Route choice models

- From recursive logit to perturbed utility

Mogens Fosgerau

UNIVERSITY OF COPENHAGEN

Outline

- The route choice challenge
- ARUM discrete choice
- Recursive logit

- Background
  - ARUM and duality
  - Perturbed utility

- Perturbed utility route choice (PURC)
  - PURC in Copenhagen
  - Equilibrium assignment and welfare
  - Estimation with micro-data

- Conclusion

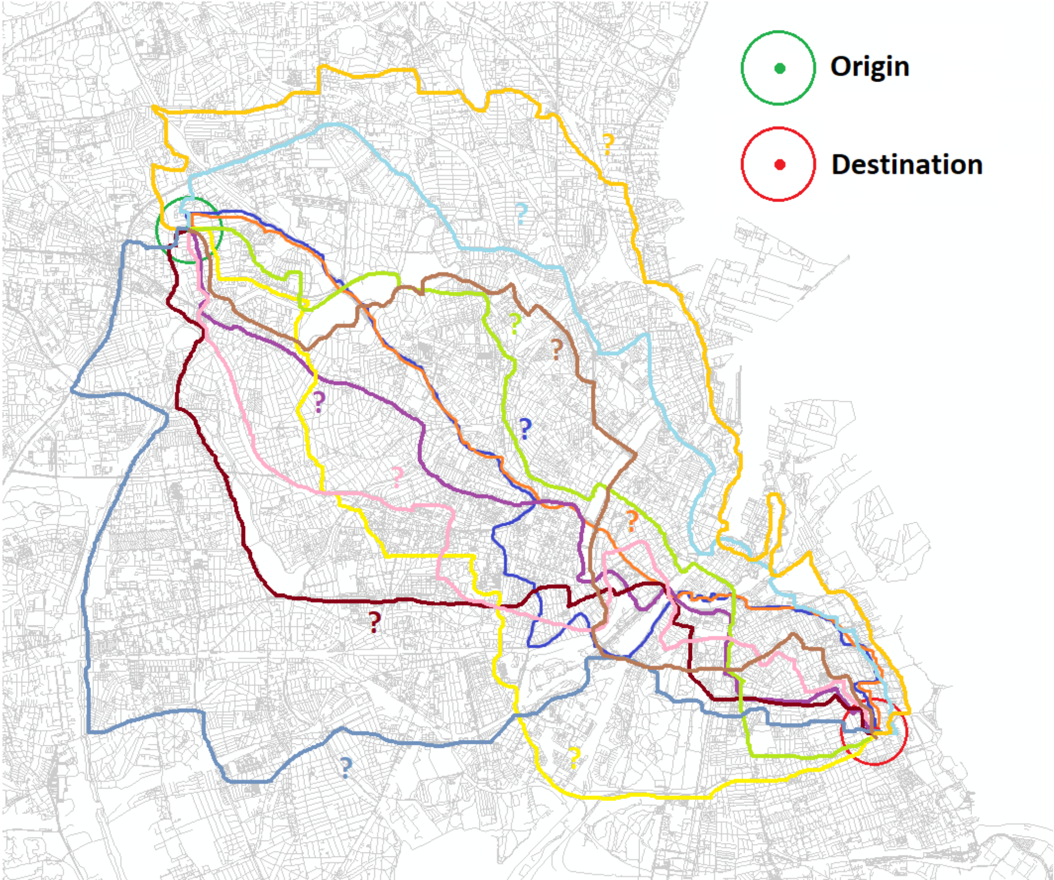
Route choice

- For trips from an origin (O) to a destination (D) through a transport network

- Determine the distribution of trips on paths

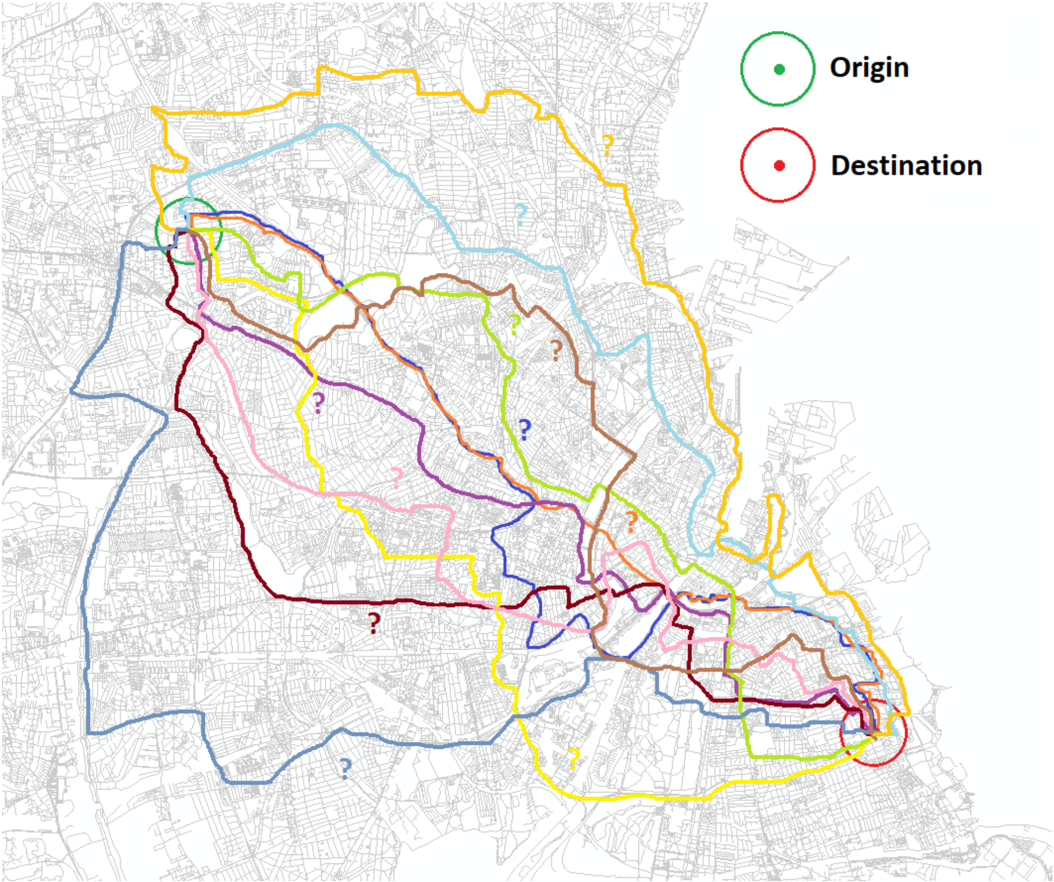
  or more fundamentally

- Determine the distribution of flow on links



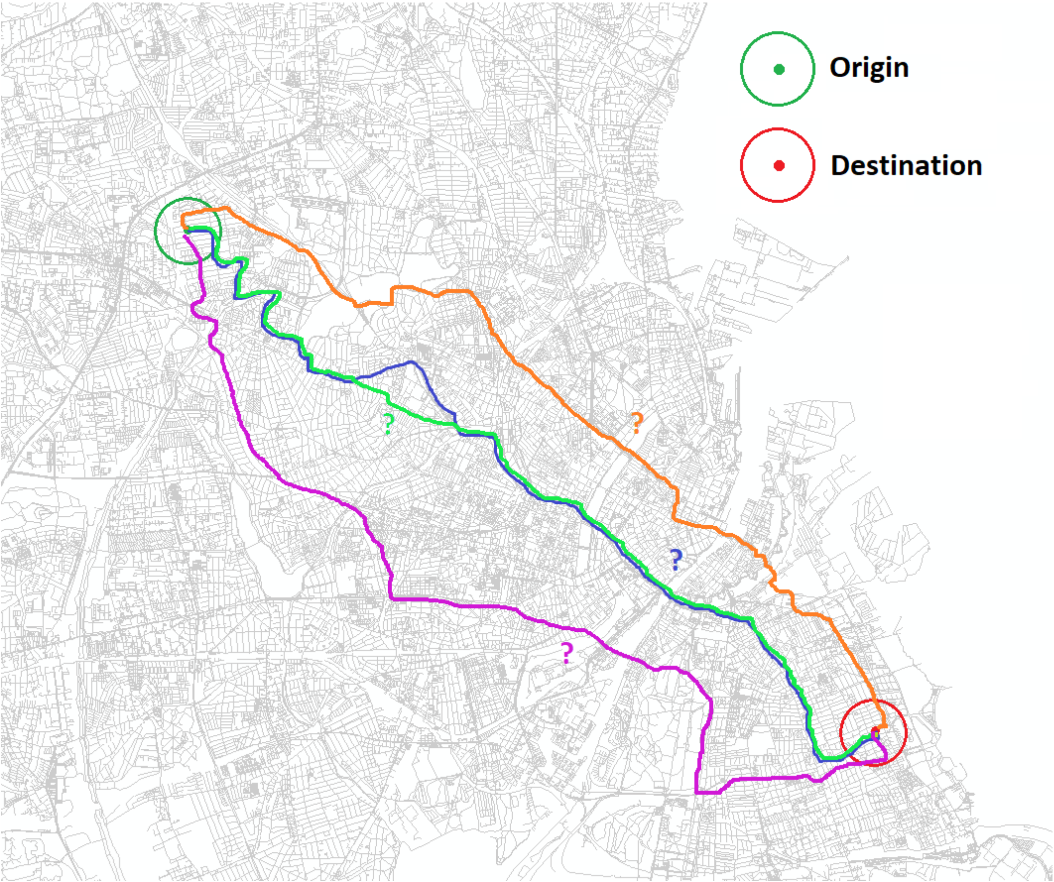
Choice alternatives?

- Not allowing paths with loops:
  - Like the number of atoms in the universe, squared!
- Common approach
  - Discrete choice model assigns probabilities to route alternatives
- Choice set generation to form a small, reasonable choice set.
  - However, there is data loss and bias



Substitution between paths

- Many alternatives share links
- "Correlation" is inherent
- Substitution patterns should incorporate network structure
- Large-scale applied models are mostly just logit models.
  - Suffer from Independence of Irrelevant Alternatives (IIA) property



The route choice challenge

We now have massive data that trace movements through complex traffic networks

- Ideal for modelling route choice!

However, we face the curse of dimensionality

- The number of routes through a real network is extremely large

It remains difficult to

