Reconstructing daily schedules of individuals: a utility maximization approach

Janody Pougala Tim Hillel Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

May 5, 2021

Outline

Introduction

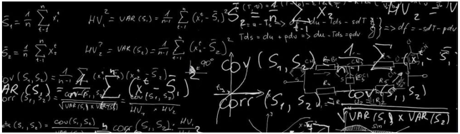
- Travel demand is derived from activity demand.
- Activity demand is influenced by socio-economic characteristics, social interactions, cultural norms, basic needs, etc. [Chapin, 1974]
- Activity demand is constrained in space and time [Hägerstraand, 1970].

TRANSP-OR

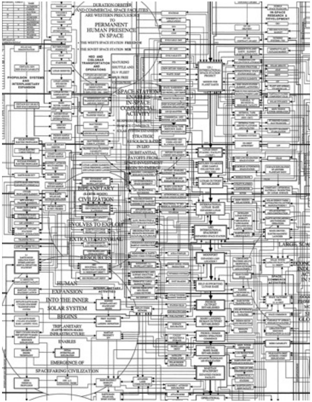
EPFL

Literature

Econometric models



Rule-based models



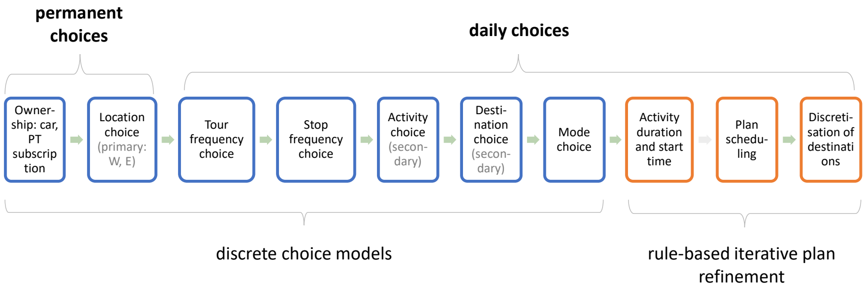
State of the art: econometric approach

[Bhat, 2005]

- Multiple Discrete Continuous Extreme Value
- Based on first principles.
- Decision-maker solves an optimization problem, with a time budget.
- Several alternatives may be chosen.
- Model derived from KKT conditions.

State of practice

Sequence of decisions Source: [Scherr et al., 2020]



Research question

Relax the *series of discrete choice models* approach

- The interactions of all decisions is complex.
- Sequence of models is most of the time arbitrary.

Integrated approach

Develop a model involving many decisions:

- activity participation,
- activity location,
- activity duration,
- activity scheduling,
- travel mode,
- travel path.

TRANSP-OR

EPFL

Research objectives

- Integrated approach based on first principles.
- Theoretical framework: utility maximization.
- Individuals solve a scheduling problem.
- Important aspects: trade-offs on activity sequence, duration and starting time.

Outline

First principles

- Each individual n has a time-budget (a day).
- Each activity a considered by n is associated with a utility U_{an}.
- Individuals schedule their activities as to **maximize** the total utility, subject to their time-budget constraint.

Further assumptions

Individuals are time sensitive

- Have a desired *start time*, *duration* and/or *end time* for each activity
- Deviations from their desired times in the scheduling process decrease the utility

TRANSP-OR

EPFL

Time

- Time horizon: 24 hours.
- Discretization: T time intervals.
- Trade-off between model accuracy and computational time.

Space

- Discrete and finite set S of locations, indexed by s.
- For each individual, each activity is associated with a list of potential locations.

Travel

- For each pair OD, list of possible modes.
- For each mode, list of possible paths.
- For each (O, D, m, p), $\rho(O, D, m, p)$ is the travel time.
- Exogenously given.

TRANSP-OR

EPFL

Activities

Definition: Activity

An activity is associated with a location and a trip.

Activities

Location, mode and route choices

- Lunch at location A, followed by trip by bus on path 1.
- Lunch at location A, followed by trip by bus on path 2.
- Lunch at location A, followed by trip by car on path 1.
- Lunch at location B, followed by trip by car on path 2.

Constraint

Only one of the “duplicates” can be chosen.

