A random-utility-consistent machine learning method to estimate agents' joint activity scheduling choice from a ubiquitous data set

Xiyuan Ren, Joseph Y. J. Chow*

C2SMART University Transportation Center, New York University Tandon School of Engineering, Brooklyn, USA

xr2006@nyu.edu, joseph.chow@nyu.edu

* Corresponding author

ABSTRACT

We propose an agent-based mixed-logit model (AMXL) that is estimated with inverse optimization (IO) estimation, an agent-level machine learning method theoretically consistent with a utility-maximizing mixed logit model framework. The method provides joint, individual-specific, and deterministic estimation, which overcomes the limitations of discrete choice models (DCMs) given ubiquitous datasets. A case study of the CBD in Shanghai is conducted with mobile phone data of 26,149 anonymous commuters whose whole-day activity schedule on weekdays contains three sub-choices and 1,470 alternatives. AMXL is built to estimate individual tastes and predict the activity scheduling choice in different scenarios. Multinomial logit model (MNL), mixed logit model (MXL), and their dynamic forms (DMNL, DMXL) are built as benchmarks. Prediction accuracies are calculated as the percentage consistency of observed choices and predicted choices, both at individual level (to each commuter) and aggregated level (to each alternative in the choice set). The results show that empirical coefficient distributions in AMXL are neither Gumbel nor Gaussian, i.e. capturing inter-individual heterogeneities in space that are hard for DCMs to capture. The prediction accuracy of AMXL is significantly higher than the best model (DMXL) in benchmarks, improving from 8.66% to 61.68% at aggregated level and from 1.69% to 4.33% at individual level. In a comparison scenario, AMXL predicts different while reasonable change of choices compared with benchmark models. In an optimization scenario, AMXL can be directly integrated into a binary programming (BP) problem, which optimally allocates 10 blocks to send restaurant coupons to increase population consumer surplus by 19%.

Keywords: activity scheduling choice, utility maximization, inverse optimization, big data, machine learning

To be published in Transportation Research Part B

1. Introduction

Joint activity scheduling choice is taking on greater importance in transportation planning, given that travel demand is derived from the need to perform activities (Hägerstrand, 1970; Ben-Akiva and Lerman, 1985; Recker et al., 1986a,b; Kitamura, 1988; Bowman and Ben-Akiva, 2001; Pinjari and Bhat, 2011; Chow and Nurumbetova, 2015). Most existing studies postulate that agents derive a (dis)utility from traveling and performing activities, and they schedule them to maximize the utility in total (Adler & Ben-Akiva, 1979; Ettema et al., 2007; Habib, 2018). With the emergence of information and communications technology (ICT), it is now possible to model and predict activity scheduling choice with quite a large sample size at the city or state level (Aguiléra, 2018; He et al., 2020).

Whereas discrete choice models (DCMs) were traditionally used to analyze choice behavior, their application to activity scheduling behavior under the big data context is hindered by at least three problems. First, since the scheduling of daily activities is a complicated process that covers multiple choice dimensions (e.g., which activities to perform, activity timings, locations, durations, and mode-of-travel between activities), it is hard to maximize the likelihood function for DCMs with a huge choice set comprised of all possible schedule combinations (Charypar and Nagel, 2005; Danalet et al., 2015; Pougala et al., 2021). Second, though flexible DCMs structures enable modelers to capture individual preference, the results are defined by a specific distribution of coefficients, for example, Gaussian-distributed coefficients in a mixed logit model (MXL). Many researchers have already pointed out the risk of failing to choose adequate distributions for random coefficients (Hess, 2010; Sarrias, 2020). Third, DCMs are stochastic estimation approaches, generating demand functions that are non-linear and non-convex in the explanatory variables (Ljubić and Moreno, 2018; Pacheco et al., 2021). Complex expressions restrict their further integration to optimization problems for congestion management, system design, and demand management policies. There have been a growing number of studies using dynamic DCMs (Hasnine and Habib, 2018; Västberg et al., 2020) or machine learning models (Lizana et al., 2021; Tanwanichkul et al., 2019; Wang et al., 2020a) to analyze individual decisions and reporting higher prediction accuracy than conventional DCMs. Although these models allow more complex relationships between explanatory variables and individual choices, it is at the cost of sacrificing interpretability and increasing the non-linearity of demand functions (LeCun et al., 2015; Liao and Poggio, 2018). Therefore, innovative methods dealing with these three limitations are theoretically essential and empirically critical to estimate agents' joint activity scheduling choice.

These limitations can be addressed if coefficients can be specified deterministically for each individual within a DCM framework resulting in a heterogeneous population of coefficients. We call this an agent-based mixed logit model (AMXL). Under most circumstances, an AMXL model does not make sense because estimating it from a sample data set would not be transferable to a population. However, such an approach would still be valid under a ubiquitous data (smartphone data) or synthetic population (e.g. He et al. (2020) or Replica (2022)) setting where attributes from the whole population (or segments of it) can be obtained or monitored instead of just from a sample. Under such settings, an AMXL model can be used to predict the outcome of alternative scenarios for the same population. We assume such a setting, where data for a population is sufficient to estimate the coefficients for each individual resulting in a random utility model with heterogeneous but deterministic coefficients.

The significance of AMXL is as follows. First, difficulties in estimating coefficients for high choice dimensions requiring high-dimensional integrals for random coefficients can be addressed with constrained optimization. Second, population distributions for the coefficients are based on non-parametric aggregation of the individual coefficients instead of having to assume a distributional fit. Third, since each individual's representative utility function is fully specified, AMXL can be directly integrated into system design optimization models as constraints instead of dealing with simulation-based approaches needed for MXL (see Pacheco et al., 2021).

As for the proposed methodology, we first decompose individuals' whole-day activity scheduling choices into a series of inter-related sub-choices, from which a series of utilities are derived accordingly. By ensuring shared coefficients to be the same among these sub-choices, we reduce the choice set while still modeling different choice dimensions jointly. We then formulate a unique inverse optimization (IO) problem for the set of choices of each individual, in which a Gumbel-distributed random utility and a safe boundary are added to each alternative for each sub-choice. This insight keeps the theoretical consistency with DCMs and avoids overfitting the individual-specific estimation. A Method of Self-Regulated Average (MSRA) (Liu et al., 2007) is applied to smooth the iterative convergence and obtain stable solutions, in which demand functions are linearly related to the explanatory variables.

We test the estimated AMXL model against multinomial logit (MNL), mixed logit (MXL), dynamic multinomial logit (DMNL), and dynamic mixed logit (DMXL) to model activity scheduling choice using a mobile-phone-derived dataset, referred to as SHC in this study. The SHC dataset was collected in 2019, containing two weekdays activity information of 26,149 anonymous commuters working in the CBD of Shanghai. Based on the experimental results, AMXL presents three key advantages over benchmark models: (1) For the whole-day activity schedule, AMXL provides higher prediction accuracy both at individual level (to each commuter) and aggregated level (to each alternative in the choice set); (2) AMXL produces an empirical distribution of individual preference in space which is hard for benchmark models to capture; (3) utility functions retrieved from AMXL can be directly integrated into system design optimization models. While the AMXL model is applied to the activity scheduling choice use case in this study, it can also be applied to other high dimensional choice problems where ubiquitous data are available. To facilitate future research, we uploaded the algorithms and the 1,000 randomly-selected samples to a Github repository: https://github.com/xr2006/AMXL.git.

The remainder of the paper is organized as follows. Section 2 reviews studies on DCMs for activity scheduling choice and emerging machine learning approaches for individual choice. Section 3 describes the general framework of the AMXL model, including utility derived from whole-day activity scheduling behavior, inverse optimization algorithm with random utility, and architecture of AMXL. Section 4 sets up a concrete experiment that compares the performance of MNL, MXL, DMNL, DMXL and AMXL. The experimental results, including prediction accuracy, distribution of coefficients, and scenario application, are presented in Section 5. Section 6 concludes the findings and points out future work.

2. Literature review

2.1 DCMs for activity scheduling choice

As econometric models, DCMs assume individuals to schedule activities by maximizing the overall utility they can expect to gain (Becker, 1965; Bowman and Ben-Akiva, 2001). Typically, decisions related to activity scheduling behavior include the choice of activity pattern (staying at home, working, or shopping), the destination, time-of-day, duration for each activity, and mode-of-travel between activities (Ben-Akiva & Lerman, 1985; Miller and Roorda, 2003; Ding et al., 2017; Ettema et al., 2007; Habib, 2018). Skeleton schedules, referring to schedules with fixed attributes pre-defined by modelers such as specific activity types (e.g., commuters' working activity on weekdays), start time (e.g., 9 a.m. for work, 12 a.m. for lunch), duration (e.g., a half-hour exercise), or destination locations (e.g., home and workplace), are widely used to control the model complexity and emphasize the research scopes (Ettema et al., 2007; Habib and Miller, 2006; Habib, 2018).

Existing studies using activity-based DCMs can be divided into two categories. The first category treats choice dimensions with a nested structure, in the sequence from primary activities to secondary activities, from time frames to travel modes, and according to time series (Horni et al., 2016; Bowman and Ben-Akiva, 2001). A basic form is the nested logit model (NL) while a more

advanced one follows a Markov decision process (MDP) and models activity scheduling choices as dynamic DCMs (Aguirregabiria and Mira, 2010; Västberg et al., 2020). Dynamic DCMs assume that individual $i \in P$ acts to maximizes the utility function defined by Eq. (1).

$$U_{ijt} = x_{ijt}\beta_{jt} + \varepsilon_{ijt} + \mu_t EV(i,j,t)\,, \qquad \forall i \in P, \forall j \in J, \forall t \in T \tag{1}$$

where t denotes the choice situation or time period. x_{ijt} denotes a set of observed variables of individual i choosing alternative j in situation t. β_{jt} is a set of coefficients reflecting preferences. $x_{ijt}\beta_{jt}$ and ε_{ijt} denotes the deterministic and random utility, which is aligned with conventional DCMs. $EV(i,j,t)$ is the expected utility of all possible alternatives in the remainder of the day, usually calculated using multi-dimensional integrals or backward induction with a relatively high computational cost (Västberg et al., 2020). μ_t is a coefficient defining the weight of expected utility in choice situation t. Accordingly, the probability of individual i choosing alternative j in situation t is defined as Eq. (2).

$$P_{ijt} = \frac{e^{x_{ijt}\beta_{jt} + \varepsilon_{ijt} + \mu EV(i,j,t)}}{\sum_{j' \in J} e^{x_{ij't}\beta_{jt} + \varepsilon_{ij't} + \mu EV(i,j',t)}}\,, \quad \forall i \in P, \forall j \in J, \forall t \in T \tag{2}$$

The second category focus on stochastic individual level models, considering that preference may vary across different choice situations of different individuals. Up to this point, logit mixtures incorporating inter- and intra-individual heterogeneity are estimated with maximum likelihood procedure (Becker et al., 2018; Krueger et al., 2021). For example, a mixed logit model (MXL) assumes that each individual i faces a choice among J alternatives. Then, the utility associated with each alternative $j = 1, \ldots, J$ for individual i is defined as Eq. (3).

$$U_{ij} = x_{ij}\beta + \varepsilon_{ij}\,, \qquad \forall i \in P, \qquad \forall j \in J \tag{3}$$

where x_{ij} denotes a set of observed variables of individual i choosing alternative j. ε_{ij} is the random utility. The vector of tastes β is assumed to be a variate that varies across individuals according to $g(\beta|\Omega)$, where $g(.)$ is usually the Gaussian distribution with the mean and covariance included in Ω. More recent studies have captured both inter- and intra-individual tastes based on conditional estimations (Becker et al., 2018; Krueger et al., 2021; Sarrias, 2020). Accordingly, the probability of individual i choosing alternative j is defined as Eq. (4).

$$P_{ij} = \int \frac{e^{x_{ij}\beta}}{\sum_{j' \in J} e^{x_{ij'}\beta}} g(\beta|\Omega) d\beta\,, \quad \forall i \in P, \qquad \forall j \in J \tag{4}$$

Despite a growing number of empirical studies, at least three issues are not perfectly addressed, which hinder the application of DCMs to activity scheduling behavior under a Big Data context.

