	θ_{true}	m	$\tilde{\delta}$			$\tilde{\theta}$		
			μ	σ	MSE	μ	σ	MSE
15	0.2	100	14.731	0.0572	0.0756	0.193	0.0023	5.43E-05
25	0.5	100	22.612	0.1606	5.7283	0.499	0.0024	6.76 E-06
25	0.1	100	24.850	0.0561	0.0256	0.098	0.0011	5.21E-06

The mean estimate of θ is close to the true value for all settings tested and the MSE is low. The estimate of δ is also reasonable in the settings tested. However, the accuracy varies, with decreasing accuracy as θ increases. This seems reasonable, as increasing θ corresponds to lower perception error and thus fewer simulated observations close to the bound (as illustrated in Figure 1).

4.1.2. Multiple parameters

In this section introduces the results of estimating three parameters, namely θ, δ, β_L. Assuming true parameters δ=15, θ=0.2 and β_L=0.5 when sampling the set of observed routes and applying *Algorithm 1* multiple times induces the estimates shown in Table 2.

Table 2- Estimation results for three parameters, multiple experiments Sioux Falls network.

	m	μ	σ	MSE
$\tilde{\theta}$	150	0.192	0.0024	6.976E-05
$\tilde{\delta}$	150	14.797	0.0661	0.0456
$\tilde{\beta}_L$	150	0.509	0.0070	0.0001

As can be seen, adding an additional parameter to estimate (compared to section 4.1.1) does not seem to hinder accurate reproduction of the assumed true parameters.

4.2. Anaheim application

For the Anaheim application, the cost function includes three parameters, namely δ, θ and β_t. To ensure identification, β_t is fixed at $\beta_t = 1.0$ throughout.

4.2.1. Investigating two parameters, δ and θ

Sampling a set of observed routes assuming true parameters $\delta = 10$ and $\theta = 0.2$ induces a likelihood surface as shown in Figure 3.

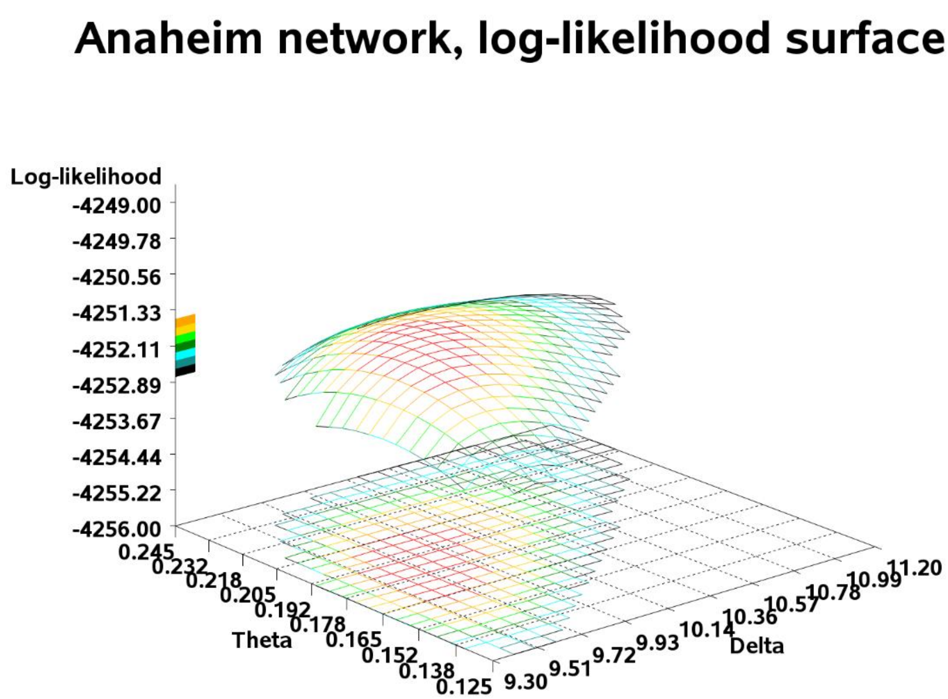

Figure 3 - Log-Likelihood surface, Anaheim network simulation experiment with δ=10 and θ=0.2.

