Recent developments in route choice modeling

Emma Frejinger

Transport and Mobility Laboratory, EPFL, `transp-or.epfl.ch`

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

- Introduction
  - Problem description
  - Existing models
- Subnetwork approach
- Latent route choice model
- Future work

Route choice problem

Given a transportation network composed of nodes, links, origin and destinations.

For a given transportation mode and origin-destination pair, which is the chosen route?

Applications

- Intelligent transportation systems
- GPS navigation
- Transportation planning

Issues

- The choice set is unknown
- There are many (feasible) alternatives available
- The alternatives are often highly correlated due to overlapping paths
- Choice data is difficult to obtain

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Existing Approaches

- Assumption: Travelers use the shortest (with regard to any arbitrary generalized cost) route among all
  - Behaviorally unrealistic
- Random utility models (discrete choice models)

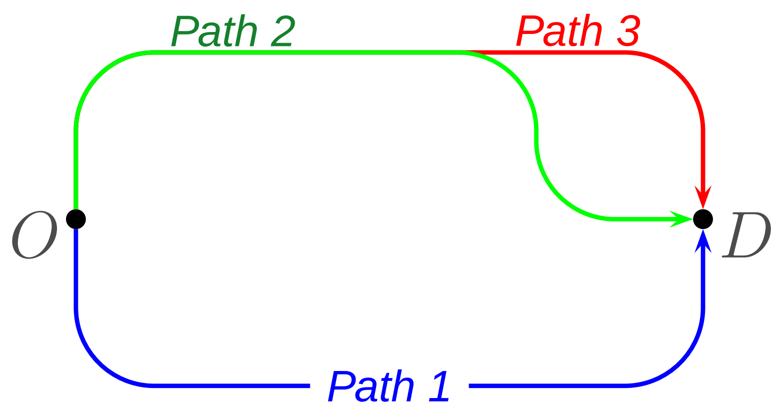
Existing Approaches - MNL

- Random terms are assumed to be i.i.d. Extreme Value

$$P(i|C_n) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}}$$

Alternatives are assumed to be independent. This assumption is (in general) not valid in a route choice context due to overlapping paths.

Existing Approaches



Travel time is the only considered attribute and

$V_1 = V_2 = V_3 = T$ then

$P(1|\{1,2,3\}) = P(2|\{1,2,3\}) = P(3|\{1,2,3\}) = \frac{1}{3}$

- Unrealistic path choice probabilities for correlated alternatives (overlapping paths)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Existing Approaches

- Few models explicitly capturing correlation have been used on large-scale route choice problems
  - C-Logit (Cascetta et al., 1996)
  - Path Size Logit (Ben-Akiva and Bierlaire, 1999)
  - Link-Nested Logit (Vovsha and Bekhor, 1998)
  - Logit Kernel model adapted to route choice situation (Bekhor et al., 2002)
- Probit model (Daganzo, 1977) permits an arbitrarily covariance structure specification but cannot be applied in a large-scale route choice context

Existing Approaches

- Link based path-multilevel logit model (Marzano and Papola, 2005)
  - Illustrated on simple examples and not estimated on real data

Subnetwork approach

Subnetworks

How can we explicitly capture the most important correlation structure without considerably increasing the model complexity?

Subnetworks

How can we explicitly capture the most important correlation structure without considerably increasing the model complexity?

- Which are the behaviorally important decisions?

Subnetworks

How can we explicitly capture the most important correlation structure without considerably increasing the model complexity?

- Which are the behaviorally important decisions?
- Our hypothesis: choice of specific parts of the network (e.g. main roads, city center)
- Concept: subnetwork

Subnetworks

- Subnetwork approach designed to be behaviorally realistic and convenient for the analyst
- Subnetwork component is a set of links corresponding to a part of the network which can be easily labeled
- Paths sharing a subnetwork component are assumed to be correlated even if they are not physically overlapping

Subnetworks - Methodology

- Factor analytic specification of an error component model (based on model presented in Bekhor et al., 2002)