Activities

Given

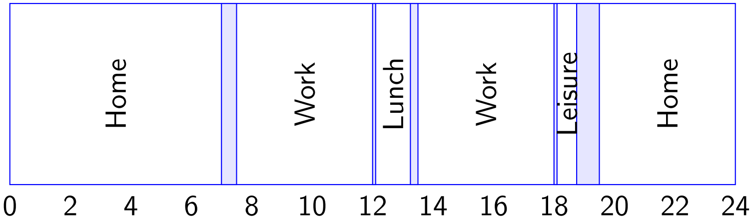
- Set A of activities.
- Location s_a.
- Feasible time interval: $[\gamma_a^-, \gamma_a^+]$ (e.g. opening hours).

Decisions

- Participation: $w_a \in \{0, 1\}$.
- Starting time x_a, $0 \leq x_a \leq T$.
- Schedule: $z_{ab} \in \{0, 1\}$.
- Duration: $0 \leq \tau_a \leq T$.

TRANSP-OR EPFL

Scheduling



Preferences

Preferences

- desired starting time x_a^*,
- desired duration τ_a^*.

Penalties

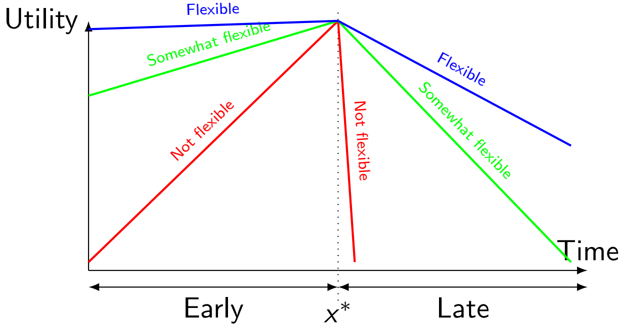
- Starting early [Small, 1982]: $\theta_e \max(x_a^* - x_a, 0)$.
- Starting late [Small, 1982]: $\theta_\ell \max(x_a - x_a^*, 0)$.
- Shorter activity: $\theta_{ds} \max(\tau_a^* - \tau_a, 0)$.
- Longer activity: $\theta_{d\ell} \max(\tau_a - \tau_a^*, 0)$.

TRANSP-OR

EPFL

Preferences

Parameters depend on activity



Disutility of travel

Each activity is followed by a trip

- Travel time from a to a^+: t_a.
- Depends on the next activity.

$$t_a = \sum_b z_{ab} \rho(s_a, s_b, m_a, p_a).$$

- Other variables can be included (cost, etc.)
- Note: If $s_a = s_b$, $\rho(s_a, s_a, m_a, p_a) = 0$
- Exception: last activity of the day (home).

Utility function

An individual n derives the following utility from performing activity a, with a schedule flexibility k:

$$\begin{aligned} U_{an} = c_{an} & \\ & + \theta_e^k \max(x_a^* - x_a, 0) \\ & + \theta_\ell^k \max(x_a - x_a^*, 0) \\ & + \theta_{ds}^k \max(\tau_a^* - \tau_a, 0) \\ & + \theta_{d\ell}^k \max(\tau_a - \tau_a^*, 0) \\ & + \varepsilon_{an}, \end{aligned}$$

where ε_{an} are error terms.

Utility function

Utility of a schedule

$$U_{sn} = \sum_a w_a U_{an} + \theta_t \sum_a \sum_b z_{ab} \rho(s_a, s_b, m_a, p_a)$$

Error terms

$$\sum_a w_a \varepsilon_{an}$$

where ε_{an} normally distributed.

TRANSP-OR

EPFL

Utility function

Error terms

- Rely on simulation.
- For each activity a, individual n,
- draw ε_{anr}, $r = 1, \ldots, R$.
- Optimization problem for each r.
- Utility: U_{anr}.

Outline

Decision variables for individual n and draw r

For each (potential) activity a:

- Activity participation: $w_{anr} \in \{0, 1\}$.
- Starting time: $x_{anr} \in \{0, \dots, T\}$.
- Duration: $\tau_{anr} \in \{0, \dots, T\}$.
- Scheduling: $z_{abnr} \in \{0, 1\}$: 1 if activity b immediately follows a.

