Generating activity-specific multimodal trips from aggregated data sources

Ariane Scheffer[1], Richard D. Connors[1], Haruko Nakao[1], Francesco Viti[1]

[1] Mobilab Transport Research Group University of Luxembourg, Luxembourg

SHORT SUMMARY

Mode choice is the result of selecting the most convenient mean(s) of transport among all options available to reach locations where daily activities are performed. Trips made by individuals are enchained, making mode choice an interconnected process, especially considering the spatial connections and temporal constraints to reach the different activity locations. This paper proposes a methodology to estimate ativity and mode-specific dynamic OD trips, based on the economic and behavioural principle of utility maximization. The proposed modelling approach aims at preserving the link between individual trips, while limiting combinatorial explosion of choices when considering individual trip chains. This approach is applied at a zonal scale and relies on observed aggregated OD trips and travel times by mode as main input data. The model parameters are estimated through Markov Chain Monte Carlo sampling and Bayesian updating techniques, hence, instead of calculating only expected values, each parameter is described by a probability distribution, which can represent and incorporate in the model the heterogeneity of population activity-travel behaviour. This model is tested on multi-day travel diary data collected in the city of Ghent, Belgium, containing around 15000 trips classified by four main activity types and four alternative modes. We show that the proposed model can properly estimate and capture the dynamics of modal split and relate these choices to activity participation.

Keywords: Dynamic demand estimation; utility maximization; Markov Chain Monte Carlo; activity-specific OD matrices.

1. INTRODUCTION

The estimation and prediction of mode-specific trips is essential to understand the potential effectiveness of policies aiming to achieve desirable sustainability targets, or to manage the demand for resource-limited systems such as transit or shared mobility services. Mode choice is a complex process that involves different determinants at various levels (e.g., mode convenience, quality of service, personal preference, and socio-demographic characteristics) (Tyrinopoulos and Antoniou, 2013). Moreover, the mode chosen for an earlier trip has an influence on subsequent decisions, for example because of the consequent (un)availability of transport alternatives at the origin or destination of trips.

Trip chaining and activity scheduling are main determinants of mode choice, with more complex tours favouring car use (Currie and Delbosc, 2011), while representing a barrier for public transport (Hensher and Reyes, 2000). Tour-based models seem to be the most suited approaches to handle correlations in sequential activity and mode choices. Tour-based models usually rely on activity-based modelling approaches, where mode choice usually follows an activity scheduling model (Miller et al., 2005). However, considering all feasible mode combinations in tour-based models is extremely complex (Vovsha et al., 2017). Besides, application of these models also requires very large and detailed data, which are often not available. On the contrary, traditional trip assignment-based models are easier to calibrate and continue to be widely adopted for estimating and forecasting the travel demand (McNally 2007). However, they are considered inadequate for incorporating daily activity scheduling and departure time choices, and they usually provide a too coarse representation of the spatial and temporal dynamics of travel demand (Cantelmo et al., 2018).

Combining trip-based and activity-based model features allows calibration and interpretability strengths to be preserved, while offering the opportunity to capture the complexity and heterogeneity of individuals' decision-making processes. Lam and Yin (2001) integrated activity-based models with time-dependent traffic assignment in a variational inequality dynamic user equilibrium model. To jointly model activity-travel decisions, different choice levels are represented as path choice through a space-time expanded network, offering a basis for joint activity-travel assignment (Liao, et al., 2013; Liu et al., 2016; Fu and Lam 2018). Dynamic Activity Travel Assignment models are arguably the most advanced methods for capturing multiple choice dimensions, including mode choice. However, because they use a very complex modelling approach, especially in terms of network representation, their computational and modelling complexity can grow extremely rapidly.

In this work we model mode-specific trips from activity-travel choices of a heterogeneous population and calibrate them against aggregated traffic data using an iterative Bayesian estimation approach. The inputs needed are aggregated time-dependent total daily trips, average travel times by origin and destination and by mode, and total activity participation, i.e. expected number of individuals performing an activity on an ordinary day. These data are used to calibrate the parameters of activity-specific marginal utility functions, coined utility primitives, which are derived from utility maximization principles (Cantelmo and Viti, 2019).

The advantages of this approach are threefold:

- First, we estimate aggregated mode-specific trips that are consistent with individual mode choices and with the observed aggregated demand flows;
- Second, the primal output can be used as input for applications such as dynamic Origin-Destination (OD) demand estimation and for transport simulation models;
- Finally, thanks to the adoption of underlying functions representing the marginal utility gained by individuals, we can relate the mode choice process to complex decisions such as when to schedule activities and the related departure and arrival times.