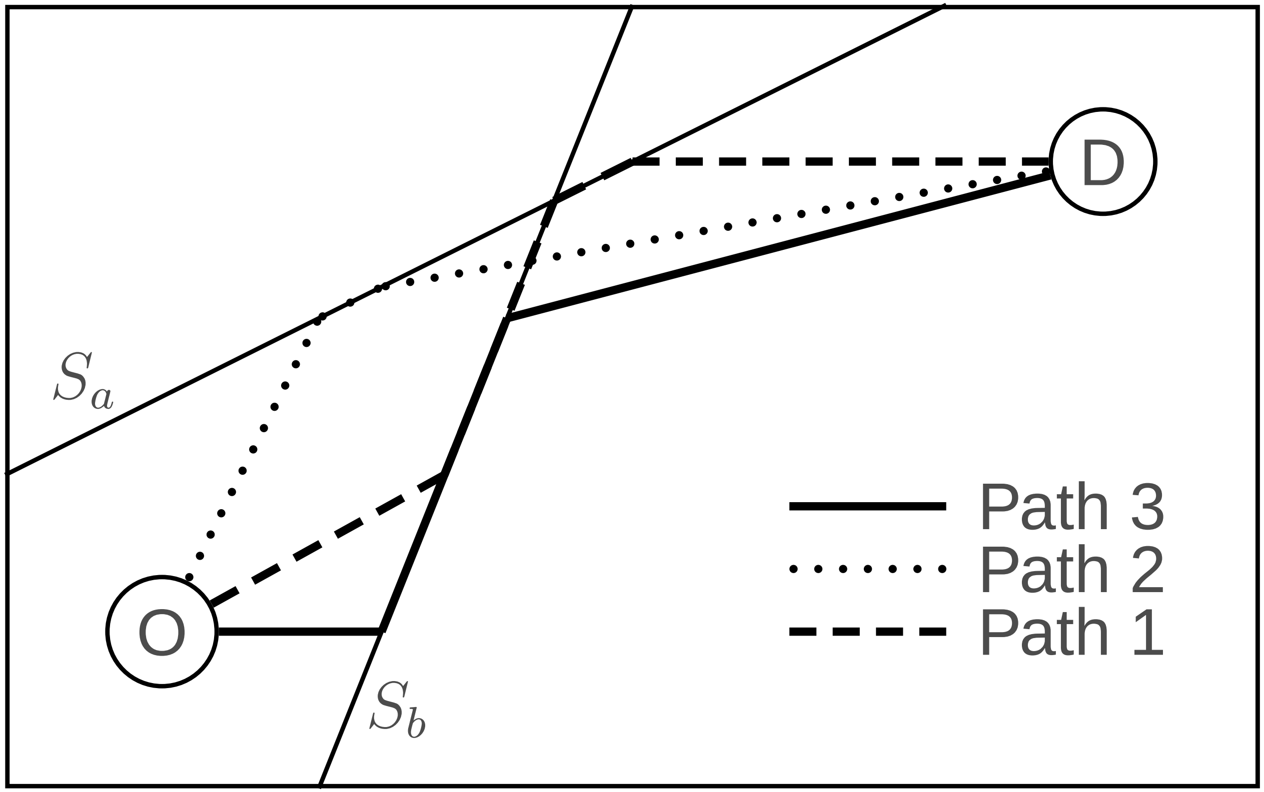
$$\mathbf{U}_n = \beta^T \mathbf{X}_n + \mathbf{F}_n \mathbf{T} \zeta_n + \nu_n$$

  - $\mathbf{F}_{n\ (J\text{x}Q)}$: factor loadings matrix
  - $(f_n)_{iq} = \sqrt{l_{niq}}$
  - $\mathbf{T}_{(Q\text{x}Q)} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_Q)$
  - $\zeta_{n\ (Q\text{x}1)}$: vector of i.i.d. N(0,1) variates
  - $\nu_{(J\text{x}1)}$: vector of i.i.d. Extreme Value distributed variates

