A perturbed utility route choice model

Fosgerau, M., Paulsen, M., & Rasmussen, T. K. (2022). *Transportation Research Part C: Emerging Technologies*, *136*, 103514.

2025 理論談話会 #9 / 29 May 2025

D1 Hsin-Kai Huang

新規性 (Novelty)、有用性 (Practicality)、信頼性 (Reliability)

新規性 (Novelty)

- Yield reasonable substitution patterns.
- The model is estimated using all route alternatives, and no choice set generation is required.

有用性 (Practicality)

- Estimation is very fast. Predictions computed using convex optimization.
- Easy to use for traffic assignment.

信頼性 (Reliability)

- Estimate and validate the model using 1,337,096 trips in large road network.

Contents

1. Introduction

Background

1. Importance of Route Choice Modeling

With the availability of large-scale vehicle trajectory data, there is now greater potential to build more behaviorally accurate route choice models, which are critical for infrastructure planning and traffic management.

2. Modeling Challenge

A major challenge in route choice modeling is the vast number of possible routes in realistic traffic networks, making traditional discrete choice approaches computationally infeasible.

3. Perturbed Utility（摂動効用） Framework

The paper frames route choice as a *perturbed utility model*, where individuals choose options that maximize a linear utility function minus a convex perturbation(Fosgerau and McFadden, 2012). This general framework includes and extends traditional random utility models and has strong microeconomic foundations.

Background

Traditional utility models

$$\mathrm{U}(x) = a^{\mathsf{T}} x$$

Here, x is the vector of chosen options (e.g., the features of a product or service).
a is a weight vector representing the importance of each option's feature.

The utility function is deterministic, meaning it does not account for any random factors.

Perturbed utility models

$$\mathrm{U}(x) = a^{\mathsf{T}} x - F(x)$$

$F(x)$ is a **convex function (凸関数)** that represents costs or risks, reflecting the **"perturbation"** or uncertainty of the choice.

Background

A new kind of route choice model!

1 Avoid concentrating flow on few links

- The perturbation function is assumed to be a positive and increasing convex function.
- The perturbation terms induce the traveler to **avoid concentrating their flow on a few links**.

2 Zero Flow in Non-Active Network Parts

- Required to satisfy $F(0) = 0$ and $F'(0) = 0$, which means that when the flow on a link is zero, the perturbation function does not contribute any additional cost or penalty.
- Only links that **contribute to overall flow balance** in the network are assigned nonzero flow.

3 Linear Reformulation Enables Ordinary Least Squares (OLS) Estimation

- Transforms route choice optimization into linear equations.
- Allows fast and simple parameter estimation using OLS.

Background

1 Path-Size Logit (PSL) and Adaptive Path-Size Logit (APSL) Models

Address substitution patterns in route choice when routes overlap, but still require path enumeration and are computationally expensive.

2 Network Generalized Extreme Value (NGEV) Model

Aims to model route choice at the network level with generalized entropy as the perturbation, but does not focus on **parameter estimation**.

3 PURC model

Unlike both PSL and the GEV model, the **PURC model** avoids the issue of path enumeration and provides a more efficient estimation approach by utilizing a perturbation function that allows for **corner solutions**. This makes it computationally feasible and more realistic in modeling route choice behavior.

Background – Contribution

1. **Not requiring choice set generation:** This addresses one of the most challenging issues in traditional models—handling the virtually infinite number of potential routes.

2. **Physically Feasible Network Flows:** The model inherently ensures **flow conservation** and a realistic distribution of flows across the network.

3. **Naturally Excludes Unreasonable Routes:** Unlike traditional random utility models where every route has a positive probability of being chosen, the PURC model predicts that some links (inactive) will not be used at all.

4. **No Likelihood Maximization:** Parameters can be estimated using **OLS regression**

5. **Highly Scalable:** Applicable to **both small toy networks and large urban networks**

2. Setup

The constrained utility maximization problem

Maximize the difference between the linear utility and the perturbation utility

$$\max_{x\in\mathbb{R}_+^{[\epsilon]}} U(x) \quad s.t\ \underline{Ax = b},$$

Flow conservation

weighted by link length

$$U(x) = l^{\top}\underbrace{(u \circ x)}_{\text{Linear term}} - l^{\top}\underbrace{F(x)}_{\text{Convex perturbation term}}.$$

(V, E)	Network
A	Incidence matrix
b	Traveler unit demand
x	Network flow vector

Assumption

$F: \mathbb{R}_+ \to \mathbb{R}$ is strictly convex, $F(0) = 0$, and $F'(0) = 0$.

Purpose

1. Ensures that when the flow $x_e = 0$, the perturbation is zero and flat, which guarantees that if the flow conservation constraint is not active at a link, the optimal flow on that link is zero.

2. The combined perturbation term $l^\top F(x)$ is strictly convex, ensuring that the utility function is strictly concave, and hence the maximization problem has a **unique solution**.

Assumption

$F: \mathbb{R}_+ \to \mathbb{R}$ is strictly convex, $F(0) = 0$, and $F'(0) = 0$.

Role of the Perturbation Term F

1 Flow Distribution

If the flow on a link becomes large, the perturbation term becomes large and negative, which incentivizes the traveler to spread their flow across more links, avoiding excessive concentration.

2 Part of the traveler's preferences

The perturbation term functions similarly to a congestion term but differs in that it is part of the traveler's preferences and depends solely on the individual's flow vector x, rather than the behavior of other travelers.

Assumption

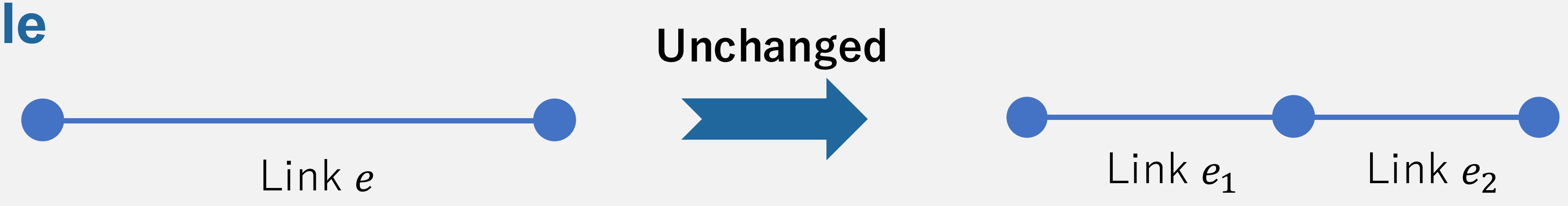
All terms in the utility function are weighted by **link length**

Invariance to Link Splitting

If we split a link e into two sub-links e_1 and e_2, the total utility contribution **remains unchanged**:

$$l_e u_e x_e - l_e F(x_e) = l_{e1} u_{e1} x_{e1} - l_{e1} F(x_{e1}) + l_{e2} u_{e2} x_{e2} - l_{e2} F(x_{e2})$$

Example



This property ensures that the model's predictions are invariant to the introduction of dummy nodes or the splitting of links, which is desirable for maintaining consistency in the model’s behavior.

Assumption

Parameterization of the Utility Rate Vector

- **Utility Rate Vector:**
The utility rate vector u is a linear function of the link characteristics matrix z and parameter vector β:

$$u = \mathbf{z}\boldsymbol{\beta}$$

- **Link Characteristics z :**
  The link characteristics $\mathbf{z}$ may include:
  1. Number of outlinks (divided by link length)
  2. Travel time per kilometer (i.e., pace)
  3. Link type dummy variables

- **Parameter Estimation (β):**
  These characteristics and the parameter vector β will be estimated within the model.

Solving the traveler's problem

Lagrangian Setup

$$\Lambda(x,\lambda) = l^{\intercal}(u \circ x) - l^{\intercal}F(x) + \lambda^{\intercal}(Ax - b).$$

First-Order Conditions (Find the optimal solution)

$$0 = \hat{x} \circ \left(l \circ \left(u - F'(\hat{x})\right) + A^{\intercal}\hat{\lambda}\right) \quad A\hat{x} = b$$

The marginal utility rate equals the marginal perturbation that is, $u_e = F'(\hat{x}_e)$

$$\hat{x} = 0 \; or \; \left(\underline{l \circ \left(u - F'(\hat{x})\right) + A^{\intercal}\hat{\lambda}}\right) = 0$$

$$l \circ \left(u - F'(\hat{x})\right) + A^{\intercal}\hat{\lambda} = 0$$

$$(l \circ u) = (l \circ F'(\hat{x})) - A^{\intercal}\hat{\lambda}$$

$\hat{x}$	The optimal flow
$\hat{\lambda}$	Lagrange multiplier values
$(l \circ u)$	The utility contribution for all active links, weighted by their lengths
$(l \circ F'(\hat{x}))$	The perturbation effect (such as congestion or other external effects) for all active links, weighted by their lengths.
$A^{\intercal}\hat{\lambda}$	The flow conservation effect for the active links. The matrix $A^{\intercal}$ comes from the flow conservation constraints, and $\hat{\lambda}$ is the vector of Lagrange multipliers associated with these constraints. This term accounts for the impact of redistributing flow across the network due to the constraints.

