Mode and Departure-Time Choice Estimates

Lucas Javaudin[1], André de Palma[1], Nathalie Picard[2]

[1]THEMA, CY Cergy Paris Université
[2]BETA, Strasbourg University

October 10, 2023
THEMA Applied Seminar

Motivations

What are the determinants of the mode of transportation and departure time chosen by the individuals for their trips?

Example of questions we want to answer:

- What is the minimum incentive amount that I must give you to convince you to switch from car to public transit?
- If road congestion decreases between 7AM and 8AM, will you leave at 07:50 instead of 08:00?
- If a road opening is decreasing your car travel time by 10 minutes, will you switch from public transit to car?

Application: evaluation of transport policies (low-emission zone, Grand Paris Express, public-transit subsidies, etc.)

Desired Arrival Time

- **Desired arrival time (t^*):** time at which the individual would choose to arrive if travel time was null.
- Desired arrival time can be different from *actual arrival time* (trade-off between travel time and schedule delay).
- Desired arrival times are *unknown* and highly *heterogeneous*.

Utility

In the standard **α-β-γ model** (Vickrey, 1969; Arnott, de Palma, Lindsey, 1990s), utility for mode i and departure time t_d is:

$$u(i, t_d) = -\underbrace{c_i}_{\text{Constant}} - \underbrace{\alpha_i \cdot tt_i(t_d)}_{\text{Travel cost}} - \underbrace{\beta \cdot [t^* - t_d - tt_i(t_d)]_+}_{\text{Early penalty}} - \underbrace{\gamma \cdot [t_d + tt_i(t_d) - t^*]_+}_{\text{Late penalty}}$$

- c_i: mode-specific constant cost
- α_i: value of time for mode i
- $tt_i(t_d)$: travel time when leaving at time t_d with mode i
- β: penalty for arriving early
- γ: penalty for arriving late
- t^*: desired arrival time
- $[x]_+ = \max(x, 0)$

Part I: Desired arrival-time distribution

In α-β-γ the model, any individual with a constant travel time will choose to arrive exactly at his / her desired arrival time:

- If my travel time is constant, I choose a departure time such that I arrive at my desired arrival time.
- If my travel time is time-dependent, I might choose to arrive early / late to reduce travel time.

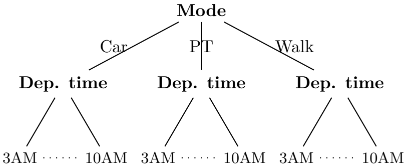
We estimate the distribution of t^* in the population from the distribution of arrival times in the subpopulation of individuals with a constant travel time (all walk trips and car trips without congestion).

Part II: Bayesian estimates

After estimating the distribution of t^*, we are able to estimate the parameters of interest (value of time, early and late penalties, etc.) using **Bayesian estimates** and a **travel survey**.

We estimate a “Mixed discrete-continuous Nested Logit model”:

- Stage 1: mode choice (Multinomial Logit)
- Stage 2: departure-time choice (Continuous Logit)
- Fixed coefficients: mode-specific constants and values of time, early and late penalties
- Random coefficients: desired arrival times (but distribution is known, from Part I)



Literature review

Departure-time choice:

- α-β-γ model using the work start time as desired arrival time (Small, 1982; Thorauge et al., 2021)
- Schedule-delay represented by time-specific constants (Zeid et al., 2006; Popuri et al., 2008; Lemp and Kockelman, 2010; Lemp et al., 2010)
- Kim and Moon, 2022: desired arrival times are estimated using a machine learning method using the arrival times of individuals facing no congestion

Joint mode and departure-time choice:

- Joint discrete-continuous model with time budget constraint (Habib, 2013; Jokubauskaité, 2019)
- Mixed Nested Logit model with stated preferences (Bajwa et al., 2008)

Results

- The desired arrival times t^* can be mostly explained by *profession category* and *workplace area*.
- Value of time is the smallest for public transit and largest for walk.
- Walking is preferred to car and public transit for trips smaller than 1.3 km.
- When there is no congestion, the odd ratio of choosing car over public transit does not depend on trip distance.