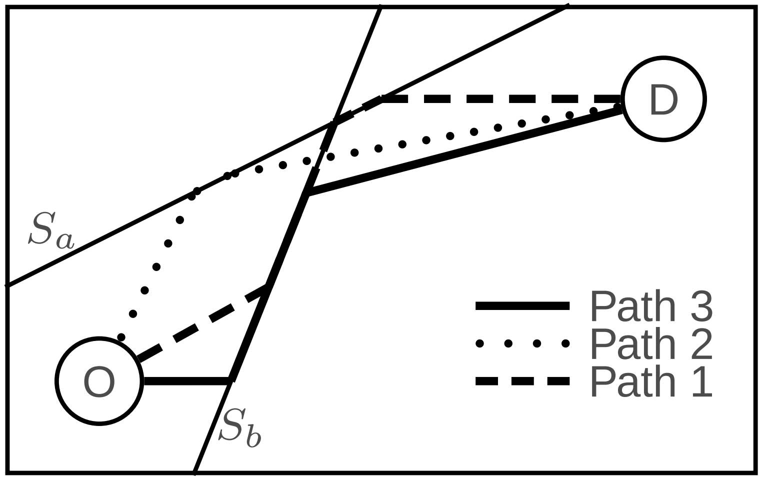
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Subnetworks - Example



Subnetworks - Example



$$U_1 = \beta^T X_1 + \sqrt{l_{1a}}\sigma_a\zeta_a + \sqrt{l_{1b}}\sigma_b\zeta_b + \nu_1$$

