Estimating the perturbed utility route choice model with individual-level data

Mogens Fosgerau*[1], Nikolaj Nielsen[1], Mads Paulsen[2], Thomas Rasmussen[2], and Rui Yao[3]

[1]University of Copenhagen, Denmark
[2]Technical University of Denmark, Denmark
[3]EPFL, Switzerland

Short summary

We provide an estimator for the perturbed utility route choice (PURC) model that works with data at the level of individual trips. The estimator, labeled microPURC, is a nested fixed point algorithm that combines a bias-corrected linear regression problem with the individual-level perturbed utility maximization problem. We establish the statistical properties of the microPURC estimator and confirm these results with an experiment using simulated data. We then [this is work in progress] apply the microPURC estimator to a dataset from the ongoing Danish national road pricing experiment. This comprises a large number of individual trip trajectories, the national road network, and a choice set that includes all possible paths through the network.

Keywords: transport network modelling; big data analytics; discrete choice modelling; perturbed utility, microPURC

1 Introduction

This paper considers the problem of estimating a perturbed utility route choice (PURC) model with trip-level data, using a large dataset of individual GPS traces through a large network.

The perturbed utility route choice model Fosgerau et al. (2022),Fosgerau et al. (2023) has several advantages over competing route choice models. In particular, it does not require any choice set generation but simply uses the complete network as it is. It generates realistic substitution patterns directly from the network structure. Most of the network is inactive for any origin-destination pair.

If one is willing to aggregate data to the level of OD flows, the perturbed utility model is also very fast and easy to estimate using linear regression. However, in general, this involves data loss. It also limits the possibilities for incorporating individual-level information in the model. This is an issue as individual-level information is often essential. For example, when dealing with road pricing, it is essential to take into account that behavioral responses are affected by individual income.

Using the data at the level of individual trips, without aggregation, allows all data to be used and makes it possible to incorporate individual-level information. To achieve this, we must solve the following problem.

The first-order condition for the perturbed utility route choice problem leads to an equation relating the optimal flow vector to the network link cost vector. This equation is the basis for estimating model parameters that determine link costs. If we could observe flows, then the model would be estimable simply by regressing a certain transformation of the flow vector on link characteristics. However, at the individual level, we observe paths and not flows.

With the microPURC estimator, we propose to resolve this issue as follows. Given a current parameter estimate, we can solve the perturbed utility maximization problem to predict the flow corresponding to each observed chosen path. Next, for each observation, we can extend the PURC first-order condition with a bias-correcting term that depends on the observed flow. Finally, we can update the parameter estimate via linear regression. The updated parameter estimate will improve on the previous, due to the bias-correction term that incorporates information from the observed chosen paths as well as correcting for the bias that results from flows being estimated and not observed.

This procedure can be repeated. On convergence, we show that it provides a consistent $\sqrt{N}$ asymptotically normal estimate of the model parameters. This is the microPURC estimator.

Brief literature review

Route choice models based on the additive random utility discrete choice model (McFadden, 1981) for the choice between alternative routes are generally estimated by maximum likelihood. However, it is a challenge for these models that the number of potential routes is extremely large. As a consequence, it has been a priority to find ways to generate choice sets with good coverage to reduce the bias that results from excluding choice alternatives (Prato, 2009). The perturbed utility route choice model operates at the network level and does not require a choice set as input.

Another branch of route choice models is recursive models. In these models, the traveler is seen as choosing a path link by link in a Markovian fashion. A recent series of papers has considered estimation by maximum likelihood of what they term the recursive logit model and generalizations building on the multivariate extreme value distribution (Fosgerau et al., 2013; Mai, Fosgerau, & Frejinger, 2015; Mai, Frejinger, & Bastin, 2015; Mai, 2016). However, this estimation procedure becomes computationally hard with large networks. The problem is that solving the Bellman equations involves a large number of inversions of matrixes that inherit their size from the network.