- Represent data well
- Achieve models with realistic substitution patterns
- Estimate models and compute equilibrium with large networks and data

The route choice challenge

In conclusion, we

- Must be able to
  - Estimate model parameters from observed route choice data
  - Compute predictions
  - Compute equilibrium
  - Deal adequately with substitution patterns (correlation)
- Must be able to deal with
  - Massive data
  - Large networks

Next, we take a first look at the ARUM discrete choice model

The ARUM discrete choice model

The model of choice for most transportation problems since McFadden (1978)

- A finite set of options $J = \{1, \dots, J\}$
- A utility associated with each option $u_j = v_j + \epsilon_j$
- Consumer observes chooses the "best" option $\mathrm{argmax}_{j \in J}\{u_j\}$

- Options are treated as distinct entities
- Framework fits well with small to moderate choice sets
  - Computationally tractable
  - Can represent substitution patterns

Problems applying discrete choice route choice

- It is impossible to enumerate all alternatives
  - Does not agree well with the discrete choice model

  - Logit: IIA property is unrealistic
  - Network MEV: Computationally challenging?
  - Probit: Even more challenging

The consideration set

- Claim that people only “consider” a small subset of alternatives
  - Somehow outside the model…
  - Ad hoc device to make the problem tractable for a discrete choice model

- What controls consideration?
  - Should be part of the model
  - Now, there are papers that, rightly, treat the formation of consideration sets as part of the model ☺