$$U_2 = \beta^T X_2 + \sqrt{l_{2a}}\sigma_a\zeta_a + \nu_2$$

$$U_3 = \beta^T X_3 + \sqrt{l_{3b}}\sigma_b\zeta_b + \nu_3$$

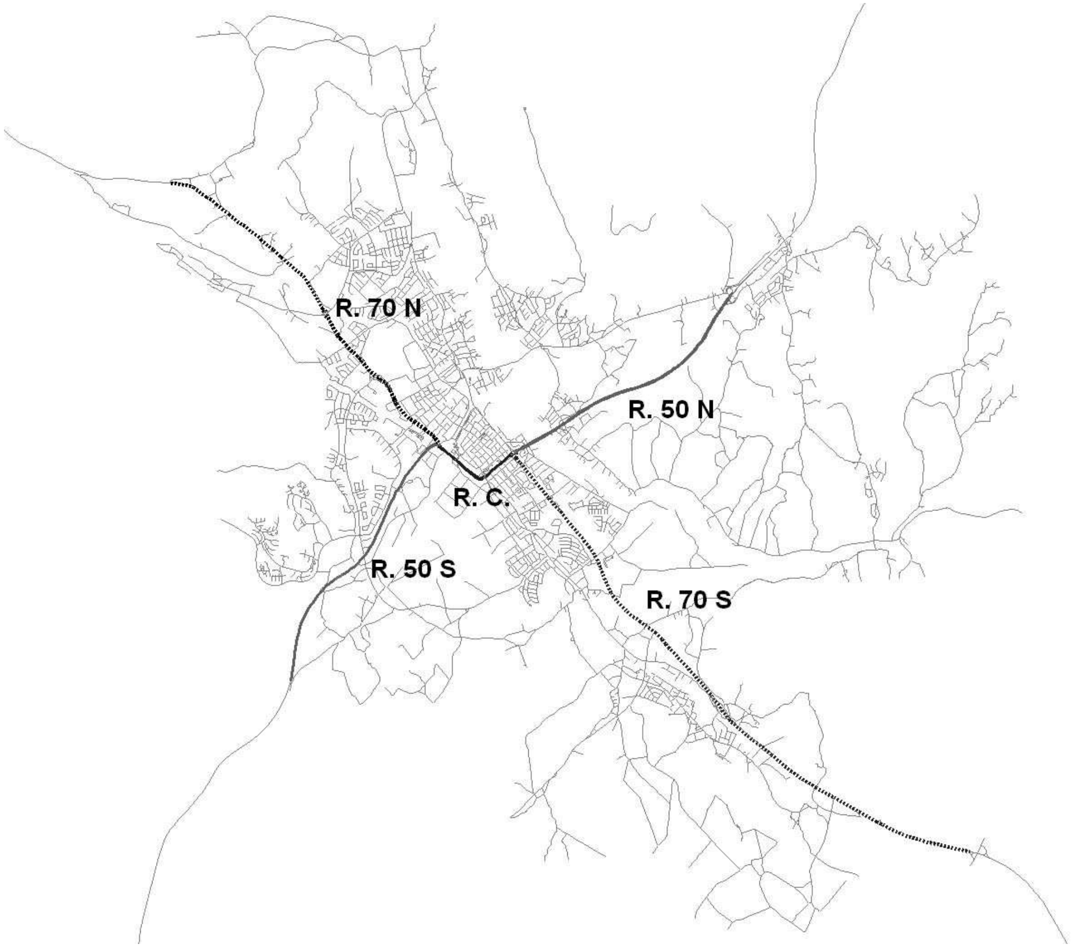
$$\mathbf{FTT}^T\mathbf{F}^T =$$

$$\begin{bmatrix} l_{1a}\sigma_a^2 + l_{1b}\sigma_b^2 & \sqrt{l_{1a}}\sqrt{l_{2a}}\sigma_a^2 & \sqrt{l_{1b}}\sqrt{l_{3b}}\sigma_b^2 \\ \sqrt{l_{1a}}\sqrt{l_{2a}}\sigma_a^2 & l_{2a}\sigma_a^2 & 0 \\ \sqrt{l_{3b}}\sqrt{l_{1b}}\sigma_b^2 & 0 & l_{3b}\sigma_b^2 \end{bmatrix}$$

Empirical Results

- The approach has been tested on three datasets: Boston (Ramming, 2001), Switzerland, and Borlänge
- Deterministic choice set generation
  Link elimination
- GPS data from 24 individuals
  2978 observations, 2179 origin-destination pairs
- Borlänge network
  3077 nodes and 7459 links
- BIOGEME (biogeme.epfl.ch, Bierlaire, 2003) has been used for all model estimations

Borlänge Road Network



Subnetwork Components

	R.50 S	R.50 N	R.70 S	R.70 N	R.C.
Component length [m]	5255	4966	11362	7028	1733
Nb. of Observations	173	153	261	366	209
Weighted Nb. of Observations (N_q) $N_q = \sum_{o \in O} \frac{l_{oq}}{L_q}$	36	88	65	73	116

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Model Specifications

- Six different models: MNL, PSL, EC_1, EC'_1, EC_2 and EC'_2
- EC_1 and EC'_1 have a simplified correlation structure
- EC'_1 and EC'_2 do not include a Path Size attribute
- Deterministic part of the utility

$$
\begin{aligned}
V_i = \; & \beta_{\text{PS}} \ln(\text{PS}_i) + \beta_{\text{EstimatedTime}} \text{EstimatedTime}_i + \\
& \beta_{\text{NbSpeedBumps}} \text{NbSpeedBumps}_i + \beta_{\text{NbLeftTurns}} \text{NbLeftTurns}_i + \\
& \beta_{\text{AvgLinkLength}} \text{AvgLinkLength}_i
\end{aligned}
$$

Estimation Results

- Parameter estimates for explanatory variables are stable across the different models
- Path size parameter estimates

Parameter	PSL	EC_1	EC_2
Path Size	-0.28	-0.49	-0.53
Scaled estimate	-0.33	-0.53	-0.56
Rob. T-test 0	-4.05	-5.61	-5.91

- All covariance parameters estimates in the different models are significant except the one associated with R.50 S

Estimation Results

Model	Nb. σ Estimates	Nb. Estimated Parameters	Final L-L	Adjusted Rho-Square
MNL	-	12	-4186.07	0.152
PSL	-	13	-4174.72	0.154
EC_1 (with PS)	1	14	-4142.40	0.161
EC'_1	1	13	-4165.59	0.156
EC_2 (with PS)	5	18	-4136.92	0.161
EC'_2	5	17	-4162.74	0.156

1000 pseudo-random draws for Maximum Simulated Likelihood estimation

2978 observations

Null log likelihood: -4951.11

BIOGEME (biogeme.epfl.ch) has been used for all model estimations.

Forecasting Results

- Comparison of the different models in terms of their performance of predicting choice probabilities
- Five subsamples of the dataset
  - Observations corresponding to 80% of the origin destination pairs (randomly chosen) are used for estimating the models
  - The models are applied on the observations corresponding to the other 20% of the origin destination pairs
- Comparison of final log-likelihood values

