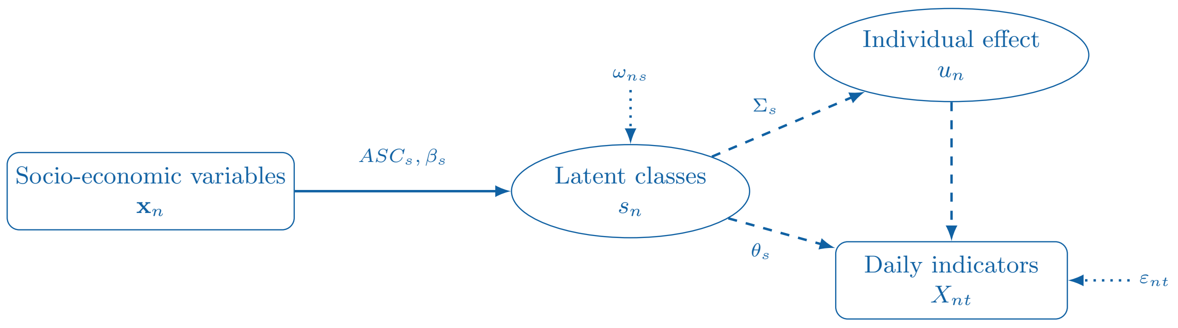

Heterogeneity in Mobility Behavior: Latent Classes from Long-Term Tracking Data

Anne-Valérie Preto

Andrea Schneider

Michel Bierlaire

STRC Conference Paper 2026 **April 30, 2026**

STRC | **26th Swiss Transport Research Conference**
Monte Verità / Ascona, May 20-22, 2026

Heterogeneity in Mobility Behavior: Latent Classes from Long-Term Tracking Data

Anne-Valérie Preto
TRANSP-OR
EPFL
`anne-valerie.preto@epfl.ch`

Andrea Schneider
Swiss Federal Railways (SBB)
`andrea.schneider@sbb.ch`

Michel Bierlaire
TRANSP-OR
EPFL
`michel.bierlaire@epfl.ch`

April 30, 2026

Abstract

Understanding mobility behavior requires going beyond mode choice and trip counts to capture how individuals structure their daily and weekly activities in time and space. Routine, flexibility, and variability in travel patterns influence public transport demand and peak congestion. We address this challenge by exploiting long-term passive tracking data from the Continuous Mobility Panel dataset, comprising approximately 2,000 individuals observed continuously between 2023 and 2025. However, tracking data at daily resolution are noisy: without careful modeling, we risk estimating noise rather than structure. To extract meaningful patterns, we use a hierarchical latent class model with individual random effects that explicitly separates between-class structure, persistent individual heterogeneity, and day-to-day variability.

The latent classes reveal distinct and interpretable mobility patterns, including routine schedule-constrained commuters, high-mobility users and hybrid workers combining regular anchors with flexible schedules. Rather than relying on a priori segmentation, class membership is inferred from observed mobility behavior. This framework allows the introduction of structural equations linking class membership probabilities to socio-economic characteristics.

The results provide insights for public transport planning. Class-specific demand profiles identify which user groups contribute to peak loads and which exhibit greater temporal or spatial flexibility, and are therefore more likely to adjust departure times or routes in response to crowding. Overall, the proposed latent class approach offers a behavioral alternative to traditional marketing personas, enabling a deeper understanding of mobility routines and their implications for public transport system design and operation.

Keywords

Mobility behavior; Latent class analysis; Passive tracking data;

Contents

List of Tables

List of Figures

1 Introduction

Passive tracking technologies provide more details into individual mobility behavior. We now observe thousands of people continuously over weeks and months, with GPS data capturing daily trips, modes, distances, and timing. However, this wealth of data presents a fundamental challenge: the noise. Each individual's mobility varies significantly from day to day. A person works from home on Monday but commutes to an office on Tuesday.