ARUM discrete choice

In conclusion,

- ARUM discrete choice models are not well suited to route choice, due to
  - The complexity of networks
  - The curse of dimensionality

- Next, we review the recursive logit model
- A model that addresses these problems

How I got started on route choice - a little personal history

- I was exposed to route choice papers at the IATBR in Kyoto (2006), talking about choice set generation and consideration sets
  - I thought the problem of choice set generation consideration was a self-inflicted problem, due to the choice of an ARUM discrete model

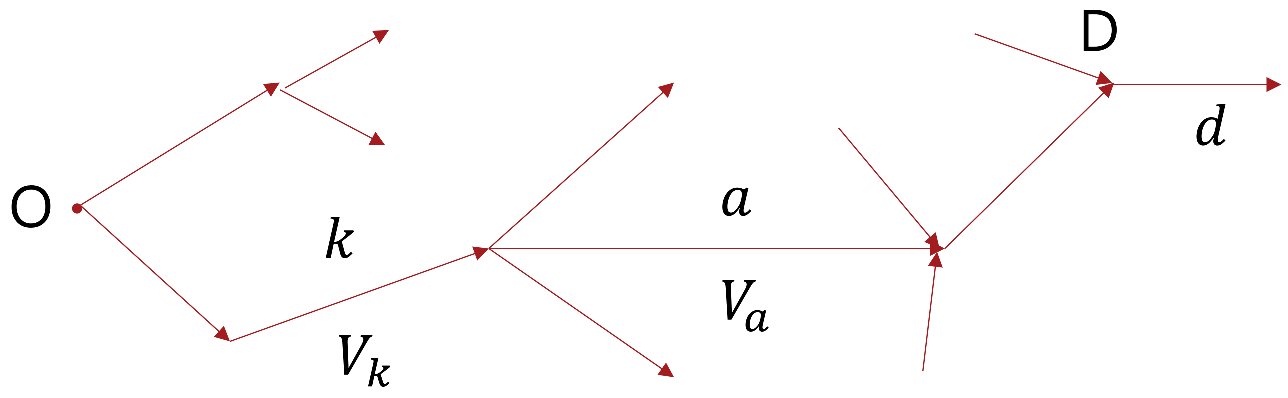
- But what could be done instead?
  - Emma Frejinger and I spent much time walking and bicycling around Kyoto, discussing route choice problems quite a bit!
  - Later, when we met in Stockholm, we developed the recursive logit model together with Anders Karlström

The recursive logit model

- Dynamic discrete choice of the next link

$$u(a|k) = v(a|k) + \epsilon_a + V_a$$

$$V_k = \mathbb{E}[\max_{a \in A_k} v(a|k) + \epsilon_a + V_a]$$



With EV1 distributed errors, the next link probability is

$$P(a|k) = \frac{e^{v(a|k)+V_a}}{\sum_{a' \in A_k} e^{v(a'|k)+V_{a'}}}$$

Fosgerau, M., Frejinger, E., Karlstrom, A. (2013) A link-based network route choice model with unrestricted choice set. Assignment same as Akamatsu (1996)

The recursive logit model

- The choice of the next link is a choice problem with a small choice set
- Probability of next link is easily computed using logit formula

$$P(a|k) = \frac{e^{v(a|k)\ + V_a}}{\sum_{a' \in A_k} e^{v(a'|k)\ + V_{a'}}}$$

- Hence the probability of any chosen path is easily computed
  - We can estimate parameters in $v(a|k)$
- However, we need first to compute the continuation values V_a by solving the Bellman equation

$$V_k = \mathbb{E} \max_{a \in A_k} [v(a|k) + V_a + \epsilon_a]$$

Solving the Bellman equation

- In the case of logit, the continuation value has an easy expression

$$V_k = \mathbb{E} \max_{a \in A_k} [v(a|k) + V_a + \epsilon_a] = \ln \sum_{a \in A_k} e^{v(a|k) + V_a}$$

- The Bellman equations can be cast in matrix form and solved inverting a matrix or using value iterations

- Can then use NFXP (Rust, 1987) to estimate parameters. Iterate until convergence:
  - Given current parameters, solve Bellman equation to obtain V_a
  - Given V_a, compute likelihood function and update parameters

A lively literature! Some notable papers

- A nested recursive logit model for route choice analysis. Mai, Fosgerau, Frejinger (2015)
- A method of integrating correlation structures for a generalized recursive route choice model. Mai (2016)
- A discounted recursive logit model for dynamic gridlock network analysis. Oyama, Hato (2017)
- Prism-based path set restriction for solving Markovian traffic assignment problem. Oyama, Hato (2019)
- A tutorial on recursive models for analyzing and predicting path choice behavior. Zimmerman, Frejinger (2020)
- Estimation of a recursive link-based logit model and link flows in a sensor equipped network. van Oijen, Daamen, Hoogendorn (2020)
- Markovian traffic equilibrium assignment based on network generalized extreme value model. Oyama, Hara, Akamatsu (2022)
- Capturing positive network attributes during the estimation of recursive logit models: A prism-based approach. Oyama (2023)
- A dynamic discrete choice modelling approach for forward-looking travel mode choices. Leong, Nassir, Mohri, Sarvi (2024)

Recursive logit

Recursive logit has significant advantages

- No consideration set, allows whole network
- Can be estimated without bias with micro-data
- Handles substitution (well?)
    - Mai, Fosgerau, Frejinger (2015), Mai (2016), Oyama, Hara, Akamatsu (2022)

Remaining challenges

- How to handle computational complexity with large network, computing continuation values very many times?
- Is Markov assumption ok empirically?