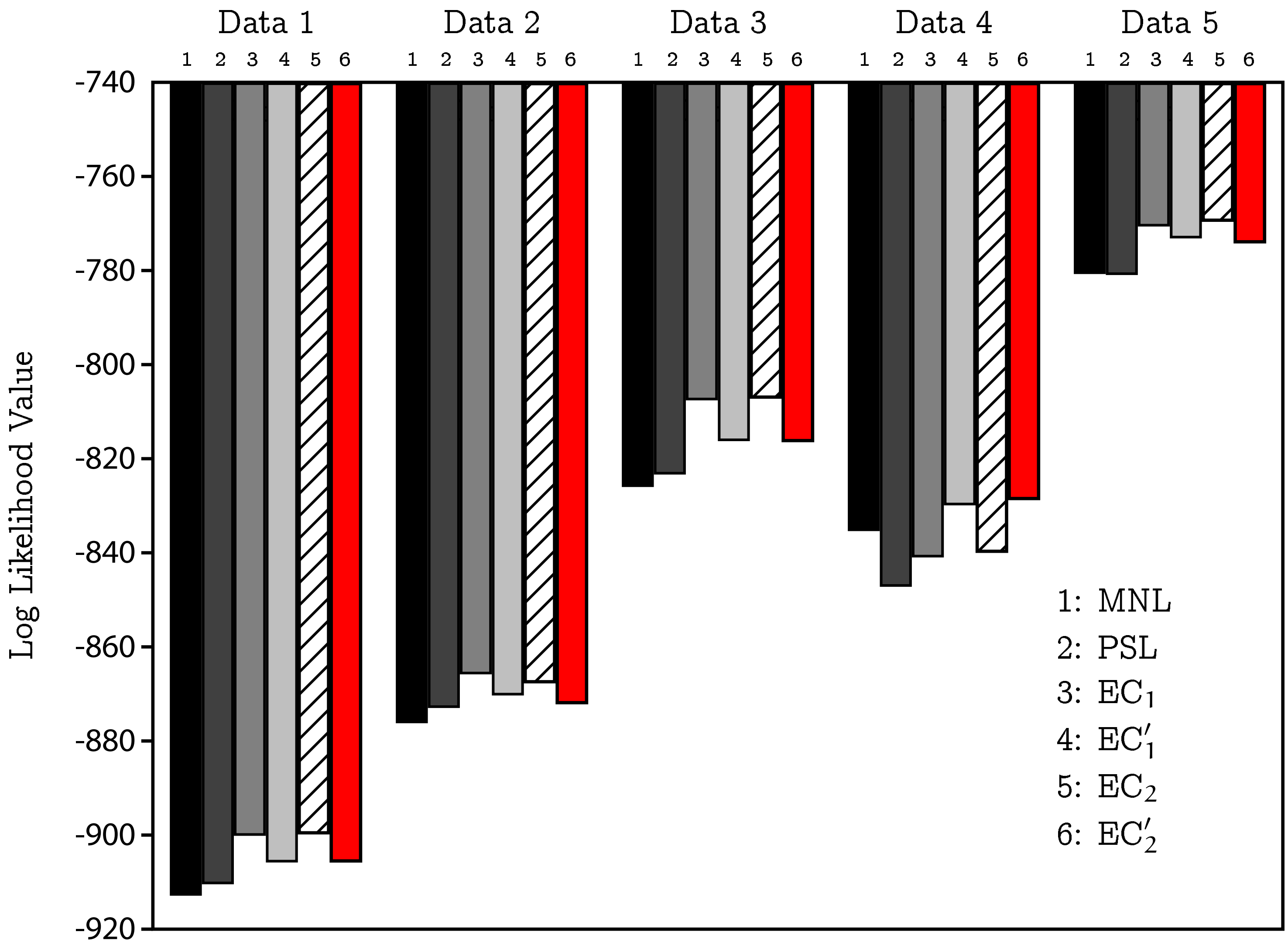
Forecasting Results

- Same specification of deterministic utility function for all models
- Same interpretation of these models as for those estimated on the complete dataset
- Coefficient and covariance parameter values are stable across models

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Forecasting Results



Conclusion - Subnetworks

- Models based on subnetworks are designed for route choice modeling of realistic size
- Correlation on subnetwork is explicitly captured within a factor analytic specification of an Error Component model
- Estimation and prediction results clearly shows the superiority of the Error Component models compared to PSL and MNL

Conclusion - Subnetworks

- The subnetwork approach is flexible and the model complexity can be controlled by the analyst
- Paper to appear in Transportation Research Part B

A latent route choice model

Mobility Pricing

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Swiss Mobility Pricing Project