Public transport planning requires understanding not just who uses the system, but how they use it and understand users' flexibility. Consider two individuals who travel 50 km daily and both use public transport. A long-distance routine commuter leaves home at 6:45 AM, takes the train, arrives by 7:45 AM, and repeats this Monday–Friday; she reliably contributes to peak demand. Another traveller departs at varying times, and make multiple stops throughout the day. She has temporal flexibility. These two individuals appear similar in aggregate statistics (mode share, overall distance) but represent fundamentally different demand profiles. Identifying these structural differences in mobility patterns has direct operational implications for SBB and other transport operators. First, understanding the amount of users' flexibility versus rigid schedules enables targeted pricing strategies: customers with temporal flexibility could be incentivized to shift their trips away from peak hours through pricing. Routine commuters on rigid schedules require adequate capacity during their peak periods. Second, by understanding the profiles that use public transport, SBB can identify which customer segments would be most responsive to specific marketing or policies. This enables more precise customer segmentation than traditional marketing personas, grounded in behavioral structure rather than assumptions.

We propose a hierarchical Bayesian latent class model with individual random effect that tackles this problem. The model recognizes that variation in mobility data occurs at two distinct levels: between-class heterogeneity, meaning distinct behavioral class (low-mobilty, routine commuters by mode and distance, high-mobility users) that truly separate individuals into different mobility lifestyles; and within-class heterogeneity, meaning individual variations within each class, or day-to-day schedule flexibility, personal preferences, occasional deviations from routine. By explicitly modeling both levels, we avoid the false choice between oversimplified clustering and averaging away behavioral reality. Moreover, we use informative priors, domain knowledge from the industry about what we expect from each mobility class, to regularize the model. The model also connects classes to socio-economic characteristics with a logit class membership model. It helps to predict class probabilities as a function of education, age, employment, and mobility ressources. This structure allows behavioral forecasting.

We use continuous passive tracking data from approximately 2,000 individuals in the Continuous Mobility Panel, observed from 2023 to 2025. Our research question is: Can we identify structurally distinct mobility patterns in long-term tracking data using latent classes, and how do these classes relate to socio-economic characteristics? We investigate whether a hierarchical Bayesian approach that simultaneously estimates latent classes and individual-level variation can reveal stable mobility patterns despite substantial day-to-day fluctuations in observed behavior.

2 Literature Review

Travel behavior shows both regularity and day-to-day variation. Schlich and Axhausen (2003) show that while travel patterns are habitual, they are not completely repetitive. Measuring the similarity between multi-day patterns is difficult and depends on the choice of distance measure. The theory for modeling behavioral heterogeneity comes from hybrid choice models (Ben-Akiva *et al.*, 2002). These models integrate discrete choice analysis with latent variables that represent unobserved preferences and behavioral types. This approach allows different sources of heterogeneity to be modeled simultaneously. There is empirical evidence that latent behavioral preferences structure mobility decisions. Vij *et al.* (2013) identify latent modal preferences, which they call modality styles. These preferences persist across different choice situations and are related to life-cycle characteristics. Walker and Li (2007) show that latent lifestyle preferences influence household location choices. By extension, these preferences also determine the mobility patterns. These results show that behavioral heterogeneity reflects stable underlying types, not just random noise.

Clustering methods have been used to identify behavioral mobility patterns from trajectory data. Ben-Gal *et al.* (2019) propose Lifestyle-Based Clustering. They represent individual mobility using Markov models of semantic location transitions. Users are clustered based on Jensen-Shannon distances between their Markov models. This approach successfully identifies user groups from large datasets. However, this two-step approach has limitations. The distance measure is chosen separately from the behavioral model. Cluster membership is deterministic. There is no probabilistic framework for forecasting or policy analysis. A better approach integrates segmentation with probabilistic choice modeling. Tsoleridis *et al.* (2025) show that latent class models are superior when welfare analysis or forecasting is needed. These models assign individuals probabilistically to classes and estimate class-specific parameters simultaneously.

Methodologically, the literature establishes latent classes or clusters as a sound foundation to investigate travel behavior heterogeneity. However it lacks a probabilistic solution to separate structure from noise in long-term tracking data. Many existing studies use two-step or distance-based clustering: they extract features or distances, then assign users to deterministic clusters. This approach does not distinguish between individual-level variation within clusters or noise in the data. In this paper, we use a hierarchical latent class model that integrates measurement equations for daily mobility indicators, class membership probabilities, and individual-level random effects. We try to solve the common confusion between noise and structure in long-term tracking data, producing interpretable, robust classes and actionable insights for operators such as SBB.