Introduction

Data

Part I: Desired arrival-time distribution

Part II: Bayesian estimates

Conclusion

Introduction

Data

Part I: Desired arrival-time distribution

Part II: Bayesian estimates

Conclusion

Enquête Globale Transport (EGT)

- 2010 **transport survey** for Île-de-France (Paris' region, with 12 millions inhabitants)
- 14 855 households, 35 175 individuals surveyed
- Observations: households characteristics, individual characteristics, trips of the previous day (including, mode, departure time, purpose)

Details

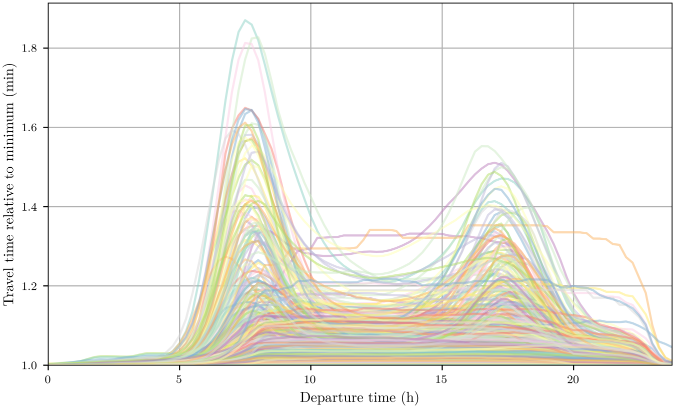
Car Travel-Time

- Source: HERE, Q1 2016
- Historical link-level speed for 15-minute intervals (typical day)
- 977 618 links in the Île-de-France area (18.51 % with a non-constant travel time)
- OD-level travel-time functions computed using a routing algorithm (Time-dependent Contraction Hierarchies)
- Link-level and OD-level travel time functions are piecewise linear functions

▸ OLS regression

Car Travel-Time

Random sample of 500 OD pairs.



Public-Transit / Walk Travel-Time

Public transit:

- Source: OpenStreetMap (walking network) and IDF Mobilités GTFS (timetables)
- Methodology: Least cost path given by OpenTripPlanner (with departure time 2023-06-26 at 8AM)
- 1937 lines, 53 199 stops
- Public-transit travel time is assumed to be constant with departure time (for now)

▸ OLS regression

Walk

- Source: OpenStreetMap
- Methodology: Distance of the shortest path given by a routing algorithm
- Walking travel time is computed assuming a speed of 4.14 km/h (estimated speed based on observed travel times)

▸ OLS regression

Introduction

Data

Part I: Desired arrival-time distribution

Part II: Bayesian estimates

Conclusion

Scope

- Home-to-work trips
- Modes: car (as a driver alone), public transit, walk
- Time window: 3AM – 10AM
- Sample size: 7881 trips

Basic Principle

Claim: When travel-time function is constant, the individual arrives at his / her t^*.

$$u(t_d) = -c - \alpha \cdot tt(t_d) - \beta \cdot [t^* - t_d - tt(t_d)]_+ - \gamma \cdot [t_d + tt(t_d) - t^*]_+$$

$$\text{if } tt(t_d) = \bar{tt} \quad \Rightarrow \quad \arg\max_{t_d} u(t_d) = t^* - \bar{tt}$$

Consequence: For individual facing no congestion, their arrival time reveal their t^* value.

Basic Principle

Claim: When travel-time function is constant, the individual arrives at his / her t^*.

$$u(t_d) = -c - \alpha \cdot \bar{tt} - \beta \cdot [t^* - t_d - \bar{tt}]_+ - \gamma \cdot [t_d + \bar{tt} - t^*]_+$$

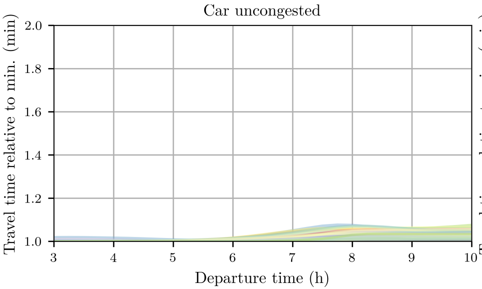
$$\text{if } tt(t_d) = \bar{tt} \quad \Rightarrow \quad \arg\max_{t_d} u(t_d) = t^* - \bar{tt}$$

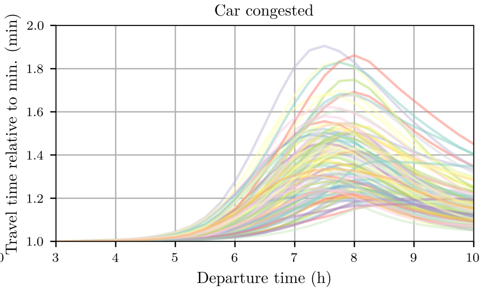
Consequence: For individual facing no congestion, their arrival time reveal their t^* value.