The paper is structured as follows. Section 2 describes the methodological approach. Section 3 describes the data used and the result of the estimation process. Section 4 concludes the paper and presents future research directions.

2. METHODOLOGY

The core of the model in this work relies on utility maximization principles and the concept of *utility primitives*, which represent the set of activity-specific marginal utility functions driving the utility maximization process, at an aggregate (e.g., zonal) level.

The main idea is to apply the utility for an individual travelling to and performing an activity in a certain zone in an aggregated way, i.e. we model the overall utility accumulated by all individual performing the said activity in a specific zone or region. Assuming individuals to be utility maximisers, they will seek to optimise their travel choices to accumulate the utility of performing the daily activities while reducing the costs of travelling to the locations where the activities are to be performed. At an aggregated level, this phenomenon can be captured by departure time probabilities, resulting in emergent trip rates, which can in this way be activity and mode-specific.

This idea allows revealed or observable preferences, such as departure time and mode choices, to be formally linked with underlying unobservable characteristics such as (marginal) utility gains/losses when performing the activity. The functional relationship representing the marginal utility accumulated in time by an individual performing a specific type of activity at a certain location is our definition of utility primitive, which is characterised by a number of calibration parameters to be estimated to define the form of the marginal utility function.

In the following we briefly describe this modelling and estimation process, leaving the details to Scheffer et al. (2024). We consider the utility accumulated by an individual travelling from any zone y to perform specific activity type a into zone z to be simplified into four components: 1) the expected utility of any activity that was being performed at y before departing towards zone z; 2) the disutility of travelling from zone y to zone z to perform activity a; 3) the utility accumulated while performing activity a in zone z; and 4) the utility of performing any other activity after stopping to perform activity a. Hence, the (daily) utility $\widehat{U}_a^i$ accrued by an individual i engaging in activity a is formulated as:

$$\begin{aligned}\widehat{U}_a^i\big(t_d, t_e, m|(y,z)\big) \\ &= \sum_{t=t_0}^{t_d-\Delta t} u_{a-}^i(t,t_0)\,\Delta t \quad + \quad u_t tt_{\to a}^{yz,m} \quad + \quad \sum_{t=t_s}^{t_e} u_a^i(t,t_s)\,\Delta t \\ &+ \sum_{t=t_e+\Delta t}^{t_N} u_{a+}^i(t,t_e+\Delta t)\,\Delta t\end{aligned} \tag{1}$$

Where t_s and t_e are the starting and ending time of activity a, respectively; u_a^i is the marginal utility function related to activity a for individual i, Δt is the discrete unit of time and $T = \{t_0, t_0+\Delta t, t_0+2\Delta t, \dots, t_0+(N-1)\Delta t\}$ the total evaluation period. Finally, $tt_{\to a}^{yz,m}$ represents the travel time (cost) from origin y to destination z when departing at time t_d using mode m. Simply stated, when estimating the utility primitive specific to activity a, we must also consider the marginal utility of any/all activities performed before a (denoted by $a-$), the trip to access the activity (denoted by $\to a$), the marginal utility specific to a, and the marginal utility of any/all activities performed after a (denoted by $a+$) in order to capture the tradeoff between performing activity a as opposed to perform any other activity during a certain time period. This is exemplified in Figure 1, where the total accumulated utility (shaded areas) given a set of departure, starting and ending times is depicted. In the figure, the loss of utility due to travelling is the time between the red and the blue vertical lines.