First, most of the choice scenarios in DCMs involve a small number of alternatives. However, the choice set in the case of activity scheduling behavior will be quite large if we include all possible schedule combinations (Pougala et al., 2021). For example, a single activity starts in t possible time blocks, at l possible locations, and with m available mode-of-travel will generate $(t \times l \times m)$ possible alternatives. Existing computational tools (e.g., MATLAB, R, Status) are generally unable to maximize a likelihood function associated with hundreds of coefficients and alternatives (Chen et al., 2005; Lemp and Kockelman, 2012). Though dynamic DCMs decompose the whole-day

scheduling choice into a sequence of sub-choices and calibrating them jointly, it is hard to calculate multi-dimensional integrals especially when alternatives vary across choice situations. Also, conditional probability can only capture general dependencies among sub-choices. It remains understudied whether a conditional probability is valid from an individual perspective.

Second, just knowing that a coefficient varies across individuals is not enough in the case of activity scheduling choices, which combines segments of individuals (Richter and Pollitt, 2018), disaggregated willingness-to-pay (WTP) (Dumont et al., 2015), spatial dependence of tastes (Budziński et al., 2018), and individual-specific strategies (Hess and Hensher, 2010). Logit mixtures incorporating inter- and intra-individual heterogeneity might be a good attempt to estimate individual level models. However, the complex assumption of conditional distribution also has limitations. Hess (2010) has argued that failing to choose adequate distributions for random coefficients might lead to misleading conclusions. Sarrias (2020) has stated that the better fit in terms of signs and values might be an artifact of the statistical behavior of the conditional estimations when the number of choice scenarios is not large enough. Under the context of location-based Big Data, attributes from the whole population can be obtained instead of just from a sample (Ahas et al., 2009), and the individual tastes might not be normally distributed due to lacking personal information (Zhao et al., 2018). To this end, modelers should consider individual-specific estimations without complex assumptions of the conditional distribution.

Third, stochastic estimations in DCMs result in non-linear or even non-convex mathematical formulations, particularly for more advanced variations like MXL or dynamic DCMs, which are difficult to embed in optimization models governing the supply-related decisions (Ljubić and Moreno, 2018; Pacheco et al., 2021; Robenek et al., 2018). For instance, mixed-integer linear programming (MILP) models have been widely used in congestion management (Qiu and Wang, 2015), transit timetabling (Cordone and Redaelli, 2011), toll setting (Gilbert et al., 2015), and vehicle routing (Kancharla and Ramadurai, 2020; Dong et al., 2022), among others. Unfortunately, MILP models require linearity and convexity of the involved functions, which is generally not the case in DCMs. Whereas Pacheco et al. (2021) have presented the feasibility of integrating mixed logit models into MILP models via a simulation-based linearization approach, longer computational time compared with conventional MILP still hinders the interaction between individual choices (demand) and operational strategies (supply). Given a ubiquitous dataset, the key idea is to develop a deterministic approach that overcomes the stochastic nature of the random component and thus expresses the demand in terms of linear, convex functions.

2.2 General-purpose machine learning methods for individual choice

In recent years, there has been an emerging trend of using general-purpose machine learning models (MLs) to analyze individual choices (Wang et al., 2020b). In the transportation field, existing studies have applied support-vector machines (SVMs), classification trees (CTs), random forests (RFs), and deep neural networks (DNNs) to analyze many choice scenarios such as automobile ownership, travel mode, vehicle route, and parking location (Lizana et al., 2021; Shaaban and Pande, 2016; Tanwanichkul et al., 2019; Tribby et al., 2017; Chow, 2018; Ma et al., 2017). General-purpose MLs for individual choice have both pros and cons. The pros are that these models allow flexible relationships between individuals' choices and explanatory variables, resulting in higher prediction accuracy than classical DCMs (Hagenauer and Helbich, 2017; Omrani, 2015; Pulugurta et al., 2013). The cons are that MLs are often criticized as "black-boxes" that are sensitive to hyperparameters and lack interpretability for modelers to explain the behavioral mechanism (Liao and Poggio, 2018; Sun et al., 2019; Wang et al., 2020b).

Besides these pros and cons widely discussed in existing studies, we would like to emphasize that general-purpose machine learning models do not generally address the three mentioned limitations of DCMs. On the one side, MLs treat individual choices as a classification task, in which cross-entropy is often used to formulate the cost functions (Kline and Berardi, 2005). Similar to the

likelihood functions in DCMs, cross-entropy-based cost functions in MLs are also inefficient to optimize, given a huge choice set of all possible schedule combinations. Hence, the performance of general-purpose MLs might decrease with the increase in size of choice sets. On the other side, though the powerful automatic learning of MLs can capture complex behavior realism, it is at the cost of local irregularity and non-linearity of demand functions (LeCun et al., 2015; Liao and Poggio, 2018). Wang et al. (2020a) have pointed out the impacts of local irregularity on individual tastes. They found that the exploding and vanishing gradients in neural networks can result in extremely high or low sensitivities at the individual level that are opposite to domain knowledge. Moreover, with hundreds of parameters in deep learning models, it is almost infeasible to formulate the utility function, let alone generate demand functions and integrate them into optimization models. An innovative, domain-specific machine learning approach is necessary to deal with the large choice set, capture individual-level tastes, and build the link between demand and supply.

2.3 Inverse optimization (IO) for individual choice

Inverse optimization (IO) is initially used to impute missing optimization model coefficients from data that represents sub-optimal solutions of that optimization problem (Ahuja and Orlin, 2001; Burton and Toint, 1992). Given an optimization problem, an IO can be formulated to impute its left-hand-side constraint parameters and feasible regions (Chan and Kaw, 2020; Ghobadi and Mahmoudzadeh, 2021). A typical IO problem is defined as follows: for a given prior θ_0 of missing coefficients and observed decision variables x^*, determine an updated coefficient set θ such that x^* is optimal while minimizing its L_1 norm from the prior, as shown in Eq. (5).

$$\min_{\theta} |\theta_0 - \theta| : x^* = \arg\min \{\theta^T x : Ax \leq b, x \geq 0\} \tag{5}$$

where A is the constraint matrix b is the vector of side constraint values. $Ax \leq b$ are constraints ensuring x^* is optimal (or the best choice). L_1 norm is used to regularize what would otherwise be an ill-posed problem with infinite solutions. Ahuja and Orlin (2001) proved that Eq. (5) can be reformulated as a linear programming (LP) problem.

Since IO imputes coefficient values from data, it can be viewed as a machine learning approach (Iraj and Terekhov, 2021; Tan et al., 2019) and shares similarities with inverse reinforcement learning (IRL) (Iskhakov et al., 2020). Though IO is less popular than general-purpose machine learning models, it has already been applied to traffic assignment, route choice, and activity scheduling problems (Chow and Recker, 2012; Hong et al., 2017; Chow, 2018; Xu et al., 2018). For instance, Chow and Recker (2012) proposed a multiagent framework for IO where a sample of individuals' trip scheduling data is obtained and used to infer parameters of individual activity scheduling. Xu et al. (2018) formulated the multiagent inverse transportation problem to estimate heterogeneous route preferences, and proved that the IO approach could obtain heterogeneous link cost coefficients even when multinomial or mixed logit models would not be meaningfully estimated. Moreover, the potential of IO in modeling individual choice has been noticed by existing studies. Iraj and Terekhov (2021) emphasized the need for stochastic IO models in scenarios where constraints, objective, and prior parameters can be defined with domain knowledge. This holds in the problem of activity scheduling choice, given existing results obtained through econometric models (e.g., DCMs).

We propose a hybrid machine learning/econometric approach designed to estimate agents' joint activity scheduling choice from a ubiquitous data set. The approach is based on the IO method of estimating a random utility model that considers different choice dimensions jointly and coefficients that vary by individual. The utility function is linear, in order to ensure its compatibility with optimization models.

3. Proposed model

The proposed Agent-based Mixed Logit model (AMXL) is a random utility model with individual-specific coefficients resulting in a heterogeneous distribution of the population coefficients. The AMXL is applied in this study to one use case (and other use cases can be explored in future research): it represents individuals' whole-day activity scheduling behavior based on a series of utility items associated with decomposed sub-choices. Notations used in this section are shown in Table 1.

Table 1
Notations used in the proposed model

$\mathcal{V}$	Total utility derived from the activity schedule of an individual
$\mathcal{V}_A$	Total utility derived from participating activities in set A
$\mathcal{V}_T$	Total utility derived from conducting trips in set T
$V_{a,i}$	Utility derived from activity a conducted by individual i
$V_{a,i}^{SD}$	Utility item of activity a conducted by individual i, related to activity schedule delay
$V_{a,i}^{Dur}$	Utility item of activity a conducted by individual i, related to activity duration
$V_{a,i}^{Des}$	Utility item of activity a conducted by individual i, related to activity destination
$\theta_{a,i}^{e}, SDE_{a,i}$	Coefficient and observed value of starting an activity earlier than schedule (min)
$\theta_{a,i}^{l}, SDL_{a,i}$	Coefficient and observed value of starting an activity later than schedule (min)
$\theta_{a,i}^{pl}, PL_{a,i}$	Coefficient and observed value of starting activity late (binary)
$\theta_{a,i}^{duration}, D_{a,i}$	Coefficient and observed value of activity duration (min)
$\theta_{a\_a',i}^{duration}$	Coefficient of the interactive item of two activity durations
$\theta_{a,i}^{des,k}, des_{a,i}^{k}$	Coefficient and observed value of a specific feature k related to the activity location
$V_{t,i}$	Utility derived from trip t conducted by individual i
$V_{t,i}^{Time}$	Utility item of trip t conducted by individual i, related to trip time
$V_{t,i}^{Cost}$	Utility item of trip t conducted by individual i, related to trip cost
$V_{t,i}^{Mode}$	Utility item of trip t conducted by individual i, related to trip mode
CU	The choice-utility incidence matrix
$\mathcal{V}'$	A vector containing decomposed utilities
ε_{ijm}	Random utility for individual i choosing alternative j in sub-choice m
θ_{0m}	Fixed-point prior of coefficients which can be divided into sub-choices $m \in M$
θ_{im}	Individual specific coefficients which can be divided into sub-choices $m \in M$

3.1 Architecture of Agent-based Mixed Logit model (AMXL)

The architecture of AMXL model is determined by how to derive a total utility from the whole-day activity schedule and how to decompose it into several inter-linked subitems.

3.1.1 Utility derived from the whole-day activity schedule

A whole-day activity schedule can be divided into a sequence of activities and trips. In line with the study of Ettema et al. (2007) and MATSim (Rieser et al., 2014), the AMXL model assumes individuals in a population P maximize their total utility derived from the activity schedule by solving Eq. (6).

$$\max \mathcal{V} = \max(\mathcal{V}_A + \mathcal{V}_T) = \max\left(\sum_{a \in A} V_{a,i} + \sum_{t \in T} V_{t,i}\right), \qquad \forall i \in P \tag{6}$$

where $\mathcal{V}_A$ is the total utility derived from participation of activity set A (assumed positive) and $\mathcal{V}_T$ is the total utility derived from trip set T (assumed negative). These two utilities are the sums of utilities of individuals' activities $\sum_{a\in A} V_{a,i}$ and trips $\sum_{t\in T} V_{t,i}$.

The utility derived from an activity depends on three elements: (1) the duration of activities, (2) the schedule deviation relative to a preferred or scheduled activity start time, and (3) the destination of activity relative to a set of attributes of the location. To this end, if individual i has conducted an activity p, the utility can be defined as Eq. (7).

$$V_{a,i} = V_{a,i}^{SD} + V_{a,i}^{Dur} + V_{a,i}^{Des}, \qquad \forall i \in P, \forall a \in A \tag{7}$$

where $V_{a,i}^{SD}$, $V_{a,i}^{Dur}$, and $V_{a,i}^{Des}$ are utilities related to the schedule deviation, the activity duration, and the activity destination. Utility linked to the activity category is not included since we mainly focus on the activity scheduling, where the number and type of activities are pre-defined in skeleton schedules (Habib and Miller, 2006).

