MRC Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Technical University of Munich

Welcome

Incorporating day-to-day stability into agent-based models

3rd Symposium on Activity-Based Modeling, 12 December 2024

Corin Staves
Qin Zhang
James Woodcock

MRC Epidemiology Unit
School of Clinical Medicine
University of Cambridge

Professorship for Travel Behaviour
School of Engineering and Design
Technical University of Munich

Agenda

1 Stability in Travel Behavior

Relevance for public health applications,
state of practice in ABMs

2 Fixed Error Terms

Stabilising the unobserved part of utility for decision makers

3 Mode Set

Restricting modes available to each decision maker

4 Discussion

Advanced approaches and potential applications

MRC Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Technic
Universi
of Munic

Day-to-day stability in Travel Behavior

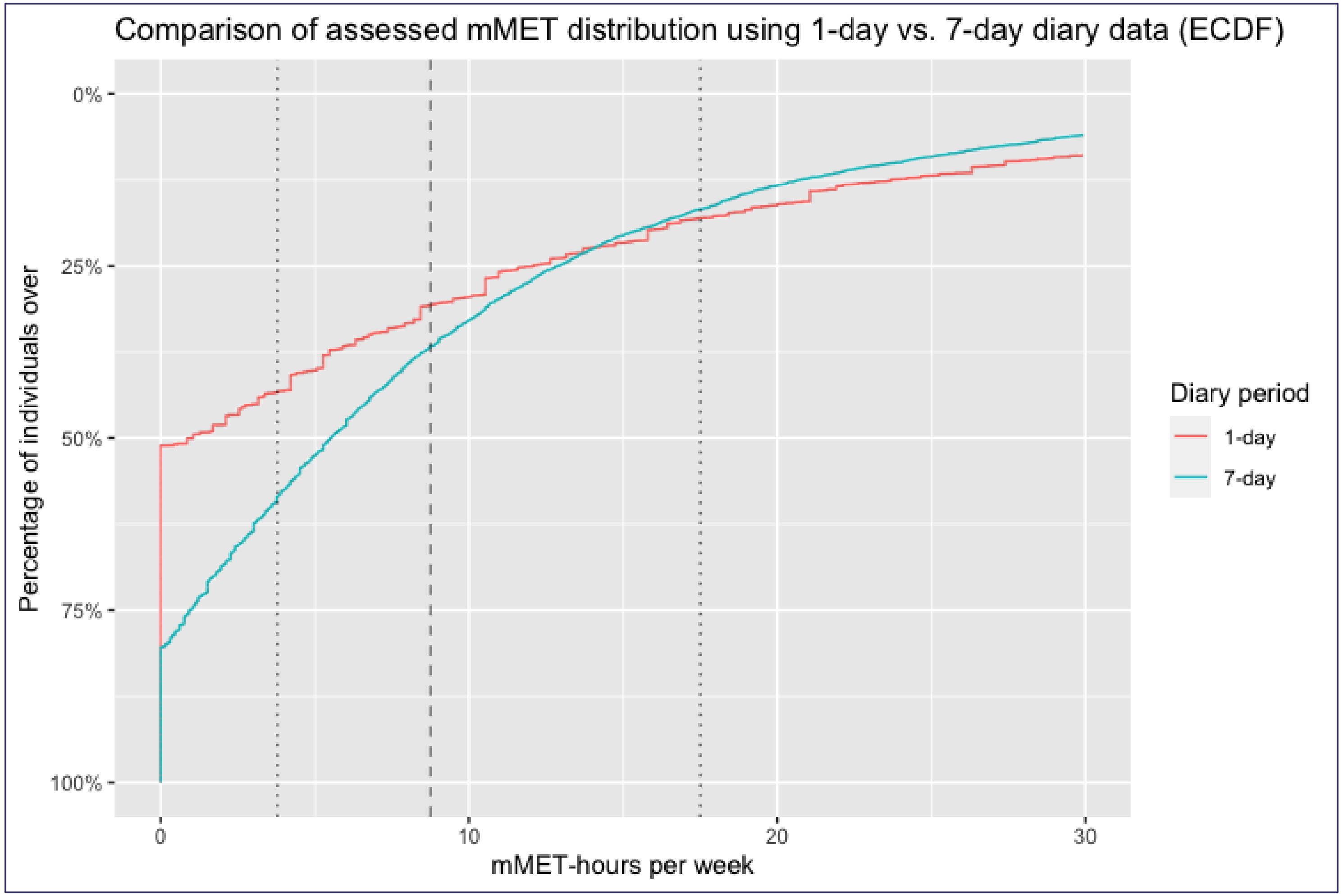
- Many transport analyses, including health, aim to understand ***habitual behaviour***
  - *How much **physical activity** does a person perform through walking and cycling?*
  - *How much **air pollution** is somebody exposed to as they travel?*
  - *How at risk is somebody to **traffic injuries**?*

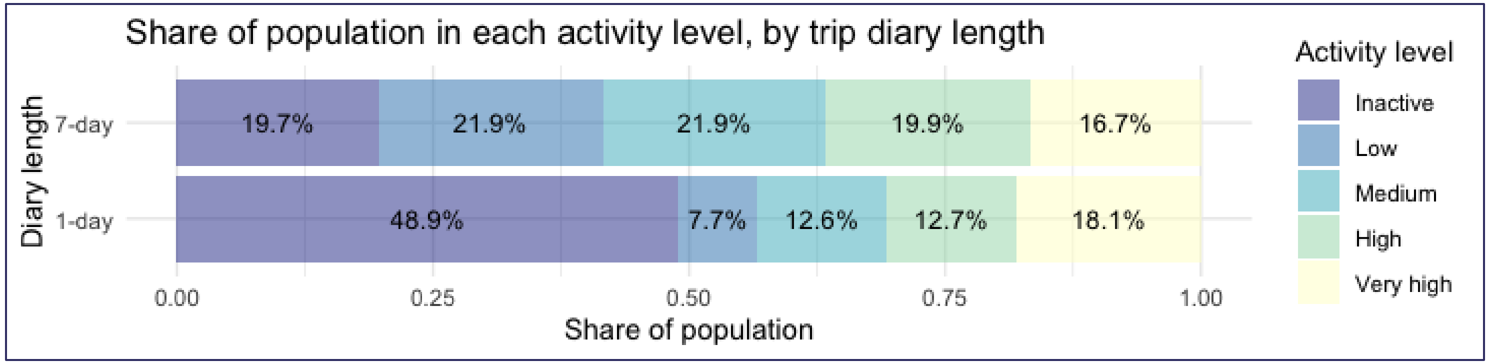
- Epidemiological studies, which relate behaviour to health, consider ***one week*** at minimum

- However, most transport surveys and models consider ***one day***
  - Exceptions: Mobilitätspanel, Mobidrive, mobiTopp