- A part of a major study on various mobility pricing scenarios in Switzerland
- A collaboration with ETH Zurich and USI Lugano
- Revealed Preferences (RP) and Stated Preferences (SP) data has been collected
- RP data concern long distance route choice by car
  - Route descriptions are approximative
  - Route choices are latent

Objective

- Estimate route choice models based on latent chosen routes
- Literature on latent choice models
  - Ben-Akiva et al. (1984), label path approach
  - Ben-Akiva and Lerman (1985), destination choice
  - Toledo et al. (2003), Ben-Akiva et al. (2006) lane choice

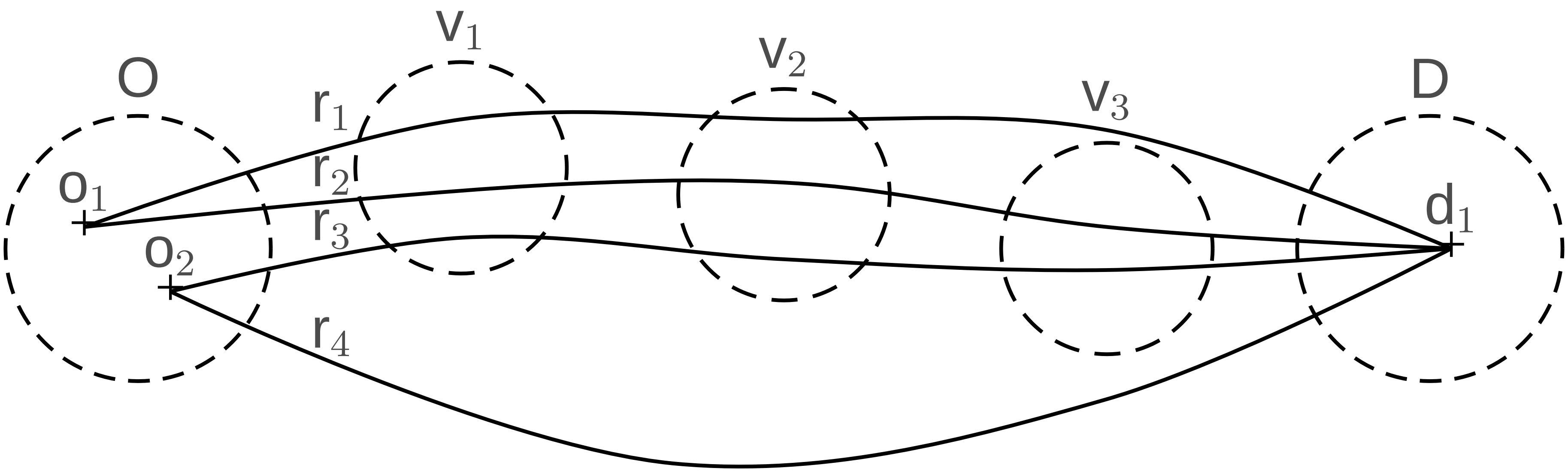
Observations

- Exact descriptions of chosen routes are difficult and expensive to obtain
- The concept of path and network as we need for modeling is abstract for respondents
- Here, a chosen route is described by a sequence of cities and locations
- *Aggregate observations* (several paths in the network can correspond to the same observation)

Observations

- Better quality of the observations
- Travelers do not need to refer to the network used by the analyst
- Exact origin-destination pairs are not necessarily known
- Exact route is not known

Observations - Example



Modeling Approach

- Several possible modeling approaches
  - Construction of paths from the aggregate observations
    - Involves subjective judgments and generate noise
  - Alternatives in the model are aggregates instead of physical paths
    - Estimated model is of little use in practice
- Our approach: compute the likelihood of an aggregate observation for a classical route choice model

Modeling Approach

- Probability of an aggregate observation i:

$$P(i) = \sum_{s \in S_i} P(s|S_i) \sum_{r \in C_s} \delta_{ri} P(r|C_s)$$

  - s: origin-destination pair
  - S_i: set of all origin-destination pairs for observation i
  - r: route
  - C_s: set of all routes for origin-destination pair s
  - $\delta_{ri} = \begin{cases} 1 & \text{if } r \text{ corresponds to } i \\ 0 & \text{otherwise} \end{cases}$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modeling Approach