The schedule deviation related utility of activity p conducted by individual i, $V_{a,i}^{SD}$, is defined in Eqs. (8) – (10).

$$V_{a,i}^{SD} = \theta_{a,i}^{e} SDE_{a,i} + \theta_{a,i}^{l} SDL_{a,i} + \theta_{a,i}^{pl} PL_{a,i}, \qquad \forall i \in P, \forall a \in A \tag{8}$$

$$SDE_{a,i} = \max(0, s_{a,i}^{*} - s_{a,i}), \qquad \forall i \in P, \forall a \in A \tag{9}$$

$$SDL_{a,i} = \max(0, s_{a,i} - s_{a,i}^{*}), \qquad \forall i \in P, \forall a \in A \tag{10}$$

where $SDE_{a,i}$ and $SDL_{a,i}$ are the early and late schedule delay, $s_{a,i}^{*}$ is individual i's targeted start time for activity a, $s_{a,i}$ is individual i's actual start time for activity a. $PL_{a,i}$ is an additional penalty for starting an activity late (independent of the actual amount of 'lateness' (Lizana et al., 2021)), which equals to 1 if $SDL_{a,i} > 0$, 0 otherwise. $\theta_{a,i}^{e}$, $\theta_{a,i}^{l}$, $\theta_{a,i}^{pl}$ are coefficients for these items.

The activity duration related utility of activity a conducted by individual i, $V_{a,i}^{Dur}$, is defined in Eq. (11).

$$V_{a,i}^{Dur} = \theta_{a,i}^{duration} \ln(D_{a,i}) + \theta_{a\_a',i}^{duration} \ln(D_{a,i}) \ln(D_{a',i}), \qquad \forall i \in P, \forall a, a' \in A \tag{11}$$

where $D_{a,i}$ is the duration of activity a performed by individual i, $D_{a',i}$ is the duration of the related activity a'. The implication of the former item is that marginal utility decreases with increasing duration ($\theta_{a,i}^{duration}$ is assumed to be positive). The latter item allows the utility of different activities to be dependent on each other. For instance, if one activity gains a higher utility with more time spent on another activity, the coefficient $\theta_{a\_a',i}^{duration}$ will have a positive sign.

The utility of activity a performed by individual i, $V_{a,i}^{Des}$, is defined in Eq. (12).

$$V_{a,i}^{Des} = \sum_{k=1}^{K} \theta_{a,i}^{des,k} des_{a,i}^{k}, \qquad \forall i \in P, \forall a \in A \tag{12}$$

where $des_{a,i}^{k}$ is the k^{th}attribute related to the destination of activity a performed by individual i and $\theta_{a,i}^{des,k}$ is the corresponding coefficient.

In the AMXL model, the utility derived from a trip depends on three elements: (1) trip duration, (2) trip cost, and (3) trip mode. To reduce the complexity, we do not include route choice in our

model, and we assume that individuals tend to choose the route with the shortest trip duration. Hence, if individual i has made a trip t, the derived utility, $V_{t,i}^{T}$, can be defined as Eq. (13).

$$V_{t,i}^{T} = V_{t,i}^{Time} + V_{t,i}^{Cost} + V_{t,i}^{Mode} = \theta_{t,i}^{time} time_{t,i} + \theta_{t,i}^{cost} cost_{t,i} + \sum_{d=1}^{d} \theta_{t,i}^{mode,d} mode_{t,i}^{d}, \quad \forall i \in P, \forall t \in T \tag{13}$$

where $time_{t,i}$, $cost_{t,i}$ are the duration and cost of trip t made by individual i. Trip mode is denoted by $mode_{t,i}^{d}$, which equals to 1 if the individual chooses the d^{th} mode, 0 otherwise. $\theta_{t,i}^{time}$, $\theta_{t,i}^{cost}$, $\theta_{t,i}^{mode,d}$ are the coefficients of these three items. Note that $time_{t,i}$ and $cost_{t,i}$ can be extended to $time_{t,i}^{d}$ and $cost_{t,i}^{d}$, which adds mode specific time and cost to utility function for trips. However, we do not recommend the extended form since we focus on the whole-day utility and want to keep the number of coefficients in each utility item at the same level.

In general, the total utility derived from individual i's whole day activity schedule can be defined as Eq. (14). For an activity schedule containing $|A|$ activities and $|T|$ trips, there are $(3|A| + 3|T|)$ items in total.

$$\mathcal{V} = \sum_{a=1}^{|A|} \left(V_{a,i}^{SD} + V_{a,i}^{Dur} + V_{a,i}^{Des}\right) + \sum_{t=1}^{|T|} \left(V_{t,i}^{Time} + V_{t,i}^{Cost} + V_{t,i}^{Mode}\right), \quad \forall i \in P, \forall a \in A, \forall t \in T \tag{14}$$

3.1.2 Whole-day scheduling choice decomposition

To control the size of the choice set, we decompose the whole-day scheduling choice into a series of inter-related sub-choices, with each sub-choice containing at least one choice dimension of the scheduling joint-choice. The choice-utility incidence matrix is introduced to fulfill this. To be specific, the choice-utility incidence matrix, CU, has M rows and N columns, where M is the number of sub-choices, N is the number of utility items, which equals to $(3|A| + 3|T|)$ in our model. The value in the m^{th} row and n^{th} column equals to 1 if the n^{th} utility item is influenced by the m^{th} sub-choice, 0 otherwise. If we convert utility items in $\mathcal{V}$ to a vector $\boldsymbol{\mathcal{V}}$, with $\boldsymbol{\mathcal{V}} = [V_{1,i}^{SD}, \dots, V_{|A|,i}^{Des}, V_{1,i}^{Time}, \dots, V_{|T|,i}^{Mode}]$ (in the shape of $(1, 3|A| + 3|T|)$), the total utility $\mathcal{V}$ can be decomposed as a series of utility one-to-one associated with the sub-choices, as shown in Eq. (15).

$$\boldsymbol{\mathcal{V}}' = CU * \boldsymbol{\mathcal{V}} = \begin{bmatrix} cu_{1,1} & cu_{1,2} & \cdots & cu_{1,N} \\ cu_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ cu_{M,1} & \cdots & \cdots & cu_{M,N} \end{bmatrix} * \begin{bmatrix} V_{1,i}^{SD} \\ \vdots \\ V_{|A|,i}^{Des} \\ V_{1,i}^{Time} \\ \vdots \\ V_{|T|,i}^{Mode} \end{bmatrix} = \begin{bmatrix} \mathcal{V}^1 \\ \mathcal{V}^2 \\ \vdots \\ \mathcal{V}^M \end{bmatrix}, \quad \forall i \in P \tag{15}$$

where $\boldsymbol{\mathcal{V}}'$ is a vector containing decomposed utilities $\mathcal{V}^1$, $\mathcal{V}^2$, …, $\mathcal{V}^M$. Each decomposed utility is the total utility derived from a sub-choice, which individuals are assumed to maximize. The benefit of decomposing the whole-day scheduling choice is that it separates different choice dimensions to make the size of the total choice set equal to a summation of alternatives in each choice dimension rather than a product.

CU matrix in our study is a pre-defined binary matrix, determining if a sub-choice depends on a utility item. However, CU matrix has the potential to be extended to a real-number matrix with a range of 0 to 1. In that case, $cu_{m,n}$ denotes the degrees of dependencies between sub-choice m and utility item n, and CU matrix should be estimated instead of pre-defined, which will introduce a series of coefficients and complicate the model.

3.1.3 Observed variables, normal coefficients, and shared coefficients

To sum up, $SDE_{a,i}$, $SDL_{a,i}$, $PL_{a,i}$, $D_{a,i}$, $des_{a,i}^{k}$, $time_{t,i}$, $cost_{t,i}$, $mode_{t,i}^{d}$ are observed variables calculated based on individuals' activity schedules, which are concatenated into a vector X_i. $\theta_{a,i}^{e}$, $\theta_{a,i}^{l}$, $\theta_{a,i}^{pl}$, $\theta_{a,i}^{duration}$, $\theta_{a\_a',i}^{duration}$, $\theta_{a,i}^{des,k}$, $\theta_{t,i}^{time}$, $\theta_{t,i}^{cost}$, $\theta_{t,i}^{mode,d}$ are a series of individual-specific coefficients one-to-one related to observed variables, which are concatenated into a vector θ_i. Due to the co-impacts of different choice dimensions (e.g., the duration of an activity is influenced by its start time and the start time of the next activity), some observed variables might be influenced by more than one sub-choice. In other words, some of the column-summations of the choice-utility incidence matrix CU should be larger than 1 in order to include all related utility items into the total utility derived from each sub-choice. To this end, coefficients in the AMXL model can be divided into normal parameters and shared coefficients. For observed variables influenced by only one sub-choice (related column-summations of CU equals to 1), we define their associated coefficients as normal coefficients. For observed variables influenced by more than one sub-choice, we define their associated coefficients as shared coefficients. θ_i^{normal} and θ_i^{shared} are used to differentiate normal and shared coefficients.

By keeping θ_i^{shared} to be the same in all sub-choices, we can jointly estimate the whole model while dealing with a much smaller choice set. Compared with nested structure in dynamic DCMs (Västberg et al., 2020), our method get rid of complex conditional probabilities by adding an additional constraint that ensures shared coefficients are fixed among sub-choice situations. There are several benefits: (1) we avoid calculating high-dimensional integrals for random coefficients; (2) dependencies among sub-choice situations are captured at individual level since θ_i^{shared} is individual-specific; (3) pre-defined choice sequence is not required since sub-choice situations can be modeled in parallel. Also, we suggest a balanced number of coefficients among each sub-choice considering the parallel structure.

3.2 Multiagent IO estimation framework for AMXL

To estimate the AMXL model, we propose to formulate a multiagent inverse utility maximization problem for each sub-choice scenario and solve them jointly with additional constraints related to shared coefficients.

3.2.1 Multiagent inverse utility maximization (MIUM) problem for coefficient estimation

Inspired by the works of Chow and Recker (2012) and Xu et al. (2018), our study formalizes the multiagent inverse utility maximization (MIUM) problem to estimate individual level coefficients in each sub-choice situation. Consider a choice set J from which a population P of agents behaviorally seek to select to maximize their overall utilities. The AMXL model decomposes the whole-day scheduling choice into a series of inter-related sub-choices M, i.e. $J_m, m \in M$. Each agent $i \in P$ has a specific preference for a sub-choice $m \in M$ reflected by a coefficient set θ_{im}, and chooses an alternative $j \in J$ based on the principle of utility maximization. In line with discrete choice models (DCMs), the total utility derived from agent i choosing alternative j is defined in Eq. (16).

$$U_{ijm} = V_{ijm} + \varepsilon_{ijm} = \theta_{im} X_{ijm} + \varepsilon_{ijm}, \qquad \forall i \in P, j \in J, m \in M \tag{16}$$

where U_{ijm} is the total utility, which is composed of a deterministic utility V_{ijm} and a Gumbel-distributed random utility ε_{ijm}. X_{ijm} is a set of observed variables related to individual i choosing alternative j for sub-choice $m \in M$. The set of attributes is defined as K, which are divided into attributes that vary by sub-choice, K_m, with corresponding coefficients θ_i^{normal}, and attributes that have shared common coefficients θ_i^{shared} across sub-choices, K_0. The individual-level coefficient set θ_{im} can be jointly estimated by solving a MIUM problem under L_2-norm as a convex quadratic programming (QP) problem, as illustrated in Eq. (17) – (20).

$$\min_{\theta_{0m}, \theta_{im}} \sum_{i \in P} (\theta_{0m} - \theta_{im})^2 \tag{17}$$