2 Methodology

The perturbed utility route choice model

The PURC model describes the choice of route through a network for a traveler who makes a single trip from an origin to a destination. The network is connected with nodes and links $(\mathcal{N}, \mathcal{L})$. Links are indexed by ij, where i and j are the start and end nodes. We also use v to index nodes. The network structure is specified by a node-link incidence matrix $A = \{a_{v,ij} | v \in \mathcal{N}, ij \in \mathcal{L}\}$ with entries

$$a_{v,ij} = \begin{cases} -1, & v = i \\ 1, & v = j \\ 0, & \text{otherwise.} \end{cases}$$

For a flow vector $x \in \mathbb{R}_+^{\mathcal{L}}$, flow conservation for a trip is expressed as $Ax = b$, where $b \in \mathbb{R}^{\mathcal{N}}$ has entries that are zero except $b_v = -1$ when v is the origin of the trip and $b_v = 1$ when v is the destination of the trip.

For each trip, we observe (y, Z, b), where $y \in \{0,1\}^{|\mathcal{L}|}$ is a vector indicating the links used and Z is a matrix with a row for each network link and columns for network characteristics, such that $c = Z\beta$ is a link cost vector. The matrix Z comprises network characteristics but can also include interactions of traveler or trip-specific information with link characteristics. The link cost vector is assumed to be positive at the true value of β, i.e. $Z\beta_0 \gg 0$.

Given an origin-destination pair, the traveler chooses flow x that minimizes a perturbed cost function. For a flow vector $x \in \mathbb{R}_+^{|\mathcal{L}|}$, we define the perturbation function as the sum of link-specific perturbation functions

$$F(x) = \sum_{ij \in \mathcal{L}} F_{ij}(x_{ij}). \quad (1)$$

Each link-specific perturbation function $F_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be continuously differentiable, strictly convex, and strictly increasing, with $F_{ij}(0) = F'_{ij}(0) = 0$ and range equal to $\mathbb{R}_+$.

The optimal flow vector for the traveler given cost vector c is the solution to the the convex minimization problem

$$\hat{x}(c, b) = \underset{x \in \mathbb{R}_+^{\mathcal{L}}}{\operatorname{argmin}} \left\{ c^\top x + F(x) \,|\, Ax = b \right\}.$$

Observed individual trips y are random, with expectation conditional on b, c equal to the optimal flow at the true parameter β_0. i.e.,

$$\mathbb{E}[y|Z,b] = \hat{x}\left(Z\beta_0, b\right).$$

This assumption implies that $\hat{x}_{ij}\left(Z\beta_0, b\right)$ is the probability that a trip with origin-destination pair b travels along link ij.

Optimizing behavior

To derive an estimator for β_0, we begin by analyzing the generic traveler's cost minimization problem. Given cost vector $c = Z\beta$, the Lagrangian for the perturbed cost minimization problem is

$$\Lambda(x,\eta) = x^\top c + F(x) + \eta^\top\left(Ax - b\right),\ x \in \mathbb{R}_+^{|\mathcal{L}|}. \tag{2}$$

Let $\hat{B} = \hat{B}(\hat{x}) = \mathrm{diag}(1_{\hat{x}>0})$ be the matrix with ones on the diagonal corresponding to interior solutions for the link flows $\hat{x}_{ij}$. The first-order condition for the cost minimization problem for the positive entries of $\hat{x}$ can then be written

$$\hat{B}\left(c + \nabla F\left(\hat{x}\right) + A^\top\hat{\eta}\right) = 0, \tag{3}$$

where $\nabla F(\hat{x})$, the gradient of F, has entries $F'_{ij}(\hat{x}_{ij})$.

Let $\hat{P} = \hat{P}(\hat{x})$ be the projection matrix

$$\hat{P} = \hat{B} - (A\hat{B})^+ A\hat{B},$$

where $(A\hat{B})^+$ denotes the Moore-Penrose inverse of $A\hat{B}$. Pre-multiplying Eq. (3) by $\hat{P}$ then yields the following projected first-order condition, which is the starting point for our estimation strategy:

$$\hat{P}(c + \nabla F(\hat{x})) = 0. \tag{4}$$

Estimation

Consider the residual sum-of-squares function

$$RSS(\beta) = \mathbb{E}\left[\left\|y - \hat{x}(Z\beta, b)\right\|^2\right], \tag{5}$$

At least in principle, we can recover β_0 by minimizing this function, according to

Lemma 1 *Under some regularity assumptions specified in the full paper, β_0 is the unique minimum of RSS.*

However, the RSS is a very difficult function to minimize. We therefore seek an approach that utilizes the problem structure to make estimation easier and not least much faster.