Solving the traveler's problem

Simplification with B

$$B(l \circ u) = B(l \circ F'(\hat{x})) - BA^{\top}\hat{\lambda}$$

This equation is useful because it allows us to **focus only on the links with non-zero flow**, reducing the dimensionality of the problem and focusing only on the active elements.

Example

Let's assume we have a network with 4 links, and the flow vector is:

$$x = [0.5, 0.3, 0, 0.2]$$

Matrix B will look like this:

$$B: \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow Bx = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.3 \\ 0 \\ 0{,}2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix}$$

Solving the traveler’s problem

Regression Model

$$B(l \circ u) = B(l \circ F'(\hat{x})) - BA^{\top}\hat{\lambda}$$

Dependent variables

The traveler's problem is transformed into a **regression problem** where the flow on each link is determined by a weighted sum of link characteristics, adjusted for **perturbations** and subject to **flow conservation constraints**.

$$B(l \circ u) = B(l \circ z\beta) = B(l \circ z)\beta$$

Independent variables

where z is the matrix of link characteristics and β is the parameter vector to be estimated.

Fixed Effects

The Lagrange multipliers $\hat{\lambda}$ can be treated as fixed effects

→ Become **computationally difficult** as the network size grows.
An alternative formulation will be sought to avoid this issue.

We will therefore seek an alternative regression equation that eliminates $\hat{\lambda}$

Eliminating the Lagrange multipliers

The matrix $BA^\top$ is not invertible.

Let $\boldsymbol{C} = (BA^\top)^*$ be a **Moore–Penrose inverse** of $BA^\top$. We may then utilize that $BA^\top \boldsymbol{C} BA^\top = BA^\top$.

$$BA^\top C\, B(l \circ u) = BA^\top C B(l \circ F'(\hat{x})) - BA^\top C BA^\top \lambda = BA^\top C B(l \circ F'(\hat{x})) - BA^\top \lambda$$

▼

$$BA^\top \lambda = BA^\top C\, B(l \circ (F'(\hat{x}) - u)).$$

Substituting this back into the reduced first-order condition, we find that

$$B(l \circ u) = B(l \circ F'(\hat{x})) - BA^\top C\, B(l \circ (F'(\hat{x}) - u)).$$

▼

$$(I - BA^\top C) B(l \circ u) = (I - BA^\top C)\ B(l \circ F'(\hat{x}))$$

Eliminate the Lagrange multipliers from the first-order condition!

A regression equation for model estimation

Construction of Variables

Given a flow vector $\hat{x}$ corresponding to demand b

$$(I - BA^{\mathsf{T}}C)B(l \circ u) = (I - BA^{\mathsf{T}}C)\ B(l \circ F'(\hat{x})) \qquad u = \mathbf{z}\boldsymbol{\beta}$$

The **vector** y is defined as:

$$y = (I - BA^{\mathsf{T}}C)\ B(l \circ F'(\hat{x}))$$

The **matrix** ω is defined as:

$$\omega = (I - BA^{\mathsf{T}}C)\ B(l \circ z)$$

Regression setup

$$y_{bi} = \omega_{bi}\beta + \epsilon_{bi}$$

For each demand vector b, pairs (y_{bi}, ω_{bi}) are constructed, where:

- $b \in \mathrm{B}$ indexes the set of demand vectors (e.g., different origin-destination pairs).
- i indexes the individual elements of each vector y_b.
- ϵ_{bi} indicates mean-zero noise terms.

- **Parameters $\boldsymbol{\beta}$** can be recovered from the regression
- Computational complexity: Linearly with the size of the set of demand vectors β

Routes are substitutes

The model predicts that routes in the network are **substitutes**, meaning increasing the utility of one route will reduce the likelihood of choosing other routes.

Complementarities

Car and petrol or linked routes in a transportation network

Substitutes

Coffee and pearl milk tea

Perturbed utility models: Unlike traditional discrete choice models, perturbed utility models can potentially allow for **complementarities**, but in this case, the routes are **substitutes**.

Proposition 1

The perturbed utility route choice model is equivalent to a discrete choice model for the choice among all loop-free routes connecting origin to destination. The routes are substitutes in the sense that if the utility of just one route is increased while the utility of all other routes is unaffected, then the choice probability weakly decreases for all other routes.

Proof

1. Utility Representation for Routes

- $\sum$: The set of all **routes** connecting the origin to the destination.
- Each route $\sigma \in \sum$ consists of a sequence of **links** $\boldsymbol{e}$, so each route σ is a list of links:

$$\sigma = \{e_{\sigma(1)}, e_{\sigma(2)}, \dots\}$$

- The **flow vector** $\boldsymbol{p}$ represents the allocation of flow (or probabilities) to each route, and can be thought of as a vector of route choice probabilities.

Proposition 1 – proof

The total utility $\mathrm{U}(p)$ can then be written as:

Path utility: The sum of the product of the flow on each route p_σ and the utility of that route μ_σ :

$$\mathrm{U}(p) = \sum_{\sigma \in \Sigma} \mathrm{p}_\sigma \mu_\sigma - G(p) \qquad \mu_\sigma = \sum_{e \in \sigma} \mu_e \qquad G(p) = \sum_{e \in \epsilon} l_e F\Big(\sum_{\sigma \supset e} p_\sigma\Big)$$

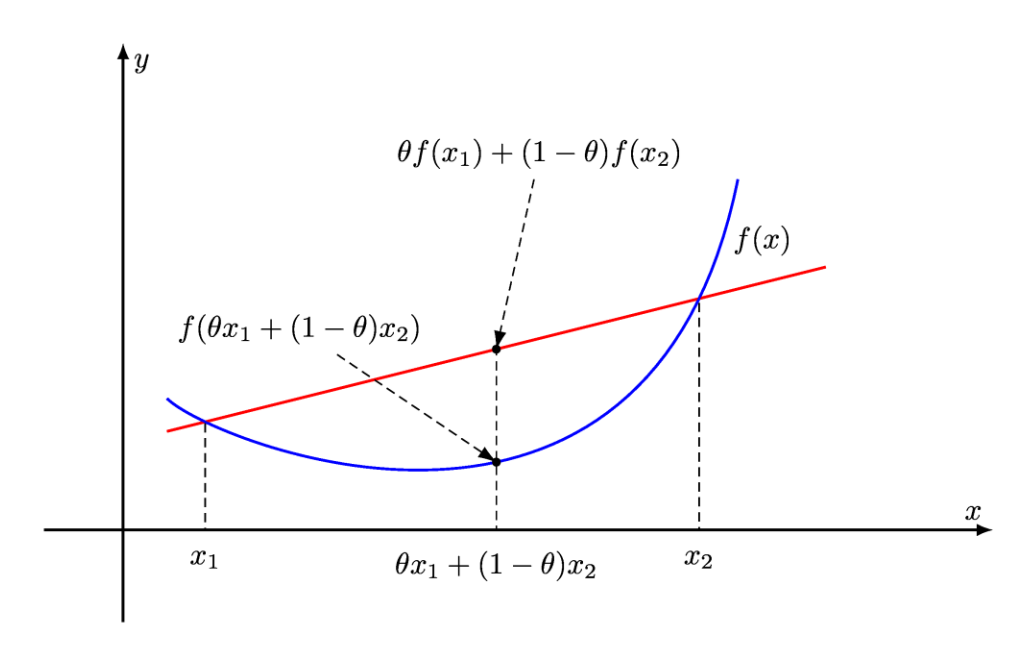
Flow loss function $G(p)$: This function captures the congestion or loss on each link due to the flow. It is a sum over all links e, where the cost of each link is based on the total flow passing through it (which is the sum of the flow from all routes using that link)

Proposition 1 – proof

Convexity and Submodularity

Prove that G(p) is **strictly convex**, which is a crucial step because it will support the argument for the substitutability of routes.

$$f(\theta x_1 + (1-\theta)x_2) < \theta f(x_1) + (1-\theta) f(x_2)$$



http://tisp.indigits.com/convex_sets/convex_functions

Proposition 1 – proof

Convexity and Submodularity

Assume that p^1 and p^2 are two distinct flow vectors (i.e., two different sets of route choice probabilities).