- Probability of an aggregate observation i:

$$P(i) = \sum_{s \in S_i} P(s|S_i) \sum_{r \in C_s} \delta_{ri} P(r|C_s)$$

  - $P(s|S_i)$ can be modeled in several ways
  - $P(r|C_s)$: route choice model that is identifiable if
    1. at least one of the routes in C_s crosses the observed zones, and
    2. at least one route in C_s does not cross the observed zones.
- This type of models can be estimated with BIOGEME

Empirical Results

- Simplified Swiss network (39411 links and 14841 nodes)
- RP data collection through telephone interviews
- Long distance car travel
- The chosen routes are described with the origin and destination cities as well as 1 to 3 cities or locations that the route pass by
- 940 observations available after data cleaning and verification

Empirical Results

Empirical Results

- This application is one of few presented in the literature that are based on RP data
- The network is to our knowledge the largest one used for evaluation of route choice modeling approaches

Empirical Results

- No information available on the exact origin destination pairs

$$P(s|i) = \frac{1}{|S_i|} \; \forall s \in S_i$$

- Two origin-destination pairs are randomly chosen for each observation

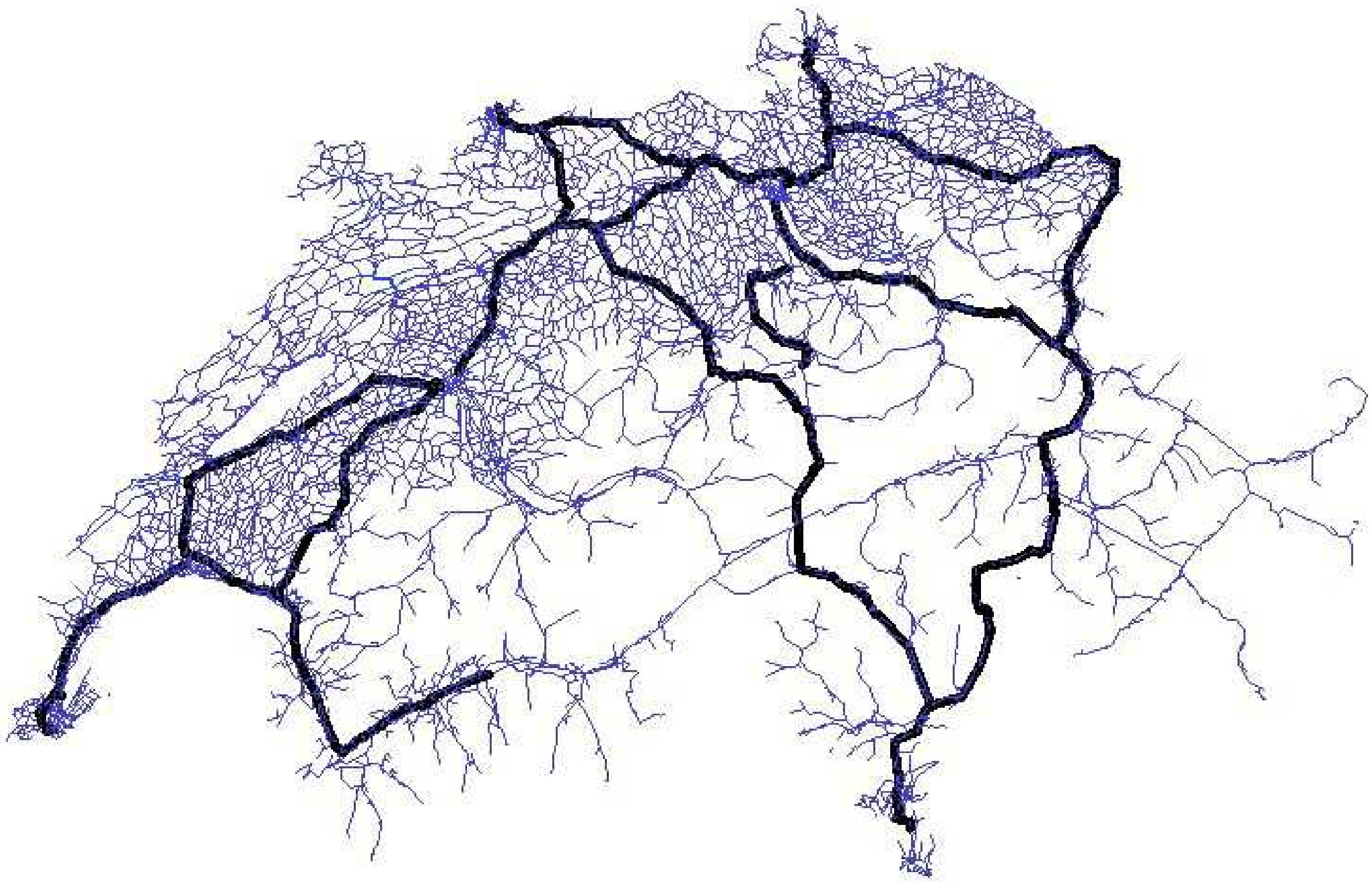
Empirical Results

- 46 routes per choice set are generated with a choice set generation algorithm
- After choice set generation 780 observations are available
  - 160 observations were removed because either all or none of the generated routes crossed the observed zones
- We estimate Path Size Logit (Ben-Akiva and Bierlaire, 1999) and Subnetwork (Frejinger and Bierlaire, 2006) models

Empirical Results - Subnetwork

- Subnetwork: main motorways in Switzerland
- Correlation among routes is explicitly modeled on the subnetwork
- Combined with a Path Size attribute
- Linear-in-parameters utility specifications

Empirical Results - Subnetwork



Parameter	PSL		Subnetwork	
ln(path size) based on free-flow time	**1.04**	(0.134) 7.81	**1.10**	(0.141) 7.78
Scaled Estimate	*1.04*		*1.04*	
Freeway free-flow time 0-30 min	**-7.12**	(0.877) -8.12	**-7.45**	(0.984) -7.57
Scaled Estimate	*-7.12*		*-7.04*	
Freeway free-flow time 30min - 1 hour	**-1.69**	(0.875) -1.93	**-2.26**	(1.03) -2.19
