Capturing Correlation in Route Choice Models using Subnetworks

Emma Frejinger and Michel Bierlaire

Transport and Mobility Laboratory (TRANSP-OR), EPFL

Outline

- Issues of route choice analysis
- Modelling correlation with subnetworks
    - Methodology
    - Example
- Empirical results
    - Borlänge GPS dataset
    - Estimation results
    - Forecasting results
- Conclusion and 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?

- Issues:
  - Universal choice set very large
  - Correlated alternatives due to overlapping paths
  - Data collection issues

Route Choice Modelling

- Deterministic utility maximisation e.g. shortest path assumption is behaviourally unrealistic
- Random utility models
  Utility U_{in} an individual n associates with alternative i:

$$U_{in} = V_{in} + \varepsilon_{in}$$

where $V_{in} = \beta^T X_{in}$ is the deterministic part and ε_{in} is the random term

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Route Choice Models

- Few models explicitly capturing correlation have been used on route choice problems of real size
  - 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 can rarely be applied in a real size route choice context

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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 behaviourally important decisions?

Subnetworks

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Subnetworks

- Subnetwork approach designed to be behaviourally realistic and convenient for the analyst
- Subnetwork component is a set of links corresponding to a part of the network which can be easily labelled
- 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)

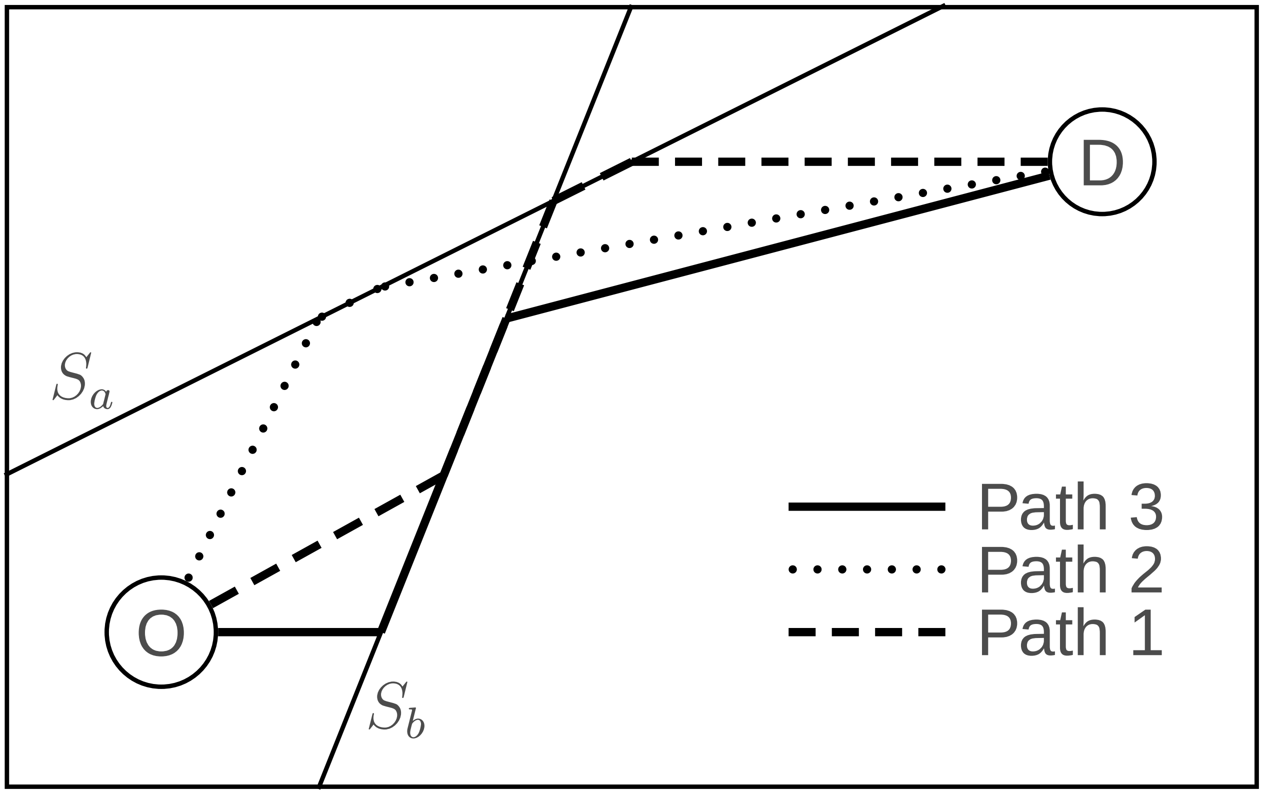
$$\mathbf{U}_n = \beta^T \mathbf{X}_n + \mathbf{F}_n \mathbf{T} \zeta_n + \nu_n$$