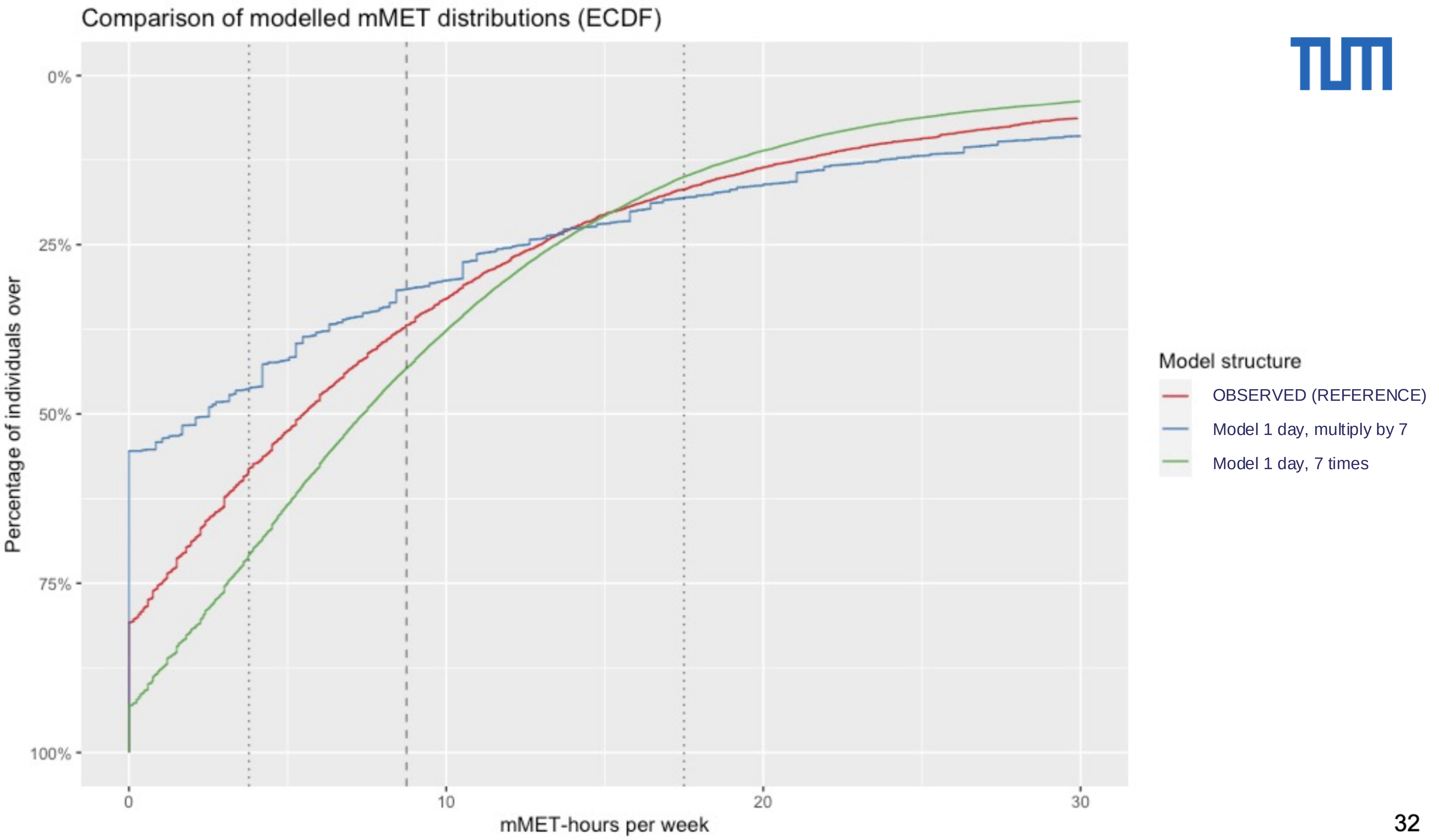
Physical Activity Demonstration

- Population's walking and cycling can be used to assess physical activity
- Units: mMET-hours
  - Walking ≈ 3.6 mMET
  - Cycling ≈ 5.4 mMET
- Distribution looks different when you use a 1-day vs. 7-day diary