The surface is smooth and increases to a top point reasonably close to the value of the true parameters, the estimates being $\tilde{\delta} = 9.8$ and $\tilde{\theta} = 0.190$.

Repeating *Algorithm 1* 500 times produces the distribution of estimated parameters illustrated in Figure 4. As can be seen, there is some variation in the estimates, but the majority of the set of estimated parameters are within vicinity of the assumed true parameters.

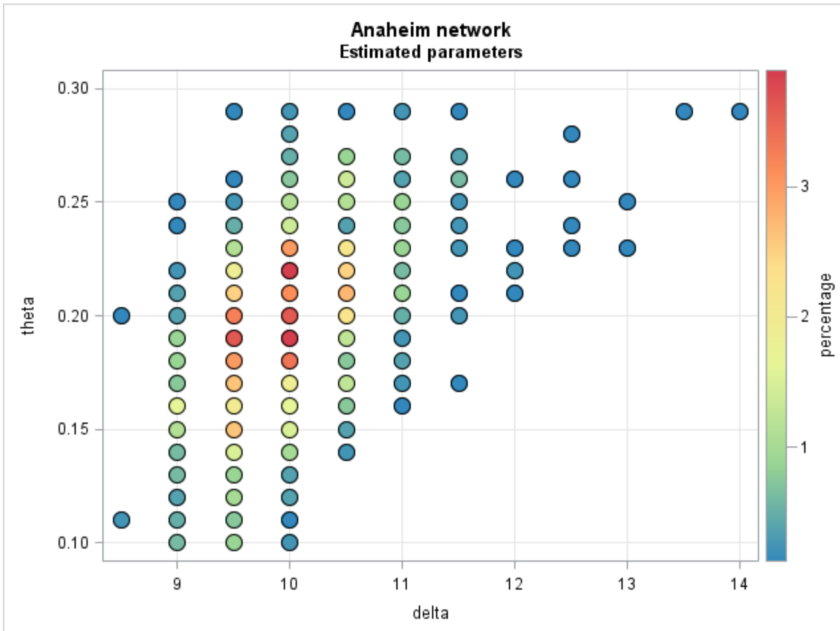

Figure 4 - Distribution of estimated parameters across multiple applications, Anaheim network simulation experiment with δ=10 and θ=0.2.

The data underlying the figure above and corresponding data for two other sets of assumed true parameters induces the statistics shown in Table 3.

Table 3 - Estimation results, multiple experiments Anaheim network.

δ_{true}	θ_{true}	m	$\tilde{\delta}$			$\tilde{\theta}$		
			μ	σ	MSE	μ	σ	MSE
10	0.2	500	0.196	.00142	2.02E-05	9.994	0.0239	6.04E-4
20	0.2	500	0.195	.00079	2.61E-05	20.232	0.1325	7.13E-2
25	0.4	500	0.399	.00059	7.04E-07	24.600	0.1764	0.19

The table demonstrates that it is also possible to reproduce true parameters on a network being considerably larger than the Sioux Falls network. Rather small MSE values are produced, and again it is indicated that the smaller the perception error is, the less 'stable' the estimate of the bound becomes – the estimate is however still rather stable at the smallest level (highest θ).

4.2.2. Varying number of observations

This section investigates the effect of varying the number of observations. It is expected that increasing the number of observations should induce mean estimates being closer to the true value of the parameters. For example, increasing the number of simulated routes should produce more routes with a cost difference to the minimum cost path that are close to the bound. This should improve the estimate of the bound, as the experiences from varying the scale (see above) indicating that it is easier to estimate the bound with observations close to the bound. Assuming true parameters $\delta = 10$ and θ=0.2, Figure 5 illustrates the mean estimates across 100 simulation experiments (applications of *Algorithm 1)* for various values of the number of simulated observed routes in each (*N*). I.e., each point in Figure 5 is based on 100 applications of *Algorithm 1*.