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

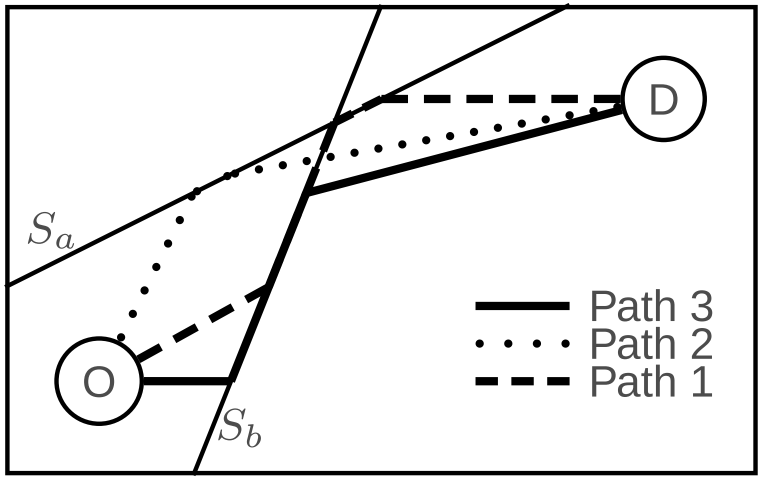
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Subnetworks - Example



Subnetworks - Example



$$U_1 = \beta^T X_1 + \sqrt{l_{1a}}\sigma_a\zeta_a + \sqrt{l_{1b}}\sigma_b\zeta_b + \nu_1$$
$$U_2 = \beta^T X_2 + \sqrt{l_{2a}}\sigma_a\zeta_a + \nu_2$$
$$U_3 = \beta^T X_3 + \sqrt{l_{3b}}\sigma_b\zeta_b + \nu_3$$

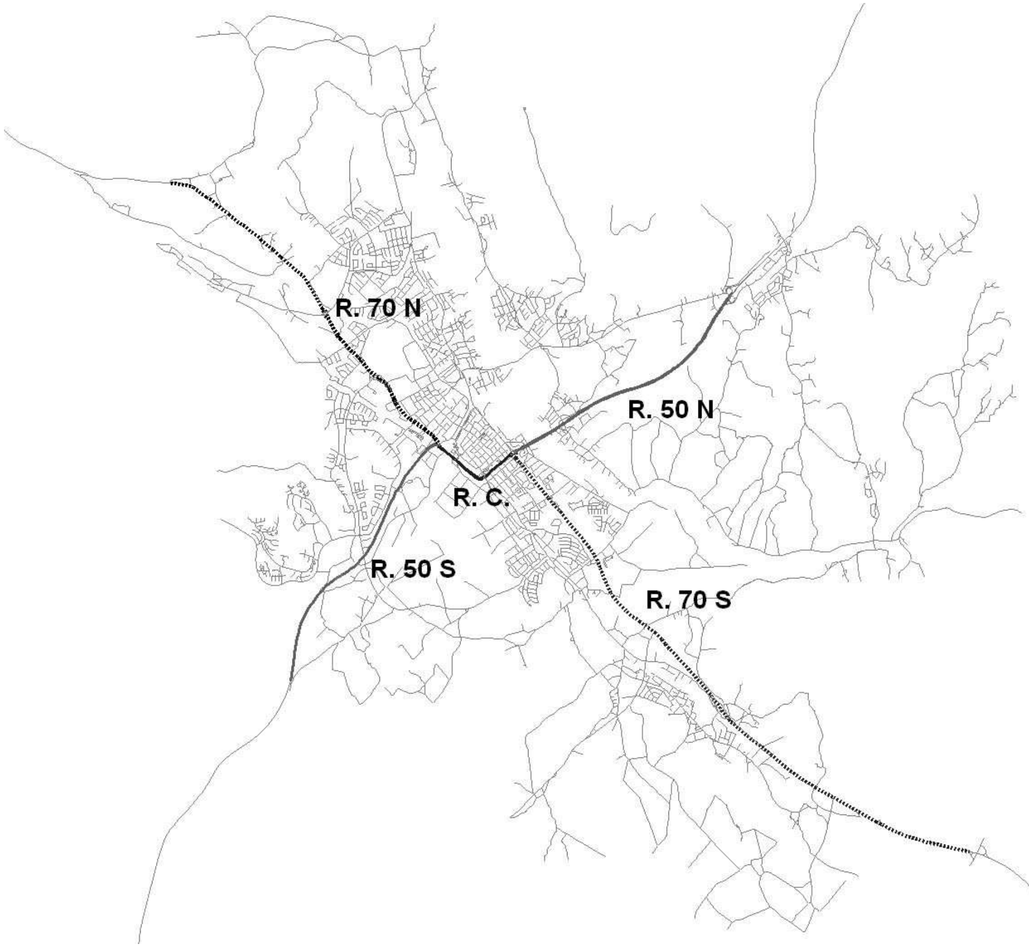
$$\mathbf{FTT}^T\mathbf{F}^T =$$

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

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

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

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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.				