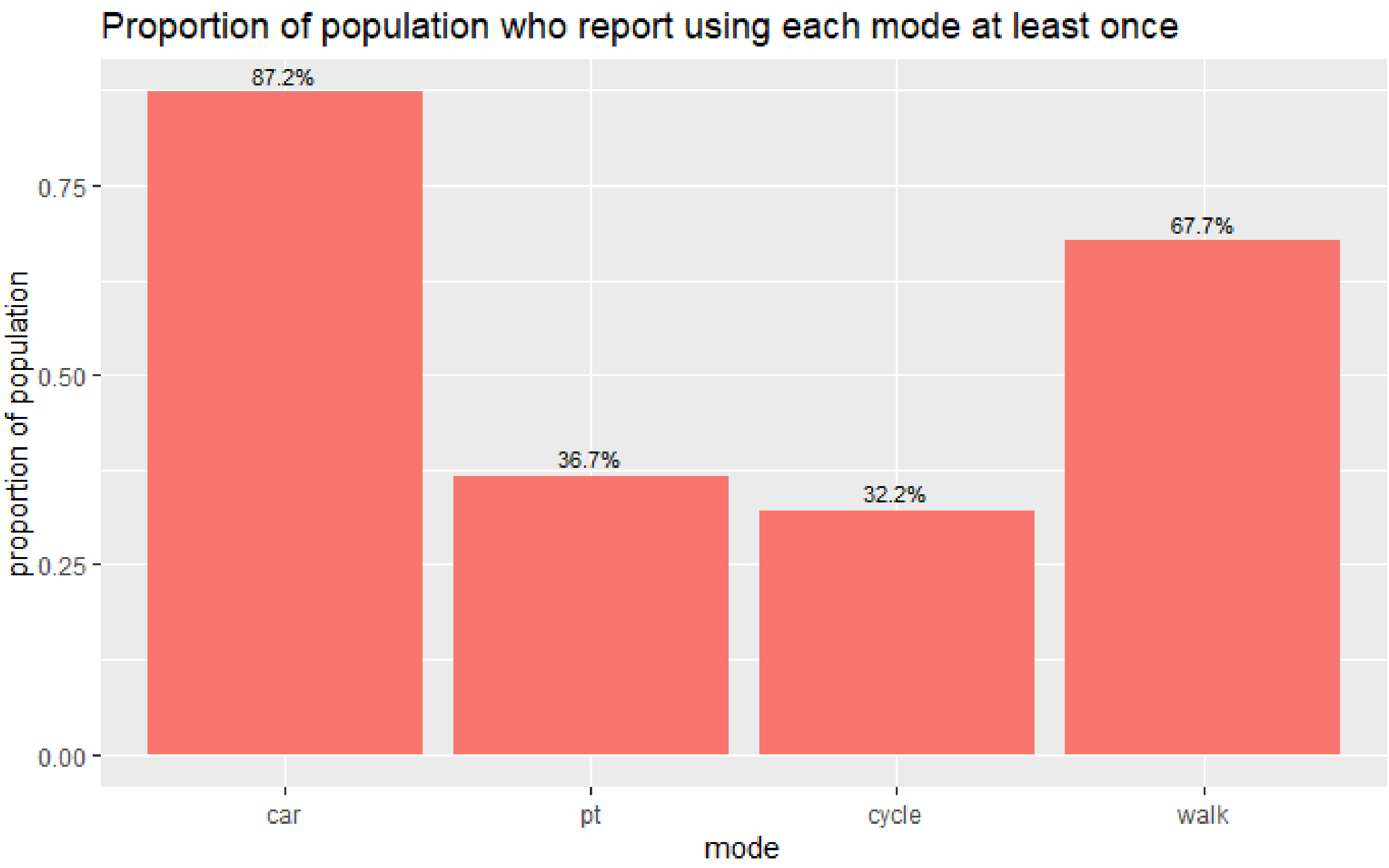




Modelling 7-day physical activity with a 1-day model



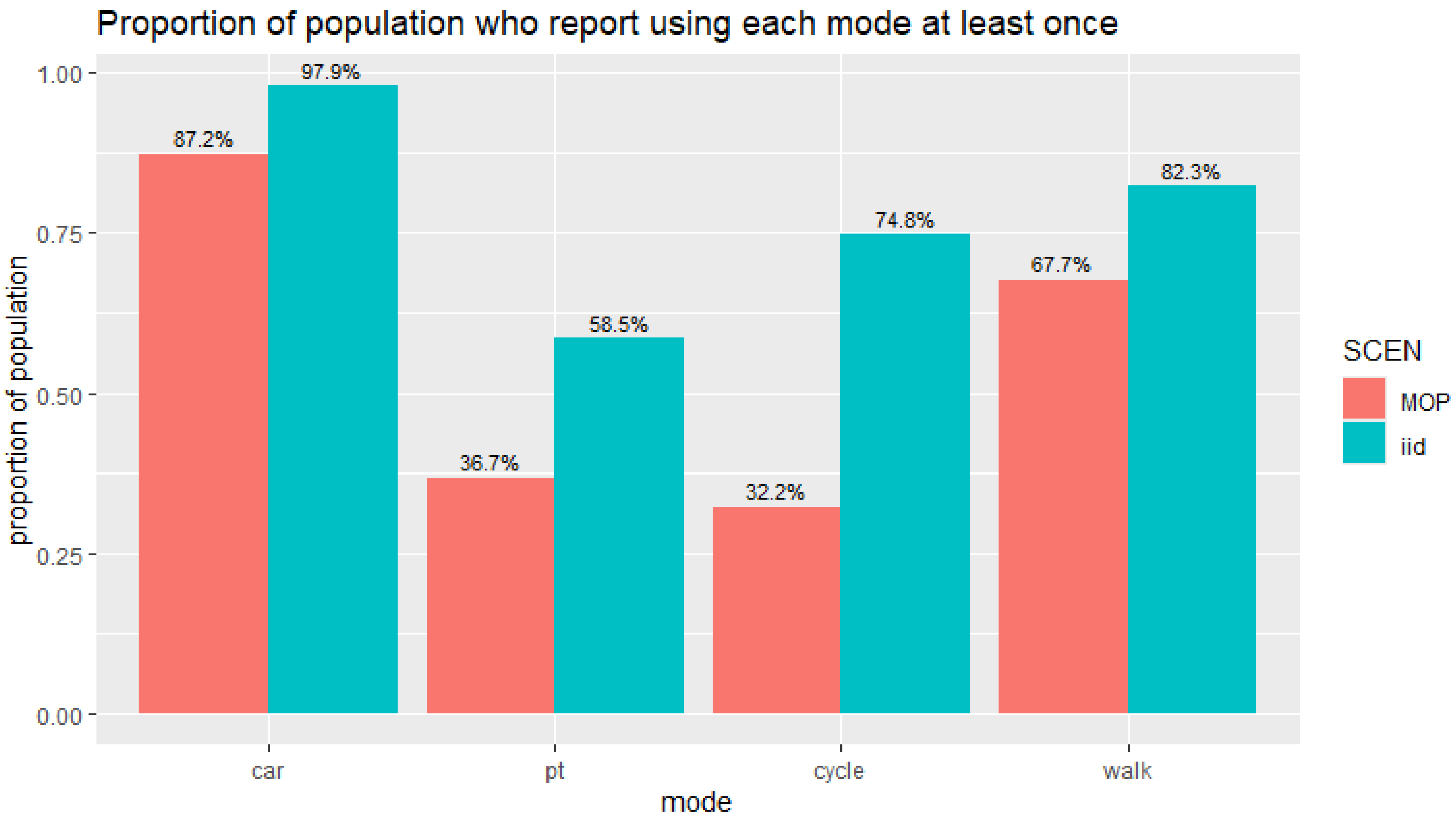
Findings from 7-day *Mobilitätspanel*



Other findings:

- 21% of respondents report trips by car only (no other modes)
- 75% of respondents report at least one active trip (cycle or walk)

Mobilitätspanel vs. 7-days Modelled (Multinomial Logit)



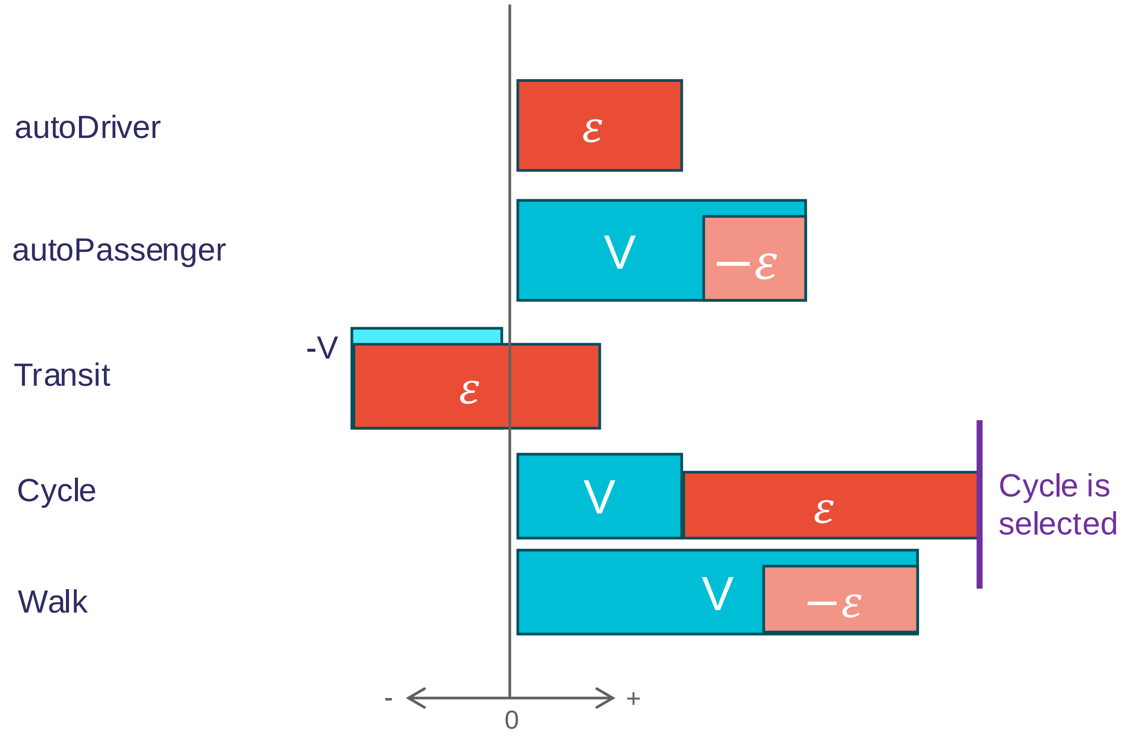
i.i.d. Assumption in Discrete Choice Models

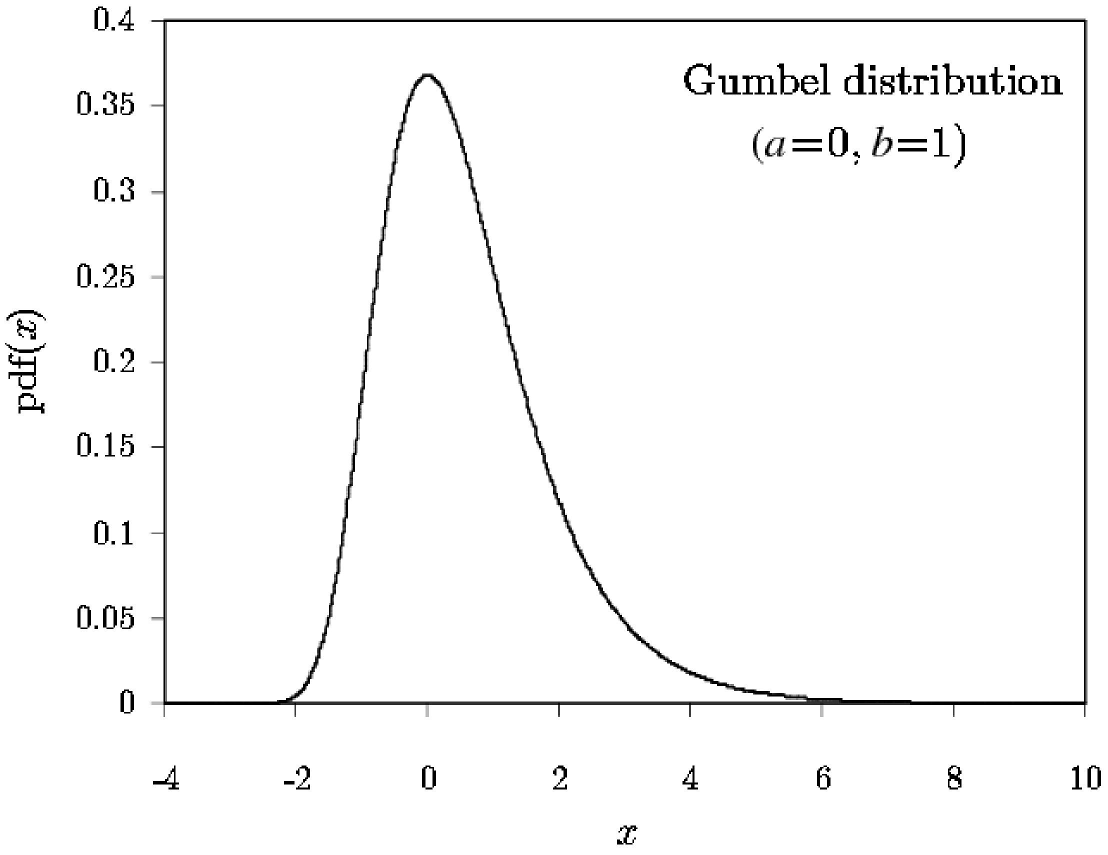
Utility contains an **observed** and **unobserved** component