Trip categories

Three trip categories are analyzed:

- **Group A:** Walk (643 trips)
- **Group B:** Car uncongested (1169 trips)
- **Group C:** Car congested (1169 trips)





▸ Congestion

Trip categories

(A: Walk, B: Car uncongested, C: Car congested)

- *Desired arrival time* t^* distribution (unobserved) is F_A, F_B, F_C
- *Arrival time* t_a distribution (observed) is G_A, G_B, G_C

- Goal: estimate F_A, F_B, F_C
- Previous claim: $F_A = G_A$, $F_B = G_B$
- Can we infer F_C from F_A and F_B?

Endogeneity

Variables that could explain t^*:

- Occupation
- Workplace
- Number of children
- Gender
- ...

All these variables also explain mode choice: Desired arrival time distribution is mode-dependent ($F_A \neq F_B$)

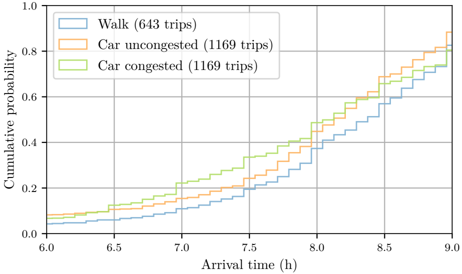
Comparing distribution

- The **two-sample Kolmogorov-Smirnov test** can be used to compare two samples and assert if they come from the same probability distribution.
- Null hypothesis: "The values in the two samples are drawn from the same probability distribution".

All trips, by mode

The null hypothesis is always rejected at the 1 % level (the distributions are different).

	KS statistic	p-value
Walk / Car uncong.	**0.1308**	**0.0000**
Walk / Car cong.	0.1439	0.0000
Car cong. / Car uncong.	0.0915	0.0001

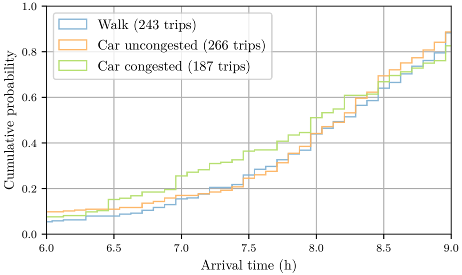


If we split population by profession category (employee, intermediate category, upper category, blue-collar workers), are the arrival-time distributions still explained by mode choice?

Employees

The null hypothesis that *Walk* and *Car uncongested* have the same distribution **cannot** be rejected.

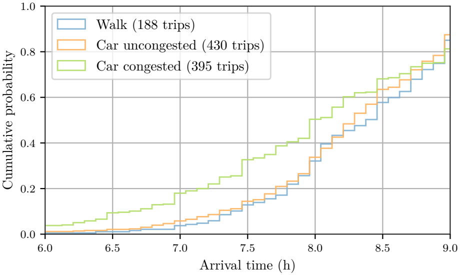
	KS statistic	p-value
Walk / Car uncong.	**0.0678**	**0.5709**
Walk / Car cong.	0.1151	0.1096
Car cong. / Car uncong.	0.1206	0.0738



Intermediate category

The null hypothesis that *Walk* and *Car uncongested* have the same distribution **cannot** be rejected.

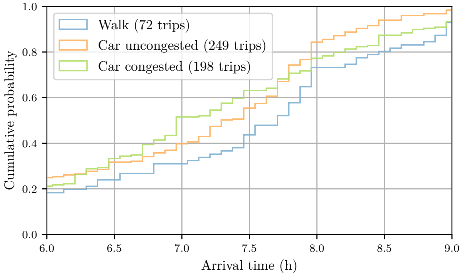
	KS statistic	p-value
Walk / Car uncong.	**0.0711**	**0.4957**
Walk / Car cong.	0.2171	0.0000
Car cong. / Car uncong.	0.1830	0.0000



Blue-Collar Workers

The null hypothesis that *Walk* and *Car uncongested* have the same distribution **cannot** be rejected.