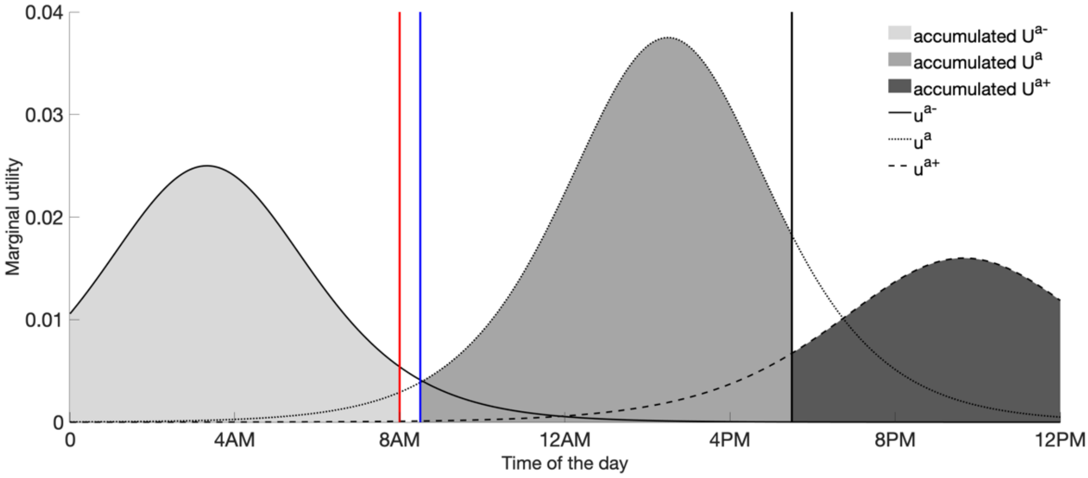

Figure 1 Accumulated utility for a sequence of activities

While equation (1) is written for the i-th individual, it will be applied and estimated in our model for a generic individual, hence the superscript i will no longer appear, i.e. we will use the terms $\widehat{U}_a$ and u_a from now on.

We then adopt a time-dependent functional form for the marginal utility that is flexible enough to represent multiple activity types. The form chosen for such marginal utility formulation, which is able to represent any activity type bringing positive utility, is taken from the work of Ettema and Timmermans (2003):

$$u_a(t, t_s) = \frac{\gamma_a \beta_a (U_a^{max})}{exp[\beta_a(t - (\alpha_a + t_s \tau_a))].(1 + exp[-\beta_a(t - (\alpha_a + t_s \tau_a))])^{\gamma_a + 1}} \quad (2)$$

where U^{max} represents the maximal marginal utility for a certain activity; α is a parameter related to the location of the saturation point if $(\gamma = 1, \tau = 1)$ and β determines the dispersion around the saturation point; γ controls the symmetry of the functional form; τ controls whether the saturation is reached at a fixed time of day or is relative to activity duration. When $\tau = 0$, the utility is determined by time of day, regardless of activity start time. This represents activity types such as work. $\tau = 1$ describes a duration-based utility function. In this case α describes an optimal duration instead of a time of the day. Therefore, the 5 parameters allow to have the required flexibility in specifying functional forms representing different activity types (work, shopping, leisure, etc.).

We consider marginal utilities to be zone independent, i.e. the same utility can be accumulated by performing the same type of activity in different zones. For this reason, the mode and destination choices are dependent only on the disutility of travelling, in particular on the expected travel time when selecting a mode and a zone of destination.

The joint mode-departure time choice for each pair of zones (y,z) is expressed as a discrete probabilistic choice process using a simple multinomial logit model. The probability of choosing the pair of activity starting and ending times (t_s, t_e) and mode m is computed as follows:

$$P_a(t_s, t_e, m|(y, z)) = \frac{\exp\left(U(t_s, t_e, m) + c_t tt_a^{yz,m}\right)}{\sum_{m'} \sum_{t'_e > t_{s'}} \sum_{t_{s'}} \exp\left(U_a^i(t'_s, t'_e, m') + c_t tt_a^{yz,m}\right)} \quad (3)$$

with t'_d and t'_e being any feasible departure and activity end time. The denominator includes only those zones where activity a is available (which may include the current zone of departure y).

The estimation process used for calibrating the marginal utility function (2) parameters is a Markov Chain Monte Carlo (MCMC) sampling and an iterative Bayesian updating process. The inputs of the MCMC are the generated total demand by time of the day and the total travelled time distribution by mode. For every parameter to be calibrated, a starting value is selected. At each iteration, the complete set of parameter values is assessed based on a score, which is composed of the likelihood of data generated by the model on one side and on the plausibility of parameter values on the other side. A conventional log-likelihood function is used to compare observed and estimated data. For a more detailed description one can refer to Scheffer et al (2024).

3. CASE STUDY AND RESULTS

To test the methodology, we used individual trips data obtained from a multiday travel survey collected in the province of Ghent in 2008, including multiple respondents, days, and tour types. A total of 15397 trips from 707 individuals is used in this application, for which we know the origin, destination, starting time, ending time, travel time, mode(s) used, activity at origin and activity at destination. The survey includes 12 activity types, regrouped in the following categories: 1) Home, 2) Work, 3) Shopping & other mandatory activities and 4) Leisure & other secondary activities.