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

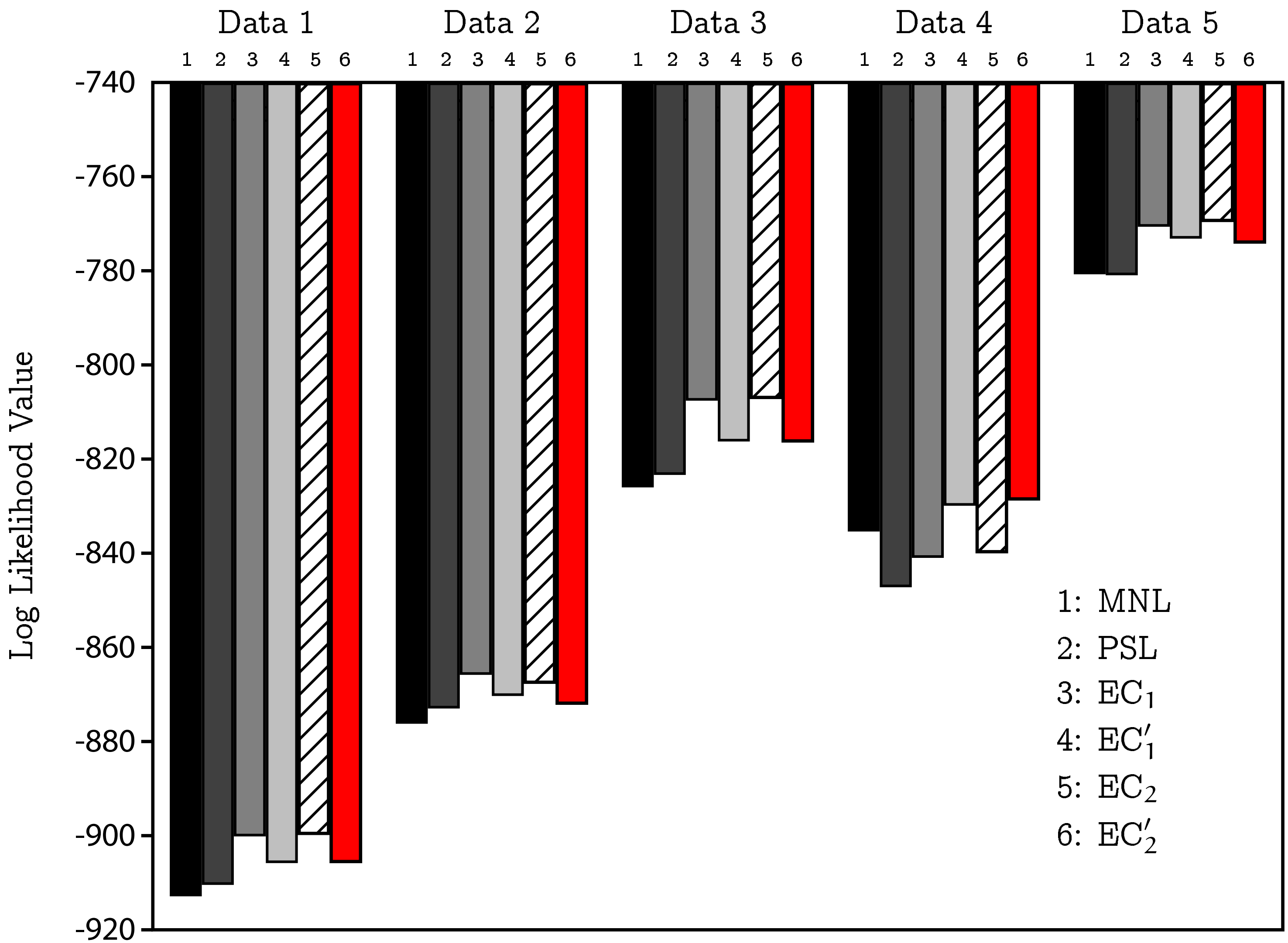
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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

Forecasting Results



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusion

- Models based on subnetworks are designed for route choice modelling 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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Conclusion

- The subnetwork approach is flexible and the trade-off between complexity and behavioural realism can be controlled by the analyst
- Paper to appear in Transportation Research Part B
- Future work
  - Analysis of the sensitivity of the results regarding the definition of the subnetwork
  - Influence of choice set generation algorithm