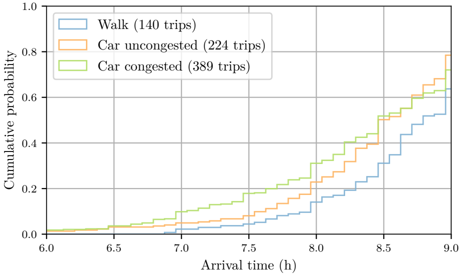
	KS statistic	p-value
Walk / Car uncong.	**0.1735**	**0.0602**
Walk / Car cong.	0.2210	0.0097
Car cong. / Car uncong.	0.1176	0.0856



Upper category

The null hypothesis that *Walk* and *Car uncongested* have the same distribution **can** be rejected.

	KS statistic	p-value
Walk / Car uncong.	**0.2071**	**0.0010**
Walk / Car cong.	0.2213	0.0001
Car cong. / Car uncong.	0.0970	0.1269



Upper category: By workplace area

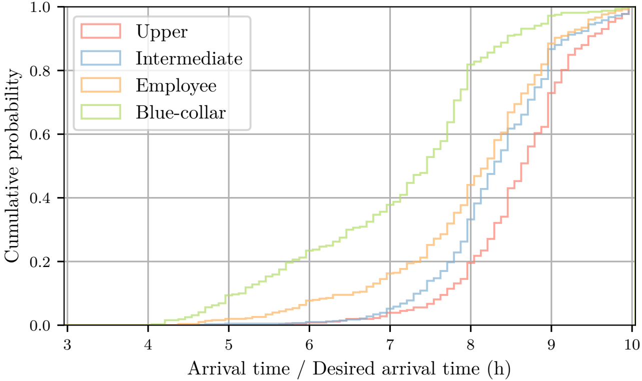
Average arrival time (trip count)

	Paris	Inner suburbs	Outer suburbs
Walk	9:10 (70)	8:50 (51)	8:38 (29)
Car uncongested	9:26 (3)	8:46 (51)	8:34 (185)
Car congested	8:45 (51)	8:32 (222)	8:38 (127)
Average	8:56	8:37	8:34

Conclusion: Ideally, t^* distributions should be split by profession category and workplace area but sample size is too small.

Summary

Conclusion: After controlling by profession category and workplace area, the desired arrival time is no longer explained by mode $\Rightarrow F_A = F_B = F_C$



▸ Robustness Checks

Introduction

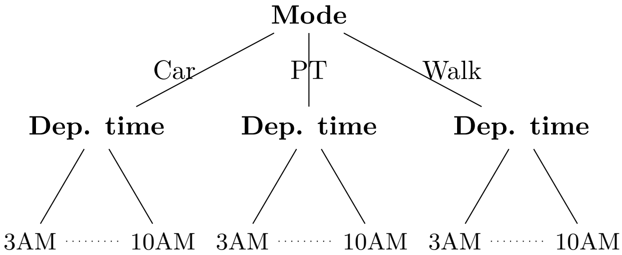
Data

Part I: Desired arrival-time distribution

Part II: Bayesian estimates

Conclusion

Decision Tree



Likelihood (lower level)

- Observed choice for individual n: $y_n = \{i, [\underline{t}, \bar{t}]\}$, where i is the mode chosen and $[\underline{t}, \bar{t}]$ is the departure-time interval chosen.
- Let $\theta = \{\boldsymbol{c}, \boldsymbol{\alpha}, \beta, \gamma, \boldsymbol{\mu}\}$ be the set of parameters to be estimated.
- The likelihood to observe departure-time interval $[\underline{t}, \bar{t}]$ given that mode i is chosen is (**Continuous Logit** assumption)

$$L_n^{\text{dep}}([\underline{t}, \bar{t}]; i, \theta, t_n^*) = \frac{\int_{\underline{t}}^{\bar{t}} e^{u_n(t; i, \theta, t_n^*)} \, \mathrm{d}\, t}{\int_{t_0}^{t_{M+1}} e^{u_n(t; i, \theta, t_n^*)} \, \mathrm{d}\, t}$$

where

$$u_n(t; i, \theta, t_n^*) = \frac{-c_i - \alpha_i \cdot tt_{n,i}(t) - \beta \cdot [t_n^* - t - tt_{n,i}(t)]_+ - \gamma \cdot [t + tt_{n,i}(t) - t_n^*]_+}{\mu_i}.$$

Likelihood (upper level)

- The log-sum of the departure-time choice for mode i is

$$V_{n,i}(\theta, t_n^*) = \mu_i \ln \int_{t_0}^{t_{M+1}} e^{u_n(t;i,\theta,t_n^*)} \,\mathrm{d}\,t + \mu_i EC$$

  where $EC \approx 0.5772$ is the Euler-Mascheroni constant.