The function F is strictly convex by assumption and hence G is convex.

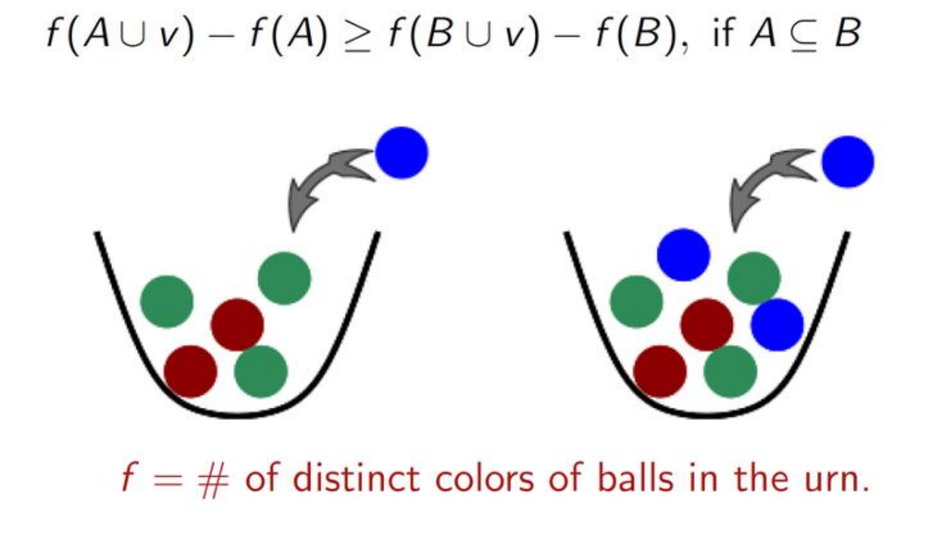
$$G(\eta p^1 + (1-\eta)p^2) = \sum_{e \in \epsilon} l_e F\left(\eta \sum_{\sigma \supset e} p_\sigma^1 + (1-\eta) p_\sigma^2\right) \leq \eta \sum_{e \in \epsilon} l_e F\left(\sum_{\sigma \supset e} p_\sigma^1\right) + (1-\eta) l_e F\left(\sum_{\sigma \supset e} p_\sigma^2\right) = \eta G(p^1) + (1-\eta)\mathrm{G}(\mathrm{p}^2)$$

The right-hand side shows that the convex combination of the flow vectors results in a total loss that is less than or equal to the weighted sum of the individual losses for p^1 and p^2. This is a direct application of the **convexity** of the function G(p).

Since G(p) is strictly convex, it means that **increasing the flow on one route decreases the utility for other routes**, which leads to a **substitutable** relationship between the routes.

Proposition 1 – proof

Submodularity



- Diminishing marginal returns
- Goods/choices are **substitutes**

Supermodularity

- Higher choices by one's opponents make one's own higher strategies look relatively more desirable (Game theory)
- Strategic complementarities

Figure: Dou, J. X., Calvin, J. Y., Jiang, Y., Wang, W. Z., & Fu, Q. Coreset Optimization by Memory Constraints, For Memory Constraints.

Proposition 1 – proof

Convex Conjugate (凸共役) and Supermodularity

$$\mathrm{U}^*(\mu) = \sup_{p\Delta(\sum)} \{U(p)\} = \sup_{p\Delta(\sum)} \left\{ \sum_{\sigma\in\sum} \mathrm{p}_\sigma \mu_\sigma - G(p) \right\}$$

Indirect Perturbed Utility
(Restricted to the set of probability vectors)

$$\frac{\partial^2 G(p)}{\partial p_{\sigma_1} \partial p_{\sigma_2}} = \sum_{e\in\epsilon} l_e \frac{\sigma F'(\sum_{\sigma\ni e} p_\sigma) 1_{\{e\in\sigma_1\}}}{\partial p_{\sigma_2}} = \sum_{e\in\epsilon} F''(\sum_{\sigma\ni e} p_\sigma) 1_{\{e\in\sigma_1\cap\sigma_2\}}$$

The mixed partial derivatives of the convex perturbation

The convex perturbation G is submodular and U^* is supermodular

Moreover, U^* satisfies the definition of a choice welfare function in Feng et al. (2018). Hence, by Feng et al. (2018, Thm. 1), **if U^* is differentiable, then all routes are substitutes.**

Proposition 1 – Summary

Strict Convexity

The **strict convexity** of the utility function and the perturbation $G(p)$ ensures that routes behave as substitutes, which is a crucial characteristic of many transportation and economics models.

Substitutes

The model predicts that routes are substitutes, meaning that increasing the utility of one route decreases the probability of choosing other routes.

Mathematical Foundation

The analysis uses **supermodularity** and **submodularity** to explain why routes behave as substitutes. The utility function is shaped by the interaction between route flows and their associated link costs.

Functional form for the perturbation function

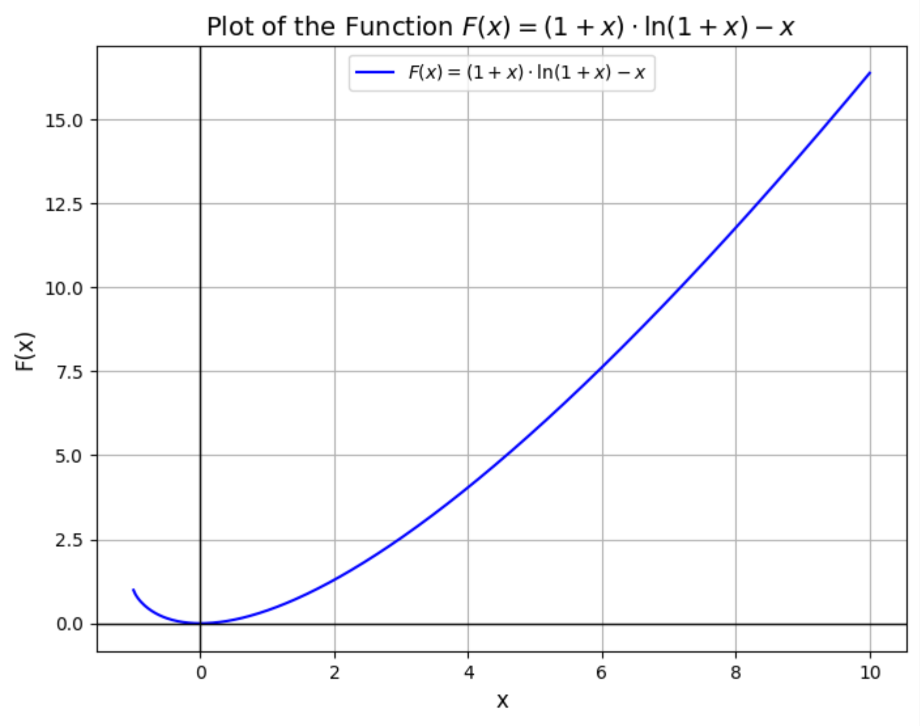
Entropy-Based Perturbation Function

Inspired by the connection between the **Multinomial Logit (MNL)** model and **Shannon entropy**, the authors propose:

$$F(x) = (1 + x) \cdot \ln(1 + x) - x$$

where $F'(x) = \ln(1 + x)\ and\ F''(x) = (1 + x)^{-1}$

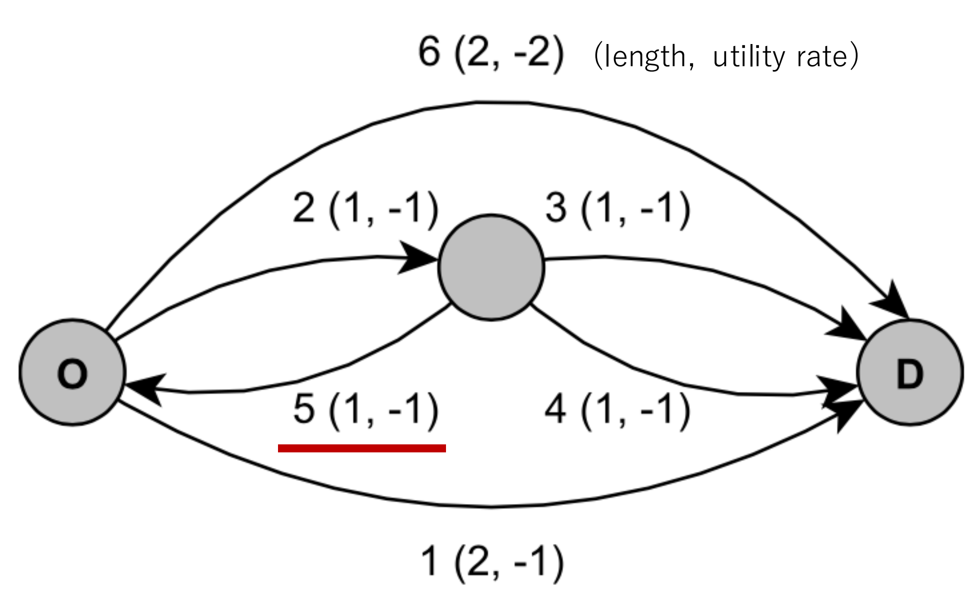
x: Network flow vector



3. Experimenting with the model

Experimenting with the model
– Toy network

- A simple network setup involving **six links** and **four loop-free routes** that connect a single **origin** to a single **destination**
- Route 1: Link 6. Route 2: Link 2, Link 3. Route 3: Link 2, Link 4. Route 4: Link 1



Introduction of a loop

- **Link 5** is added:
  - It is **identical to Link 2** but flows in the **opposite direction**
  - This allows for a **loop** between **Links 2 and 5** under flow conservation

Utility Settings

- Utility rate for most links:

$$u_e = -1$$

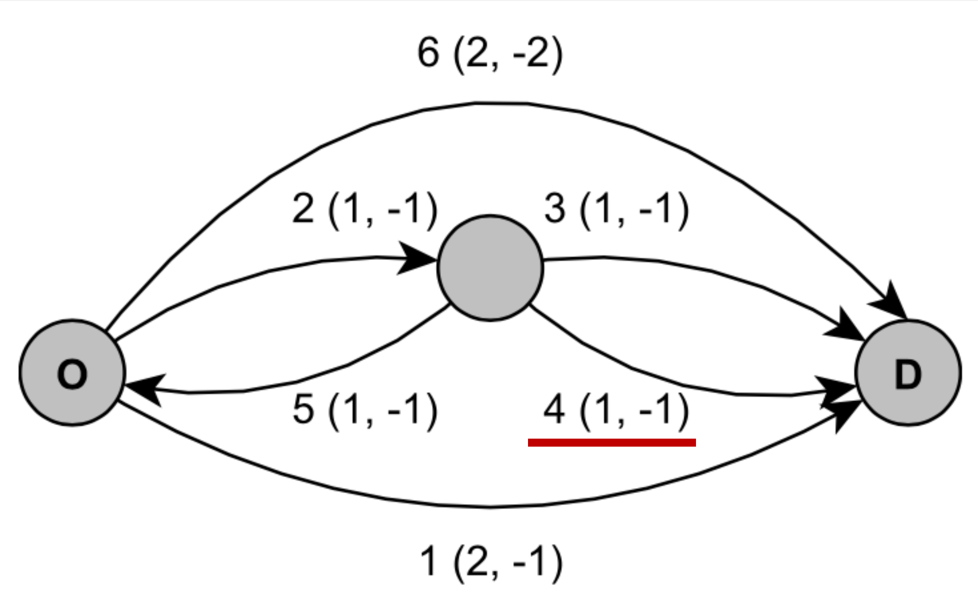
- Exception:
  Link 6 has a higher disutility:

$$u_6 = -2$$

Table 1
Utility maximizing flows in the toy network. Numbers in parentheses indicate flows relative to base.

	Link	l_e	u_e	PURC flow		MNL flow		PSL flow	
Base	1	2	−1	0.424		0.331		0.404	
	2	1	−1	0.576		0.663		0.589	
	3	1	−1	0.288		0.331		0.294	
	4	1	−1	0.288		0.331		0.294	
	5	1	−1	0		0[a]		0[a]	
	6	2	−2	0		0.006		0.007	
Increase unit cost on link 4 *(rel. to base)*	1	2	−1	0.445	*(1.047)*	0.352	*(1.064)*	0.427	*(1.056)*
	2	1	−1	0.555	*(0.965)*	0.641	*(0.967)*	0.566	*(0.961)*
	3	1	−1	0.342	*(1.187)*	0.352	*(1.064)*	0.311	*(1.056)*
	4	1	−1.1	0.214	*(0.743)*	0.289	*(0.871)*	0.255	*(0.865)*
	5	1	−1	0	–	0[a]	–	0[a]	–
	6	2	−1	0	–	0.006	*(1.064)*	0.008	*(1.056)*
Move node at the end of link 2 *(rel. to base)*	1	2	−1	0.381	*(0.897)*	0.331	*(1)*	0.364	*(0.902)*
	2	0.5	−1	0.619	*(1.076)*	0.663	*(1)*	0.629	*(1.069)*
	3	1.5	−1	0.310	*(1.076)*	0.331	*(1)*	0.315	*(1.069)*
	4	1.5	−1	0.310	*(1.076)*	0.331	*(1)*	0.315	*(1.069)*
	5	0.5	−1	0	–	0[a]	–	0[a]	–
	6	2	−2	0	–	0.006	*(1)*	0.007	*(0.902)*

[a]Zero flow is purely a consequence of the choice set generation method.

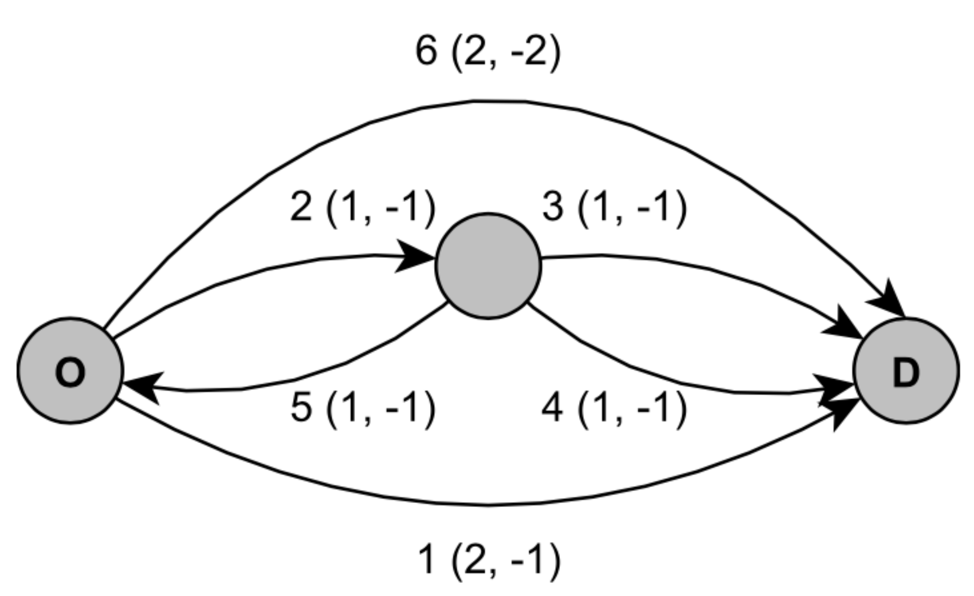


- In the **PURC model**, flow increases more on the route using link 3 than on the one using link 1.
- This means **routes 3 and 4 act as closer substitutes** (because they share **link 2**).
- The **Independence of Irrelevant Alternatives (IIA)** property **does not hold**, which is **desirable** and reflects real-world substitution effects.

Table 1
Utility maximizing flows in the toy network. Numbers in parentheses indicate flows relative to base.