Subject to

$$V_{ijm^*}(\theta_{im}) + \varepsilon_{ijm^*} \geq V_{ijm}(\theta_{lm}) + \varepsilon_{ijm} + b, \qquad j \neq j^*, \forall j \in J, i \in P, m \in M \tag{18}$$

$$\theta_{0m} = \frac{1}{|P|} \sum_{i \in P} \theta_{im}, \qquad \forall m \in M \tag{19}$$

$$\theta_{imk} = \theta_{im'k}, \qquad \forall i \in P, k \in K_0, m, m' \in M \tag{20}$$

where θ_{0m} is a common prior corresponding to a sub-choice $m \in M$. The objective is quadratic while the constraints are linear. Compared with Ahuja and Orlin (2001)'s study, we replace L_1-norm with L_2-norm since L_2-norm results in a smaller variance of individual-level coefficients (by making θ_{im} closer to θ_{0m}), which is ideal for inverse optimization at a large scale. For Eq. (18), ε_{ijm^*} and ε_{ijm} are random utilities that should be randomly generated for each agent. This is a bit different from DCMs, in which the random utility is used to calculate the cumulative density function (CDF), but is not really drawn for each individual. In case ε_{ijm^*} is much larger than ε_{ijm} in a single draw (making the comparison of deterministic utilities meaningless), we add a safe boundary b ($b \geq 0$) to the linear inequalities. In general, the safe boundary b acts as a hyperparameter in the MIUM. From a machine learning perspective, this is similar to adding noise when training the model and thus helps to avoid overfitting. A proposed value of b around 75th quantile of $(\varepsilon_{ij^*} - \varepsilon_{ij})$ is recommended, which ensures the utility of chosen alternative should at least be larger than 75% of the rest of the alternatives. The proposed value is based on a series of experiments focusing on the balance between feasible solutions and prediction accuracy. It is noted that few studies have included random items (ε_{ij^*} and ε_{ij}) into IO problems. Eq. (19) makes sure that the estimated individual coefficients have a fixed point consistency with the common priors. For the attributes that are meant to have shared coefficient, Eq. (20) ensures that the estimated coefficient is identical across all sub-choices where it's present.

Solving the model in Eqs. (17) – (20) as a single QP would be computationally costly as it would lead to highly diagonal sparse matrix. Instead, the model is decomposed. In each iteration, we first break down our sub-choices into $|M|$ independent $MIUM_m$ problems to obtain $|M|$ posteriors, which includes both normal and shared coefficients at the individual level. Then we calculate the mean values of shared coefficients in these posteriors, fix these shared coefficients, and re-solve the $MIUM_m$ problems, which can be fulfilled by adding a constraint ensuring the shared coefficients are equal to the fixed values (let us denote the new $MIUM_m$ problem as $MIUM'_m$). Finally, we update the fixed-point prior using a convergent iterative algorithm, and check if the stopping criteria has been reached. If reached, then we concatenate posteriors obtained by $MIUM'_m$ problems and output the estimated individual-specific coefficient θ_{im}. Otherwise, we use the

updated fixed-point prior and go to the next iteration. The iterations continue until a set of coefficients θ_{0m} stabilizes (see Xu et al. (2018) for an example of this kind of decomposition for the $|M| = 1$ case). The $MIUM_m$ problem can be solved using any optimizer software or package that can handle QP like Gurobi, CVXPY, etc. The estimation process is illustrated in Fig. 1.

In Fig. 1, boxes in light-green color represent $MIUM_m$ problems, one for each sub-choice. Boxes in light-red color represent the procedures related to the joint estimation. Based on the illustrated architecture, coefficients in different sub-choices can be estimated in parallel (horizontal structure) while shared coefficients that are invariant among these sub-choices ensure the inter-relationship (vertical structure). The MIUM is a QP and thus the whole structure can reach a unique solution. Given an initial guess of a fixed-point prior θ_0, a convergent iterative algorithm (e.g., method of successive averages (MSA)) would reach a unique and statistically consistent fixed point with respect to that guess (see Chow and Recker 2012).

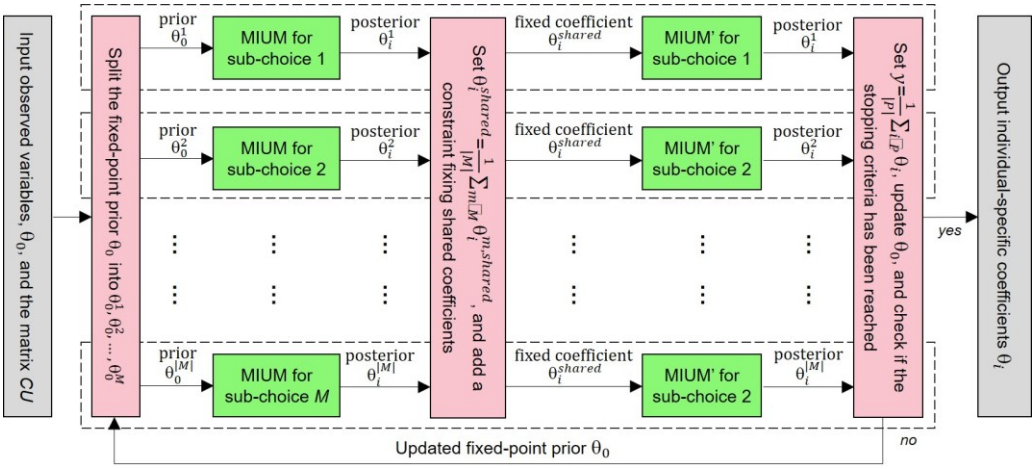


Fig. 1. Illustration of joint estimating several MIUM problems

Note that AMXL in this study can only capture inter-individual heterogeneity, since it models activity scheduling choice based on one observation per individual. To deal with intra-individual heterogeneity with multiple observations per individual, multiday inverse utility maximization problems should be formulated and solved before multiagent inverse utility maximization problems. In that case, there would be multiple levels of priors (population priors from which individual priors are drawn, and then individual and time-specific parameters that are drawn from those individual priors. Future research will explore this further.

3.2.2 Proposed algorithm

The convergent iterative algorithm used in our study is the Method of Self-Regulated Average (MSRA). Compared with the conventional Method of Successive Average (MSA), MSRA adjusts the weight of optimal solution in each iteration to speed up the convergence rate (see Liu et al. (2007)). The MSRA algorithm is specified in Algorithm 1, in which $x^{(n)}$ is equivalent to the fixed-point prior in the n^{th} iteration ($\theta_0^{(n)}$) and $y^{(n)}$ is equivalent to the averaged posterior coefficients of $|P|$ individuals ($y^{(n)}$= F($\theta_0^{(n)}$) = $\frac{1}{|P|}\sum_{i \in P} \theta_i^{(n)}$). We adopt the value of some hyperparameters from Liu et al. (2007)'s study, including Γ, γ, and $\beta^{(0)}$. Using these parameters, we found that the algorithm reduced the convergence time by 35% relative to using MSA for the examples in the study.

Algorithm 1. Method of Self-Regulated Averages (Liu et al., 2007)

1. Set the initial point $x^{(0)}$, n=1, Γ>1, 0<γ<1, $\beta^{(0)}$=0, and the stop criteria ε'>0
2. Calculate $y^{(0)}$= F($x^{(0)}$), $x^{(1)}$= ($x^{(0)}$+$y^{(0)}$)/2, and $y^{(1)}$=F($x^{(1)}$)
3. Do while $\left\|x^{(n)} - y^{(n)}\right\| \geq \varepsilon'$:
   if $\left\|x^{(n)} - y^{(n)}\right\| \geq \left\|x^{(n-1)} - y^{(n-1)}\right\|$:
   $\beta^{(n)} = \beta^{(n-1)} + \Gamma$
   else:
   $\beta^{(n)} = \beta^{(n-1)} + \gamma$
   $\alpha^{(n)} = 1/\beta^{(n)}$
   $x^{(n+1)} = x^{(n)} + \alpha^{(n)}.(y^{(n)} - x^{(n)})$
   $y^{(n+1)} = \mathrm{F}(x^{(n+1)})$
   $n = n + 1$
4. Output $x^{(n)}$

An additional issue is that if we use different initial points or different standard deviations of ε_{ijm}, it would lead to a different fixed-point solution set (Xu et al., 2018). This is similar to DCMs, in which the estimated coefficients are only unique relative to each other, but the overall values can be scaled up or down. To ensure comparable results between different scenarios, we normalize the observed variables and set the std. of the Gumbel error ε_{ijm} to 1. In that case, $b = 1$ is a recommended value of safe boundary, which equals to the 75% quantile of ($\varepsilon_{ij^*} - \varepsilon_{ij}$). The whole estimation approach is summarized in Algorithm 2. The stopping rule is set to ε′=0.001 considering time to converge.

Algorithm 2. Proposed multiagent IO estimation for the AMXL model

1. Given observed variables X_{ij} and the choice-utility incident matrix CU, initialize with n=1, b=1, and the fixed-point prior $\theta_0^{(n)}$=[0,0, …,0]
2. Solve $|M|$ decomposed $MIUM_m$ problems with the fixed-point priors $\theta_0^{1(n)}, \theta_0^{2(n)}, \dots, \theta_0^{|M|(n)}$:
   $$\min_{\theta_0^{m(n)}, \theta_i^{m(n)}} \left\{(\theta_0^{m(n)} - \theta_i^{m(n)})^2 \; s.t. \; V_{ij^*}^m + \varepsilon_{ij^*}^m \geq V_{ij}^m + \varepsilon_{ij}^m + b, \;\; j \neq j^*, \forall j \in J^m\right\}, \quad \forall i \in P, \forall m \in M$$
3. Set average to $\theta_i^{shared(n)} = \frac{1}{|M|} \sum_{m \in M} \theta_i^{m,shared(n)}$
4. Solve $|M|$ $MIUM'_m$ problems, with an additional constraint fixing coefficients $\theta_i^{shared(n)}$:
   $$\min_{\theta_0^{m(n)}, \theta_i^{m(n)}} \left\{(\theta_0^{m(n)} - \theta_i^{m(n)})^2 \; s.t. \; V_{ij^*}^m + \varepsilon_{ij^*}^m \geq V_{ij}^m + \varepsilon_{ij}^m + b, \;\; j \neq j^*, \forall j \in J^m\right\}, \quad \forall i \in P, \forall m \in M$$
   subject to $\theta_i^{m,shared(n)} = \theta_i^{shared(n)}$
5. Concatenate $\theta_i^{1(n)}, \theta_i^{2(n)}, \dots, \theta_i^{|M|(n)}$ to $\theta_i^{(n)}$, set average to $y^{(n)} = \frac{1}{|P|} \sum_{i \in P} \theta_i^{(n)}$
6. MSRA: set Γ=1.8, γ=0.3, $\beta^{(0)}$=0, ε′=0.001, calculate $\theta_0^{(n+1)}$
7. If MSRA stopping criteria reached, stop and output $\theta_i^{(n)}$, else let n = n+1 and go to step 2

The computational time is proportional to the total number of iterations and the time spent at each iteration. In each iteration, ($2 \times |M|$) $MIUM_m$ problems are estimated in parallel with $|P|$ individuals in each $MIUM_m$. For each IO problem in a $MIUM_m$ problem $m \in M$, the computational time is proportional to the number of constraints decided by the size of the choice set $|S^m|$. Hence, the computational time of Algorithm 2 would increase proportionally by ($|P| \times \left(|S^1| + |S^2| + \cdots +\right.$

$\left|S^{|M|}\right|$)). As a comparison, the computational time without schedule decomposition would increase proportionally by ($|P| \times |S^1| \times |S^2| \times \ldots \times \left|S^{|M|}\right|$).