- Next, we review ARUM and duality
- Aim to find a basis for formulating a new kind of model

Outline

- The route choice challenge
- ARUM discrete choice
- Recursive logit

- Background
  - ARUM and duality
  - Perturbed utility

- Perturbed utility route choice (PURC)
  - PURC in Copenhagen
  - Equilibrium assignment and welfare
  - Estimation with micro-data

- Conclusion

ARUM and duality

- The surplus function $W(v) = \mathbb{E}\, max_j [v_j + \epsilon_j]$ is a convex function.
  - (Assume ϵ has a density with full support)
- Define the conjugate surplus $W^*(q) = \sup_v \{v'q - W(v)\}$
- Williams-Daly-Zacchary Theorem:

the choice probability vector is the gradient of the surplus

$$P(v) = \nabla W$$

- Hotz-Miller Theorem:

the inverse choice probability function is the gradient of the conjugate surplus

$$P^{-1}(p) = \nabla W^*(p)$$

See Sørensen & Fosgerau (2022) for theory under maximally general assumptions

An alternative representation of ARUM

- The surplus function is the conjugate, conjugate surplus!

$$W(v) = \sup_{p \in \Delta}\{v'p - W^*(p)\}$$

$$\nabla W^*(P(v)) = v$$

- The ARUM model can be represented as a model of optimal choice probabilities

ARUM and Duality

- The ARUM model can be represented as a model of optimal choice probabilities

$$W(v) = \sup_{p\in\Delta}\{v'p - W^*(p)\}$$

$$\nabla W^*(p(v)) = v$$

- Next, we will use this structure to define a broader class of models:
  - the perturbed utility model

The general perturbed utility model (PUM)

- A consumer maximises perturbed utility

$$\max_{x \in B} v'x - F(x)$$

  - x: consumption vector
  - $v \in \mathbb{R}^n$: vector of subutilities
  - $F: \mathbb{R}^n \to \mathrm{R}$: convex function (perturbation)
  - $B \subseteq \mathbb{R}^n$: compact budget set

- PUM generalises a range of models, including standard consumer theory, ARUM, Gentzkow bundles, choice under risk, stochastic choice, moral hazard, and matching

Fosgerau and McFadden (2012), Fudenberg et al (2015), Allen and Rehbeck (2019)

Logit as perturbed utility

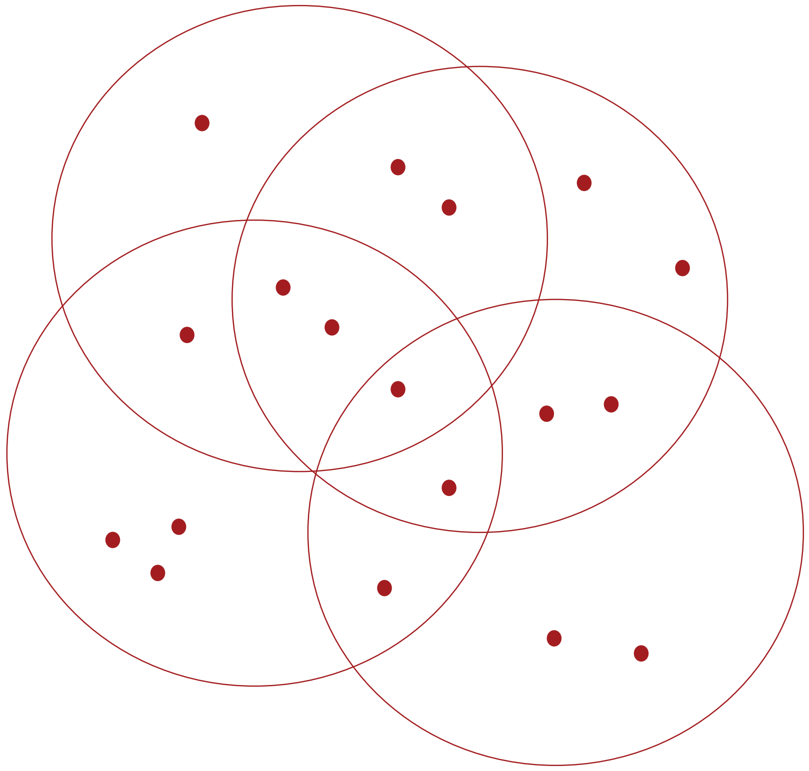
Perturbed utility representation $$\max_{p\in\Delta} u(p) = v'p - p'\ln p$$

- Δ is the probability simplex.
- Perturbation is entropy
- The optimal choice probability vector p maximizes utility given v
- $y \in \{0,1\}^I$ choice indicator
- The observed discrete choice is random $$\mathbb{E}[y|v] = P(v)$$

Inverse choice probabilities: $$v = \ln P(v) - \underline{1}$$

Inverse product differentiation logit (IPDL)

- Generalizes nested logit
- Allows arbitrarily overlapping nests
- This form allows non-ARUM models
  - Alternatives can be complements



Fosgerau, Monardo, de Palma (2024)

Estimation of perturbed utility model via inverse demand

Consider consumers who maximize perturbed utility

$$U(p) = v'p - W^*(p) \qquad s.t.\, p \in \Delta$$

Interior solution for perturbed utility demand $\quad \mathrm{v} = \nabla W^*(p) + \lambda \underline{1}$

Specify $v = z\beta$ to obtain $\quad \nabla W^*(p) + \lambda \underline{1} = z\beta$

- Nonlinearity is confined to ∇W^*.
- β can be estimated regressing $\nabla W^*(p)$ on z

Fosgerau; Monardo, de Palma (2024)

Perturbed utility

In conclusion,

- Perturbed utility is a general framework for modelling consumer choice
- Incorporates discrete choice and much more
- Parameters can be estimated via the inverse demand function

Now use perturbed utility framework to build a route choice model

- No explicit routes
- No choice set
- Just a straightforward convex optimization problem

Outline

- The route choice challenge
- ARUM discrete choice
- Recursive logit

- Background
  - ARUM and duality
  - Perturbed utility

- Perturbed utility route choice (PURC)
  - PURC in Copenhagen
  - Equilibrium assignment and welfare
  - Estimation with micro-data