	Link	l_e	u_e	PURC flow		MNL flow		PSL flow	
Base	1	2	−1	0.424		0.331		0.404	
	2	1	−1	0.576		0.663		0.589	
	3	1	−1	0.288		0.331		0.294	
	4	1	−1	0.288		0.331		0.294	
	5	1	−1	0		0[a]		0[a]	
	6	2	−2	0		0.006		0.007	
Increase unit cost on link 4 *(rel. to base)*	1	2	−1	0.445	*(1.047)*	0.352	*(1.064)*	0.427	*(1.056)*
	2	1	−1	0.555	*(0.965)*	0.641	*(0.967)*	0.566	*(0.961)*
	3	1	−1	0.342	*(1.187)*	0.352	*(1.064)*	0.311	*(1.056)*
	4	1	−1.1	0.214	*(0.743)*	0.289	*(0.871)*	0.255	*(0.865)*
	5	1	−1	0	–	0[a]	–	0[a]	–
	6	2	−1	0	–	0.006	*(1.064)*	0.008	*(1.056)*
Move node at the end of link 2 *(rel. to base)*	1	2	−1	0.381	*(0.897)*	0.331	*(1)*	0.364	*(0.902)*
	2	0.5	−1	0.619	*(1.076)*	0.663	*(1)*	0.629	*(1.069)*
	3	1.5	−1	0.310	*(1.076)*	0.331	*(1)*	0.315	*(1.069)*
	4	1.5	−1	0.310	*(1.076)*	0.331	*(1)*	0.315	*(1.069)*
	5	0.5	−1	0	–	0[a]	–	0[a]	–
	6	2	−2	0	–	0.006	*(1)*	0.007	*(0.902)*

[a]Zero flow is purely a consequence of the choice set generation method.



- In the **PURC model**, routes using link 2 become **more attractive**, increasing flow there.
- In the extreme:
    - If link 2's length = 0 → flow splits equally among 3 routes (1/3 each).
    - If link 2's length = 2 → flow splits 50/50 between the remaining routes.

→This shows the model **responds to overlap and dissimilarity** between routes.

MNL model: **No change in flows**, as it doesn't consider overlapping structure—only total route cost.
PSL model: Responds similarly to PURC, since it includes a <u>correction for path overlap</u>.

Experimenting with the model
– Toy network

	PURC	MNL	PSL
Choice set generation necessary	😊	😢	😢
Loops	😊	😢	😢
Assign zero flow	😊	😢	😢
IIA	😊	😢	😢
Overlap	😊	😢	😊

The **PURC (Perturbed Utility Route Choice)** model clearly outperforms the traditional MNL and PSL models due to:

1. No need for predefined choice sets.
2. Natural exclusion of illogical options (like loops)(MNL and PSL need to adjust the choice set).
3. Capability to assign **zero flow** to unattractive options.
4. Avoidance of IIA, reflecting more realistic choice behavior.
5. Demonstrate substitution patterns that depend on **route overlap**.

Experimenting with the model
– Copenhagen metropolitan area road network

Fig. 2. Network of the Copenhagen metropolitan area.

- The model is applied to a **large real-world road network** in the **Copenhagen metropolitan area**.

  30,773 links   12,876 nodes

- A sample trip is modeled from **Copenhagen Airport (SE)** to **Technical University of Denmark (North)**.

  Link utility $= \beta \times$ **pace**

  where "pace" is **minutes per kilometer**.

Experimenting with the model
- Copenhagen metropolitan area road network

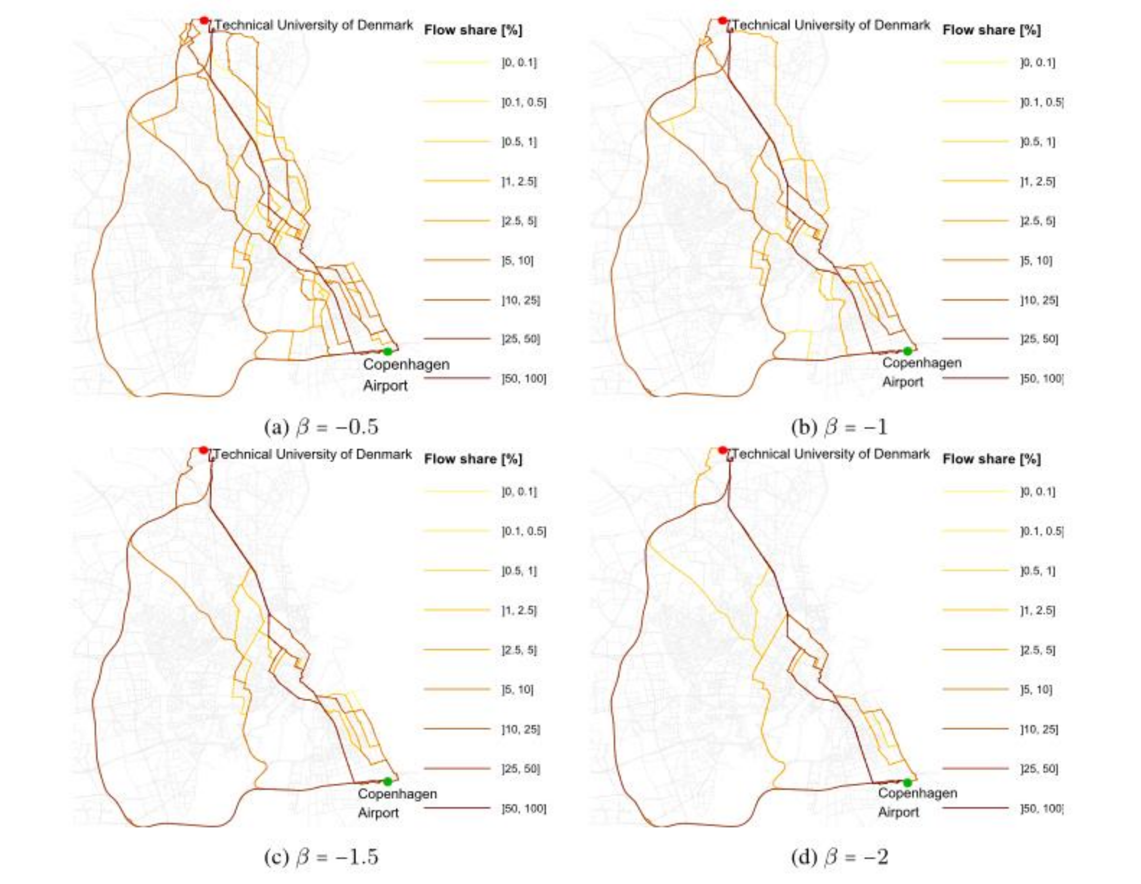


Fig. 3. Link flows using the PURC model with unit demand from Copenhagen Airport to Technical University of Denmark for different values of β.

Key observation

1. Sparse Use of Links

- **Most links show zero predicted flow**—they are inactive.
- This supports earlier findings (e.g., Rasmussen et al. 2017), even though they required full **path enumeration**.
- The **PURC model avoids path enumeration** by solving at the link level and still identifies **a limited set of active routes**.

2. Two clear main alternatives

- One route uses the **Kalvebod Bridge** (western motorway bypass).
- The other set of routes goes through **central Copenhagen**.

Experimenting with the model
– Copenhagen metropolitan area road network

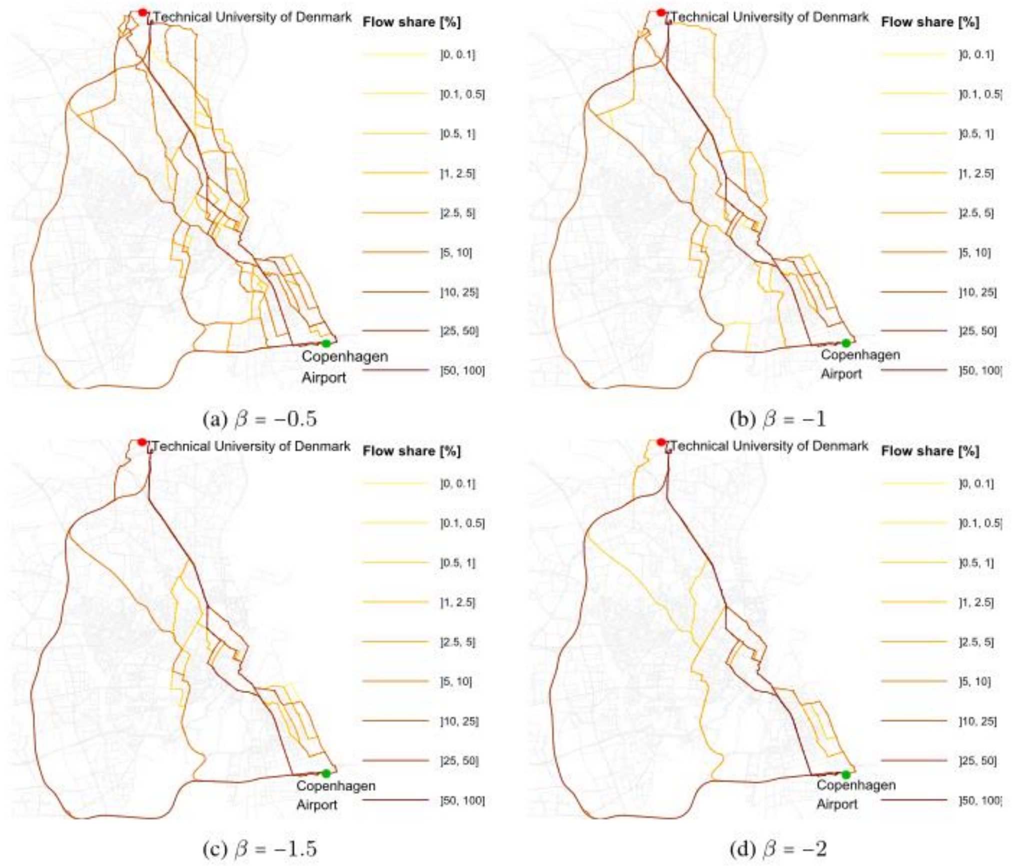


Fig. 3. Link flows using the PURC model with unit demand from Copenhagen Airport to Technical University of Denmark for different values of β.