3 Methodology

3.1 Data and Indicators

We use continuous tracking data from the Continuous Mobility Panel, with approximately 2,000 individuals observed up to two years (2023–2025). For each individual n and observation day t, we process daily behavioral indicators: binary indicators (home-only day, private vehicle use, active/public transport use), count indicator (daily trip frequency), and continuous indicators with bounded support (daily distance, time spent out-of-home, first departure time). We also observe socio-economic characteristics at the individual level: education (categorical), age (continuous), car availability (binary), public transport subscription type (categorical), and employment category (categorical).

3.2 Hierarchical Latent Class Model

In order to have a meaningful interpretation of the data, we propose to combine latent classes and individual effects: we use the classes to capture persistent mobility patterns, and the individual effects to capture stable deviations within each pattern. This lets us decompose the variation observed in tracking data into three components: (i) between-class structure, (ii) within-class differences across individuals, and (iii) residual day-to-day noise.

Estimating these components jointly is what stabilizes class identification and avoids mistaking noisy daily fluctuations for structural differences.

We estimate three different layers of parameters from the observed indicators and socio-economic characteristics. First, lifestyle classes (s_n): each individual has a probability of belonging to each of the six classes representing distinct mobility lifestyles. Second, individual random effects (u_n): persistent heterogeneity within each class, capturing how each person deviates from their class average. Third, class-specific parameters (θ_s, Σ_s): the mean behavior and correlation structure in each class. As in standard latent class models, we also add a class membership model by linking it to socio-economic variables.

The joint posterior distribution over all unknowns, for individual n, daily observation t, is:

$$
\begin{aligned}
&p(s_n, u_n, ASC, \beta, \theta, \Sigma \mid X_{nt}, \mathbf{x}_n) \\
&\propto \prod_n \left[\prod_t \underbrace{p(X_{nt} \mid s_n, u_n, \theta_{s_n})}_{\text{measurement model}} \cdot \underbrace{p(u_n \mid s_n, \Sigma_{s_n})}_{\text{random effects}} \cdot \underbrace{p(s_n \mid \mathbf{x}_n, ASC, \beta)}_{\text{class membership model}} \right] \cdot \underbrace{p(ASC, \beta, \theta, \Sigma)}_{\text{priors}}.
\end{aligned} \tag{1}
$$

Class membership is modeled as a function of observed socio-economic characteristics $\mathbf{x}_n$ (education, age, car availability, transit subscription, employment type) through a logit specification. The systematic utility of class k is:

$$
V_{n,k} = \begin{cases} 0 & \text{if } k = 0 \text{ (baseline: low-mobility)} \\ \text{ASC}_k + \beta_k^\top \mathbf{x}_n & \text{otherwise} \end{cases} \tag{2}
$$

This framework allows us to predict how class composition changes under different socio-economic scenarios.

Given class s_n and individual effects u_n, the observed indicators follow class-specific distributions. The latent class is identified through the joint distribution across all indicators. Table 1 summarises the indicator types, their distributions and whether a random effect is included.

Binary indicators have no individual random effects, the count indicator includes a

Figure 1: Hierarchical latent class model with individual effects.

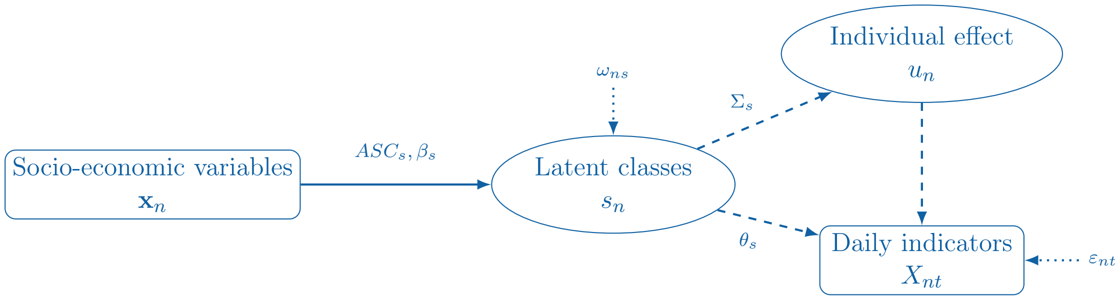

Table 1: Indicator types, distributions, and random effect specifications.