- The likelihood to observe mode i is (**Multinomial Logit** assumption)

$$L_n^{\text{mode}}(i; \theta, t_n^*) = \frac{e^{V_{n,i}(\theta, t_n^*)}}{\sum_j e^{V_{n,j}(\theta, t_n^*)}}.$$

- The likelihood to observe the choice $y_n = \{i, [\underline{t}, \overline{t}]\}$ is

$$L_n(y_n; \theta, t_n^*) = L_n^{\text{mode}}(i; \theta, t_n^*) \cdot L_n^{\text{dep}}([\underline{t}, \overline{t}]; i, \theta, t_n^*).$$

Likelihood

- The previous likelihood is *conditional on* t_n^*.
- The unconditional likelihood is

$$L_n(y_n;\theta) = \int L_n(y_n;\theta,t^*)\,f(t^*)\,\mathrm{d}\,t^*$$

  where f is the probability distribution of t^*.
- Maximum Likelihood is not feasible, instead we can use Maximum Simulated Likelihood or **Bayesian estimates**.

Bayesian estimates

Goal: Find the estimates of $\theta = \{\boldsymbol{c}, \boldsymbol{\alpha}, \beta, \gamma, \boldsymbol{\mu}\}$ and $\boldsymbol{t^*} = \{t^*_n\}_n$ which best "explain" the observed choices $\boldsymbol{y} = \{y_n\}_n$.

- Density of the prior distribution of θ and $\boldsymbol{t^*}$ is $k(\theta, \boldsymbol{t^*})$ (assumed to be diffuse).
- Density of the posterior distribution is

$$K(\theta, \boldsymbol{t^*}; \boldsymbol{y}) \propto \prod_n L_n(y_n; \theta, t^*_n) f(t^*_n) k(\theta, \boldsymbol{t^*}).$$

Values are drawn from the posterior distribution by combining Gibbs sampling and Metropolis-Hastings algorithm.

▸ Details

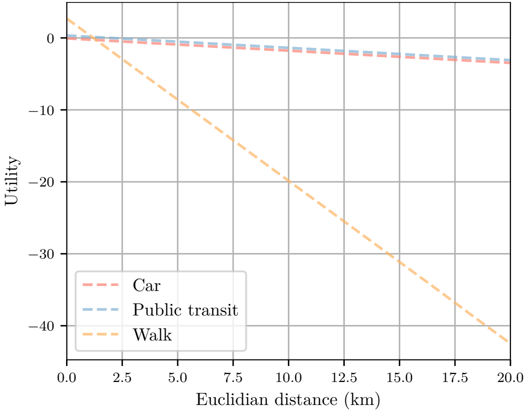
Results: Intermediate category

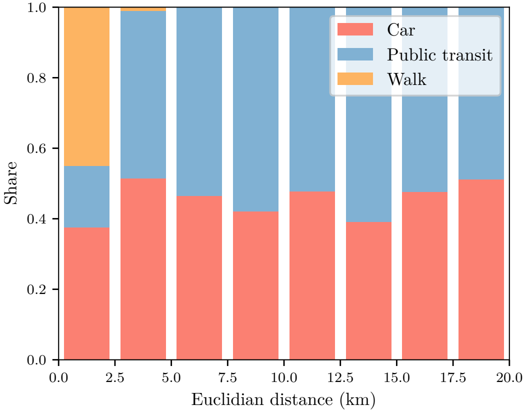
Variable	Estimate	CI [1 %, 99 %]
Const:Public_transit c_{PT}	-0.40	[-0.65, -0.15]
Const:Walk c_{Walk}	-2.76	[-3.38, -2.15]
VOT:Car α_{Car}	3.47	[2.91, 4.03]
VOT:Public_transit α_{PT}	2.18	[1.80, 2.60]
VOT:Walk α_{Walk}	8.02	[6.66, 9.48]
Early penalty β	0.97	[0.82, 1.14]
Late penalty γ	0.82	[0.73, 1.06]
Scale μ	0.16	[0.13, 0.20]

▸ Convergence

Utility as function of distance

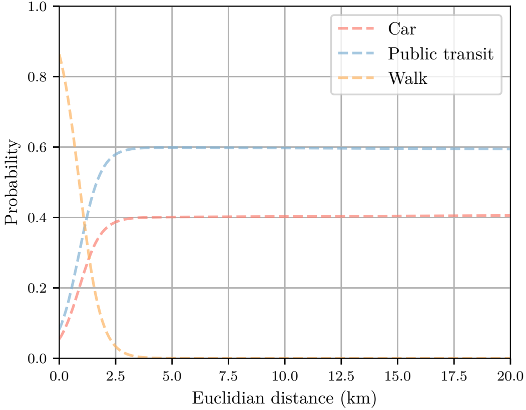
Average observed speeds: Car 20.37 km/h, Public-transit 12.75 km/h, Walking 3.56 km/h