$$U = V + \varepsilon$$

We commonly assume the error terms to be **independent and identically distributed (i.i.d.)**

Example decision (Multinomial logit model)



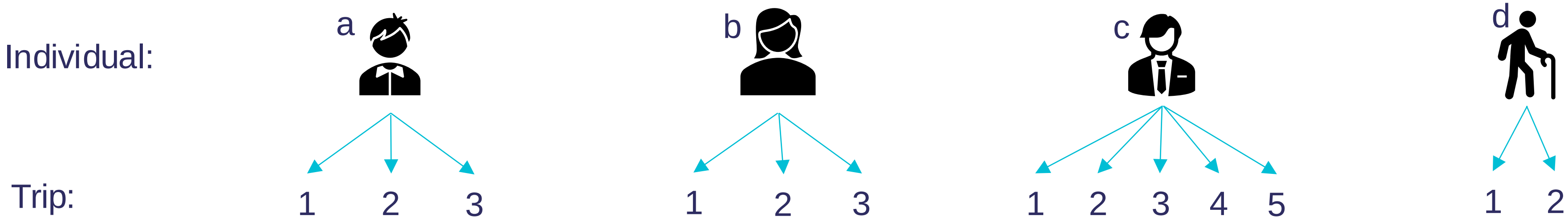


In many cases, unobserved utility is correlated

Between alternatives (e.g., red bus / blue bus problem)

(Image from Rolf Moeckel's Lecture Notes)

Between choices (e.g., of the same person)



Standard Mode Choice Implementation in ABMs

$$U_i = V_i + \varepsilon_i$$

Utility is commonly converted to a probability….

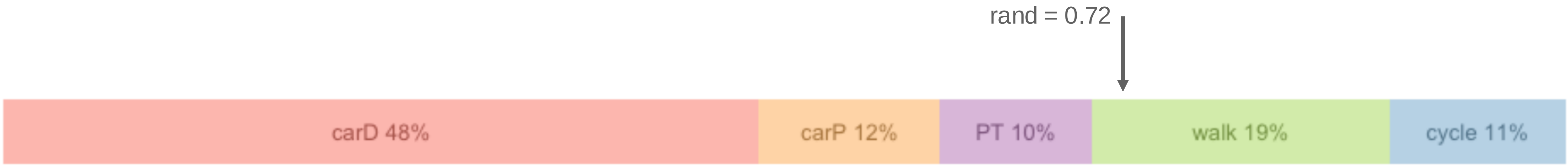
Multinomial logit

$$p_i = \frac{e^{V_i}}{\sum_{j \in M} e^{V_j}}$$

Nested logit

$$p_i = \frac{e^{V_i/\lambda_k} \left(\sum_{j \in B_k} e^{V_i/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^{k} \left(\sum_{j \in B_l} e^{V_j/\lambda_k} \right)^{\lambda_l}}$$

…. and then the choice is sampled from a Uniform (0,1) Distribution



Equivalent Implementation by Sampling ε_i

$$U_i = V_i + \varepsilon_i$$

ε_i for each alternative is sampled from a distribution…

Multinomial logit

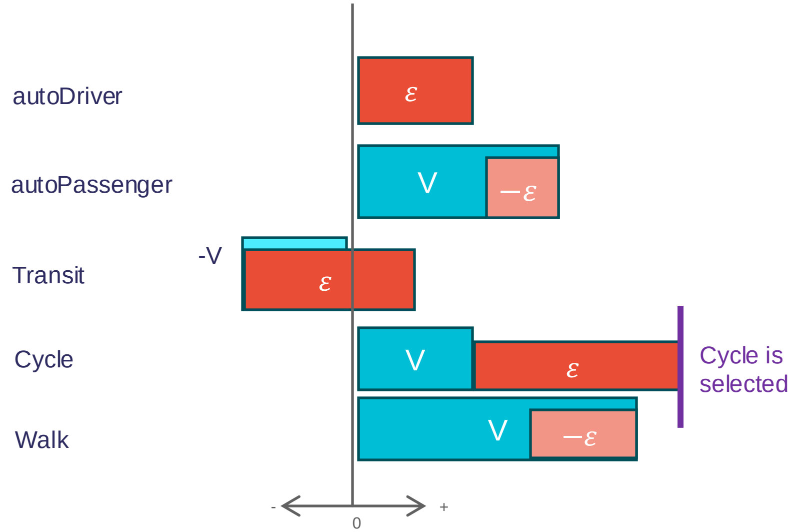
Sample i.i.d. from Gumbel (0,1)

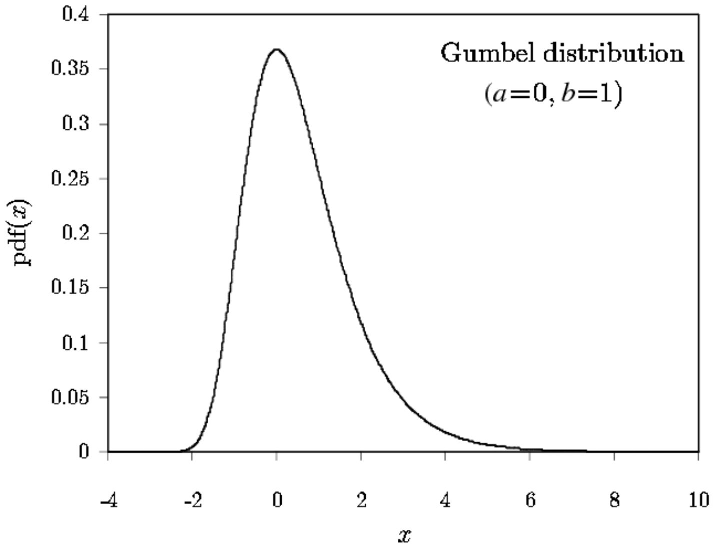
Nested logit

Outside nests: Sample i.i.d. from Gumbel (0,1)

Within nests: Sample from joint CDF $F = \exp\left(-\left(\sum e^{-\frac{\varepsilon}{\lambda}}\right)^{\lambda}\right)$

…. and then choose the highest





Potential solutions to improve stability

Multinomial / nested logit

$$U = V + \varepsilon$$

V systematic
ε sampled per ***trip***

Fixed error terms

$$U = V + \varepsilon$$

V systematic
ε sampled per ***person***

Preliminary Mode Set Choice

Mode Set

RESTRICTED CHOICE SET

$$U = V + \varepsilon$$

V systematic
ε sampled per ***trip***

Random Parameters

$$U = V + W + \varepsilon$$

V systematic
W sampled per ***person***
ε sampled per ***trip***

[Standard practice] **Simple** **Complicated**

Agenda

1 Stability in Travel Behavior

State of practice for agent-based modelling, relevance for public health applications