Trade-Off: Time vs. Flow Distribution

•Based on the first-order condition, the model balances:
- **Minimizing travel time** (via β).
- **Distributing flow more broadly** (via perturbation function F).

•When $\boldsymbol{\beta}$ **is small**, the model favors **flow spreading** → more active links/routes.

•When $\boldsymbol{\beta}$ **is large**, the model favors **shorter paths** → concentrates flow on fastest routes.
- As β increases, **fewer routes** are used and city center routes become **less attractive**.

Experimenting with the model
– Copenhagen metropolitan area road network

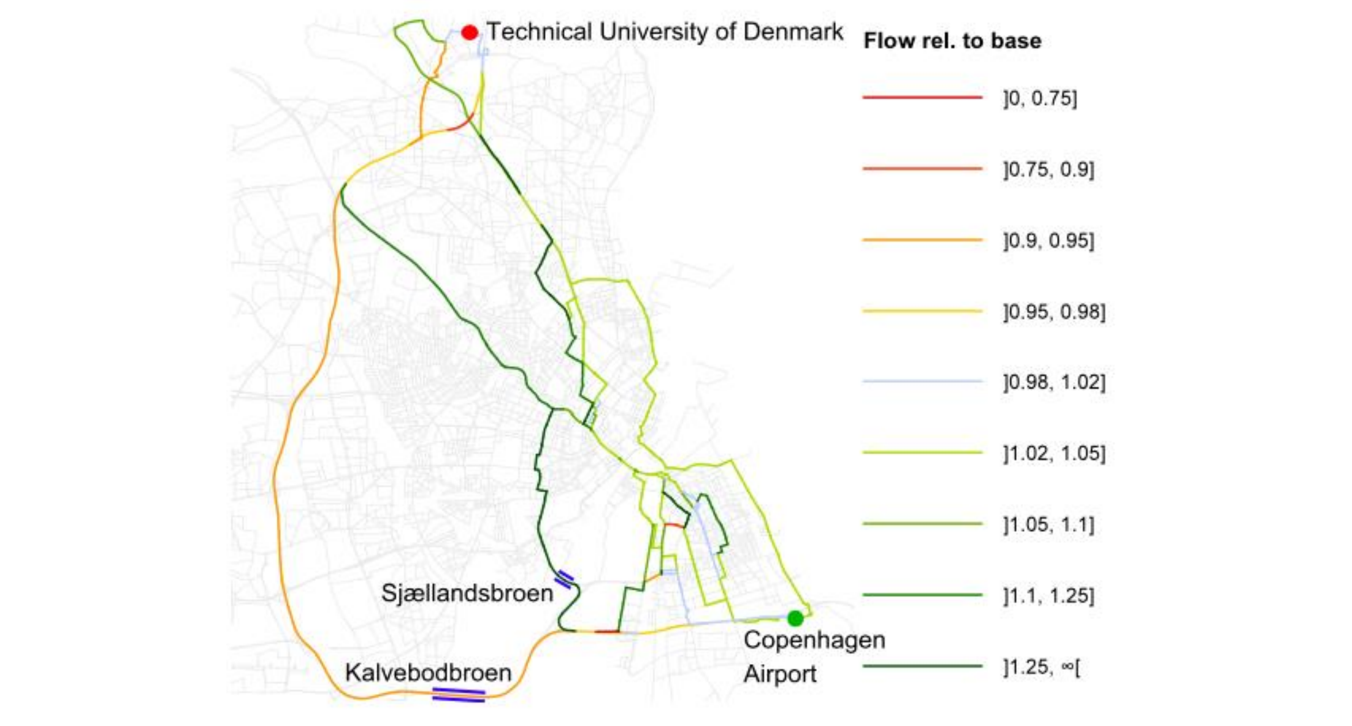


Fig. 4. Substitution patterns for trips between Copenhagen Airport and Technical University of Denmark using the PURC model and parameters estimated in model C (see Section 5.1) when increasing travel time on Kalvebodbroen by two minutes, relative to the base scenario with no alterations of travel times.

Policy Simulation: Travel Time Increase on Kalvebodbroen

•A +**2 minute delay** on Kalvebodbroen increases disutility for all routes using that bridge.

•As expected:

- **Flow shifts** to alternative routes.
- Nearby routes (e.g., those using **Sjællandsbroen**) absorb more flow.

•This confirms the model's **substitution behavior**:

Routes that **overlap** are **closer substitutes**— consistent with theory from **Proposition 1**.

Testing the estimator on simulated data for Copenhagen

Test the estimator's ability to **recover known parameters** using the Copenhagen metropolitan area network.

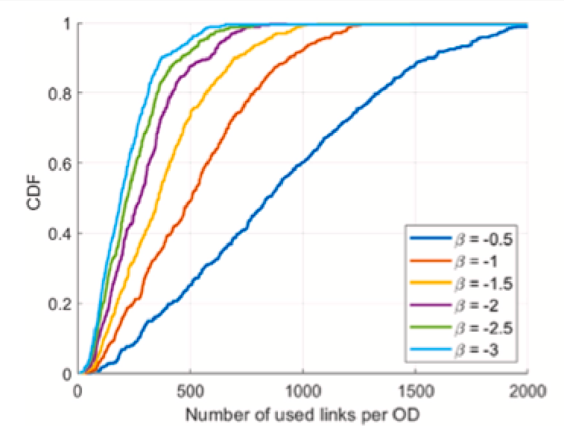
Simulation setup

22 nodes selected as origins/destinations (OD).

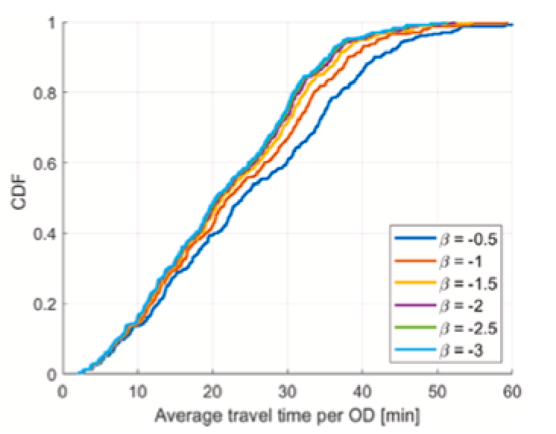
1 / 5 / 20 / 100 OD pairs

Pace parameter (β): Different values of β from $\{-3, -2.5, \ldots, -0.5\}$ to represent travel time per kilometer.

Testing the estimator on simulated data for Copenhagen



(a) Number of active links per OD



(b) Travel time per OD

Statistical Analysis & Results

•**96 datasets created** (4 OD sizes × 4 β values × 6 parameter settings).

•For each dataset:

- Calculate flow vector x for each OD.
- Transform data as per Section 2.3.
- Estimate β using Ordinary Least Squares (OLS).

Key Results

(a) **Active Links**: As β approaches 0, the influence of travel time decreases, leading to more active links.

(b) **Average Travel Time**: A smaller β (lower travel time weight) results in higher average travel time.