3.3 Illustrative example

We built a simple example with 10 individuals, 2 home locations, 3 shopping malls, and 4 dinner spots to illustrate how the AMXL model and the multiagent IO estimation algorithm works. In this example, each individual departs from a home location, conducts a shopping activity and a dinner activity in sequence, and finally goes back home. Only two choice dimensions are considered: location of shopping activity and location of dinner activity. For the shopping choice, the alternatives are 3 shopping malls, and the derived utility is related to the number of stores in the shopping mall and the travel distance between the home location and the shopping mall. For the dinner choice, the alternatives are 4 dinner spots, and the derived utility is related to the number of restaurants in the dinner spot, the travel distance between the shopping mall and the dinner spot, and the travel distance between the dinner spot and the home location. The coordinates of the locations and synthetic schedules of the first three individuals are shown in Fig. 2. The synthetic schedule of the 10 individuals are listed in Table 2. The derived utilities of two sub-choices are defined in Eqs. (21) – (22).

$$V_{ij}^{shopping} = \beta_{n\_store} * n\_store + \beta_{dist} * dist\_hs, \qquad \forall i \in P, \forall j \in 1,2,3 \tag{21}$$

$$V_{ij}^{dinner} = \beta_{n\_rest} * n\_rest + \beta_{dist} * (dist\_sd + dist\_dh), \quad \forall i \in P, \forall j \in 1,2,3,4 \tag{22}$$

where $V_{ij}^{shopping}$ and V_{ij}^{dinner} are utilities derived from shopping and dinner activities. n_store and n_rest are the number of stores in the shopping mall and number of restaurants in the dinner spot. $dist\_hs$, $dist\_sd$, $dist\_dh$ are the travel distance from home to shopping mall, from shopping mall to dinner spot, and from dinner spot to home. β_{n_store} and β_{n_rest} are normal coefficients. β_{dist} is a shared coefficient.

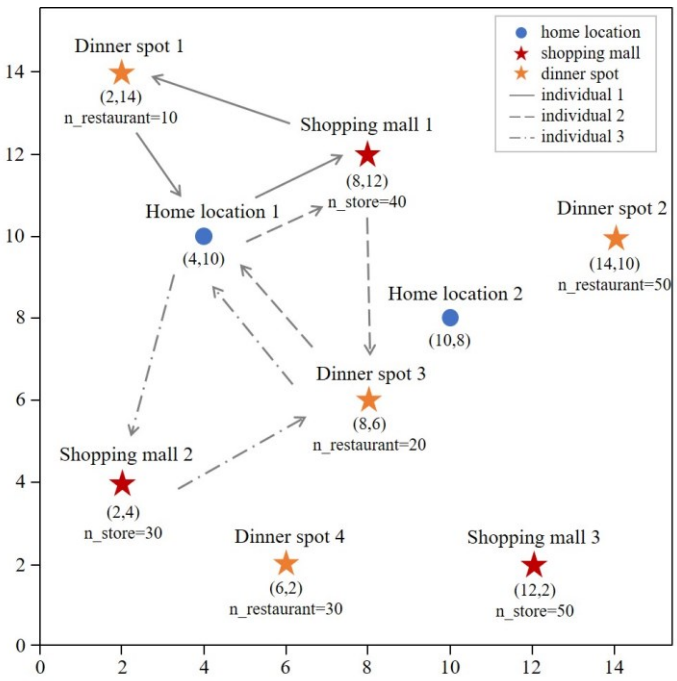


Fig. 2. Locations and activity schedules for illustration.

Table 2
Synthetic activity schedules of 10 individuals

Individual id	Synthetic activity schedule
Individual 1	Home location 1--Shopping mall 1--Dinner spot 1--Home location 1
Individual 2	Home location 1--Shopping mall 1--Dinner spot 3--Home location 1
Individual 3	Home location 1--Shopping mall 2--Dinner spot 3--Home location 1
Individual 4	Home location 1--Shopping mall 2--Dinner spot 4--Home location 1
Individual 5	Home location 1--Shopping mall 3--Dinner spot 3--Home location 1
Individual 6	Home location 2--Shopping mall 1--Dinner spot 2--Home location 2
Individual 7	Home location 2--Shopping mall 1--Dinner spot 3--Home location 2
Individual 8	Home location 2--Shopping mall 2--Dinner spot 3--Home location 2
Individual 9	Home location 2--Shopping mall 3--Dinner spot 2--Home location 2
Individual 10	Home location 2--Shopping mall 3--Dinner spot 3--Home location 2

We run Algorithm 2 with the above settings. The algorithm converged after 9 seconds at the 78th iteration under the average coefficient change tolerance of 0.001. The final results are shown in Table 3. The mean values of β_{n_store} and β_{n_rest} are positive, indicating that on average, individuals prefer larger shopping malls and dinner spots. The mean value of β_{dist} is negative, indicating that travel distance on average has negative effects on the utility. Moreover, the results reflect diverse tastes at the individual level: (1) individuals {1,3,8} have negative β_{n_rest} values, while the others have positive ones; (2) individual 2 has the shortest total travel distance. Accordingly, the β_{dist} value of individual 2 is the lowest (-1.297); and (3) the total utility derived from the synthetic schedule has a large standard deviation (6.048) compared to its mean value (-3.886), though each individual has already made the best choice assumed in our model.

Table 3
Results obtained from Algorithm 2

Individual id	β_{n_store}	β_{n_rest}	β_{dist}	$V^{shopping}$	V^{dinner}	V^{Total}
Individual 1	0.025	-0.042	-0.673	-2.270	-7.683	-9.953
Individual 2	0.127	0.162	-1.297	-2.006	-11.890	-13.897
Individual 3	0.221	-0.052	-0.654	4.710	-8.876	-4.166
Individual 4	0.220	0.148	-0.647	4.702	-3.799	0.903
Individual 5	0.399	0.046	-0.599	13.165	-5.861	7.305
Individual 6	0.011	0.077	-0.664	-2.624	-3.322	-5.946
Individual 7	0.012	0.010	-0.666	-2.628	-5.672	-8.299
Individual 8	0.051	-0.052	-0.189	0.335	-2.772	-2.437
Individual 9	0.127	0.126	-0.654	2.190	-2.036	0.154
Individual 10	0.127	0.046	-0.659	2.155	-4.679	-2.525
Mean value	0.132	0.047	-0.670	1.773	-5.659	-3.886
Standard deviation	0.122	0.081	0.265	4.930	3.067	6.048

4. Setup of experiments

4.1 Datasets

Our experiments are based on the SHC dataset, a processed mobile phone dataset containing two weekdays activity information of 26,149 anonymous commuters working in the CBD of Shanghai. The original mobile phone data were provided by Wisdom Footprint Data Technology Co., LTD.,

generated by China Telecom mobile subscribers from May 1st to 31st in 2019. We use a processed SHC dataset, which contains the activity schedule of each commuter on two weekdays. The two weekdays are May 7th and May 14th, both on Tuesday. Table 4 shows a sample of the SHC dataset containing four activity schedules of two commuters on two weekdays. Each row contains several fields: id (a unique number to differentiate a commuter), date (to differentiate a weekday), home (commuter's home place id, in 500m*500m grids), work (commuter's workplace id, in street blocks), lunch (commuter's lunch spot id, in 500m*500m grids), c_mode (trip mode for commute), time_lh (timestamp when leaving home, in minutes), time_aw (timestamp when arriving workplace, in minutes), time_sl (the start time of lunch activity), time_el (the end time of lunch activity), time_lw (timestamp when leaving workplace), and time_ah (timestamp when arriving home). We aggregate home and work location of commuters before mapping them in space so that there will not be any privacy issue.

Table 4
A sample of the SHC dataset

id	date	home	work	lunch	c_mode	time_lh	time_aw	time_sl	time_el	time_lw	time_ah
26	0507	1342	79	6789	transit	8:03	9:05	12:15	13:13	17:50	18:54
26	0514	1342	79	539	driving	8:35	9:02	12:00	12:30	17:23	21:25
78	0507	945	14	345	driving	7:25	8:48	11:47	12:56	18:32	20:01
78	0514	945	14	345	driving	7:37	9:02	12:05	12:52	18:22	20:05

To understand the impacts of different sample sizes and contexts, we construct three datasets from the SHC dataset for our experiments: (1) the dataset with randomly sampled 1,000 observations on May 7th (1K-SHC07), (2) the dataset with full observations on May 7th (26K-SHC07), and (3) the dataset with full observations on May 14th (SHC14). Fig. 3 visualizes the study area and the commuting origin-destination matrix of these datasets. We use the 1K-SHC07 and 26K-SHC07 datasets to build models and the SHC14 dataset to check their performance. The comparison between the 1K-SHC07 and 26K-SHC07 datasets reveals the effect of the sample size, and the comparison between the 26K-SHC07 and SHC14 reveals the transferability of our models to the same population on a different day scenario.

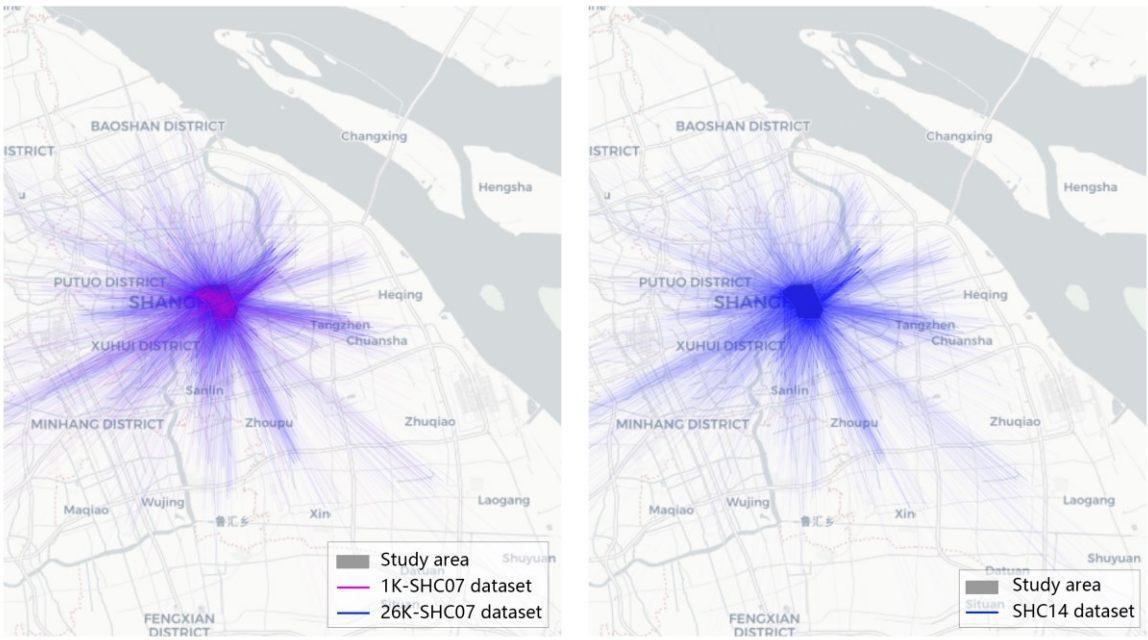


Fig. 3. Visualization of the study area and the datasets.

4.2 Model specification

Based on the dataset, our experiments consider five choice dimensions in commuters' whole-day activity schedule: time to leave home for work, commute mode, time to have lunch, lunch location, and time to finish work. To control the size of the choice set, we use 30-minute discrete time periods. Table 5 lists the five choice dimensions and the alternatives in them. In general, there are 1,470 ($7 \times 2 \times 5 \times 3 \times 7$) possible activity schedules for each individual. Our experiments do not consider the trips and activities after work since they vary from individual to individual. Instead, only the total duration of afterwork activity is included. Focusing on such a skeleton schedule make sense because: (1) it is hard to model all possible activities, especially those performed randomly; (2) travel demands during non-peak hours contribute less to congestion (compared with commuting).