2 Fixed Error Terms

Stabilising the unobserved part of utility for decision makers

3 Mode Set

Restricting modes available to each decision maker

4 Discussion

Advanced approaches and potential applications

UKRI MRC Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Technic Universi of Munic

1) Fixed Error Terms

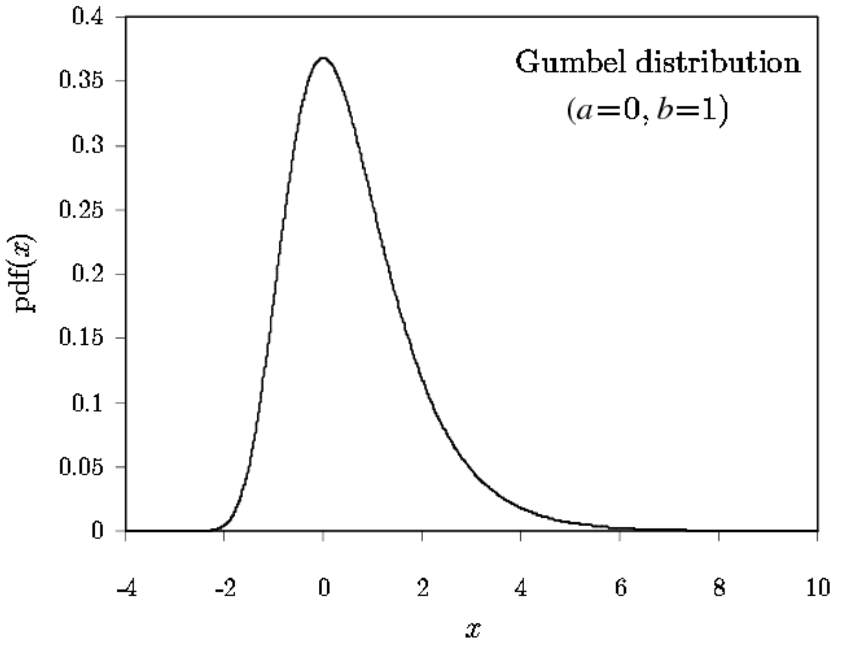
$$U = V + \varepsilon$$

Instead of sampling ε once per ***decision***, we sample ε once per ***person***

Assumption: 100% of unobserved utility is due to personal characteristics

		Without Fixed Error Terms				
Person	**Trip**	ε_{drive}	ε_{pax}	ε_{pt}	ε_{bike}	ε_{walk}
1	**1**	1.85	2.01	0.22	1.42	1.38
1	**2**	-0.73	-1.02	0.6	-0.07	0.73
2	**1**	-0.26	-0.09	-0.61	-0.02	0.27
2	**2**	0.06	0.76	1.04	-1.13	-0.22
2	**3**	0.71	0.79	-0.83	-0.36	-1.4

$$\underbrace{\varepsilon_{drive} \quad \varepsilon_{pax}}_{\lambda_{car} = 0.25}$$



1) Fixed Error Terms

$$U = V + \varepsilon$$

Instead of sampling ε once per ***decision***, we sample ε once per ***person***

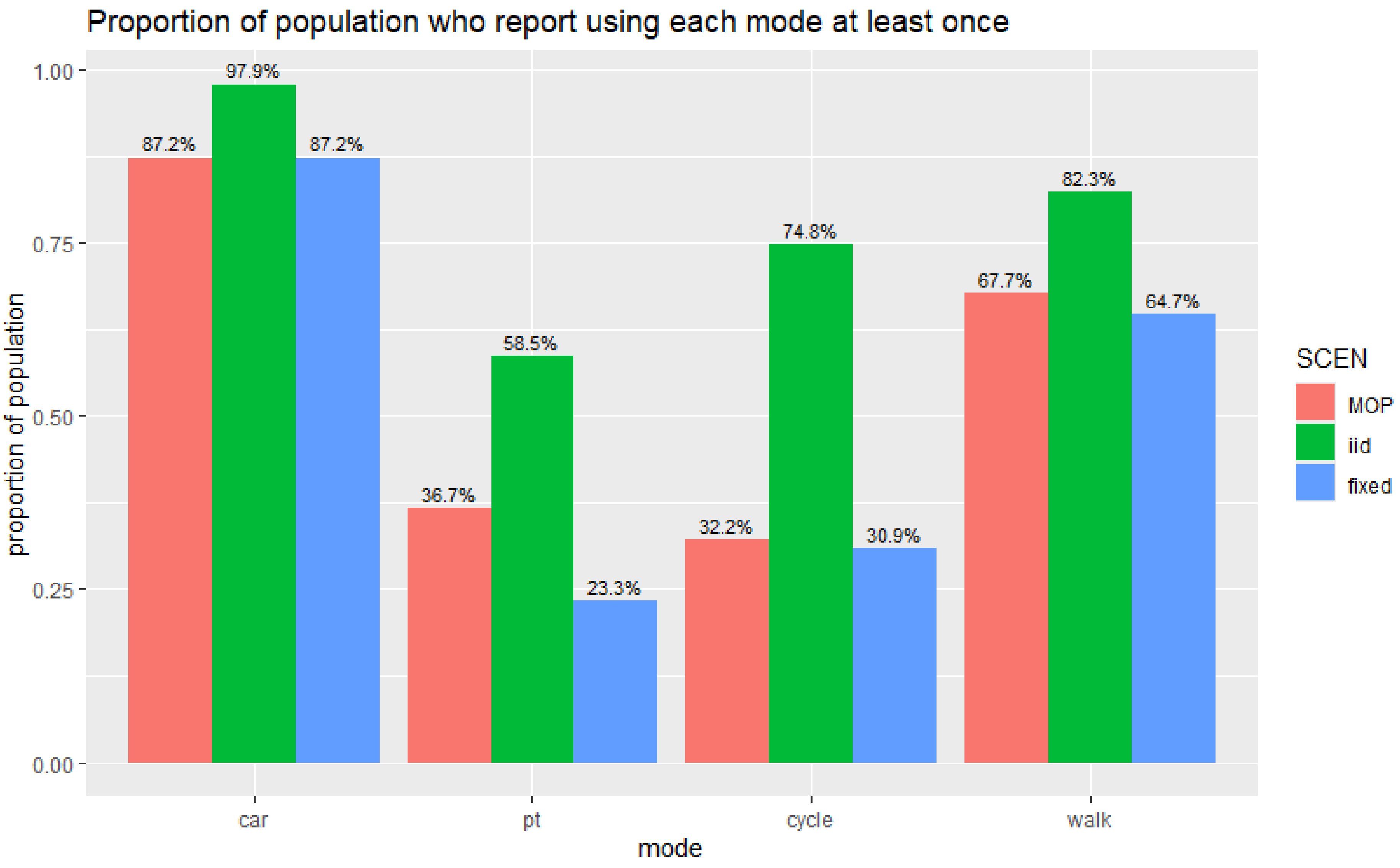
Assumption: 100% of unobserved utility is due to personal characteristics