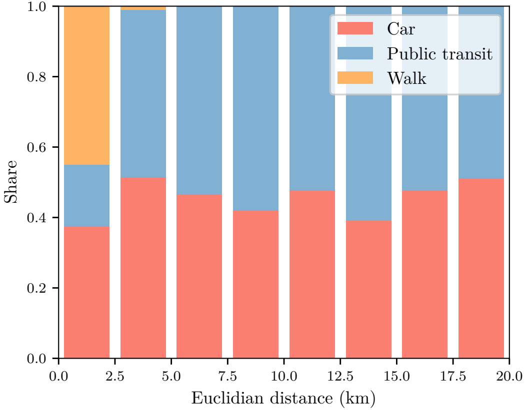




Mode-choice probabilities as function of distance

Average observed speeds: Car 20.37 km/h, Public-transit 12.75 km/h, Walking 3.56 km/h





Introduction

Data

Part I: Desired arrival-time distribution

Part II: Bayesian estimates

Conclusion

Takeaways

We estimate the $\mathbf{t}^*$ **distribution** using *arrival-time* distribution of individuals facing no congestion.

- When controlling by *profession category* and *workplace area*, the desired arrival-time distribution cannot be explained by mode (walk vs uncongested car).

We estimate preference parameters using a **Mixed discrete-continuous Nested Logit model** (for intermediate categories).

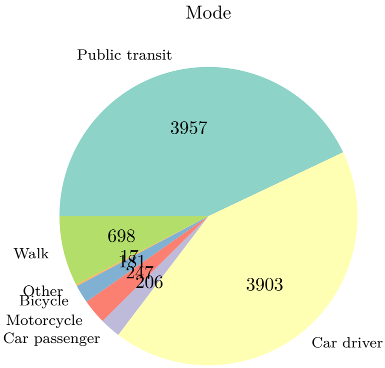
- Value of time is the smallest for public transit and largest for walk.
- Walking is preferred to car and public transit for trips smaller than 1.3 km.
- When there is no congestion, trip distance does not affect the choice between car and public transit.

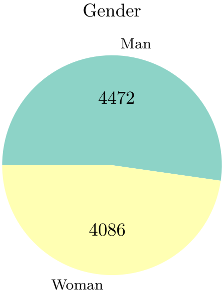
Future works

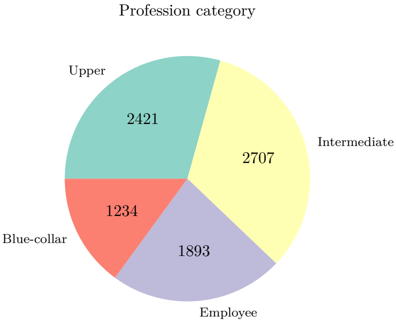
- **Fuel cost** / public-transit fare
- **Evening commute** (desired departure time from origin)
- **Trip chaining** (with t^* at intermediate stop and at destination)
- Car ownership
- Day-to-day travel-time variability

Thank you

Characteristics of home-to-work trips







Back

Car Travel-Time

Reported travel time for car trips in the travel survey can be well predicted by the computed travel time with HERE data ($R^2 = 66\,\%$).

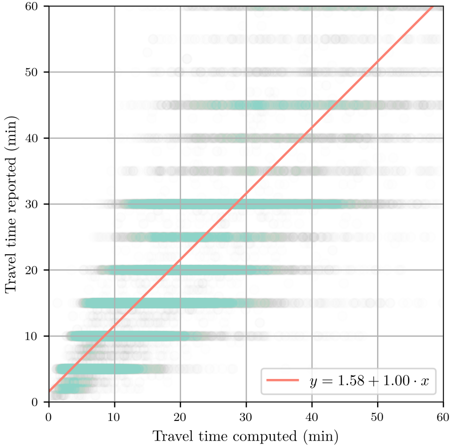
Dep. Variable:	EGT tt	**R-squared:**	0.664
Model:	OLS	**F-statistic:**	8.274e+04
No. Observations:	41934	**Prob (F-statistic):**	0.00

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
cst	1.5758	0.087	18.061	0.000	1.405	1.747
HERE tt	1.0003	0.003	287.650	0.000	0.993	1.007

Note: Travel-time penalties at intersections are calibrated to reach a slope close to 1.