- Conclusion

Perturbed utility route choice (PURC): network

- Directed network with vertices (nodes) $\mathcal{V}$ and directed edges (links) $\mathcal{E}$
- The network is connected, it is possible to go from any node to any other node
- $ij \in \mathcal{E}$ is a link from node i to node j
- The node-link incident matrix $A \in \mathbb{R}^{|\mathcal{V}||\mathcal{E}|}$ has entries

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

More definitions

- A unit demand vector $b \in \mathbb{R}^{|\mathcal{V}|}$ is given as

$$b_v = \begin{cases} -1 & v \text{ is the origin} \\ 1 & v \text{ is the destination} \\ 0 & \text{otherwise} \end{cases}$$

- A network link flow vector $x \in \mathbb{R}_+^{|\mathcal{E}|}$ is vector of node net outflows
- Flow conservation: $Ax = b$
- For a trip with OD b satisfying flow conservation, the link flow x_e is the probability of observing the traveller on link e

Perturbed utility route choice

The traveller is assumed to minimize cost (negative perturbed utility)

$$x^*(c) = \underset{x \in \mathbb{R}_+^{|\mathcal{E}|}, Ax=b}{\operatorname{argmin}} \; c'x + F(x)$$

- Budget ensures flow conservation from O to D
- Demand is a flow vector $x^*(c)$, not a probability vector
- Observed choice $y \in \{0,1\}^{\mathcal{E}}$ satisfies $\mathbb{E}[y|c] = x^*(c)$
- Perturbation function F constructed to incorporate network structure and allow zero flows on many links

Utility specification

- Link costs $c \in \mathbb{R}_{++}^{|\mathcal{E}|}$.
  - In the route choice context, it is more natural than utilities

- Define network perturbation function to incorporate network structure

$$F(x) = \sum_{e \in \mathcal{E}} F_e(x_e)$$

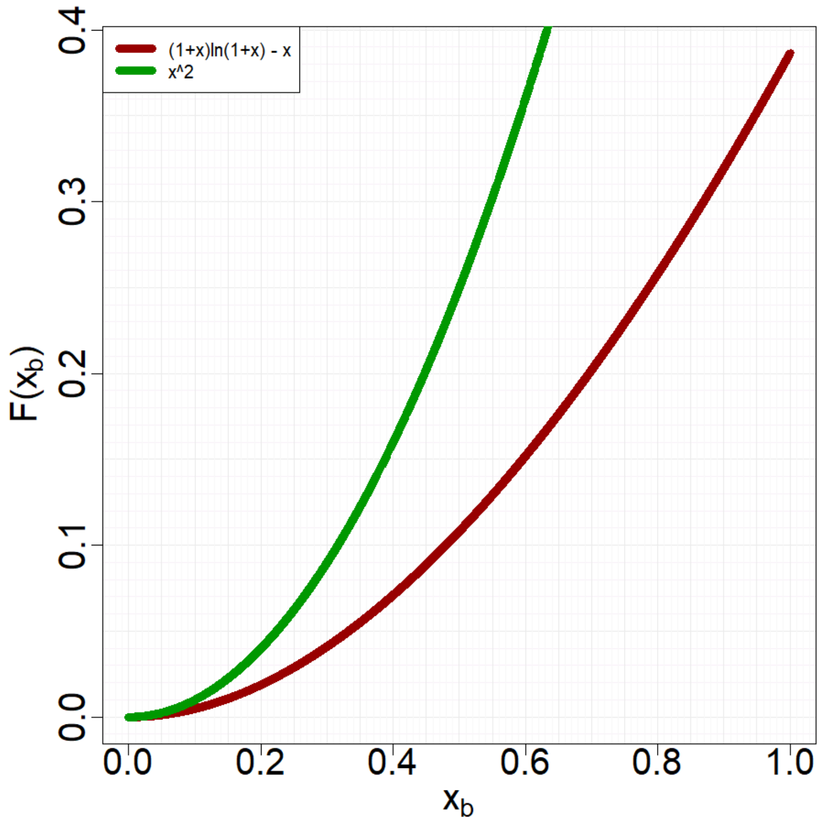
- Link-specific convex perturbation functions $F_e: \mathbb{R}_+ \to \mathbb{R}_+, \forall e \in \mathcal{E}$
  - Strictly convex and strictly increasing with $F_e(0) = F_e'(0) = 0$

The role of the perturbation function

$$\min_{x \in \mathbb{R}_+^{|\mathcal{E}|}, Ax=b} \quad c'x + F(x)$$

- Induces travellers to distribute flows across multiple paths
  - Without it, assignment would be shortest path
  - With it, flow spreads on a few shortest paths

- Examples
  - $F(x_e) = \text{length}_e \ (1+x)ln(1+x) \ - x$ or
  - $F_e(x_e) = \text{length}_e x_e^2$, e.g.