Based on the projected first-order condition (4), we will construct a map that has the true parameter as a fixed point. In case the map has multiple fixed points, it is straightforward to determine which point minimizes the residual sum-of-squares.

Define first a function T as a first-order approximation to the projected first-order condition in the direction of the choice variable y.

$$T\left(\beta, x, Z, y\right) = \hat{P}(x)\left(Z\beta + \nabla F\left(x\right) + \nabla^2 F\left(x\right)\cdot\left(y - x\right)\right).$$

Next, define an iterative map $\beta \to \Phi(\beta)$ by

$$\Phi(\beta) \in \operatorname*{argmin}_{\beta'} \mathbb{E}\left[\left\|T(\beta', \hat{x}(Z\beta, b), Z, y)\right\|^2\right]. \tag{6}$$

The mechanics of this map are the following. Based on a β, we compute the corresponding predicted flow $\hat{x}(Z\beta, b)$ Then $T(\beta', \hat{x}(Z\beta, b), Z, y)$ is the approximate projected first-order condition at the candidate parameter β'. The function value $\Phi(\beta)$ is the value of β' that minimizes the sum of squares of $T(\beta', \hat{x}(Z\beta, b), Z, y)$. Under our regularity conditions, the value $\Phi(\beta)$ is uniquely determined for β in a neighborhood of β_0.

The function $T(\beta', \hat{x}(Z\beta, b), Z, y)$ is linear as a function of the β', which makes the iterative map available in closed-form as

$$\Phi(\beta) = -\left(\mathbb{E}\left[Z^\top \hat{P} Z\right]\right)^{-1} \mathbb{E}\left[Z^\top \hat{P}\left(\nabla F(\hat{x}) + \nabla^2 F(\hat{x})(y - \hat{x})\right)\right], \tag{7}$$

where $\hat{x} = \hat{x}(Z\beta, b)$ is the optimal flow corresponding to β, the projection $\hat{P}$ is constructed based on $\hat{x}$, and $\mathbb{E}[Z'\hat{P}Z]$ is assumed to be invertible.

In the full paper we give a straightforward proof of the following

Lemma 2 *Any point β^* is a fixed point of Φ if and only if it satisfies the first-order condition*

$$\mathbb{E}\left[Z'P^*\nabla^2 F(x^*)(\hat{x}(Z\beta_0, b) - x^*)\right] = 0, \tag{8}$$

where $x^* = \hat{x}(Z\beta^*, b)$ and $P^* = \hat{P}(x^*)$.

Lemma 2 shows in particular that the true parameter β_0 is a fixed point of Φ. This observation leads us to the following

Theorem 1 *The true parameter β_0 minimizes the residual sum-of-squares over the set of fixed points of Φ, i.e.*

$$\beta_0 = \operatorname*{argmin}_\beta RSS(\beta) \quad s.t. \quad \Phi(\beta) = \beta. \tag{9}$$

The iterative map Φ has a particularly elegant form when F is a quadratic function. We would favor using

$$F(x) = \frac{1}{2} x^\top L x, \tag{10}$$

where L is a diagonal matrix of positive link lengths since this makes the perturbation function invariant with respect to link splitting. In that case,

$$\nabla F(x) + \nabla^2 F(x)(y - x) = Lx + L(y - x) = Ly,$$

such that the iterative map reduces to

$$\Phi(\beta) = -\mathbb{E}\left[Z^\top \hat{P} Z\right]^{-1} \mathbb{E}\left[Z^\top \hat{P} L y\right].$$

The flow $\hat{x}$ enters only with the set of active flows into the projection matrix $\hat{P}$.