Back

Car Travel-Time



◂ Back

Public-Transit Travel-Time

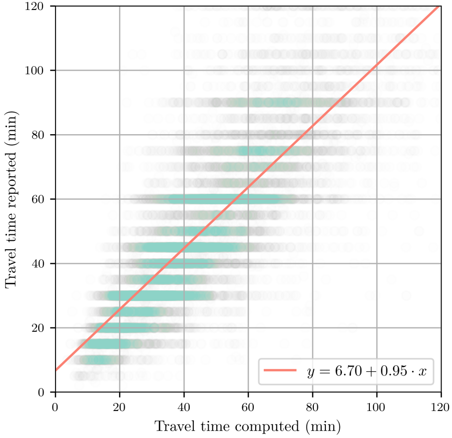
Reported travel time for public-transit trips in the travel survey can be well predicted by the computed travel time with OpenTripPlanner ($R^2 = 65\,\%$).

Dep. Variable:	EGT tt	**R-squared:**	0.653
Model:	OLS	**F-statistic:**	3.472e+04
No. Observations:	18453	**Prob (F-statistic):**	0.00

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
cst	6.7029	0.257	26.095	0.000	6.199	7.206
OTP tt	0.9499	0.005	186.330	0.000	0.940	0.960

Back

Public-Transit Travel-Time



Back

Walk Travel-Time

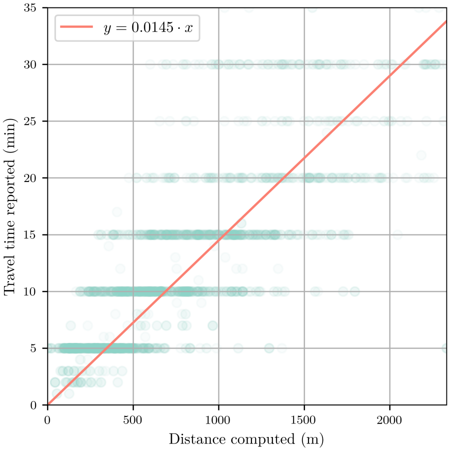
Reported travel time for walking trips in the travel survey can be well predicted by the computed distance with OpenStreetMap.

Dep. Variable:	EGT tt	**F-statistic:**	1.313e+04
Model:	OLS	**Prob (F-statistic):**	0.00
No. Observations:	2141		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
OSM dist	0.0145	0.000	114.571	0.000	0.014	0.015

Back

Walk Travel-Time



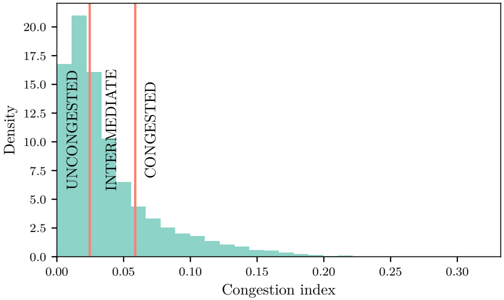
◂ Back

Congestion Index

For each OD pair, we have a travel-time function defined by breakpoints $\{(td_i, tt_i)\}_i$. We compute a congestion index as

$$c = \sigma_{tt}/tt_0,$$

where $\sigma_{tt} = \sqrt{(1/n)\sum_i (tt_i - \bar{tt})^2}$ is the standard-deviation of the travel times and $tt_0 = \min_i tt_i$ is the minimum travel time.



Car trips are split in three categories of equal size based on the congestion index (uncongested, intermediate and congested).

Back

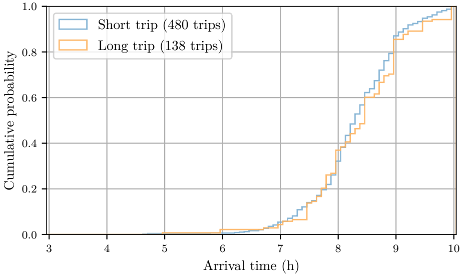
Robustness check: travel time

Comparing long / short trips (intermediate category; walk and car uncongested only).

Long trip: Travel time is longer than 30 minutes.

The null hypothesis that *Short trip* and *Long trip* have the same distribution **cannot** be rejected.