Table 5
Five choice dimensions and related alternatives

Time to leave home for work	Commute mode	Time to have lunch	Lunch location	Time to finish work
6:30-7:00	Transit	11:00-11:30	Inside the CBD	17:30-18:00
7:00-7:30	Driving	11:30-12:00	Outside the CBD	18:00-18:30
7:30-8:00		12:00-12:30	In workplace	18:30-19:00
8:00-8:30		12:30-13:00		19:00-19:30
8:30-9:00		13:00-13:30		19:30-20:00
9:00-9:30				20:00-20:30
9:30-10:00				20:30-21:00

In the operational AMXL model, the total utility of a whole-day activity schedule consists of utilities related to work activity, lunch activity, afterwork activity, and the trips between them, as shown in Eqs. (23) – (28), which differs from conventional DCMs in that each coefficient is indexed by individual, e.g. $\theta^{e}_{work,i}$.

$$\mathcal{V} = V^{A}_{work,i} + V^{A}_{lunch,i} + V^{A}_{afterwork,i} + V^{T}_{commute,i} + V^{T}_{work-lunch,i}, \qquad \forall i \in P \tag{23}$$

$$\begin{aligned} V^{A}_{work,i} &= V^{SD}_{work,i} + V^{Dur}_{work,i} \\ &= \theta^{e}_{work,i} SDE_{work,i} + \theta^{l}_{work,i} SDL_{work,i} + \theta^{pl}_{work,i} PL_{work,i} \\ &\quad + \theta^{dur}_{work,i} \ln\left(D_{work,i}\right), \qquad \forall i \in P \end{aligned} \tag{24}$$

$$V^{A}_{lunch,i} = \theta^{e}_{lunch,i} SDE_{lunch,i} + \theta^{l}_{lunch,i} SDL_{lunch,i} + \theta^{des,1}_{lunch,i} des^{1}_{lunch,i} + \theta^{des,2}_{lunch,i} des^{2}_{lunch,i}, \quad \forall i \in P \tag{25}$$

$$V^{A}_{afterwork,i} = \theta^{dur}_{afterwork,i} \ln\left(D_{afterwork,i}\right) + \theta^{dur}_{inter,i} \ln\left(D_{work,i}\right) \ln\left(D_{afterwork,i}\right), \quad \forall i \in P \tag{26}$$

$$V^{T}_{commute,i} = \theta^{t}_{commute,i} t_{commute,i} + \theta^{c}_{commute,i} c_{commute,i} + \theta^{m}_{commute,i} m_{commute,i}, \quad \forall i \in P \tag{27}$$

$$V^{T}_{work-lunch,i} = \theta^{t}_{work-lunch,i} t_{work-lunch,i}, \qquad \forall i \in P \tag{28}$$

where $\mathcal{V}$ is utility of individual i derived from the whole-day activity schedule; $V^{A}_{work,i}$ is work activity utility, depending on schedule deviation $V^{SD}_{work,i}$ (schedule early $SDE_{work,i}$, schedule delay $SDL_{work,i}$, and additional late-for-work penalty $PL_{work,i}$), and log-formed work duration $V^{Dur}_{work,i}$; $V^{A}_{lunch,i}$ is lunch activity utility, depending on schedule deviation (schedule early $SDE_{lunch,i}$ and schedule delay $SDL_{lunch,i}$), and two dummy variables reflecting lunch spot ($des^{1}_{lunch,i}$=1 denotes inside the CBD, $des^{2}_{lunch,i}$=1 denotes outside the CBD, having lunch in workplace is the reference

group); $V^{A}_{afterwork,i}$ is afterwork activity utility, here we only consider its total duration in log form, $\ln(D_{afterwork,i})$, and an interaction item with work duration, $\ln(D_{work,i})\ln(D_{afterwork,i})$; $V^{T}_{commute,i}$ is commute utility, depending on travel time $t_{commute,i}$, travel cost $c_{commute,i}$, and trip mode $m_{commute,i}$ (0 for driving; 1 for public transit); $V^{T}_{work-lunch,i}$ is the utility of traveling between workplace and lunch spot, here we only consider the travel time $t_{work-lunch,i}$. All these observed variables can be calculated from the SHC dataset and Gaode Direction API, and $\theta^{e}_{work,i}$, $\theta^{l}_{work,i}$, $\theta^{pl}_{work,i}$, $\theta^{dur}_{work,i}$, $\theta^{e}_{lunch,i}$, $\theta^{l}_{lunch,i}$, $\theta^{des,1}_{lunch,i}$, $\theta^{des,2}_{lunch,i}$, $\theta^{dur}_{afterwork,i}$, $\theta^{dur}_{inter,i}$, $\theta^{t}_{commute,i}$, $\theta^{c}_{commute,i}$, $\theta^{m}_{commute,i}$, $\theta^{t}_{work-lunch,i}$ are 14 coefficients per individual to be calibrated as a mixed logit model with deterministic individual tastes.

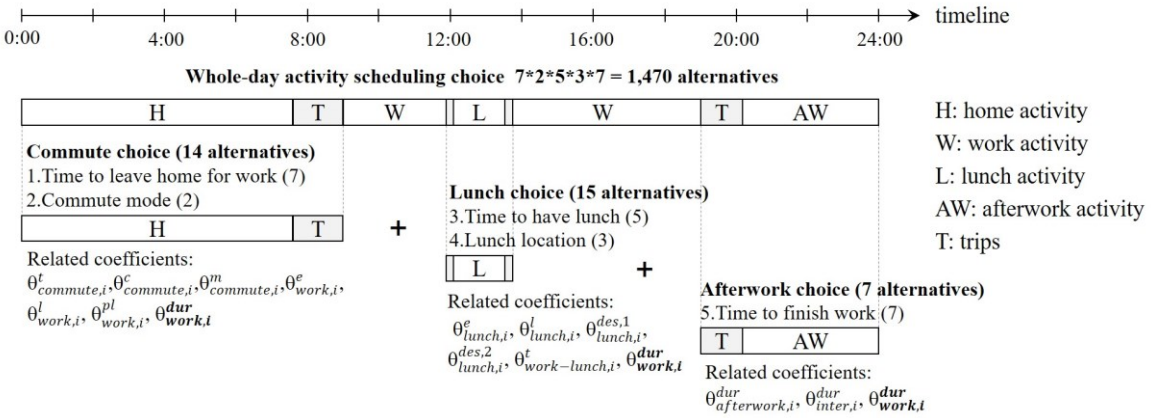


Fig. 4. An illustration of whole-day activity schedule and its decomposition

We decompose the whole-day scheduling choice into three sub-choices (Fig. 4): (1) joint-choice of time to leave home for work and commute mode (Commute choice), (2) joint-choice of time to have lunch and lunch location (Lunch choice), and (3) choice of time to finish work (Afterwork choice). Table 6 lists their related coefficients, in which the work duration related coefficient, $\theta^{dur}_{work,i}$, is the shared coefficient (since work duration is co-influenced by time to leave home for work, time spent on lunch, and time to finish work). By doing this, we reduce the choice set from 1,470 alternatives to 14, 15, and 7 alternatives, respectively, with an additional constraint ensuring that the shared coefficient $\theta^{dur}_{work,i}$ should be the same (and hence, jointly estimated).

Table 6
Five choice dimensions and related alternatives

	Choice dimension	**Alternatives**	**Related coefficients**
Commute choice	Time to leave home for work & Commute mode	14 (7*2)	$\theta^{t}_{commute,i}$, $\theta^{c}_{commute,i}$, $\theta^{m}_{commute,i}$, $\theta^{e}_{work,i}$, $\theta^{l}_{work,i}$, $\theta^{pl}_{work,i}$, $\boldsymbol{\theta^{dur}_{work,i}}$
Lunch choice	Time to have lunch & Lunch location	15 (5*3)	$\theta^{e}_{lunch,i}$, $\theta^{l}_{lunch,i}$, $\theta^{des,1}_{lunch,i}$, $\theta^{des,2}_{lunch,i}$, $\theta^{t}_{work-lunch,i}$, $\boldsymbol{\theta^{dur}_{work,i}}$
Afterwork choice	Time to finish work	7 (7)	$\theta^{dur}_{afterwork,i}$, $\theta^{dur}_{inter,i}$, $\boldsymbol{\theta^{dur}_{work,i}}$

4.3 Benchmarking and scenario design

Our experiments are divided into two major parts. The first part builds the AMXL model and compares it with benchmark models, including multinomial logit model (MNL), mixed logit model (MXL), dynamic multinomial logit model (DMNL), and dynamic mixed logit model (DMXL). Since the choice set of the whole-day schedules is quite large, we build MNL and MXL for each sub-

choice. The utility functions for each sub-choice are defined as Eqs. (29) – (31). We assume the utility of work duration $V_{work,i}^{Dur}$ is considered in all sub-choice situations.

$$U_{ij}^{commute} = V_{commute,i,j}^{T} + V_{work,i}^{SD} + V_{work,i}^{Dur} + \varepsilon_{ij}, \qquad \forall i \in P, \forall j \in J^{commute} \tag{29}$$

$$U_{ij}^{lunch} = V_{lunch,i,j}^{A} + V_{work-lunch,i,j}^{T} + V_{work,i}^{Dur} + \varepsilon_{ij}, \qquad \forall i \in P, \forall j \in J^{lunch} \tag{30}$$

$$U_{ij}^{afterwork} = V_{afterwork,i,j}^{A} + V_{work,i}^{Dur} + \varepsilon_{ij}, \qquad \forall i \in P, \forall j \in J^{afterwork} \tag{31}$$

DMNL and DMXL are in line with the basic idea of dynamic DCMs (Västberg et al., 2020), in which the utilities of work duration in commute and lunch choices are modified to include expected utilities of downstream choice situations, as illustrated in Eqs. (32) – (34). The only difference is that the expected utility in our study is defined as the log-sum of utilities of choosing all possible alternatives in downstream choice situations. To be more specific, for DMNL, we first build a MNL for afterwork choice and get estimated coefficients. Then we calculate the log-sum of utilities of choosing alternatives in the afterwork choice set, and build another MNL for lunch choice with a new variable called $EV(j, lunch)$. Finally, we calculate the log-sum of utilities of choosing alternatives in the lunch and afterwork choice set and build a MNL for commuting choice with a new variable $EV(j, commute)$. Similarly, we replace MNL with MXL in each stage to build DMXL. The reason for doing so is that the choice set varies across choice scenarios, making it hard to calculate multi-dimensional integrals. Though the Västberg et al. (2020)'s study overcomes this issue by approximating expected utility with backward induction, it's at the cost of complicating the algorithm and it's unnecessary in our case since the choice set in each sub-choice has been pre-defined.

$$U_{ij}^{commute} = V_{commute,i,j}^{T} + V_{work,i}^{SD} + \mu_{lunch} EV(j, commute) + \varepsilon_{ij}, \quad \forall i \in P, \forall j \in J^{commute} \tag{32}$$

$$U_{ij}^{lunch} = V_{lunch,i,j}^{A} + V_{work-lunch,i,j}^{T} + \mu_{afterwork} EV(j, lunch) + \varepsilon_{ij}, \quad \forall i \in P, \forall j \in J^{lunch} \tag{33}$$

$$U_{ij}^{afterwork} = V_{afterwork,i,j}^{A} + V_{work,i}^{Dur} + \varepsilon_{ij}, \qquad \forall i \in P, \forall j \in J^{afterwork} \tag{34}$$

Comparisons of AMXL and benchmark models are conducted in several ways: (1) basic statistics of model results; (2) distribution of estimated coefficients, and; (3) prediction accuracy. 1K-SHC07 and 26K-SHC07 datasets are used as training data to build models. SHC07 is used as test dataset. Prediction accuracy is calculated both at individual level and aggregated level. Individual level accuracy is defined as the percentage of commuters whose predicted choices are the same as observed choices. Aggregated level accuracy is defined as the percentage overlap between observed alternative share and predicted alternative share.

In the second part, we applied these models to two scenarios. The first one is a scenario for comparison, in which driving durations from 7:30 a.m. to 9:30 a.m. are reduced by 10% and 20% to reflect the relief of peak-hour congestion. AMXL and all of the benchmark models can be used to predict the choice shift. The second scenario is unique to AMXL, in which utility functions derived from the AMXL are directly integrated into a BP problem for a restaurant to select which blocks to give coupons to commuters to maximize profits. The aim of this scenario is to showcase the benefits of deterministic estimation in AMXL in integrating with optimization models.