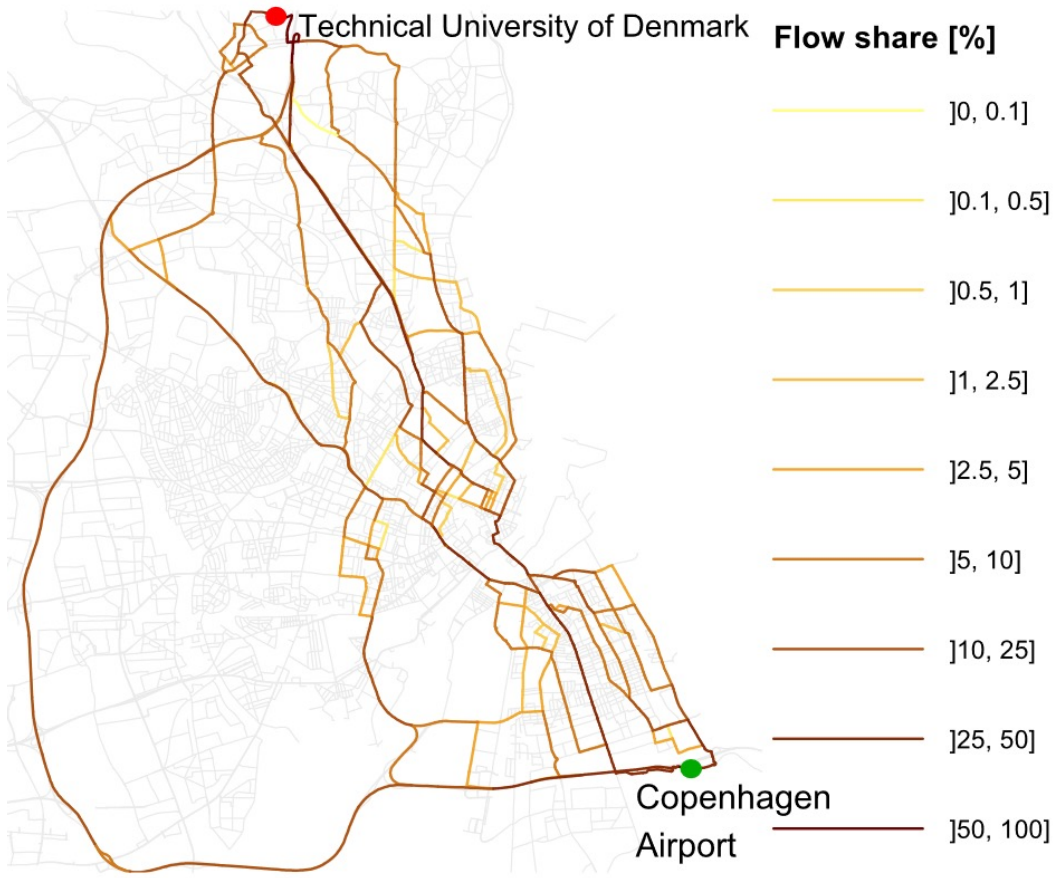
PURC addresses some challenges of route choice modelling

- No choice set generation required, uses the complete network as it is
- Allows that most links are unused in any given OD pair
- Correlation induced directly by network structure

Will show illustrations of these points

A small and plausible set of active routes for each OD

- In PURC, corner solutions occur naturally
- Flow on the shortest paths, usually more than one

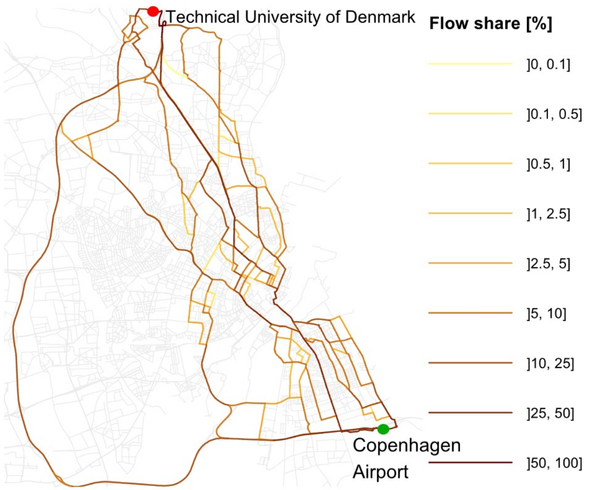


Correlation / substitution

In PURC, correlation comes directly from the complete network structure

Not from correlation of random terms

So better to call it substitution



Let x be optimal flow for an OD pair

Let $\sigma \subset \mathcal{E}$ be an active path in x, i.e. $e \in \sigma \Leftrightarrow x_e > 0$.

Then

$$\sum_{e \in \sigma} c_e + F_e'(x_e)$$

does not depend on σ

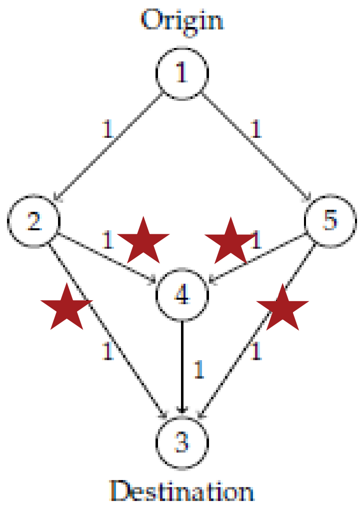
All active paths have same marginal cost

Example where paths are substitutes or complements

This example network contains 4 paths, where path flows are equal to the flows on links (23), (24), (54), (53)

Matrix shows derivatives of link flows (vertical) by link costs (horizontal)

We observe both substitutes and complements



$\nabla x^*(c) =$

	(12)	(15)	(23)	(24)	(43)	(54)	(53)
(12)	−0.333	0.333	−0.167	−0.167	0	0.167	0.167
(15)	0.333	−0.3330	0.167	0.167	0	−0.167	−0.167
(23)	−0.167	0.167	−0.458	0.292	0.25	−0.0417	0.208
(24)	−0.167	0.167	0.292	−0.458	−0.25	0.208	−0.0417
(43)	0	0	0.25	−0.25	−0.5	−0.25	0.25
(54)	0.167	−0.167	−0.0417	0.208	−0.25	−0.458	0.292
(53)	0.167	−0.167	0.208	−0.0417	0.25	0.292	−0.458

The perturbed utility model has potential

In conclusion,

- No choice set generation required, uses the complete network as it is
- Allows that most links are unused in any given OD pair
- Substitution induced directly by network structure

- But how can we estimate parameters for the model from observed route choice data?

Estimation with aggregate data

- Consider first case where flows are regarded as observed
- Complementarity condition for an OD pair
  - With many Lagrange multipliers λ^* corresponding to flow conservation $Ax = b$

$$\forall e \in \mathcal{E}, \text{ either } x_e^* = 0 \text{ or } c + F_e'(x_e^*) + (A'\lambda^*)_e = 0$$

- Must deal with very many zeros (most link flows are zero) and lots of Lagrange multipliers λ^* (one for each network node)
- Will remove the Lagrange multipliers!