Indicator	**Type**	**Distribution**	**Random effect**
Home-only day	Binary	Bernoulli($p^{H}_{s_n}$)	–
Private vehicle use	Binary	Bernoulli($p^{V}_{s_n}$)	–
Active/PT use	Binary	Bernoulli($p^{PT}_{s_n}$)	–
Daily trip rate	Count	Poisson($\lambda_{s_n} \cdot \exp(u_{n,\text{trips}})$)	multiplicative
Daily distance	Continuous	Trunc. Normal($\mu_{s_n} + u_{n,\text{dist}}, \sigma^2_{s_n}$)	additive

multiplicative (positive) random effect, and continuous indicators include an additive random effect as shown in Table 1.

$$x^{\text{continuous}}_{nt} \sim \text{TruncatedNormal}(\theta^{\text{continuous}}_{s_n} + u_{n,\text{continuous}}, \sigma^{\text{continuous}}_{s_n}, \text{bounds}) \tag{3}$$

We specify informative, class-specific priors on θ_s reflecting domain knowledge about mobility types. These priors act as regularization for noisy data, preventing noisy fragmentation into artificial clusters. For the multinomial logit: $\text{ASC}_k \sim \mathcal{N}(0,1)$ and $\beta_k \sim \mathcal{N}(0,1)$. Covariance structures use LKJ priors (concentration 2) and standard deviations use HalfNormal priors.

Missing data is handled via masking: unobserved indicators do not contribute to the likelihood for that day, but observed data for each individual still inform class and effects estimation.

4 Results

The results presented below are still exploratory: they illustrate what the proposed measurement framework can reveal in principle, and they provide first indications that are coherent. Several estimates remain numerically fragile, so interpretations are cautious and framed as hypotheses to be tested in follow-up work.

4.1 Why the simple single-layer model struggles

Estimating latent classes directly from daily indicators (without individual random effects) gives clear hints of classes, but the measurement noise is large. Empirically, for indicators such as daily distance the within-class variance is of the same order as the between-class variance (50-50), and for some indicators (e.g., departure time) within-class variability dominates. As a result, although class-specific distributions appear different they are not distinguishable. The class distributions overlap, and the noise is too large. Assigning individuals to classes from daily observations is difficult and class membership probabilities are noisy.

To address this, the hierarchical model introduces persistent individual random effects that capture stable deviations from class means. Conceptually, this separates (i) between-class structure, (ii) within-class persistent heterogeneity, and (iii) residual day-to-day noise, which stabilizes class identification and reduces misattribution to separate classes (see Section 3.2 for the full specification).

4.2 Main findings (single-layer indications and class profiles)

Initial results estimated class-specific parameters using maximum likelihood without individual random effects. Table 2 presents these baseline results. Parameters are indexed by class: 0 (low-mobility), 1 (short-distance car), 2 (long-distance car), 3 (short-distance multimodal), 4 (long-distance PT), 5 (high-mobility). This single-layer structure reveals which classes can be marginally separated by the daily indicators alone. However, the overlapping distributions prevents reliable class assignment before adding the within-class variance structure.

Table 2: Maximum likelihood estimates of class-specific parameters in single-layer latent class model.