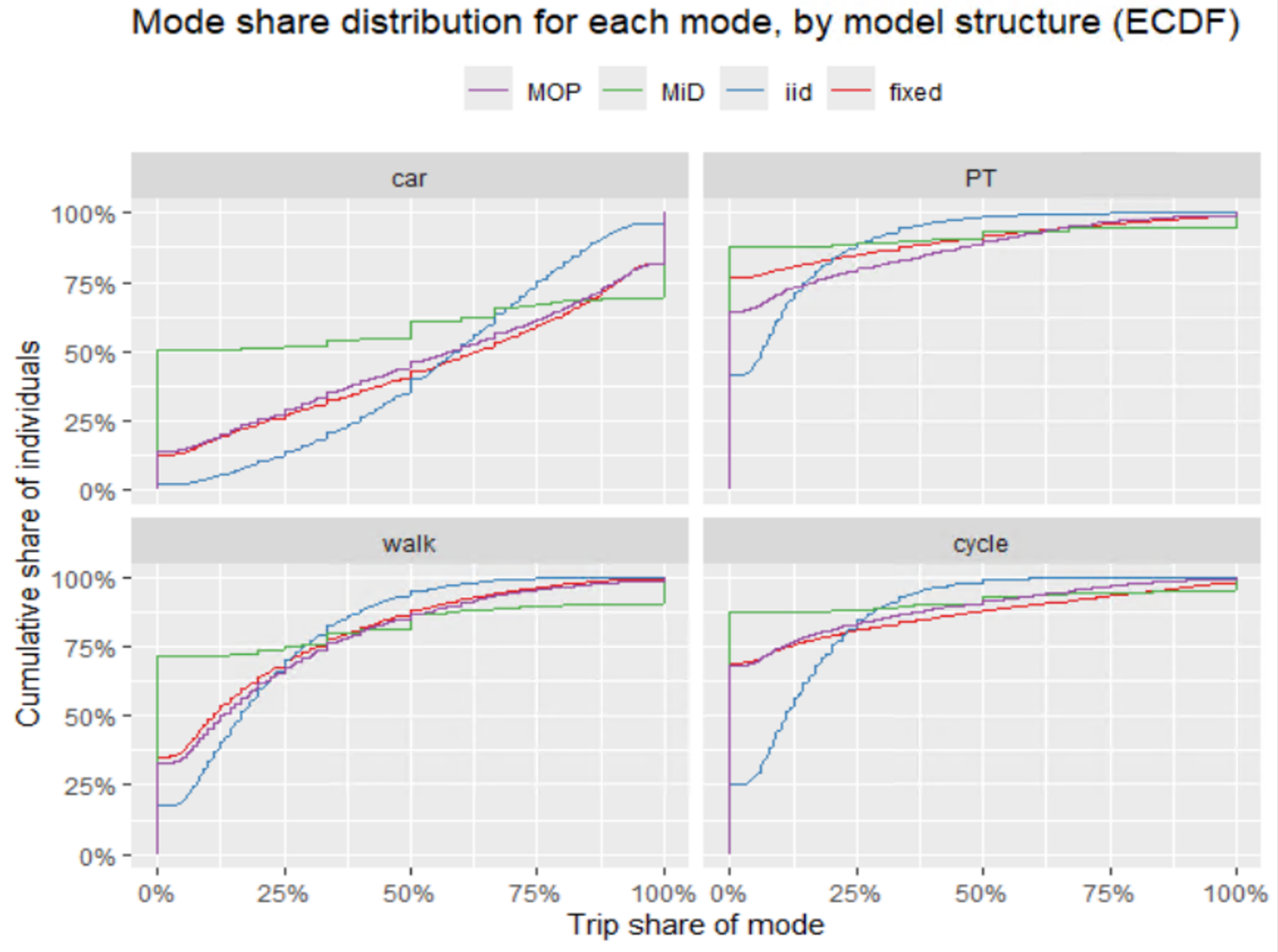
		Without Fixed Error Terms					With Fixed Error Terms				
Person	**Trip**	ε_{drive}	ε_{pax}	ε_{pt}	ε_{bike}	ε_{walk}	ε_{drive}	ε_{pax}	ε_{pt}	ε_{bike}	ε_{walk}
1	**1**	1.85	2.01	0.22	1.42	1.38	1.85	2.01	0.22	1.42	1.38
1	**2**	-0.73	-1.02	0.6	-0.07	0.73	1.85	2.01	0.22	1.42	1.38
2	**1**	-0.26	-0.09	-0.61	-0.02	0.27	-0.26	-0.09	-0.61	-0.02	0.27
2	**2**	0.06	0.76	1.04	-1.13	-0.22	-0.26	-0.09	-0.61	-0.02	0.27
2	**3**	0.71	0.79	-0.83	-0.36	-1.4	-0.26	-0.09	-0.61	-0.02	0.27

$\lambda_{car} = 0.25$

$\lambda_{car} = 0.25$

Implementing Fixed Error Terms in MITO




Mode share distribution for each mode, by model structure (ECDF)
MOP
MiD
iid
fixed
car
PT
walk
cycle
Cumulative share of individuals
0%
25%
50%
75%
100%
Trip share of mode

Further development: correlated error terms

		With Fixed Error Terms					With Correlated Error Terms				
Person	Trip	ε_{drive}	ε_{pax}	ε_{pt}	ε_{bike}	ε_{walk}	ε_{drive}	ε_{pax}	ε_{pt}	ε_{bike}	ε_{walk}
1	1	1.85	2.01	0.22	1.42	1.38	1.85	2.01	0.22	1.42	1.38
1	2	1.85	2.01	0.22	1.42	1.38	1.99	1.62	0.28	1.47	1.22
2	1	-0.26	-0.09	-0.61	-0.02	0.27	-0.26	-0.09	-0.61	-0.02	0.27
2	2	-0.26	-0.09	-0.61	-0.02	0.27	-0.03	-0.15	-0.69	0.01	0.29
2	3	-0.26	-0.09	-0.61	-0.02	0.27	-0.30	-0.11	-0.59	-0.09	0.23

λ_p

λ_p

$\lambda_{car} = 0.25$

$\lambda_{car} = 0.25$

1) Fixed Error Terms (Discussion)

- Simple to implement
- Works with models estimated on 1-day data
- Implementation contradicts i.i.d. assumption
- For a 7-day model, results are closer to reality (for a given input) than assuming i.i.d. error terms.

$$U = V + \varepsilon$$

V systematic
ε sampled per ***person***

Agenda

1 Stability in Travel Behavior

State of practice for agent-based modelling, relevance for public health applications

2 Fixed Error Terms

Stabilising the unobserved part of utility for decision makers

3 Mode Set

Restricting modes available to each decision maker

4 Discussion

Advanced approaches and potential applications

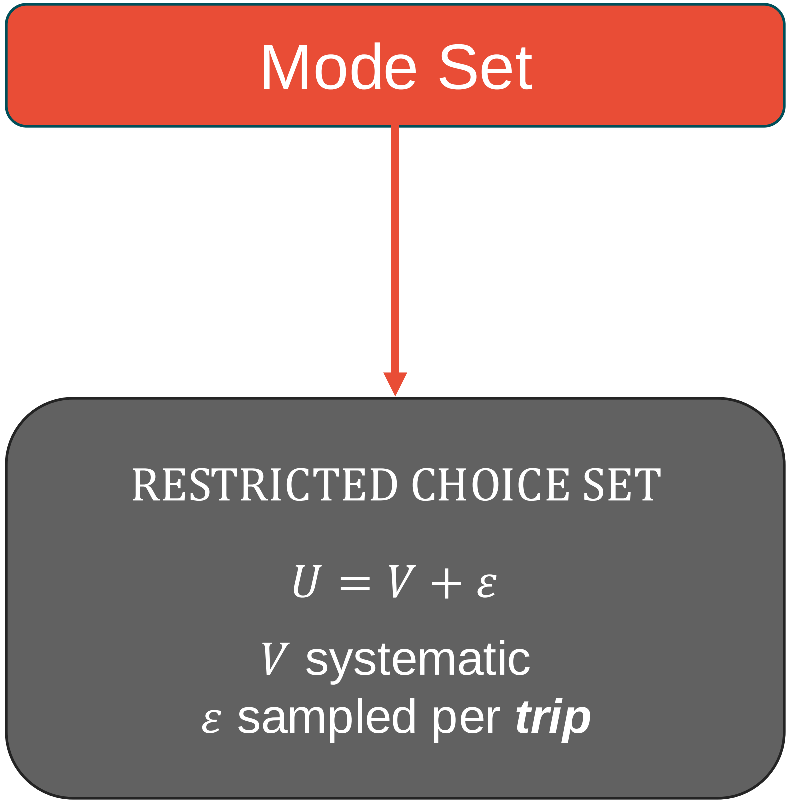
MRC Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Technic
Universi
of Munic