5. Experimental results

This section presents the model results and scenario applications of AMXL and benchmark models. All of the experiments are conducted on a local machine with Intel(R) Core(TM) i7-10875H CPU and 32GB installed RAM. The Gurobi package is used to solve the QP problems. The Xlogit package is used to estimate discrete choice models. Codes are written in Python.

5.1 Model results

We present the model results from three aspects: (1) basic statistics; (2) coefficient distribution; and (3) prediction accuracy.

5.1.1 Basic statistics

Table 7 and Table 8 summarize the basic statistics of models built with 1K-SHC07 and 26K-SHC07, from which we can compare our AMXL model to benchmark models under the same context. Several interesting points were found.

(1) The results of AMXL and benchmark models show great consistency in signs: most of the activity duration-related coefficients (work duration, afterwork-activity duration) have positive signs, and most of the trip-related coefficients (commuting time, commuting cost, work-lunch travel time) and activity schedule delay related coefficients (work schedule early and delay, lunch schedule early and delay) have negative signs.

(2) The shared coefficient estimated by MNL and MXL varies greatly in three choice scenarios, while DMNL, DMXL, and AMXL gets rid of such a problem. DMNL and DMXL conduct joint estimation by including expected utilities of downstream choices. AMXL conduct joint estimation by fixing the shared coefficient in each iteration.

(3) Although the AMXL model cannot report the standard error and significance level of coefficients due to its deterministic feature, it ensures coefficients that are insignificant in MNL and MXL are also close to 0 (the initial value), implying that these coefficients are relatively useless in utility maximization.

(4) The means of coefficients in AMXL change slightly from 1K sample to 26K sample, similar to benchmark models. This indicates the stability of AMXL given a relatively small sample size, though AMXL is designed for ubiquitous datasets.

(5) AMXL took 28.9 hours to converge given the 26K-SHC07 dataset, which is obviously higher than DMXL (14.1 hours in total), DMNL (13.5 hours in total), MXL (2.0 hours in total), MNL (17 seconds in total). The longer computational time of AMXL is partly due to "for" loops solving the QP problems in our codes, which can be parallelized in the future.

Table 7

Basic statistics of models built with 1K-SHC07 (each entry represents the average value of one estimated coefficient, and the number in the parenthesis is the standard error)

Commute choice	MNL (1K)	MXL (1K)	DMNL (1K)	DMXL (1K)	AMXL (1K)
Commuting time ($\theta^{t}_{commute,i}$)	-8.608*** (0.561)	-12.956*** (2.091)	-7.373*** (0.469)	-8.912*** (0.739)	-4.537
Commuting cost ($\theta^{c}_{commute,i}$)	--	--	--	-2.756* (1.257)	-0.314
Commuting trip mode ($\theta^{m}_{commute,i}$)	--	--	--	--	0.301
Work schedule early ($\theta^{e}_{work,i}$)	-6.340*** (0.841)	-11.124*** (2.097)	-9.214*** (1.570)	-10.573*** (1.682)	-3.707
Work schedule delay ($\theta^{l}_{work,i}$)	-4.566*** (1.187)	--	-7.262* (2.891)	-7.766* (3.299)	-6.460
Late-for-work penalty	--	--	--	--	-0.354

($\theta_{work,i}^{pl}$)					
Shared coefficient: Work duration ($\theta_{work,i}^{dur}$)	20.860*** (3.261)	37.962*** (7.838)	--	--	7.014
Expected utility (μ_{lunch})	--	--	--	--	--
Converging time	1 s	2 min	27 min	29 min	54 min
Log-likelihood value	-2348.5	-2311.1	-2345.3	-2340.2	--
Lunch choice	**MNL (1K)**	**MXL (1K)**	**DMNL (1K)**	**DMXL (1K)**	**AMXL (1K)**
Lunch schedule early ($\theta_{lunch,i}^{e}$)	-0.867*** (0.105)	-1.164*** (0.193)	-0.840*** (0.020)	-1.227*** (0.044)	-3.722
Lunch schedule delay ($\theta_{lunch,i}^{l}$)	-1.759*** (0.129)	-2.122*** (0.224)	-1.802*** (0.025)	-1.972*** (0.030)	-7.103
Eating inside the CBD ($\theta_{lunch,i}^{des,1}$)	-1.900*** (0.257)	--	-1.882*** (0.049)	-2.060*** (0.323)	-1.180
Eating outside the CBD ($\theta_{lunch,i}^{des,2}$)	--	--	--	--	-0.182
Work-lunch travel time ($\theta_{work-lunch,i}^{t}$)	-0.474*** (0.073)	-1.179** (0.434)	-0.330*** (0.018)	-0.324*** (0.045)	-1.165
Shared coefficient: Work duration ($\theta_{work,i}^{dur}$)	1.661** (0.533)	-4.233*** (0.083)	--	--	7.014
Expected utility ($\mu_{afterwork}$)	--	--	0.854*** (0.111)	1.008** (0.321)	--
Converging time	1 s	2 min	4 min	5 min	54 min
Log-likelihood value	-1841.6	-1807.6	-1836.2	-1820.1	--
Afterwork choice	**MNL (1K)**	**MXL (1K)**	**DMNL (1K)**	**DMXL (1K)**	**AMXL (1K)**
Afterwork-activity duration ($\theta_{afterwork,i}^{dur}$)	5.568*** (0.476)	--	6.398*** (0.096)	6.289*** (0.656)	3.112
Work-afterwork interaction ($\theta_{inter,i}^{dur}$)	10.393*** (0.903)	4.515*** (0.500)	11.892*** (0.182)	16.429*** (1.399)	3.568
Shared coefficient: Work duration ($\theta_{work,i}^{dur}$)	1.973*** (0.164)	-0.497*** (0.024)	2.277*** (0.033)	1.758*** (0.212)	7.014
Converging time	1 s	1 min	1 s	1 min	54 min
Log-likelihood value	-1817.2	-1797.7	-1817.2	-1792.5	--

Note: ***p-value<0.001, **p-value<0.01, *p-value<0.05, all of the observed variables were normalized before modeling

Table 8

Basic statistics of models build with 26K-SHC07 (each entry represents the average value of one estimated coefficient, and the number in the parenthesis is the standard error)

Commute choice	**MNL (26K)**	**MXL (26K)**	**DMNL (26K)**	**DMXL (26K)**	**AMXL (26K)**
Commuting time ($\theta_{commute,i}^{t}$)	-8.877*** (0.110)	-8.852*** (0.192)	-9.650*** (0.116)	-10.399*** (0.148)	-5.407
Commuting cost ($\theta_{commute,i}^{c}$)	-1.005*** (0.185)	-4.212*** (0.368)	-0.581** (0.188)	-1.230*** (0.214)	-0.296
Commuting trip mode ($\theta_{commute,i}^{m}$)	0.374* (0.156)	-2.124*** (0.295)	0.678*** (0.159)	--	0.246
Work schedule early ($\theta_{work,i}^{e}$)	-6.933*** (0.165)	-10.913*** (0.366)	-11.228*** (0.323)	-12.139*** (0.344)	-4.315
Work schedule delay ($\theta_{work,i}^{l}$)	-5.324*** (0.335)	-5.233*** (0.500)	-10.836*** (0.693)	-11.673*** (0.779)	-7.691
Late-for-work penalty ($\theta_{work,i}^{pl}$)	--	--	--	--	-0.322
Shared coefficient: Work duration ($\theta_{work,i}^{dur}$)	22.364*** (0.632)	36.327*** (1.357)	--	--	9.269
Expected utility (μ_{lunch})	--	--	3.494*** (0.140)	1.983*** (0.077)	--
Converging time	7 s	89 min	13.0 hours	13.5 hours	28.9 hours

Log-likelihood value	-61184	-60585	-60862	-60293	--
Lunch choice	**MNL (26K)**	**MXL (26K)**	**DMNL (26K)**	**DMXL (26K)**	**AMXL (26K)**
Lunch schedule early ($\theta^{e}_{lunch,i}$)	-0.840*** (0.020)	-1.234*** (0.065)	-0.840*** (0.020)	-1.227*** (0.044)	-3.761
Lunch schedule delay ($\theta^{l}_{lunch,i}$)	-1.802*** (0.025)	-2.068*** (0.079)	-1.802*** (0.025)	-1.972*** (0.030)	-7.147
Eating inside the CBD ($\theta^{des,1}_{lunch,i}$)	-1.927*** (0.048)	-2.719*** (0.747)	-1.882*** (0.049)	-2.060*** (0.323)	-1.197
Eating outside the CBD ($\theta^{des,2}_{lunch,i}$)	--	--	--	--	-0.012
Work-lunch travel time ($\theta^{t}_{work-lunch,i}$)	-0.429*** (0.014)	-2.776*** (0.805)	-0.330*** (0.018)	-0.324*** (0.045)	-1.002
Shared coefficient: Work duration ($\theta^{dur}_{work,i}$)	1.421*** (0.099)	12.129** (3.912)	--	--	9.269
Expected utility ($\mu_{afterwork}$)	--	--	0.854*** (0.111)	1.008** (0.321)	--
Converging time	7 s	26 min	30 min	35 min	28.9 hours
Log-likelihood value	-48875	-48283	-48844	-48185	--
Afterwork choice	**MNL (26K)**	**MXL (26K)**	**DMNL (26K)**	**DMXL (26K)**	**AMXL (26K)**
Afterwork-activity duration ($\theta^{dur}_{afterwork,i}$)	6.398*** (0.096)	8.415*** (0.005)	6.398*** (0.096)	6.289*** (0.656)	4.502
Work-afterwork interaction ($\theta^{dur}_{inter,i}$)	11.892*** (0.182)	16.450*** (0.019)	11.892*** (0.182)	16.429*** (1.399)	3.106
Shared coefficient: Work duration ($\theta^{dur}_{work,i}$)	2.277*** (0.033)	2.908*** (0.001)	2.277*** (0.033)	1.758*** (0.212)	9.269
Converging time	3 s	7 min	3 s	7 min	28.9 hours
Log-likelihood value	-46156	-45753	-46156	-45540	--

Note: ***p-value<0.001, **p-value<0.01, *p-value<0.05, all of the observed variables were normalized before modeling

5.1.2 Distribution of individual-specific coefficients

The AMXL model with 26K-SHC07 dataset converged at the 84th iteration of Algorithm 2 (Fig. 5 (a)-(c)), resulting in calibrated coefficients per individual that are empirically derived, revealing them to be neither Gumbel nor Gaussian. Instead, the empirical distribution seems to be a combination of a constant (assumed in MNL, DMNL) and Gaussian distribution (assumed in MXL, DMXL).

According to Fig. 5 (d)-(f), coefficients can be divided into three categories: (1) highly-concentrated coefficients with non-zero mean values, such as commuting time (t_commute), work schedule delay (l_work), lunch schedule delay (l_lunch). These coefficients are concentrated around their mean values with small variations, reflecting homogeneous tastes among individuals; (2) even-distributed coefficients, such as work duration (dur_work), afterwork-activity duration (dur_afterwork), work schedule early (e_work), lunch schedule early (e_lunch). These coefficients have larger variations, reflecting heterogeneous tastes among individuals, and; (3) highly-concentrated coefficients with mean values close to zero, such as late-for-work penalty (pl_work), eating outside the CBD (des2_lunch). These coefficients imply that individuals are insensitive to related variables, and they are also insignificant in benchmark models. To this end, AMXL provides a flexible approach for modelers to capture inter-individual homogeneities and heterogeneities, which are infeasible in DCMs and dynamic DCMs.