Fosgerau, Paulsen, Rasmussen (2022)

Remove the Lagrange multipliers

- Define B as identity matrix with rows corresponding to $x_e^* = 0$ removed
- Define projection on the null space of BA'

$$P = B - BA'(BA')^+$$

- The projected first-order condition

$$P\left(c + \nabla F(x^*)\right) = 0$$

- This will be VERY important!

Regression equation

- We have

$$P\left(Z\beta + \nabla F(x^*)\right) = 0$$

or

$$P\ \nabla F(x^*) = -P\ Z\beta$$

- This suggests regression equation for demand b

$$\text{"y"} = \text{"x"}b + \epsilon$$

where $\text{"y"} = P\ \nabla F(x^*)$ and $\text{"x"} = -PZ$

How does it work in practice?

- Can estimate a PURC from aggregate OD data

- Examples from Copenhagen car data

Application to Copenhagen car data

GPS vehicle trajectory data: 8,046 active ODs and 1,337,096 trips HOW MUCH REDUCED

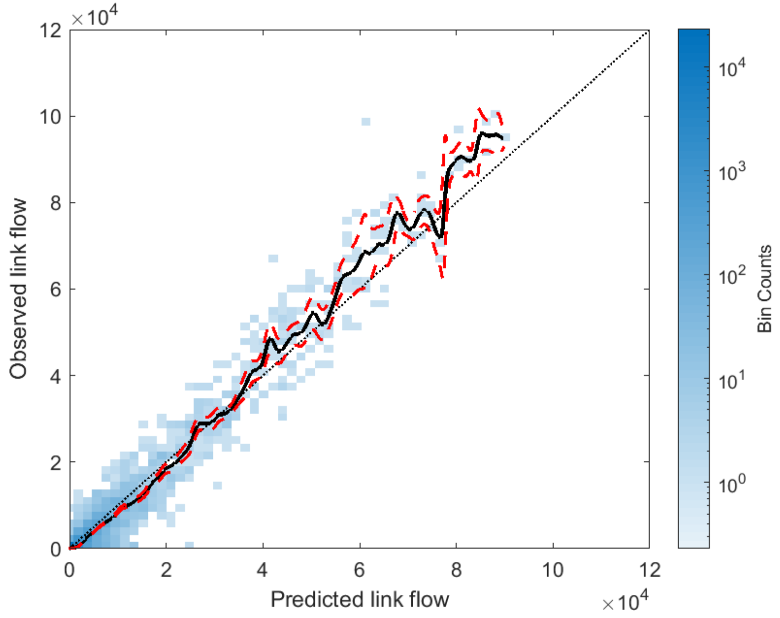
- avg observed travel time per link
- #outlinks
- road type dummies

	Model A	Model B	Model C
$\text{Constant}_{\text{Outlinks}\geq 2}$	–	-0.03428 *(0.00030)*	-0.01546 *(0.00030)*
Travel time [min.]	-0.74642 *(0.00171)*	-0.63773 *(0.00225)*	–
Travel time [min.]:			
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

Validation against data (model C)

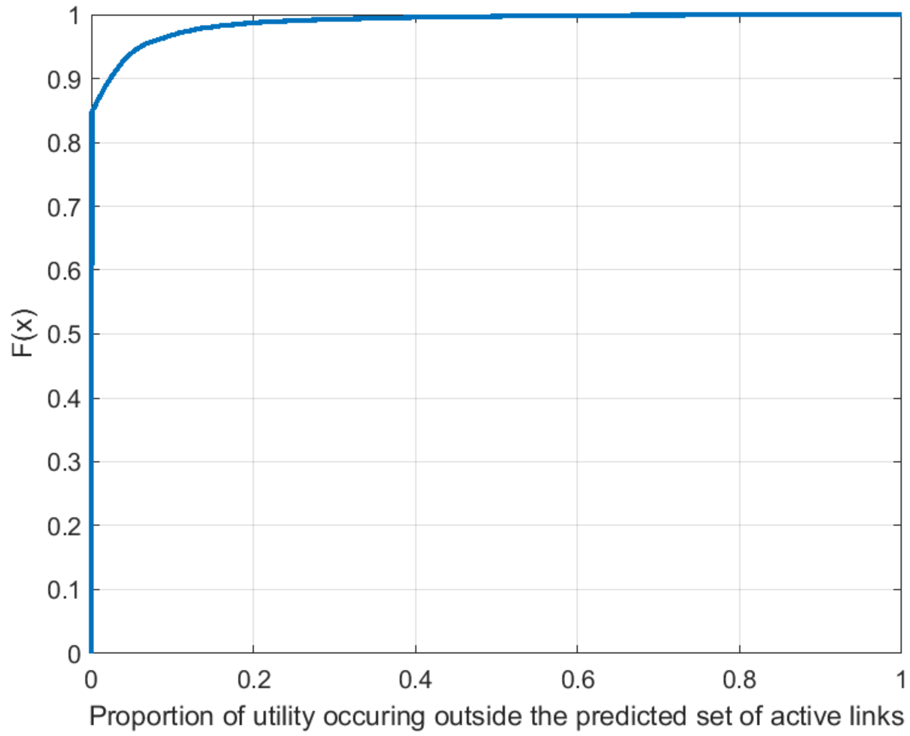
Compare predicted to observed link-flows

- Predictions are generally close to the 45°
- The fit is quite satisfactory, as we have employed no calibration whatsoever - the predictions are driven solely by the 9 estimated parameters
- Out-of-sample prediction test did not reveal any issues



How well do predicted sets of active links cover the chosen routes?