Implementation and convergence

A sample of trips would have observations $(y^n, Z^n, b^n), n = 1, \ldots, N$ that are assumed to be i.i.d. realizations of the random variables (y, Z, b). We suggest estimating β_0 by the sample analog of the procedure described above, i.e.

$$\tilde{\beta} \equiv \operatorname*{argmin}_\beta RSS_N(\beta) \quad \text{s.t.} \quad \Phi_N(\beta) = \beta, \tag{11}$$

where RSS_N and Φ_N replaces population expectations with sample averages in (5) and (6). Starting from some initial parameter vector $\tilde{\beta}_{(0)}$, we then suggest the following Algorithm 1 for iterating the sample version of Φ towards a fixed point.

Algorithm 1: microPURC iterated algorithm

Step 0: Choose starting value $\tilde{\beta}_{(0)}$;

Step 1: Given $\tilde{\beta}_{(t)}$, compute $\hat{x}^n_{(t)} = \hat{x}(Z^n\tilde{\beta}_{(t)}, b^n)$, and $\hat{P}^n_{(t)} = \hat{P}(\hat{x}^n_{(t)})$;

Step 2: Compute $\tilde{\beta}_{(t+1)} = \Phi_N(\tilde{\beta}_{(t)})$, i.e.

$$\tilde{\beta}_{(t+1)} = -\left(\mathbb{E}_N\left[Z^{n\top}\hat{P}^n_{(t)}Z^n\right]\right)^{-1}\mathbb{E}_N\left[Z^{n\top}\hat{P}^n_{(t)}\left(\nabla F(\hat{x}^n_{(t)}) + \nabla^2 F(\hat{x}^n_{(t)})(y^n - \hat{x}^n_{(t)})\right)\right]$$

- Repeat step 1 and 2 until convergence.

Step 3: If different starting values lead to different fixed points, compare the residual sum of squares. Define $\tilde{\beta}$ as the fixed point that minimizes RSS.

We go on to establish two theorems concerning the statistical properties of the microPURC algorithm. We first show that our algorithm has at least one fixed point and that there is a neighborhood from which it is guaranteed to converge to this fixed point. Convergence is quadratic, except for a small error term that goes to zero as the sample size increases.

Theorem 2 *For $\tilde{\beta}_{(0)}$ sufficiently close to $\tilde{\beta}$, the sequence $\tilde{\beta}_{(t+1)} = \Phi_N(\tilde{\beta}_{(t)})$ converges almost quadratically to $\tilde{\beta}$, i.e. there exists $C > 0$ and a random variable ϵ_N such that*

$$\left\|\tilde{\beta}_{(t+1)} - \tilde{\beta}\right\| \leq C\left\|\tilde{\beta}_{(t)} - \tilde{\beta}\right\|^2 + \epsilon_N\left\|\tilde{\beta}_{(t)} - \tilde{\beta}\right\|, \tag{12}$$

where $\epsilon_N \xrightarrow{p} 0$.

Theorem 2 only applies to the fixed point $\tilde{\beta}$ which minimizes RSS_N. The next theorem establishes consistency and asymptotic normality of $\tilde{\beta}$. This allows us to compute standard errors and conduct inference on the parameter estimates that we obtain.

Theorem 3 *The estimator $\tilde{\beta}$ is consistent, $\left\|\tilde{\beta} - \beta_0\right\| \xrightarrow{p} 0$ and $\sqrt{N}$-asymptotically normal $\sqrt{N}(\tilde{\beta}_N - \beta_0) \xrightarrow{D} \mathbf{N}(0, H^{-1}SH^{-1})$ where*

$$H = \mathbb{E}\left[Z^\top \hat{P} Z\right], \quad S = \mathbb{E}\left[Z^\top \hat{P}(y-x)(y-x)^\top \hat{P} Z\right].$$

The asymptotic variance of $\hat{\beta}$ can be consistently estimated by replacing H and S by their sample counterparts,

$$\tilde{H} = \mathbb{E}_N\left[Z^{n\top}\hat{P}^n Z^n\right], \quad \tilde{S} = \mathbb{E}_N\left[Z^{n\top}\hat{P}^n(y^n - \hat{x}^n)(y^n - \hat{x}^n)^\top \hat{P}^n Z^n\right], \tag{13}$$

where $\hat{x}^n = \hat{x}(Z^n\tilde{\beta}, b^n)$ and $\hat{P}^n = \hat{P}(\hat{x}^n)$.