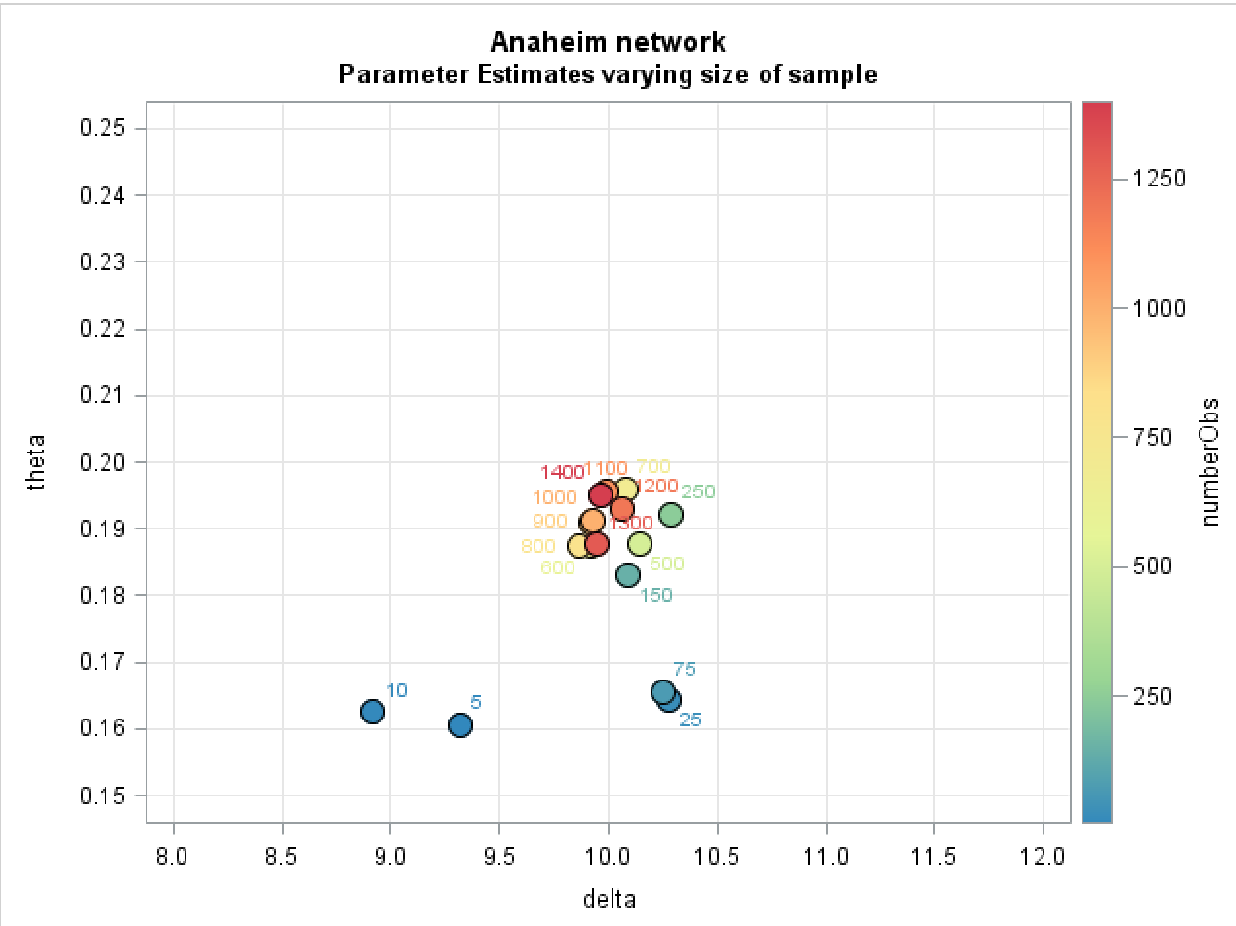

Figure 5 - Mean estimates across 100 applications for varying number of simulated observations, Anaheim network with δ=10 and θ=0.2.