- Cumulative distribution of the proportion of utility of chosen routes outside the predicted set of active links.
- A value of zero is perfect and small values are good
- About 85% of observed trips are completely inside the predicted set of active links
- Almost all observed trips have less than 20% of cost outside the predicted set of active links



Application to Copenhagen data

- It works!
- Reasonable parameter estimates
- Matches observed link-flows well
- Covers chosen routes well

- Next, we will look at equilibrium assignment with PURC
- Will utilize the convex structure and the projected first-order condition

$$P\left(Z\beta + \nabla F(x^*)\right) = 0$$

The equilibrium problem

- Real network are large
  - E.g. 50,000 links and 20,000 nodes
- Each PURC problem is large
  - 50,000 decision variables and 20,000 constraints
- Data can have millions of trips.
  - The PURC problem must be solved for each trip (in principle)

- Each link in the network is congestible
  - Travel time depends on total demand
  - Demand depends on travel time
- Altogether, finding equilibrium is a huge computational problem.
  - Brute force is completely infeasible
  - Must have very fast algorithm

Yao, Fosgerau, Paulsen, Rasmussen (2024)

Equilibrium assignment

- Set up a primal convex minimization problem
  - Beckmann et al (1995) trick: First-order conditions are individual optimality conditions
  - Solution is the Wardrop/Nash equilibrium
- Dual problem
  - Find optimal flows as functions of dual variables
  - Substitute into primal problem
- Dual problem is unconstrained
  - Use fast gradient methods for solving the dual problem
  - Mix with travel time updates by quasiNewton steps
  - Iterate until convergence
- Primal solution can be computed directly

Yao, Fosgerau, Paulsen, Rasmussen (2024)

Dual algorithms runtime performance with real networks

Table: Dual algorithm runtime performances

Network	Problem size $(\lvert\mathcal{N}\rvert \times \lvert\mathcal{W}\rvert)$	Runtime [s] qN-AGD*	qN-AGD	AGD*	AGD
Sioux Falls	1.27×10^4	0.25	0.43	1.95	5.78
Berlin-Friedrichshain	1.13×10^5	2.17	7.23	42.83	144.12
Berlin-Tiergarten	2.31×10^5	3.23	7.70	142.85	410.33
Anaheim	5.85×10^5	0.58	0.65	16.87	26.71
Berlin MPFC	9.26×10^6	68.99	72.01	1487.23	3742.31
Chicago-Sketch	8.69×10^7	94.90	125.77	7196.18	9322.30

Yao, Fosgerau, Paulsen, Rasmussen (2024)

Equilibrium assignment

- Equilibrium assignment with PURC is very fast
- We can handle very large problems

- Next, we consider how to analyse marginal changes
- What is the effect on demand and welfare of a small change in costs?
- Will again rely on the projected first-order condition

$$P\left(Z\beta + \nabla F(x^*)\right) = 0$$

Sensitivity analysis with PURC

- The activation boundary $\mathcal{C}_0 = \left\{c\colon 1_{x^*(c)} \text{ is not continuous at } c\right\}$ has Lebesgue measure 0
- x^* is continuously differentiable on the complement of $\mathcal{C}_0$ with symmetric, negative semidefinite Jacobian
$$\nabla x^*(c) = -(P\nabla^2 F(x^*)P)^+$$
  - Need only invert a matrix corresponding to the active network
- Similarly, we can compute the change in equilibrium flow for marginal changes in cost parameters, without resolving for equilibrium

Welfare analysis of marginal changes

- In the same way, we can consider the effect on welfare of marginal changes
- The welfare function

$$W(-c) = \max_{Ax=b, x\geq 0} \{-x'c - F(x)\}$$

is continuously differentiable and

$$\nabla W(-c) = x^*(c)$$

Approximate welfare analysis of network changes

- To compute the welfare change following a marginal change in the link cost vector, we may apply a first-order Taylor approximation.

$$\Delta W(-c) \simeq x^*(c)\Delta c$$

- We can also compute a second-order approximation

$$\Delta W(-c) \simeq x^*(c)\Delta c + \frac{1}{2}\Delta c' \nabla x^*(c)\Delta c$$

Sensitivity analysis

- We can compute the effect on demand and welfare of small changes without having to recompute equilibrium
- Next, we will consider estimation of PURC with micro-data

Individual-level estimation (microPURC)

- Aim to estimate a PURC model
  - many observations and large network
- Allow individual-specific information
- We seek to use all data without restriction.
  - No trimming and aggregation,
  - Large networks, no choice set

Work in progress: Fosgerau, Nielsen, Paulsen, Rasmussen, Yao (2025)

Basic idea - debiasing

- Recall the projected first-order condition for the traveller's problem

$$P(Z\beta^0 + \nabla F(x^*)) = 0$$

- We may plug an estimate $\hat{x}$ into $\nabla F(\cdot)$ and estimate β^0 by regression
  - However, this leads to bias in $\nabla F(x^*)$ since $\nabla F(\cdot)$ is nonlinear

- Can we correct the bias?

Debiasing

- Consider the function

$$T(\beta, x, Z, y) = P\left(Z\beta + \nabla F(x) + \nabla^2 F(x) \circ (y - x)\right)$$

- T is not sensitive to small deviations in x from the true flow

$$\mathbb{E}\frac{\partial T(\beta^0, x^*, Z, y)}{\partial x} = 0$$

- Moreover,

$$\mathbb{E}T(\beta^0, x^*, Z, y) = 0$$

We can estimate β by minimising $\mathbb{E}\|T(\beta, \hat{x}, Z, y\|^2$ where $\hat{x}$ is an estimate of optimal x^*

A debiasing algorithm

Repeat until convergence

- Compute prediction $\hat{x}$ given current estimate β
- Estimate new β by minimising $\mathbb{E}\|T(\beta, \hat{x}, Z, y\|^2$
  - This is just regression

On convergence, the procedure finds a local minimum of the likelihood function

- Works great on simulated data, application to real data is underway

Conclusion

Perturbed utility route choice has significant advantages

- Estimation with aggregate data is fast and easy
- No choice set
- Allows zero flows
- Correlation induced directly by network structure
- Fast equilibrium assignment
- Easy sensitivity analysis
- Estimation with microdata

Papers and presentations can be found on my homepage