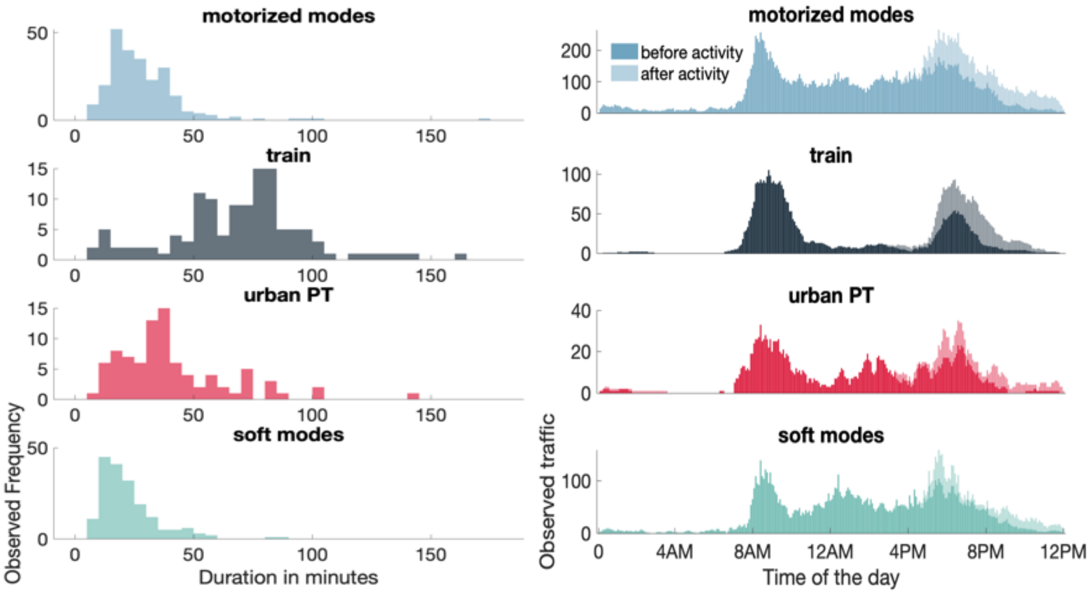

Figure 2: Travel time frequency (a) and observed arrival and departure times for work trips (b) by time of the day (5 minutes interval)

Only trips to and from work using different modes of transport are considered for this analysis. Travel modes are grouped into the following categories: motorized modes, train, urban public transport, and soft modes (Figure 2). Figure 2a shows the travel time distribution per mode, whereas Figure 2b presents the number of observed trips before and after an activity has been performed by mode and by time of the day. Time resolution of 5 mins intervals is used for both time-dependent demand and travel times. A full description of the database and analysis of the variability of daily activity-travel pattern is available in Raux, et al. (2016).

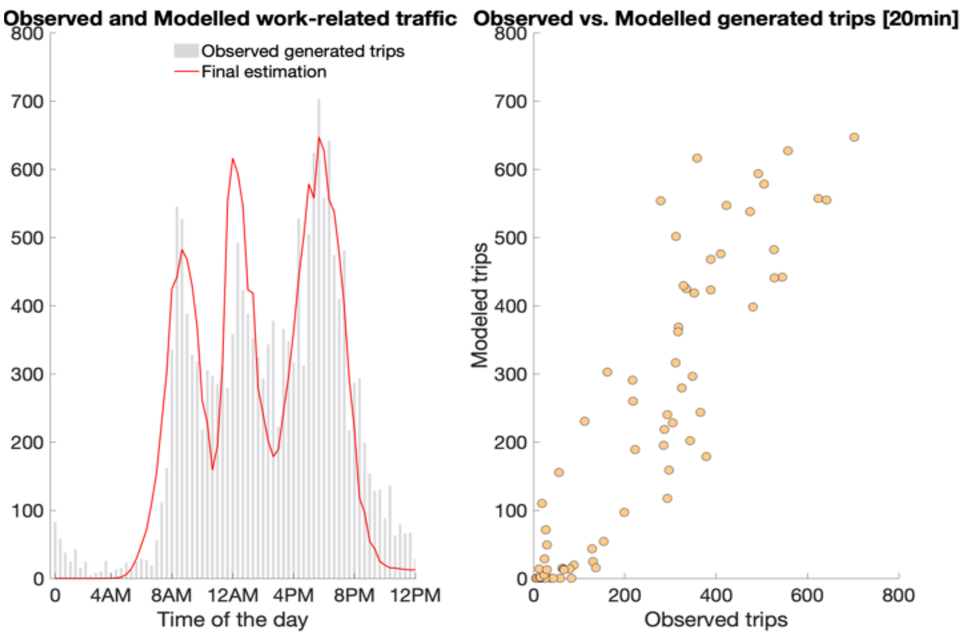

Figure 3: Estimated vs. observed trips for each time of the day

To estimate realistic travel times by mode with respect to the observed traffic, we calculate a truncated average of observed travel times using survey answers for each time period. Despite the large number of observations, there was not much data for every mode in every time interval. Missing data was added via linear interpolation of neighbouring data.