Objective function

Additive utility

$$\max \sum_{a \in A} w_{anr} U_{anr} + \theta_t \sum_{a \in A} \sum_{b \in A} z_{abnr} \rho(s_a, s_b, m_a, p_a).$$

Constraints

Time budget

$$\sum_{a \in A} w_{anr} \tau_{anr} + \sum_{a \in A} \sum_{b \in A} z_{abnr} \rho(s_a, s_b, m_a, p_a) = T, \; \forall n, r.$$

Time windows

$$0 \leq \gamma_a^- \leq x_{anr} \leq x_{anr} + \tau_{anr} \leq \gamma_a^+ \leq T, \; \forall a, n, r.$$

Constraints

Precedence constraints

$$z_{abnr} + z_{banr} \leq 1, \ \forall a, b, n, r.$$

Single successor/predecessor

$$\sum_{b \in A \setminus \{a\}} z_{abnr} = w_{anr}, \ \forall a, n, r,$$

$$\sum_{b \in A \setminus \{a\}} z_{banr} = w_{anr}, \ \forall a, n, r.$$

Constraints

Consistent timing

$$(z_{abnr} - 1)T \leq x_{anr} + \tau_{anr} + t_{anr} - x_{bnr} \leq (1 - z_{abnr})T, \ \forall a, b, n, r.$$

where

$$t_{anr} = \sum_{b \in A} z_{abnr} \rho(s_a, s_b, m_a, p_a).$$

Mutually exclusive duplicates

$$\sum_{a \in B_k} w_{anr} = 1, \ \forall k, n, r.$$

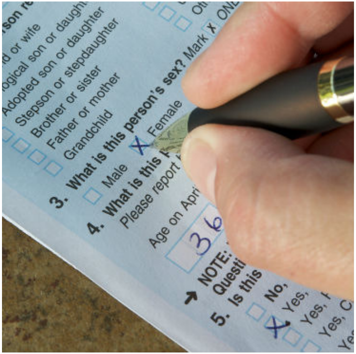
Optimization problem

Simulation-based optimization

- For each realization of the error terms, we have an optimal schedule.
- It includes all the choice dimensions (activity participation, location, duration, scheduling, and mode and route).
- We can generate an empirical distribution of chosen schedules.

Outline

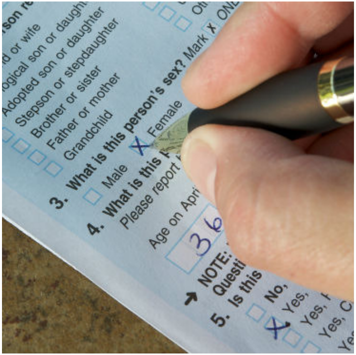
Real data



Dataset

- 2015 Swiss Mobility and Transport Microcensus.
- Daily trip diaries for 57'000 individuals.
- Records of activities, visited location, mode/path choice.

Real data

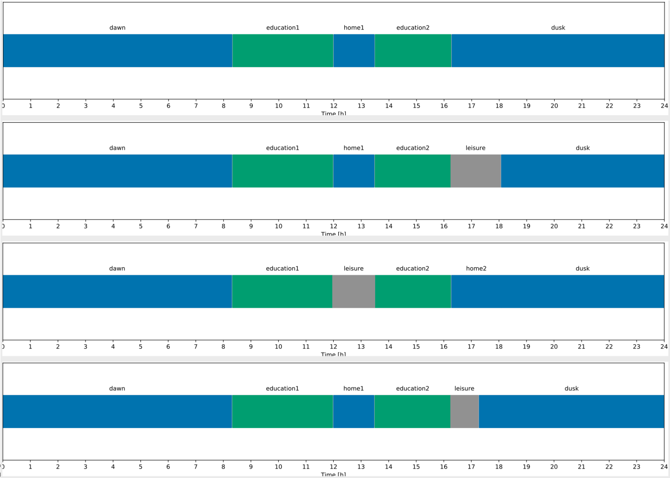


Assumptions

- Desired start times and durations are the recorded ones.
- Feasible time windows: percentiles start and end times from out of sample distribution.
- Only the recorded locations are considered.
- Uniform flexibility profile across population.