	KS statistic	p-value
Short / Long	0.0923	0.2453



Back

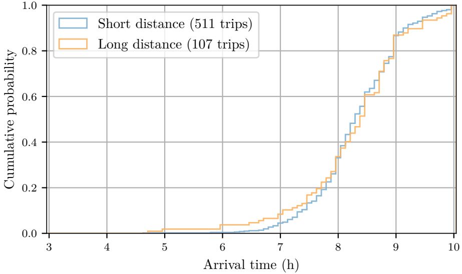
Robustness check: distance

Comparing long / short distance trips (intermediate category; walk and car uncongested only).

Long distance: Euclidian distance between origin and destination is greater than 10 kilometers.

The null hypothesis that *Short distance* and *Long distance* have the same distribution **cannot** be rejected.

	KS statistic	p-value
Short / Long	0.0588	0.8531



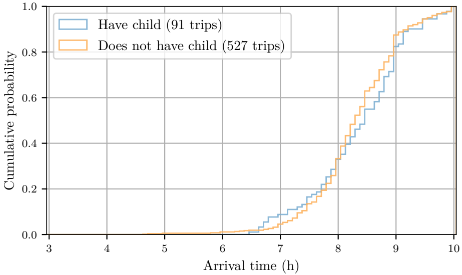
Back

Robustness check: children

Comparing trips of people with / without child (intermediate category; walk and car uncongested only).

The null hypothesis that *Male* and *Female* have the same distribution **cannot** be rejected.

	KS statistic	p-value
Child / No child	0.0914	0.4662



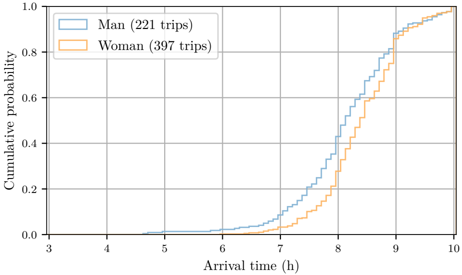
Back

Robustness check: gender

Comparing trips of men / women (intermediate category; walk and car uncongested only).

The null hypothesis that *Man* and *Woman* have the same distribution **can** be rejected.

	KS statistic	p-value
Man / Woman	0.1469	0.0029



Back

Gibbs Sampling

1. Draw $(t_n^*)^{\tau+1}$, $\forall n$ given θ^τ $\rightarrow$ Metropolis-Hastings

$$K(t_n^*|\theta; y_n) \propto L_n(y_n|\theta; t_n^*) f(t_n^*), \quad \forall n$$

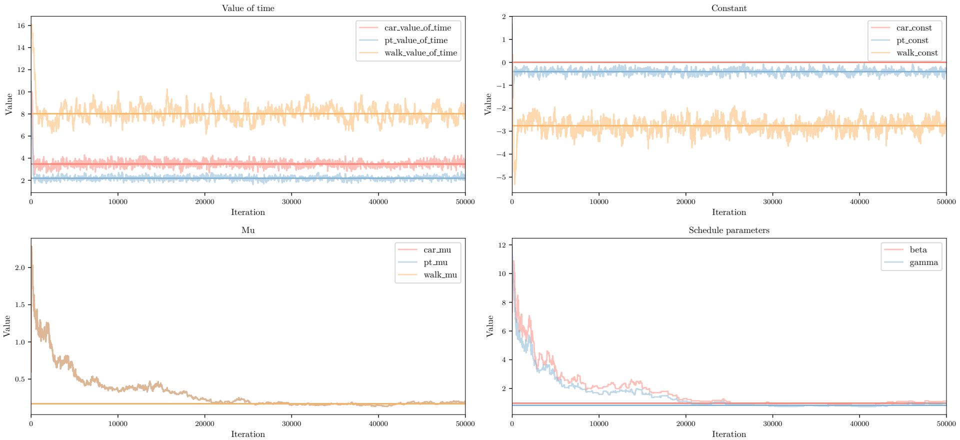
2. Draw $\theta^{\tau+1}$ given $(t_n^*)^{\tau+1}$ $\rightarrow$ Metropolis-Hastings

$$K(\theta|\boldsymbol{t}^*; \boldsymbol{y}) \propto \prod_n L_n(y_n|\theta; t_n^*)$$

Uniform random values are drawn to initialize the first iteration. Each simulation consists in 50 000 iterations of Gibbs sampling.

◂ Back

Results: Convergence of simulation



Back