The estimation procedure ran for 15000 iterations. The resulting final demand profile is rather accurate with respect to observed generated demand, with a satisfactory $r^2 = 0.82$. The main result of the estimation is the posterior distribution of estimated parameters. Figure 3 shows the final estimation for each time of the day for all trips regardless of the mode of transport.

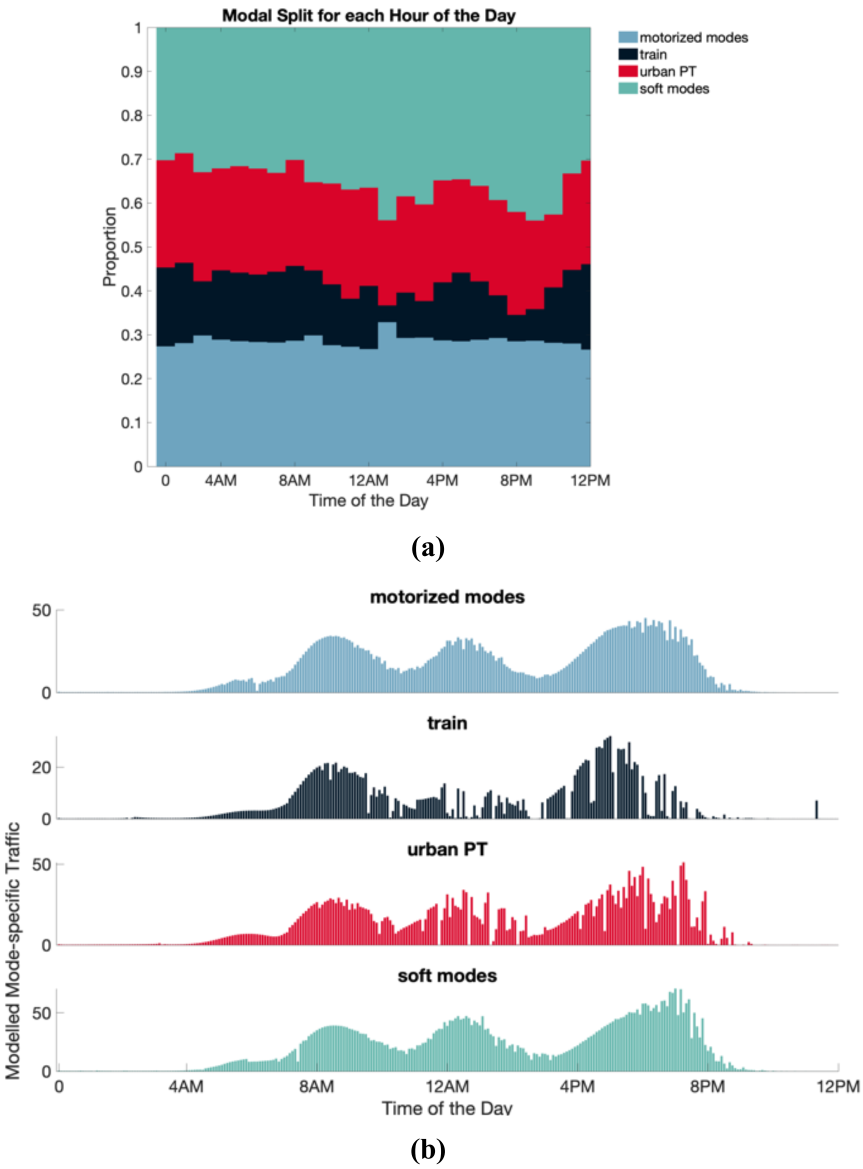

Figure 3: Estimated modal split (a) and mode specific demand profiles (b)

Including mode specific travel times in the estimation process results in a dynamic modal split (Figure 4a), which shows the ability of the model to produce such output even without resorting to advanced cost functions but only based on the positive component of the accumulated utility and a probability distribution of travelled distances. The estimation was done for a 5-minute interval, but the modal split results are shown for a 20-minute time interval. This avoids skewing the output with missing or outlying data, in terms of modal speed for example. The mode-specific departure time profiles (Figure 4b) indicate a good representation of large-scale temporal dynamics. For example, around mid-day, the peak is more visible for soft modes and car users and almost no train users appear. The estimated work-related trips can be qualitatively compared Figure 2b which represent the real values for all kind of trip purposes. The peak of soft modes around 12AM can be explained by shorter trips associated to the secondary (e.g., shopping or eating) activities. In general, there is an overrepresentation of modes for which trip duration is typically short, such as soft modes.