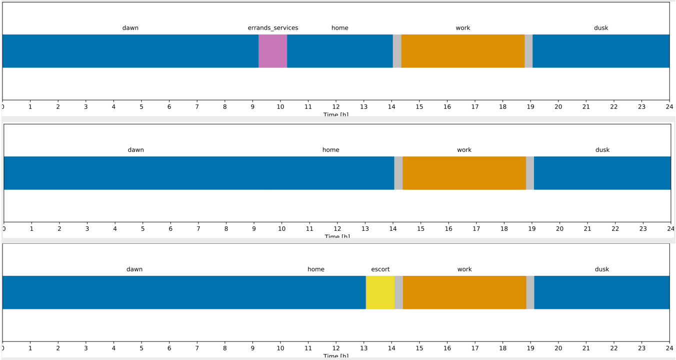
Individual 1 (weekday)

Optimal schedules generated for random draws of ε_{a_n}



Individual 2 (weekday)

Optimal schedules generated for random draws of ε_{a_n}

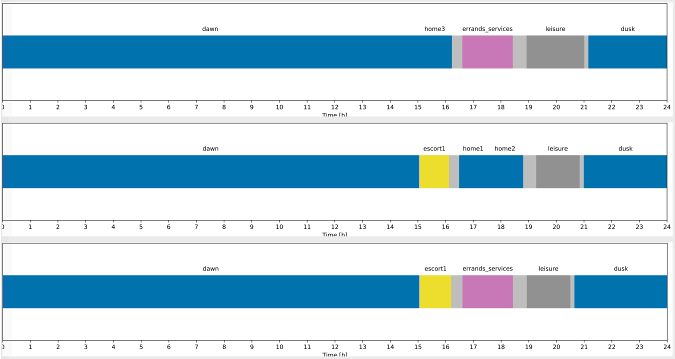


TRANSP-OR

EPFL

Individual 3 (weekday)

Optimal schedules generated for random draws of ε_{a_n}



Outline

Parameter estimation

Simulation

- Given the parameters,
- generate optimal schedules.

Parameter estimation

- Given observed schedules,
- estimate the parameters.

Parameters to be estimated

For each activity type k:

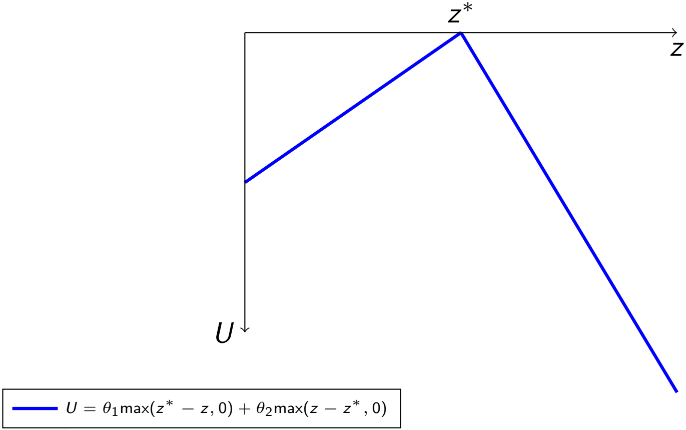
- Activity specific constant $c_k \in \mathbb{R}$.
- Desired start time $x_k^* \in [0, 24]$.
- Desired duration $\tau_k^* \in [0, 24]$.
- Early penalty $\theta_e^k \in [-\infty, 0]$.
- Late penalty $\theta_\ell^k \in [-\infty, 0]$.
- Short penalty $\theta_{ds}^k \in [-\infty, 0]$.
- Long penalty $\theta_{d\ell}^k \in [-\infty, 0]$.

Parameter estimation

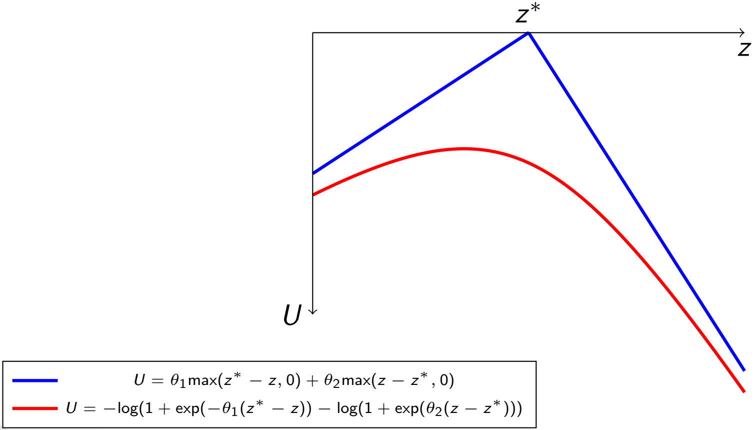
Difficulties

- Non differentiability.
- Choice set cannot be enumerated.