A consistent estimator of the standard error of $\tilde{\beta}$ is then $\tilde{\sigma} = \sqrt{\mathrm{diag}(\tilde{H}^{-1}\tilde{S}\tilde{H}^{-1})/N}$.

3 Results and discussion

Simulated experiment

We have carried out a simulation study to verify that our microPURC estimator is indeed capable of recovering known true parameters from simulated data.

We have set up an experiment using a grid network with 288 links and 81 nodes (Figure 1). We use the entropy perturbation function (14) for validation:

$$F_{ij}(x_{ij}) = (1 + x_{ij})\ln(1 + x_{ij}) - x_{ij} \tag{14}$$

The true parameter vector is $\beta = [1/2, 1/10, 1/4, 1/10]$. We draw independent variables Z representing four link attributes.

A subset of the nodes are designated as centroids, which can be endpoints for a trip. OD pairs are randomly sampled from the centroids. This ensures that randomly sampled trips are not too short. For each OD pair, we compute $\hat{x}(Z\beta, b)$ with the true parameters β_0, denoted as $\hat{x}^*$.

We then generate synthetic path choice observations y as random walks on the subnetwork with positive link flows $\hat{x}^*_{ij} > 0, \forall ij \in \mathcal{L}$. This subnetwork is necessarily acyclic since link costs are positive. Consequently, the synthetic observations can be generated by departing from the origin node and choosing the outgoing link with probabilities

$$p_{ij} = \frac{\hat{x}^*_{ij}}{\sum_{k:(i,k)\in\mathcal{L}} \hat{x}^*_{ik}}.$$

Furthermore, for each OD pair, the random walk terminates exactly at the destination node since it is the only node without any outgoing link in the active subnetwork. Each observation $y \in \{0,1\}^{|\mathcal{L}|}$ is an incident vector, where $y_{ij} = 1$ if observation traverses link ij, and $y_{ij} = 0$, otherwise.

In the simulation experiment, we test the effects of sample size. We compute the empirical mean and standard deviation of the estimated beta $\tilde{\beta}$ over 10 replications.

Table 1: Simulated experiment (10 replications)
Mean microPURC estimates (std. in the bracket)

Sample size	$\tilde{\beta}$				**RMSE**
10	1.4122 (0.3864)	0.5919 (0.2268)	1.0398 (0.4026)	0.4156 (0.2625)	7.5385
100	0.5534 (0.0097)	0.0875 (0.0100)	0.2795 (0.0112)	0.1326 (0.0089)	0.6388
200	0.5180 (0.0052)	0.0940 (0.0024)	0.2628 (0.0071)	0.0889 (0.0043)	0.4348
500	0.4923 (0.0020)	0.1132 (0.0018)	0.2387 (0.0026)	0.1218 (0.0016)	0.2895
1000	0.5002 (0.0010)	0.1153 (0.0006)	0.2368 (0.0011)	0.0984 (0.0009)	0.1880
2000	0.5047 (0.0005)	0.0945 (0.0005)	0.2447 (0.0007)	0.0875 (0.0005)	0.1559
5000	0.5016 (0.0002)	0.0984 (0.0001)	0.2540 (0.0003)	0.1016 (0.0001)	0.0815

The results show that our microPURC estimator can recover the preset parameters well when the sample size is above 100. Moreover, the standard deviation of the parameter estimates decreases as the sample size increases, demonstrating the estimator's asymptotic property. This is also illustrated by the decreasing norm (the RMSE) which decreases with the sample size at the rate predicted by our Theorem 3.

Application to real data

This is work in progress. We have access to a large database of GPS traces of car trips and corresponding network data. The data are from a national road pricing experiment that is currently ongoing in Denmark. The implementation is ongoing.