Parameter	**Estimate**	**Std. Err.**	***t*-stat**	***p*-value**
Home-only$_0$	0.0417	0.0203	2.05	0.0401
Home-only$_1$	0.111	0.00824	13.5	< 0.001
Home-only$_2$	0.00172	0.000757	2.27	0.0229
Home-only$_3$	0.001	0.0118	0.0851	0.932
Home-only$_4$	0.001	0.0112	0.0895	0.929
Home-only$_5$	0.001	0.0144	0.0694	0.945
Private vehicle$_0$	0.001	—	—	—
Private vehicle$_1$	0.995	0.00204	488	< 0.001
Private vehicle$_2$	0.983	0.00422	233	< 0.001
Private vehicle$_3$	0.001	0.0118	0.0851	0.932
Private vehicle$_4$	0.001	0.0112	0.0896	0.929
Private vehicle$_5$	0.791	0.0104	76.1	< 0.001
Active+PT$_0$	0.001	—	—	—
Active+PT$_1$	0.001	0.0154	0.0649	0.948
Active+PT$_2$	0.001	0.0144	0.0694	0.945
Active+PT$_3$	0.998	0.00262	381	< 0.001
Active+PT$_4$	0.999	0.00744	134	< 0.001
Active+PT$_5$	0.164	0.0114	14.4	< 0.001
Trip rate$_0$	0.100	0.00802	12.5	< 0.001
Trip rate$_1$	4.41	0.0688	64.1	< 0.001
Trip rate$_2$	7.47	0.0400	187	< 0.001
Trip rate$_3$	7.09	0.0722	98.2	< 0.001
Trip rate$_4$	10.1	0.0521	194	< 0.001
Trip rate$_5$	10.8	0.0115	939	< 0.001
Daily distance$_0$	-2.79×10^{-9}	1.52×10^{-7}	-0.0184	0.985
Daily distance$_1$	5.08	0.0447	114	< 0.001
Daily distance$_2$	41.7	0.00238	1.75×10^4	< 0.001
Daily distance$_3$	15.7	0.00418	3.74×10^3	< 0.001
Daily distance$_4$	82.2	0.000992	8.28×10^4	< 0.001
Daily distance$_5$	175	0.000104	1.68×10^6	< 0.001
Daily distance (SD)$_0$	1.4×10^{-8}	1.8×10^{-8}	0.78	0.435
Daily distance (SD)$_1$	3.48	0.0623	55.9	< 0.001
Daily distance (SD)$_2$	29.2	0.00118	2.47×10^4	< 0.001
Daily distance (SD)$_3$	8.78	0.00552	1.59×10^3	< 0.001
Daily distance (SD)$_4$	51.0	0.000631	8.09×10^4	< 0.001
Daily distance (SD)$_5$	109	0.000220	4.96×10^5	< 0.001

The descriptions below show patterns that are already visible in the single-layer estimates (without individual effects). These patterns provide strong indications of distinct lifestyles, but the high measurement noise means that individual class assignments from the single-layer model are uncertain.

- **Low-mobility users (baseline).** Low measured daily distance combined with a relatively high probability of home-only days; reflects limited geographic mobility but not necessarily working-from-home in all cases. This group is the modal, low-distance reference.

- **Short-distance car commuters.** Moderate daily distances and very high private-vehicle probability; frequent short trips suggest local car-based commuting patterns.

- **Long-distance car commuters.** High daily distances and private-vehicle use; these individuals generate the largest between-class means for distance and trips in the single-layer fit.

- **Short-distance multimodal commuters.** Moderate distances with high active+PT shares; mode-choice probabilities indicate substantial non-car modal share despite non-negligible trip counts.

- **Long-distance PT commuters.** Large distances coupled with strong public-transport and active-mode probabilities, suggesting long commutes served by transit.

- **High-mobility professionals / traveling salespeople.** High average distances and trip rates with mixed mode use; this class captures multi-activity itinerant patterns that differ from routine commuters.

4.3 Hierarchical Latent Class Model: Posterior Estimates

Table 3 presents the posterior means and standard deviations for class-specific behavioral parameters from the hierarchical model, which incorporates individual random effects. *Note: These results are preliminary and have not fully converged, primarily due to identifiability challenges in the class membership model.*

The six classes show distinct measurement profiles. Low-mobility travelers show the lowest daily distance (0.39 km) and moderate private vehicle usage (0.71), suggesting limited geographic mobility and balanced mode preference. Car-based commuters are differentiated by distance: short-distance (17.6 km) and long-distance (33.4 km) commuters have high private vehicle adoption (0.71 and 0.78, respectively). Multimodal commuters show strong public transport and active mode usage (0.64 and 0.86) and moderate trip distances (22 and 64 km). High-mobility professionals stand apart with mixed mode use and high average distances (46 km), indicating complex, multi-activity patterns.

A key finding is how the model's priors, informed by expert intuition about mobility patterns, were adjusted by the data. The home-only parameter for low-mobility travelers illustrates an important lesson: our prior (0.59) substantially overestimated the estimated rate. This reflects a common misconception we had: individuals with low travel distance

Table 3: Posterior estimates of class-specific behavioral parameters from hierarchical Bayesian model.