Non differentiability



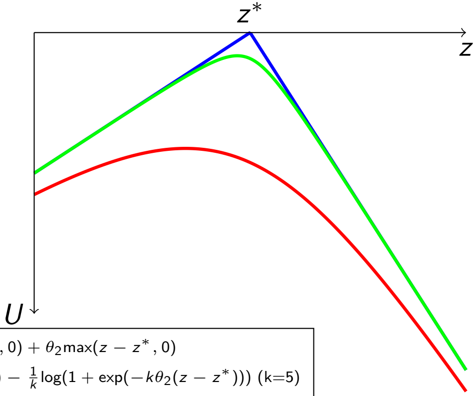
Non differentiability



TRANSP-OR

EPFL

Non differentiability



$U = \theta_1 \max(z^* - z, 0) + \theta_2 \max(z - z^*, 0)$

$U = -\frac{1}{k}\log(1 + \exp(-k\theta_1(z^* - z)) - \frac{1}{k}\log(1 + \exp(-k\theta_2(z - z^*)))$ (k=5)

$U = -\frac{1}{k}\log(1 + \exp(-k\theta_1(z^* - z)) - \frac{1}{k}\log(1 + \exp(-k\theta_2(z - z^*)))$ (k=1)

Parameter estimation

Choice set generation

- Full set of schedules C_n is combinatorial.
- Must rely on a sample of alternatives $\tilde{C}_n$.

Choice model estimation

- Include an EV error term to obtain a mixture of logit.
- Probability of choosing a schedule y for individual n is conditional on the parameters β_n, the variables x_n and the sampled choice set $\tilde{C}_n$ [Guevara and Ben-Akiva, 2013]
- Maximum likelihood estimators of the parameters:

$$\max_{\widehat{\beta}} L(y|\widehat{\beta}, X) = \prod_n P(y|x_n, \widehat{\beta}_n, \tilde{C}_n)$$

Choice set generation

Sequential approach

- Draw the number of out-of-home activities.
- Draw each activity independently.
- Draw the starting times.
- Sort activities.
- Derive the duration.

Main issue: cannot correct for sampling bias.

Integrated approach

- Draw complete and valid schedules.
- Metropolis-Hastings algorithm [Flötteröd and Bierlaire, 2013]

Some estimation results

Sample size: 1045

	leisure	work	education
constant	2.84	3.92	2.74
desired_start_time	8.91	7.12	7.57
desired_duration	1.01	10.	5.68
long	-0.162	-0.695	-0.227
short	-1.33	-0.495	-0.913
late	-0.161	-0.478	-0.725
early	-1.22	-1.2	-2.44

Some estimation results

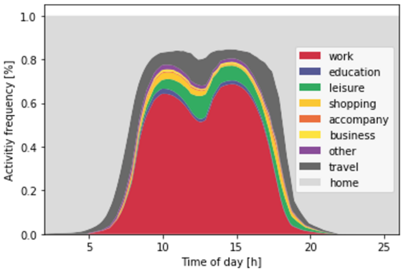
Sample size: 1045

	shopping	errands services	business trip	escort
constant	2.94	0.918	-0.127	1.39
desired_start_time	9.04	16.9	8.19	17.7
desired_duration	0.107	0.	0.245	0.
long	-1.	-0.779	0.	-0.976
short	-47.2	-0.00122	-93.	-0.00239
late	-0.21	-0.898	-0.276	-0.644
early	-1.52	0.	-1.58	-0.0328

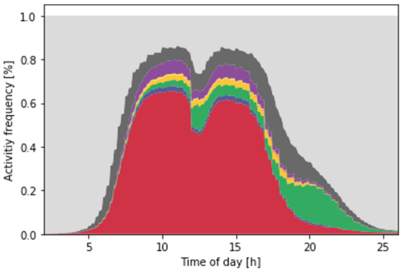
Validation

Activity profiles for full-time workers, Lausanne area

Simulation



Microcensus

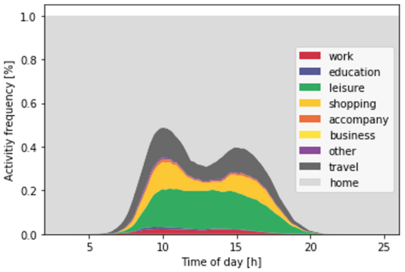


Source: SBB. Acknowledgement to Patrick Manser.