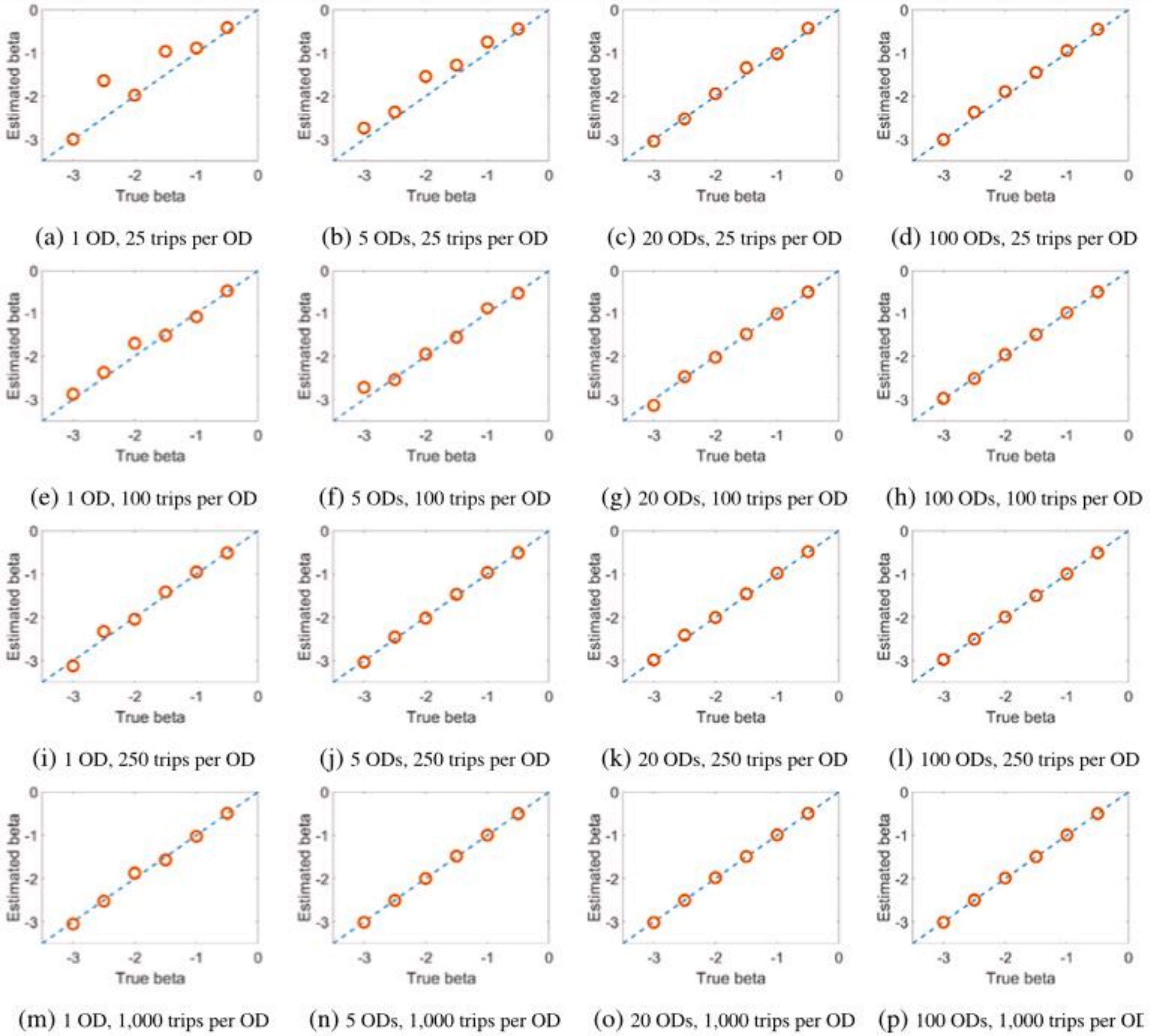


Fig. 6. Parameter estimates against the true parameters for different combinations of the number of trips per OD relation and number of OD relations.

Bias & Precision Analysis

- **All positive flow links** were included in the analysis without truncation.
- Non-linear function $\ln(1 + x)$ might introduce bias due to noisy flow estimates.
- **No noticeable bias** found when the dataset size is sufficiently large.

Reason

For small flow values, the logarithmic function behaves approximately linearly, mitigating significant bias.

Empirical application
– Estimation

Data Overview

GPS vehicle trajectory data from the Copenhagen metropolitan area.

Estimation dataset **8046** OD combinations **1,337,096** trips

Average **166.2** trips per OD relation **153.4** active links on average per OD relation

Utility Specification

Variables

1. Average observed pace (minutes per kilometer)
2. Number of outlinks
3. Dummy variables for road types

Empirical application
– Estimation

Model Comparisons

Model A (Base Model):
- **The utility rate:** Pace as the only variable.
- **Fit:** Flow is distributed across a moderate number of links.
- **Adjusted R^2** : Provides reasonable range.

Model B (Extended):
- **The utility rate** : Adds a constant for links with ≥ 2 outlinks (penalizes intersections).
- **Fit:** Expected decrease in the pace parameter.
- **Adjusted R^2** : Increases by 0.006.

Table 3
Estimation results, based on 1,234,289 observations corresponding to eq. (5). Numbers in parentheses are robust standard errors.

	Model A	Model B	Model C
$\text{Constant}_{\text{Outlinks} \geq 2}$	–	−0.03428 *(0.00030)*	−0.01546 *(0.00030)*
Pace [min/km]	−0.74642 *(0.00171)*	−0.63773 *(0.00225)*	–
Pace [min/km]:			
Motorways	–	–	−0.35011 *(0.00342)*
Motorway ramps	–	–	−0.55419 *(0.00437)*
Motor traffic roads	–	–	−0.56233 *(0.00464)*
Other national roads	–	–	−0.59969 *(0.00297)*
Urban roads	–	–	−0.56917 *(0.00219)*
Rural roads	–	–	−0.77783 *(0.00308)*
Smaller roads	–	–	−0.53003 *(0.01101)*
Other ramps	–	–	−0.44558 *(0.00448)*
Adjusted R^2	0.36288	0.36855	0.40880

Model C (Advanced):
- **The utility rate** : Interacts pace with link-type dummy variables.
- **Fit:** Parameters align with intuition for different link types.
- **Adjusted R^2** : Increases by 0.04 compared to Model B.

Empirical application
– Estimation

Model Fit

Adjusted R^2 values 0.4
(satisfactory for cross-sectional data)

The perturbation function form

1. Modified entropy function
   $F(x) = (1+x) \cdot \ln(1+x) - x$

2. The quadratic function
   $F(x) = x^2$

Table 4
Adjusted R^2 values for modified entropy and quadratic perturbation functions.

Model	A	B	C
Modified entropy	0.36288	0.36855	0.40880
Quadratic	0.36249	0.36695	0.40403

Empirical application
– Validation

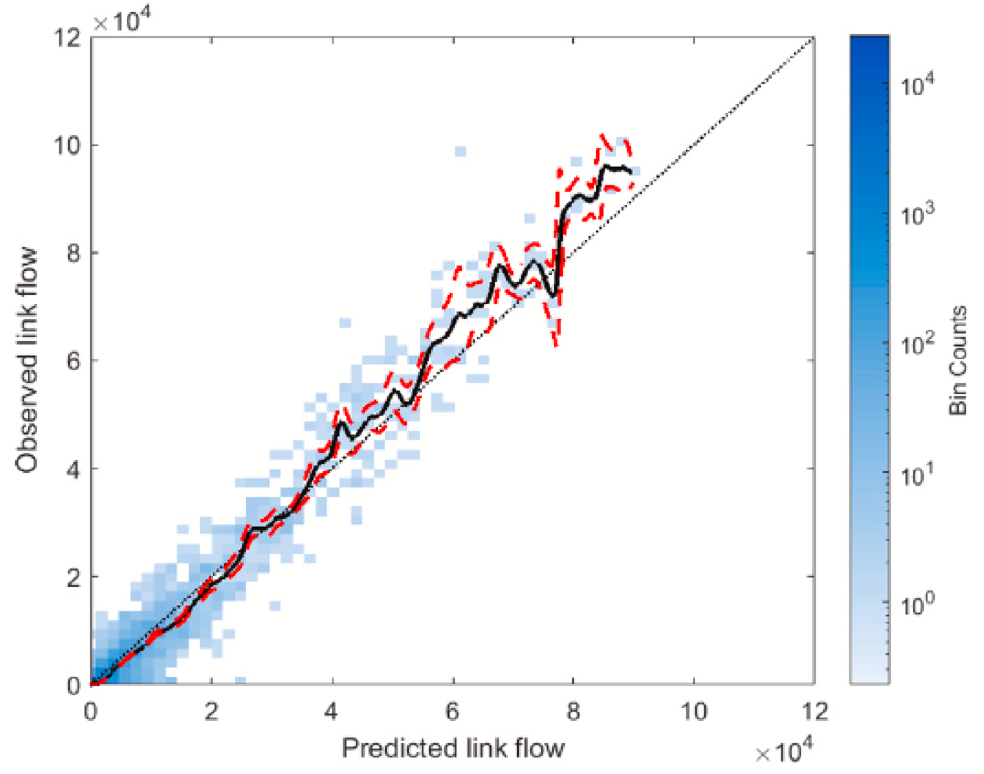


Fig. 7. Observed vs. predicted total link flows when summing across all OD combinations.