4. CONCLUSIONS

In this work we proposed a novel approach to estimate utility parameters of activity-specific marginal utilities and mode-specific time-dependent trips. The modelling approach give the possibility to combine many information sources relating to diverse aspects of travel decisions. In this paper we showed how the approach can be used to reproduce the dynamics of mode choices using aggregated trips and travel time distributions as input. The application of the proposed methodology shows a relatively good estimation, especially given the low input data requirements.

A better overall estimation is expected to be reached by combining the proposed method with speed variations by mode and time of the day and apply more detailed travel cost functions including e.g. monetary costs. Moreover, combining the demand model with traffic assignment and simulation may allow to include congestion effects, which is an approach that will be developed in future research.

ACKNOWLEDGEMENTS

This work is financially supported by the EU-FEDER project grant ODIN (n. 2023-1-8) and the EU Horizon Europe ACUMEN (n. 101103808)

REFERENCES

Antoniou C., Barceló J., Breen M., Bullejos M., Casas J., Cipriani E., Ciuffo B., et al. (2016). Towards a Generic Benchmarking Platform for Origin–Destination Flows Estimation/Updating Algorithms: Design, Demonstration and Validation. *Transportation Research Part C: Emerging Technologies*, 66: 79–98.

Cantelmo G., Viti F. (2019). Incorporating Activity Duration and Scheduling Utility into Equilibrium-Based Dynamic Traffic Assignment. *Transportation Research Part B: Methodological* 126: 365–290.

Cantelmo G., Viti F., Cipriani E., Nigro M. (2018). A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration. *Transportation Research Part A: Policy and Practice*, 114: 303–320.

Currie, G., Delbosc A. (2011). Exploring the Trip Chaining Behaviour of Public Transport Users in Melbourne. Transport Policy 18(1): 204–210.

Ettema D., Timmermans H.P.J. (2003). Modeling Departure Time Choice in the Context of Activity Scheduling Behavior. *Transportation Research Record: Journal of the Transportation Research Board* 1831 (January): 39–46. https://doi.org/10.3141/1831-05.

Fu X., Lam W.H.K. (2018). Modelling Joint Activity-Travel Pattern Scheduling Problem in Multi-Modal Transit Networks. *Transportation* 45(1): 23–49.

Hensher, D.A., Reyes A.J., (2000). Trip Chaining as a Barrier to the Propensity to Use Public Transport. *Transportation* 27 (4): 341–361.

Lam, W.H.K., Yin Y. (2001). An Activity-Based Time-Dependent Traffic Assignment Model. *Transportation Research Part B: Methodological* 35(6): 549–574.

Liu P., Liao X., Huang H.J., Timmermans H.P.J. (2016). Dynamic Activity-Travel Assignment in Multi-State Supernetworks under Transport and Location Capacity Constraints. *Transportmetrica A: Transport Science* 12(7): 572–590.

McNally, M.G. (2007). The Four-Step Model. In *Handbook of Transport Modelling*, 1: 35–53.

Miller, E.J., Roorda M.J., Carrasco J.A., (2005). A Tour-Based Model of Travel Mode Choice. *Transportation* 32(4): 399–422.

Raux, C., Ma T.-Y., Cornelis E. (2016). Variability in Daily Activity-Travel Patterns: The Case of a One-Week Travel Diary. *European Transport Research Review* 8 (4): 26.

Scheffer A., Bandiera C., Connors R., Cantelmo G., Viti F. (2024). A multi-scale utility maximisation approach to estimate within-day activity-specific demand flows. Submitted to Transportation Research Part A: Policy and Practice (under 4th round of revision).

Tyrinopoulos Y., Constantinos A. (2013). Factors Affecting Modal Choice in Urban Mobility. European Transport Research Review 5: 27–39.

Vovsha P., Hicks J.E., Vyas G., Livshits V., Jeon K., Anderson R., Giaimo G. (2017). Combinatorial Tour Mode Choice. In TRB 96th Annual Meeting Compendium of Papers.