Scaled Estimate	*-1.69*		*-2.14*	
Freeway free-flow time 1 hour +	**-4.98**	(0.772) -6.45	**-5.64**	(1.00) -5.61
Scaled Estimate	*-4.98*		*-5.33*	
CN free-flow time 0-30 min	**-6.03**	(0.882) -6.84	**-6.25**	(0.975) -6.41
Scaled Estimate	*-6.03*		*-5.91*	
CN free-flow time 30 min +	**-1.87**	(0.331) -5.64	**-2.16**	(0.384) -5.63
Scaled Estimate	*-1.87*		*-2.04*	
Main free-flow travel time 10 min +	**-2.03**	(0.502) -4.05	**-2.46**	(0.624) -3.95
Scaled Estimate	*-2.03*		*-2.33*	
Small free-flow travel time	**-2.16**	(0.685) -3.16	**-2.75**	(0.804) -3.42
Scaled Estimate	*-2.16*		*-2.60*	
Proportion of time on freeways	**-2.2**	(0.812) -2.71	**-2.31**	(0.865) -2.67
Scaled Estimate	*-2.2*		*-2.18*	
Proportion of time on CN	**0 fixed**		**0 fixed**	
Proportion of time on main	**-4.43**	(0.752) -5.88	**-4.40**	(0.800) -5.51
Scaled Estimate	*-4.43*		*-4.16*	
Proportion of time on small	**-6.23**	(0.992) -6.28	**-6.02**	(1.03) -5.83
Scaled Estimate	*-6.23*		*-5.69*	
Covariance parameter			**0.217**	(0.0543) 4.00
Scaled Estimate			*0.205*	

Empirical Results

	PSL	Subnetwork
Covariance parameter		0.217
(Rob. Std. Error) Rob. T-test		(0.0543) 4.00
Number of simulation draws	-	1000
Number of parameters	11	12
Final log-likelihood	-1164.850	-1161.472
Adjusted rho square	0.145	0.147
Sample size: 780, Null log-likelihood: -1375.851		

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Empirical Results

- All parameters have their expected signs and are significantly different from zero
- The values and significance level are stable across the two models
- The subnetwork model is significantly better than the Path Size Logit (PSL) model

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusion - Latent route choice

- Aggregate observations are convenient to report paths
- They can be used for estimating route choice models
- Care must be taken about the level of aggregation
- Parameters of the RP model are significant and meaningful
- Available in Biogeme / Bioroute

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Future work

- Choice set generation
  - Stochastic path generation algorithm
- Analysis of sensitivity of the modeling results regarding the choice set definition