2) Mode Set Model

- Prior to mode choice, **estimate the available alternatives** for each person. This is called the “mode set”
- During mode choice, **restrict availability** of alternatives to the mode set
- Ton et al., (2019) developed an empirical approach to help



Mode Set as a Multinomial Logit Model

Modes $m \in M$

Car | PT | Cycle | Walk

Alternatives $i \subseteq M, i = 1, \ldots, 15$

- Car
- Car, PT
- Car, Cycle
- Car, PT, Walk
- Car, Walk
- PT
- Car, PT, Cycle
- Car, PT, Cycle, Walk
- Car, Cycle, Walk
- PT, Cycle
- Cycle
- PT, Cycle, Walk
- Cycle, Walk
- PT, Walk
- Walk

Utility for mode *m*: $V_m = \alpha_m + \boldsymbol{\beta}_m \boldsymbol{x}$

Where $\alpha_m = mode\ specific\ constant$

$\boldsymbol{\beta}_m = vector\ of\ coefficients\ for\ mode\ m$

$\boldsymbol{x} = vector\ of\ trip\ attributes$

Utility for alternative i, person *n*:

$$V_i = ASC_i + \sum_{m \in i} V_m$$

Ton et al., (2020)

Mode Set: Empirical Results from *Mobilitätspanel*

Alternative Specific Constants

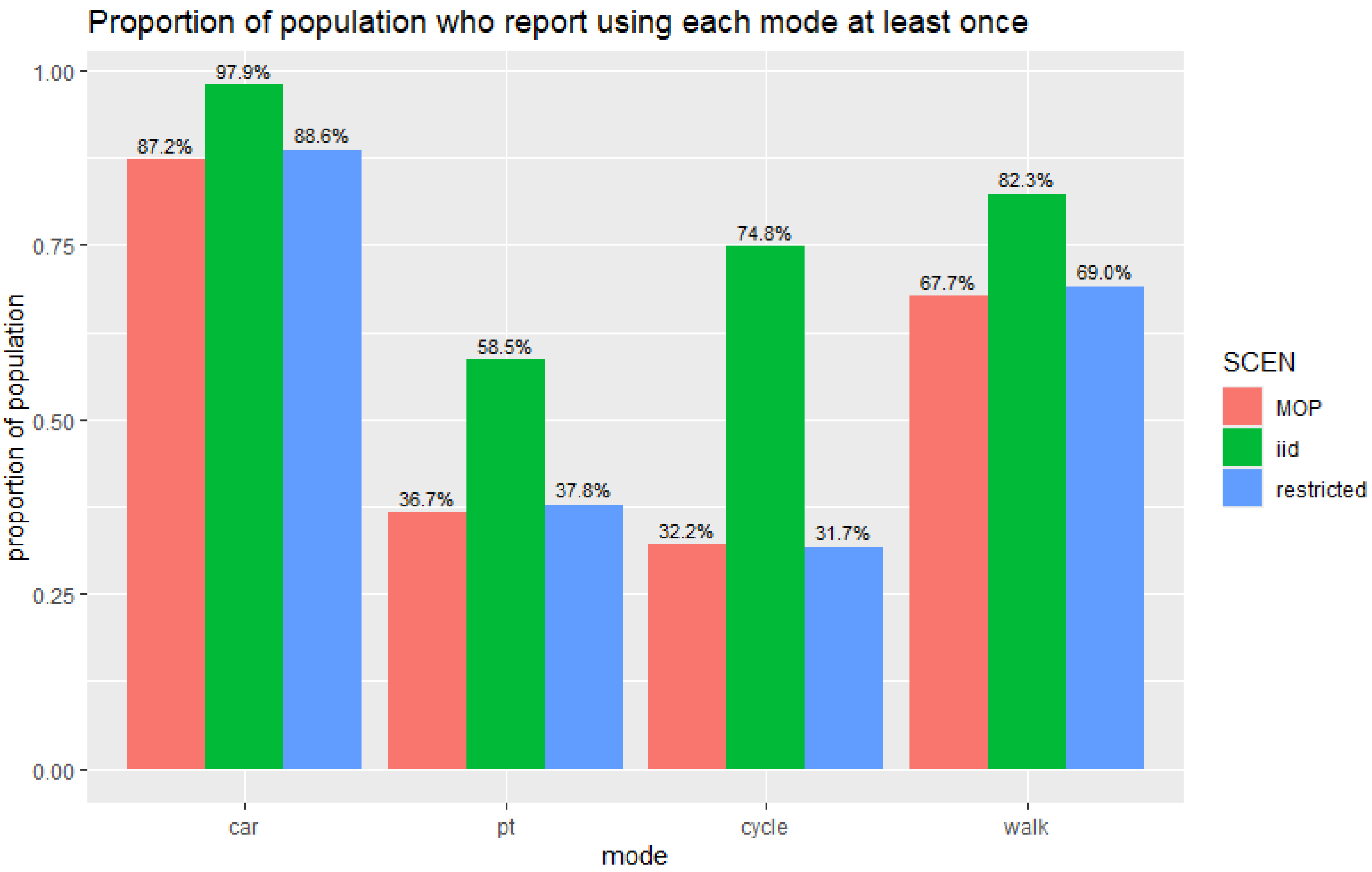
Car	Pt	Walk	Cycle	Share	ASC
√				20.53%	—
√	√			5.91%	-0.15
√		√		12.22%	0.167
√			√	3.83%	-0.721
√	√	√		2.07%	0.985
√	√		√	6.66%	-0.601
√		√	√	12.89%	-0.493
√	√	√	√	25.97%	—
	√			0.49%	—
	√	√		1.20%	1.582
	√		√	0.68%	-0.191
	√	√	√	0.44%	0.858
		√		2.15%	—
		√	√	3.89%	0.27
			√	1.06%	—