4 Conclusions

We have formulated the microPURC estimator, which is suitable for estimating the parameters of the perturbed utility route choice model with trip-level data. We have established the sta-

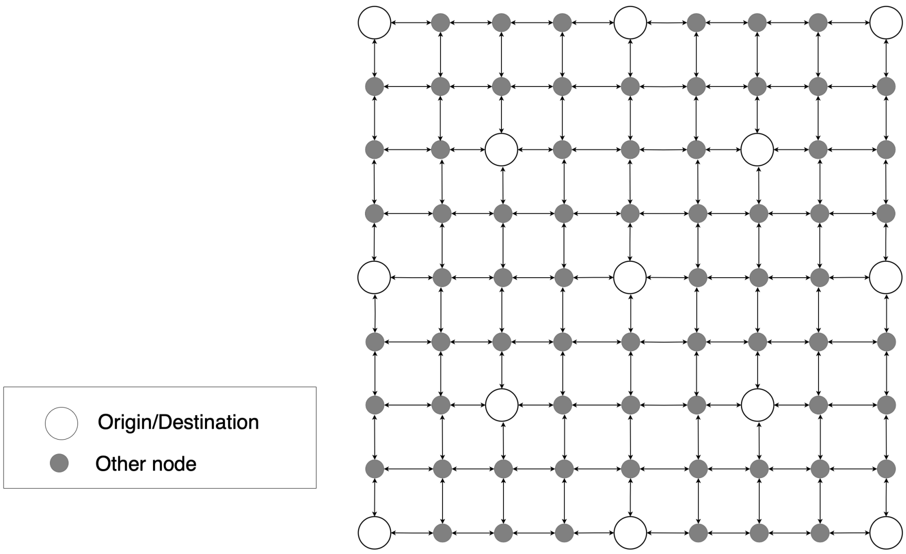

Figure 1: Grid network

tistical properties of the estimator. In particular, the microPURC estimator is consistent and $\sqrt{N}$-asymptotically normal.

Our simulation experiment is successful. It shows that the microPURC estimator is able to recover the true parameters with high precision, even with moderate sample sizes.

We expect our application to real data to demonstrate that the microPURC estimator is feasible with the large networks and data sets that are met in practice. This will make the power of the perturbed utility route choice model available for practice, allowing us to use a model that generates realistic substitution patterns directly from the network structure, using the complete network as it is while not requiring any choice set generation.

References

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70–80. doi: 10.1016/j.trb.2013.07.012

Fosgerau, M., Lukawska, M., Paulsen, M., & Rasmussen, T. K. (2023). Bikeability and the induced demand for cycling. *Proceedings of the National Academy of Sciences*, *120*(16).

Fosgerau, M., Paulsen, M., & Rasmussen, T. K. (2022). A perturbed utility route choice model. *Transportation Research Part C*, *136*(103514). Retrieved from `https://arxiv.org/abs/2103.13784` doi: https://doi.org/10.1016/j.trc.2021.103514

Mai, T. (2016). A method of integrating correlation structures for a generalized recursive route choice model. *Transportation Research Part B: Methodological*, *93*, 146-161. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0191261515302113` doi: https://doi.org/10.1016/j.trb.2016.07.016

Mai, T., Fosgerau, M., & Frejinger, E. (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, *75*, 100–112. doi: 10.1016/j.trb.2015.03.015

Mai, T., Frejinger, E., & Bastin, F. (2015, 4). A Dynamic Programming Approach for Quickly Estimating Large Scale MEV Models. *Transportation Research Part B: Methodological*(June). doi: 10.1016/j.trb.2016.12.017

McFadden, D. (1981). Econometric Models of Probabilistic Choice. In C. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 198–272). Cambridge, MA, USA: MIT Press. doi: 10.1086/296093

Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100. Retrieved from http://www.sciencedirect.com/science/article/pii/S1755534513700058http://www.sciencedirect.com/science/article/pii/S1755534513700058#http://www.sciencedirect.com/science/article/pii/S1755534513700058/pdfft?md5=047f3816cb6a3d636aa9eab4915af0f8&pid=1-s2.0 doi: 10.1016/S1755-5345(13)70005-8