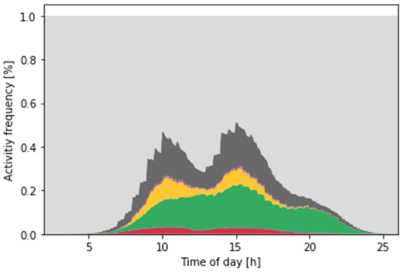
Validation

Activity profiles for individuals older than 65, Lausanne area

Simulation



Microcensus

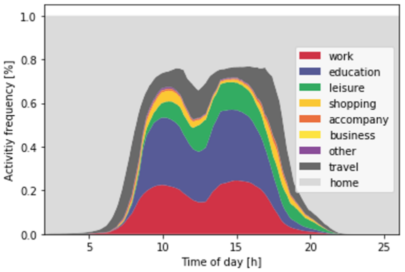


Source: SBB. Acknowledgement to Patrick Manser.

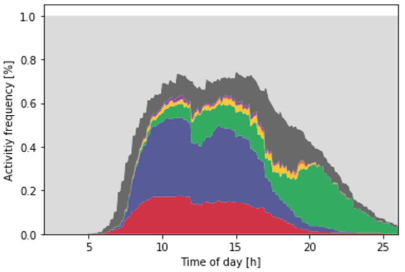
Validation

Activity profiles for students, Lausanne area

Validation



Microcensus

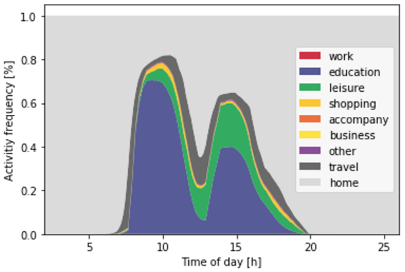


Source: SBB. Acknowledgement to Patrick Manser.

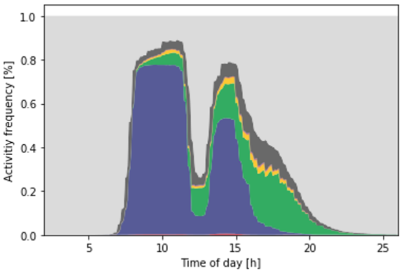
Validation

Activity profiles for primary school pupils, Lausanne area

Validation



Microcensus



Source: SBB. Acknowledgement to Patrick Manser.

Conclusions

Achievements so far

- Formulation of the model.
- Applied on real data.
- We are able to draw from a distribution of activity schedules.
- Preliminary estimation of the parameters.
- The results make sense.

Ongoing work

- Choice set generation using Metropolis-Hastings.

Bibliography I

Bhat, C. R. (2005).
A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions.
Transportation Research Part B: Methodological, 39(8):679 – 707.

Chapin, F. S. (1974).
Human activity patterns in the city: Things people do in time and in space, volume 13.
Wiley-Interscience.

Flötteröd, G. and Bierlaire, M. (2013).
Metropolis–hastings sampling of paths.
Transportation Research Part B: Methodological, 48:53–66.

Bibliography II

Guevara, C. and Ben-Akiva, M. (2013).
Sampling of alternatives in logit mixture models.
Transportation Research Part B: Methodological, 58:185 – 198.

Hägerstraand, T. (1970).
What about people in regional science?
Papers in Regional Science.

Scherr, W., Manser, P., and Bützberger, P. (2020).
Simba mobi: Microscopic mobility simulation for corporate planning.
Transportation Research Procedia, 49:30–43.

Small, K. A. (1982).
The scheduling of consumer activities: work trips.
American Economic Review.