The figure confirms the expected, namely that the accuracy of the estimates increases as the number of simulated observations increase. For the configuration of the model analysed, it seems that the estimates are reasonably close to the true values when $N \geq 150$.

4. Discussion and Conclusions

Watling et al (2018) developed a new discrete choice model that allow assigning zero probability to alternatives that has a random utility that are below some bound to the maximum random utility alternative. Such a characteristic is in particular attractive for route choice modelling, where the universal choice set is enormously large. This paper formulates a corresponding likelihood measure to use in the estimation of this model, based upon route choice information collected by e.g. GPS data. This likelihood measure was applied to analyse the ability to reproduce assumed true parameter-settings in simulation studies across two case-studies, namely the Sioux Falls and the Anaheim network. Overall, the results indicate that it is generally possible to reproduce the true parameters, even for relatively small sample sizes, and that the estimates are stable. Relating to the estimation of the bound, it is found that it is generally more difficult (albeit reasonably good estimates are still produced) to obtain a good estimate of this when the cost of none of the observed paths are close to the cost bound, e.g. due to a low perception error.

Having demonstrated promising results for the simulation study, an interesting further extension will be to apply the estimation procedure to a real-world case, where the observations are obtained from GPS data collected by trackers or mobile phones. If data allows, it should be feasible to estimate individual- or trip-purpose specific parameters and bounds, particularly if the same individuals are observed over a long period. A possible complication however arises, as one might risk observing routes with costs above the cost bound of the individual in cases where data is collected across many individuals and a long time horizon in a large complex network. Such observations might occur due to unexpected disruptions in the network, travellers getting lost, etc. In such cases, using a lower bound in the choice model induces zero probabilities of observed choices (and thus a likelihood of zero). This property of the choice model drives the estimated bound value to be above the largest cost difference, as this will generate a positive likelihood. Thereby the approach is highly sensitive to outliers, as a single rogue observation can skew the estimation considerably. To address this issue, one could perform outlier analysis to identify and remove such undesired observations from the set of observed routes. One could e.g. do this either based on the distribution of the cost differences to the corresponding minimum cost routes or based on repeated estimation of the model parameters where the observations with the largest cost differences are sequentially removed until removing additional observations does not influence the estimates.

The findings of the simulation study indicate that it is feasible to estimate the BCM in a route choice setting. The BCM is however a general discrete choice model which may also be applied to other choice situations, such as destination choice where there are also numerous alternatives. In future research, it could be interesting to explore whether the likelihood formulation and estimation procedure outlined for route choice could apply to such choice situations.

References

Bekhor, S., Ben-Akiva, M., Ramming, M. S., 2006. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144 (1), 235–247.

Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unresticted choice set. *Transportation Research Part B: Methodological*, 56 (1), 70-80.

Prato, C.G., Rasmussen, T.K., Nielsen, O.A., 2014. Estimating Value of Congestion and of Reliability from Observation of Route Choice Behavior of Car Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, 2412, 20-27.

Rasmussen, T.K., Anderson, M.K., Nielsen, O.A., Prato, C.G., 2016. Timetable-based simulation method for choice set generation in large-scale public transport networks. *European Journal of Transport and Infrastructure Research,* 16(3), 2016, 467-489.

Rasmussen, T.K., Watling, D.P., Prato, C.G., Nielsen, O.A., 2015. Stochastic user equilibrium with equilibrated choice sets: Part II –Solving the restricted SUE for the Logit Family. *Transportation Research Part B: Methodological*, 77, 166–181.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2015. Stochastic User Equilibrium with Equilibrated Choice Sets: Part I – Model Formulations under Alternative Distributions and Restrictions. *Transportation Research Part B: Methodological*, 77, 146–165.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2018. Stochastic user equilibrium with a bounded choice model. *Transportation Research Part B: Methodological*, 114, 254-280.