Prediction vs. Actual Data

Method: Predicted flow vectors for each OD combination are compared to actual link flows.

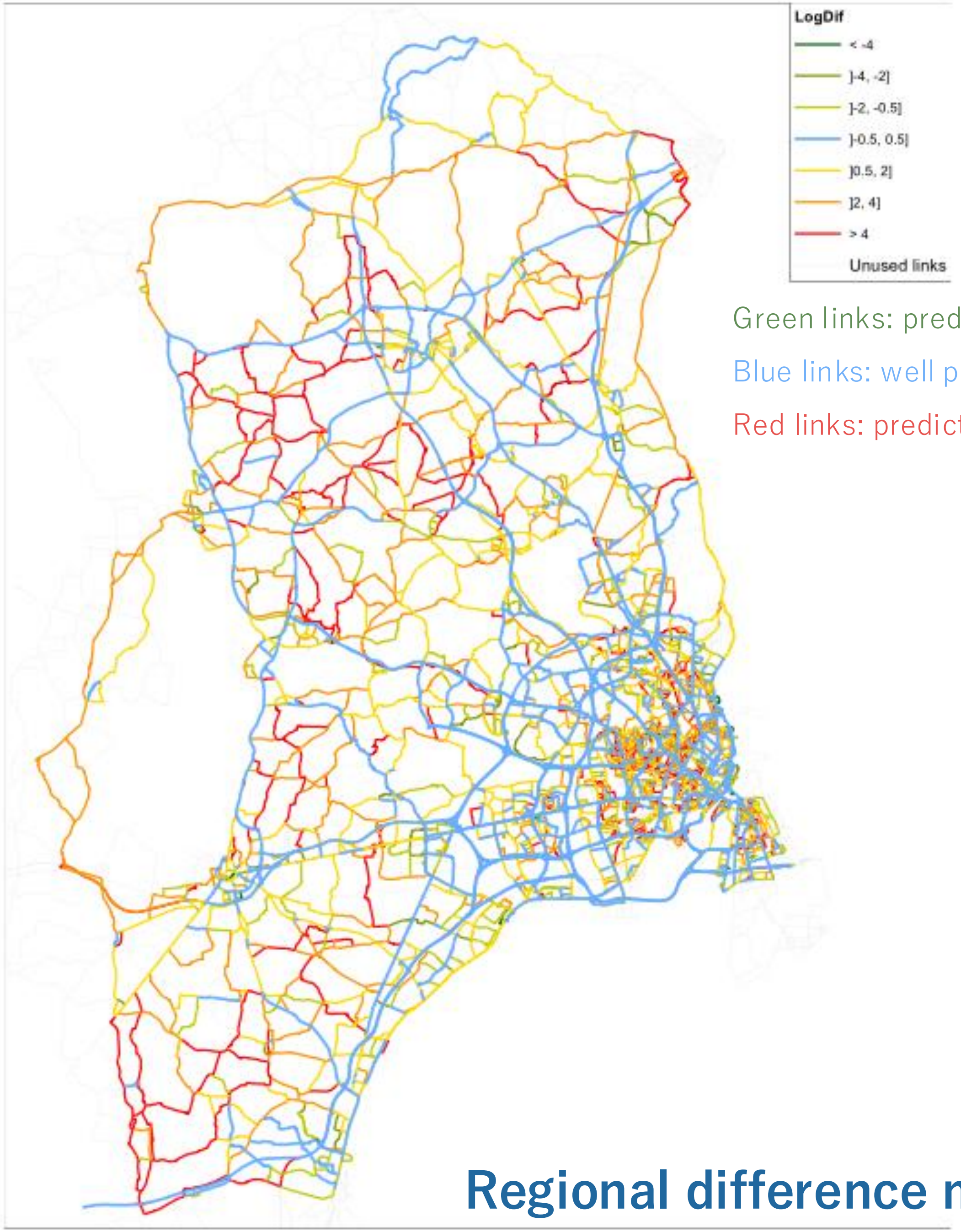
Out-of-Sample Validation: The model was tested on a split dataset, showing similar results and no significant issues.

Main Validation Result

Adjusted R^2 0.9356

Indicate a strong fit

The model predicts link flows accurately without calibration.



Green links: predicted < original

Blue links: well predicted

Red links: predicted > original

Fig. 8. Regional difference map. Differences given as $\ln(q^{predicted} + 1) - \ln(q^{observed} + 1)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 9. City-wide difference map. Differences given as $\ln(q^{predicted} + 1) - \ln(q^{observed} + 1)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Empirical application
– Validation

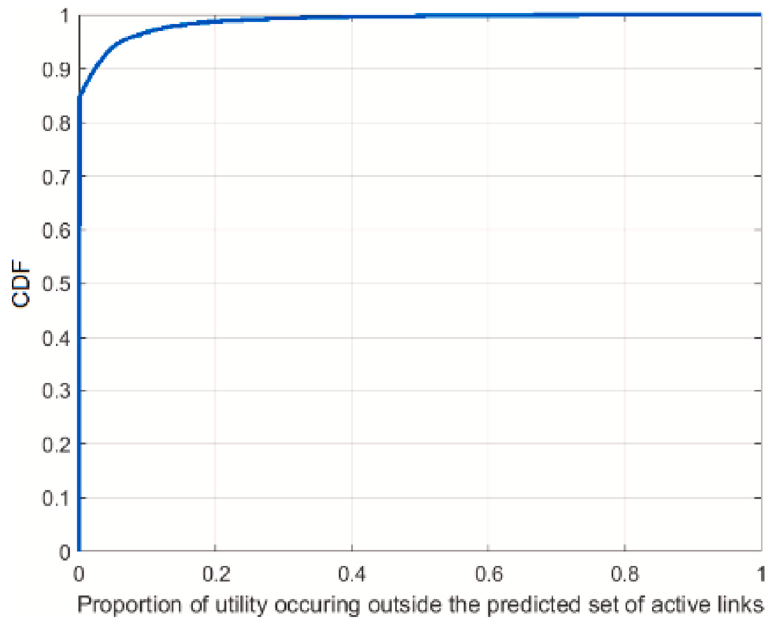


Fig. 10. Cumulative distribution function of the proportion of utility of the observed trips outside the predicted set of active links.

Active Link Prediction

Evaluate how well the model's **predicted active links** (per OD pair) **cover the actual observed routes**.

85 %

of observed trips are fully contained within the predicted active links, with minimal deviation.

Almost all trips have

<20 %

of their utility outside the predicted set.

4. Conclusion

Conclusion

Advantages of the PURC model

- It can be estimated using **simple linear regression**
- **No choice set generation** is required
- It is **easy to use for assignment**
- It yields **reasonable substitution patterns**
- Applied to a **large Copenhagen dataset**
- **Accurately predicts** observed traffic flows

Future research

- Address **sampling error** using **observation-based weighting**
- Explore **individual-level estimation** with iteratively reweighted least squares **methods**
- Extend to **dynamic or stochastic networks**
- Test alternative **perturbation functions**
- Add **cross-link interactions** for improved substitution modeling
- Integrate with **equilibrium assignment** frameworks (e.g., Oyama et al., 2022)

所感

- The content and methods of this research are highly related to our lab’s past research.
- As mentioned in the "Future Research" section, there are many directions for further exploration, and some research is currently ongoing in these areas.

- Other convex function (ex: semi-nonparametric perturbation function (TRC-30))
- Integrated with Markovian choice model （TRISTAN, 2025)

- The interaction between agents or day-to-day route choice may be applied as a potential research direction in the perturbed route choice model.
- Additionally, it can also be applied to route choice for medium- and long-distance logistics deliveries.