Mode Coefficients

McFadden R^2 = 0.43

	car	pt	walk	cycle
INTERCEPT	-1.95 [-1.816 `]	-2.968 [-1.486]	-0.337 [-1.192]	-3.211 [-1.669 `]
hh.econStatus_2	0.271 [7.276 ***]	0 [NA]	0 [NA]	0.068 [6.243 ***]
hh.econStatus_3	0.333 [4.285 ***]	0 [NA]	-0.033 [-11.113 ***]	0 [NA]
hh.econStatus_4	0.37 [3.949 ***]	0 [NA]	-0.145 [-8.842 ***]	0 [NA]
hh.econStatus_34	0 [NA]	0.279 [5.049 ***]	0 [NA]	0.119 [4.639 ***]
hh.urban	-0.436 [-6.474 ***]	0.95 [6.91 ***]	-0.021 [-13.429 ***]	0.168 [7.035 ***]
hh.homePT	0 [NA]	0.52 [6.783 ***]	0.212 [11.699 ***]	0.057 [6.709 ***]
hh.children_1	0 [NA]	0 [NA]	0 [NA]	0.098 [4.025 ***]
hh.children_2	0 [NA]	0 [NA]	-0.074 [-8.507 ***]	0.336 [3.257 **]
hh.children_3	0 [NA]	0 [NA]	-0.035 [-5.213 ***]	0.425 [1.73 `]
hh.children_123	0.483 [5.131 ***]	0.383 [4.279 ***]	0 [NA]	0 [NA]
hh.cars_1	2.722 [6.116 ***]	-1.377 [-2.667 **]	0 [NA]	-0.012 [-2.961 **]
hh.cars_2	0 [NA]	-1.508 [-2.328 *]	0 [NA]	-0.381 [-2.528 *]
hh.cars_3	0 [NA]	-1.497 [-2.114 *]	0 [NA]	-0.866 [-2.277 *]
hh.cars_23	3.5 [4.808 ***]	0 [NA]	0 [NA]	0 [NA]
hh.autosPerAdult	0 [NA]	-1.506 [-2.947 **]	-0.368 [-8.729 ***]	-0.336 [-2.842 **]
p.age_gr_1	-0.569 [-2.759 **]	1.057 [1.543]	0.007 [4.095 ***]	0.198 [2.062 *]
p.age_gr_2	-0.002 [-3.775 ***]	0.746 [3.098 **]	0.202 [6.869 ***]	0.336 [3.364 ***]
p.age_gr_4	0.316 [3.801 ***]	0.135 [4.337 ***]	-0.025 [-10.086 ***]	0.3 [4.434 ***]
p.age_gr_5	0.537 [3.697 ***]	0.289 [4.066 ***]	-0.192 [-8.625 ***]	0.619 [3.942 ***]
p.age_gr_6	0.667 [3.871 ***]	0.373 [3.888 ***]	0.06 [7.324 ***]	0.401 [3.799 ***]
p.female	0.165 [7.706 ***]	0.401 [6.658 ***]	0.01 [14.618 ***]	-0.273 [-6.462 ***]
p.driversLicense	0.209 [6.231 ***]	-0.066 [-2.992 **]	-0.393 [-7.162 ***]	0.15 [4.262 ***]
p.ownBicycle	-0.004 [-6.129 ***]	0.095 [6.688 ***]	-0.089 [-12.092 ***]	2.303 [6.846 ***]
p.km_min_T	-0.251 [-14.97 ***]	-0.049 [-12.864 ***]	-1.702 [-26.691 ***]	-0.408 [-13.502 ***]
p.km_max_T	0.525 [14.299 ***]	0.47 [11.49 ***]	0.098 [26.207 ***]	0.122 [12.369 ***]
p.workTrips_1234	0 [NA]	0.479 [2.177 *]	-0.155 [-5.416 ***]	0.196 [2.814 **]
p.workTrips_5	0 [NA]	0.257 [2.164 *]	-0.166 [-5.095 ***]	0.18 [2.682 **]
p.isMobile_HBW	0.542 [2.959 **]	0 [NA]	0 [NA]	0 [NA]
p.eduTrips_1234	0 [NA]	0.641 [1.685 `]	-0.458 [-5.374 ***]	0.219 [1.928 `]
p.eduTrips_5	0 [NA]	0.424 [1.387]	-0.539 [-3.912 ***]	0.55 [1.881 `]
p.isMobile_HBE	0.585 [2.636 **]	0 [NA]	0 [NA]	0 [NA]
p.workPT_12	0 [NA]	0.844 [4.664 ***]	-0.027 [-11.172 ***]	-0.142 [-5.215 ***]
p.isMobile_RRT	0 [NA]	0 [NA]	2.014 [12.781 ***]	0.557 [4.754 ***]
p.usualCommuteMode_carD	0 [NA]	-1.354 [-2.172 *]	0.096 [4.823 ***]	-0.406 [-2.939 **]
p.usualCommuteMode_carP	0 [NA]	-0.237 [-1.557]	0.686 [3.636 ***]	0 [NA]
p.usualCommuteMode_PT	-0.673 [-2.583 **]	0 [NA]	0.349 [4.562 ***]	0.16 [2.287 *]
p.usualCommuteMode_walk	-1.084 [-2.295 *]	-0.842 [-1.478]	0 [NA]	-0.35 [-2.278 *]
p.usualCommuteMode_cycle	-0.476 [-2.428 *]	-0.996 [-1.777 `]	0.038 [4.314 ***]	0 [NA]

Implementing Mode Set in MITO



Mode Set (Discussion)

Advantages:

- Simple model structures
- Easy to estimate
- Can precisely segment population by mode use

Limitations:

- Requires panel data
- Mode choice sensitivity split between two models

Mode Set

RESTRICTED CHOICE SET

$$U = V + \varepsilon$$

V systematic
ε sampled per ***trip***

Agenda

1 Stability in Travel Behavior

State of practice for agent-based modelling, relevance for public health applications

2 Fixed Error Terms

Stabilising the unobserved part of utility for decision makers

3 Mode Set

Restricting modes available to each decision maker

4 Discussion

Advanced approaches and potential applications

MRC Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Random Parameters (Mixed) Logit

$$U = V + W + \varepsilon$$

Where:

- V is the **systematic** component of utility (same for everyone)
- W is the **random** component of utility (varies between individuals)
- ε is the i.i.d. error term

W follows any distribution, representing **unobserved variation** in individual preferences

If W follows a discrete distribution, it's called a **latent class logit**

Empirical Findings using Random Parameters Logit

Findings from 6-week MobiDrive panel

- **Cherchi and Chirillo (2014):** Individual tastes for time and cost are relatively stable, repeated trips likely to use the same mode

- **Cherchi et al., (2017):** Intrapersonal variability significant day-to-day, but not week-to-week. Suggests a 7-day survey can be sufficient.

Findings from 4-week Dutch panel

- **Thomas et al., (2019):** Intrapersonal variation mainly for short trips (<2km) and recreation trips. Stable for longer trips and commute trips.

MRC Epidemiology Unit
UNIVERSITY OF CAMBRIDGE
Technical University of Munich

Implementation for Mixed Logit

$$U_i = V_i + W_i + \varepsilon_i$$

For each alternative,

- V_i estimated as usual (e.g., $V = \beta x$)
- W_i sampled from the specified distribution, ***once per person***
- ε_i sampled from i.i.d. Gumbel (0,1) distribution ***once per decision***

Could better capture **intrapersonal stability** and provide **more informative policy sensitivity**

Why hasn't this been done?

- Random parameters models ideally require **panel data**
- They are difficult to estimate and calibrate
- Probably won't capture all the intrapersonal stability in W

$$U = V + W + \varepsilon$$

Potential solutions to improve stability

Multinomial / nested logit

$U = V + \varepsilon$

V systematic
ε sampled per ***trip***

Fixed error terms

$U = V + \varepsilon$

V systematic
ε sampled per ***person***

- Simple to estimate
- No special data
- Limited flexibility

Preliminary Mode Set Choice

Mode Set

↓

RESTRICTED CHOICE SET

$U = V + \varepsilon$

V systematic
ε sampled per ***trip***

- Simple model structure
- Requires panel data
- Limited sensitivity to mode shift scenarios

Random Parameters

$U = V + W + \varepsilon$

V systematic
W sampled per ***person***
ε sampled per ***trip***

- Incorporates intrapersonal stability
- Requires panel data
- Difficult to estimate & calibrate
- Hasn’t been done (yet)

Discussion

- We increasingly need to model habitual behavior
- Statistical models we use in ABMs are behind empirical literature
- How can we incorporate stability into the ABMs used in practice as well as emerging frameworks (e.g., activity-based models)?

References

- Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511805271
- Ton, D., Bekhor, S., Cats, O., Duives, D. C., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2020). The experienced mode choice set and its determinants: Commuting trips in the Netherlands. *Transportation Research Part A: Policy and Practice*, *132*, 744–758. https://doi.org/10.1016/j.tra.2019.12.027
- Cherchi, E., & Cirillo, C. (2014). Understanding variability, habit and the effect of long period activity plan in modal choices: A day to day, week to week analysis on panel data. *Transportation*, *41*(6), 1245–1262. https://doi.org/10.1007/s11116-014-9549-y
- Cherchi, E., Cirillo, C., & Ortúzar, J. de D. (2017). Modelling correlation patterns in mode choice models estimated on multiday travel data. *Transportation Research Part A: Policy and Practice*, *96*, 146–153. https://doi.org/10.1016/J.TRA.2016.11.021
- Thomas, T., La Paix Puello, L., & Geurs, K. (2019). Intrapersonal mode choice variation: Evidence from a four-week smartphone-based travel survey in the Netherlands. *Journal of Transport Geography*, *76*, 287–300. https://doi.org/10.1016/J.JTRANGEO.2018.06.021

Thank you

MRC Epidemiology Unit
@MRC_EPID
mrc-epid.cam.ac.uk