Parameter	**Mean**	**SD**	$\hat{R}$	$\mathbf{ESS_{bulk}}$
Home-only$_0$	0.041	0.014	2.34	5.0
Home-only$_1$	0.024	0.012	1.69	6.0
Home-only$_2$	0.022	0.009	1.56	7.0
Home-only$_3$	0.015	0.006	1.31	14.0
Home-only$_4$	0.006	0.002	1.56	7.0
Home-only$_5$	0.014	0.005	2.08	5.0
Private vehicle$_0$	0.714	0.091	3.26	4.0
Private vehicle$_1$	0.709	0.127	2.87	5.0
Private vehicle$_2$	0.777	0.107	3.03	5.0
Private vehicle$_3$	0.301	0.144	2.35	5.0
Private vehicle$_4$	0.098	0.049	2.60	5.0
Private vehicle$_5$	0.348	0.174	2.56	5.0
Active+PT$_0$	0.223	0.084	3.27	4.0
Active+PT$_1$	0.226	0.116	2.88	5.0
Active+PT$_2$	0.168	0.100	3.12	5.0
Active+PT$_3$	0.637	0.156	2.46	5.0
Active+PT$_4$	0.858	0.057	2.79	5.0
Active+PT$_5$	0.596	0.187	2.57	5.0
Trip rate$_0$	6.257	0.397	1.89	6.0
Trip rate$_1$	7.217	0.801	1.75	6.0
Trip rate$_2$	7.086	0.708	2.17	5.0
Trip rate$_3$	7.982	0.412	1.19	28.0
Trip rate$_4$	8.689	0.298	1.12	31.0
Trip rate$_5$	8.356	0.586	2.03	5.0
Daily distance$_0$	0.390	0.437	1.16	18.0
Daily distance$_1$	17.646	2.862	1.90	6.0
Daily distance$_2$	33.411	10.733	2.14	5.0
Daily distance$_3$	21.998	2.433	1.93	6.0
Daily distance$_4$	63.983	3.029	1.56	7.0
Daily distance$_5$	45.996	13.045	2.81	5.0
Daily distance (SD)$_0$	37.598	26.653	2.38	5.0
Daily distance (SD)$_1$	263.834	211.657	2.76	5.0
Daily distance (SD)$_2$	163.093	183.855	3.26	4.0
Daily distance (SD)$_3$	162.375	186.791	3.17	5.0
Daily distance (SD)$_4$	59.769	29.599	2.89	5.0
Daily distance (SD)$_5$	53.755	29.039	2.80	5.0

are not necessarily working from home, and not moving from home; instead, they may simply live and work locally, with commutes that are short but frequent. The low-mobility travelers class better describes the empirical behavior: these individuals have short daily distances but still make multiple trips.

The model also learned that trip rates were systematically underestimated in most classes. Low-mobility travelers showed a 67% increase from prior to posterior, and multimodal classes increased by 13–14%. In contrast, high-mobility professionals were initially overestimated (prior: 13.3 trips), suggesting that truly high-mobility individuals are rarer than anticipated. Distance priors were consistently high: the model reduced prior distance estimates by 10–40% across classes, with the largest adjustment in low-mobility travelers (40% reduction). This indicates that tracking data capture more localized mobility than category-based heuristics would suggest.

4.4 Socio-Economic characteristics and Class Membership

Table 4 presents the logit coefficients linking socio-economic attributes to class membership, with low-mobility travelers as baseline. These results are preliminary, and we can see that convergence is still an issue, with almost no significant effects. The only meaningful patterns are related to car availability and transit subscription, which are strong predictors of class membership.

Table 4: Multinomial logit coefficients: effects of socio-economic covariates on class membership (HW is baseline). Shown are posterior means with 95% credible intervals.

Class	**Covariate**	**Mean**	**SD**	**95% CI Lower**	**95% CI Upper**
Short-dist car	Education	0.030	0.293	–0.544	0.604
	Age	–0.221	0.225	–0.663	0.220
	Car availability	0.766	0.392	–0.002	1.535
	Transit subscription	0.854	0.359	0.151	1.557
	Job: other	0.078	0.547	–0.995	1.152
	Job: retired	0.867	0.514	–0.141	1.874
	Job: student	–0.603	0.559	–1.699	0.492
Long-dist car	Education	0.134	0.211	–0.279	0.546
	Age	0.028	0.239	–0.440	0.496
	Car availability	1.557	0.464	0.647	2.466
	Transit subscription	–0.854	0.364	–1.568	–0.140
	Job: other	–0.499	0.559	–1.595	0.598
	Job: retired	–0.586	0.597	–1.755	0.583
	Job: student	0.081	0.658	–1.208	1.371
Short-dist modal	Education	–0.194	0.270	–0.723	0.335
	Age	0.095	0.249	–0.393	0.583
	Car availability	–0.978	0.388	–1.738	–0.217
	Transit subscription	0.523	0.504	–0.466	1.512
	Job: other	0.033	0.538	–1.022	1.088
	Job: retired	–0.350	0.690	–1.702	1.003
	Job: student	0.476	0.614	–0.727	1.679
Long-dist PT	Education	–0.170	0.268	–0.695	0.355
	Age	–0.127	0.285	–0.686	0.431
	Car availability	–0.951	0.445	–1.823	–0.079
	Transit subscription	0.351	0.607	–0.838	1.541
	Job: other	–0.282	0.669	–1.593	1.028
	Job: retired	–0.382	0.781	–1.912	1.149
	Job: student	–0.062	0.710	–1.453	1.329
High-mobility	Education	0.020	0.466	–0.895	0.934
	Age	–0.274	0.324	–0.909	0.361
	Car availability	0.658	0.573	–0.465	1.780
	Transit subscription	–0.108	0.573	–1.232	1.016
	Job: other	–0.897	0.800	–2.465	0.671
	Job: retired	–0.714	0.822	–2.324	0.897
	Job: student	0.044	0.698	–1.325	1.412

5 Conclusions and Future Work

Tracking data are very information-rich, but at a daily resolution they are fundamentally noisy: when handled naively we risk estimating noise rather than structure. Therefore, we need models that explicitly separate structural variation from persistent individual heterogeneity and day-to-day variability to extract meaningful patterns. This analysis has identified six interpretable mobility lifestyles using daily tracking indicators within a hierarchical latent class framework with individual random effects. The classes reveal differences in mobility resources, activity patterns, and behavioral flexibility. However, the current work operates at the daily level, which conceals important structure about how mobility is organized within days.

The next step is to decompose daily patterns into tours; sequences of trips. At the tour level, we gain the ability to estimate activity purpose. For instance, a long-duration tour (8+ hours) where the individual stays at the same point of interest likely represents work or other major activity commitment, whereas shorter tours reflect errands or social activities. By identifying probable working tours, we can better understand commuting patterns, which account for a large portion of overall mobility budget. Building a tour-level latent class structure will allow several improvements: first, we can validate and refine the current daily-level classes by examining their tour composition; second, we can infer activity purposes (work, social, shopping, etc.) from tour characteristics without requiring explicit survey data; third, we can model how individuals' activity patterns, not just their mobility resources, drive class membership.

In parallel, we plan to incorporate additional contextual information like the workplace locations (inferred from repeated long-duration visits). In the class membership model, we will explore household-level covariates (e.g., household size, presence of children). This extension will improve our understanding of how mobility lifestyles emerge from the data.

6 References

Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Boersch-Supan, D. Brownstone, D. S. Bunch, A. Daly, A. D. Palma, D. Gopinath,

A. Karlstrom and M. A. Munizaga (2002) Hybrid choice models: Progress and challenges, *Marketing Letters*, **13** (3) 163–175.

Ben-Gal, I., S. Weinstock, G. Singer and N. Bambos (2019) Clustering users by their mobility behavioral patterns, *ACM Transactions on Knowledge Discovery from Data*, **13** (4) 45.

Schlich, R. and K. W. Axhausen (2003) Habitual travel behaviour: Evidence from a six-week travel diary, *Transportation*, **30** (1) 13–36.

Tsoleridis, P., C. F. Choudhury and S. Hess (2025) Using probabilistic clustering techniques as a specification tool for capturing heterogeneity in choice models, *Transportation Research Part C: Emerging Technologies*, **179**, 105289.

Vij, A., A. Carrel and J. L. Walker (2013) Incorporating the influence of latent modal preferences on travel mode choice behavior, *Transportation Research Part A: Policy and Practice*, **54**, 164–178.

Walker, J. L. and J. Li (2007) Latent lifestyle preferences and household location decisions, *Journal of Geographical Systems*, **9** (1) 77–101.