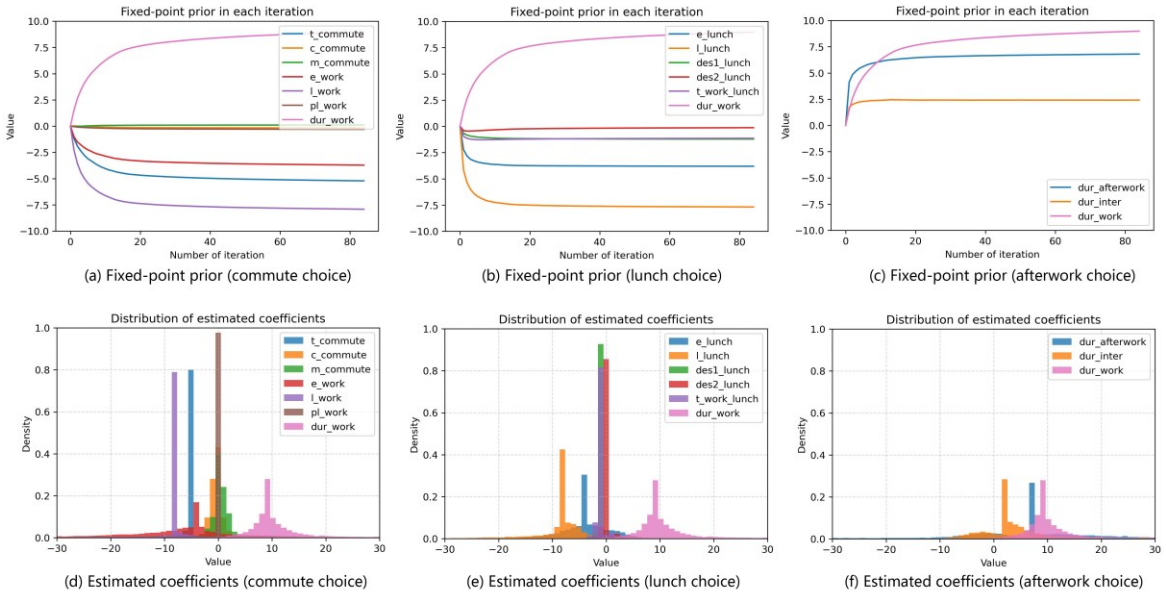


Fig. 5. Fixed-point prior in each iteration and distribution of estimated coefficients. In (a)-(c), x-axis is the number of iterations, y-axis is the value of fixed-point prior θ_0. In (d)-(f), x-axis is the value of estimated coefficients, y-axis is the probability density.

Specifically, Fig.6 presents the probability density function (PDF) of three selected coefficients: (1) a highly-concentrated coefficient, commuting time, (2) an even-distributed coefficient, work schedule early, and (3) the shared coefficient, work duration. Interestingly, MXL reported three different PDFs of work duration in commute, lunch, and afterwork choice, while the PDF of work duration obtained by AMXL is in the middle of them. One intuition from the results is that the shared coefficient might be a combination of coefficients in a series of models estimated separately.

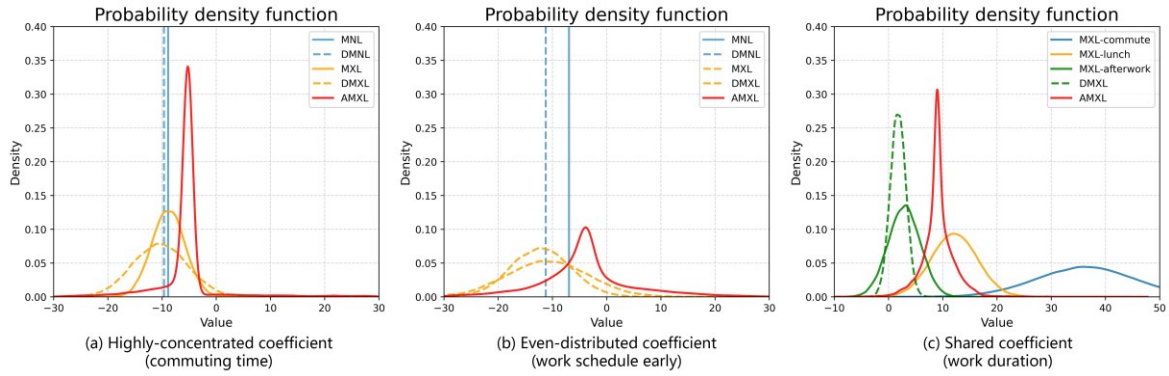


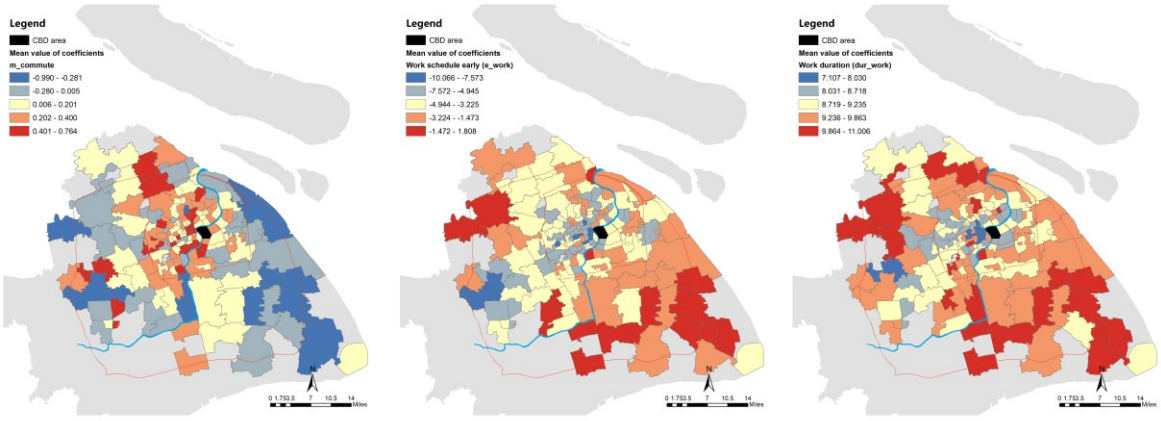
Fig. 6. Probability density function (PDF) of selected coefficients. In (a)-(c), x-axis is the value of estimated coefficients, y-axis is the probability density.

Moreover, for these even-distributed coefficients, we aggregate individuals into community tracts according to their home location and calculate the mean values of the estimated coefficients in each tract. Fig. 7 presents the spatial distribution of selected coefficients, which is infeasible for DCMs and dynamic DCMs to capture. The spatial distribution of coefficients reveals the impacts of the transit system and jobs-housing balance on individual tastes:

(1) The coefficient of mode preference is positive (prefer transit) in communities with better transit accessibility, and negative (prefer driving) in suburbs where metro lines cannot cover.

(2) The negative utilities of starting work earlier than schedule are larger in communities closer to the CBD, indicating that individuals living nearby the CBD dislike arriving early for work while individuals living in suburbs can better accept it considering the uncertainty of the travel time.

(3) The positive utilities of work duration are larger in communities farther way from the CBD, indicating that individuals living in suburbs put greater emphases on their working time, probably due to the longer time spent on commuting.



(a) Mode preference (m_commute) (b) Work schedule early (e_work) (c) Work duration (dur_work)

Fig. 7. Spatial distribution of selected coefficients.

5.1.3 Prediction accuracy

We compare the prediction accuracy of AMXL and benchmark models at the aggregated level and individual level. Aggregated-level predictions of each alternative are made by summing up individuals' probabilities of choosing that alternative (for AMXL we obtain probabilities by applying softmax function to retrieved utilities). Individual-level predictions are made by selecting the alternative with the highest utility retrieved from the models. Moreover, we use 26K-SHC07 datasets to calculate in-sample accuracy and use SHC14 dataset to calculate out-of-sample accuracy. Table 9 shows a comparison of prediction accuracy. We find that the in-sample accuracy of AMXL is generally higher than DCMs and dynamic DCMs, especially at the individual level. This is because DCMs and dynamic DCMs find coefficients to maximize the overall likelihood function, while AMXL is fitted via IO to each individual. Moreover, the accuracy of MNL and MXL drops significantly when it comes to the whole-day schedule prediction, owing to the ignorance of inter-relationship between sub-choice scenarios. Though DMNL and DMXL consider expected utilities of downstream choices, the improvement on prediction accuracy is trivial. That is probably because DMNL and DMXL link sub-choices through pre-assumed conditional probability, which is hard to fit an empirical joint distribution. In contrast, AMXL considerably improves the in-sample accuracy of whole-day schedule prediction, from 9.71% to 80.89% at the aggregated level and from 2.12% to 47.18% at the individual level.

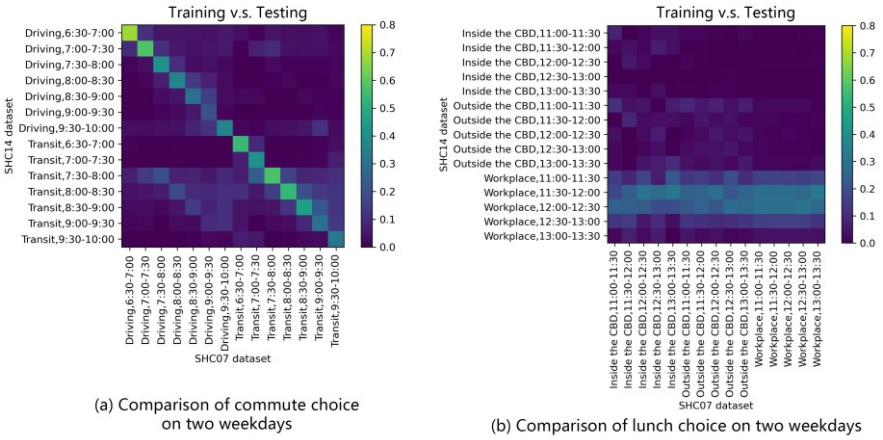
Table 9
In-sample and out-of-sample prediction accuracy

	Commute choice (14 alternatives)	Lunch choice (15 alternatives)	Afterwork choice (7 alternatives)	Whole-day schedule (1,470 alternatives)
In-sample, aggregated-level accuracy				
MNL	81.40%	85.53%	92.88%	7.50%
MXL	82.05%	89.68%	91.55%	7.37%
DMNL	82.81%	85.53%	92.90%	7.47%
DMXL	82.61%	85.97%	91.55%	9.71%
AMXL	89.61%	86.71%	98.87%	80.89%
In-sample, individual-level accuracy				
MNL	13.70%	31.16%	35.76%	1.37%
MXL	13.69%	32.75%	35.76%	1.65%

DMNL	18.21%	32.21%	35.76%	1.97%
DMXL	17.85%	34.45%	34.66%	2.12%
AMXL	77.16%	78.43%	80.93%	47.18%
Out-of-sample, aggregated-level accuracy				
MNL	82.50%	89.92%	89.75%	5.21%
MXL	81.72%	93.08%	94.81%	5.87%
DMNL	83.89%	89.90%	89.68%	5.62%
DMXL	83.51%	90.11%	93.16%	8.66%
AMXL	75.79%	86.73%	96.07%	61.68%
Out-of-sample, individual-level accuracy				
MNL	13.53%	27.93%	28.99%	1.06%
MXL	13.62%	32.02%	28.99%	1.39%
DMNL	18.63%	27.93%	28.99%	1.60%
DMXL	18.30%	32.02%	28.90%	1.69%
AMXL	30.74%	24.25%	37.71%	4.33%

Note: In-sample accuracy is calculated using 26K-SHC07 dataset. Out-of-sample accuracy is calculated using SHC14 dataset.

When it comes to out-of-sample accuracy, however, the performance of AMXL at the individual level drops significantly, indicating a high risk of overfitting though we have already added random items to avoid it. Also, AMXL does not outperform benchmark models in all sub-choice scenarios. For instance, AMXL obtains an out-of-sample accuracy of 86.73% at the aggregated level and 24.25% at the individual level, which is lower than the best performance in benchmark models (93.08% and 32.02%, respectively). To further examine how overfitting happens in our experiments, we first check individual's schedule change from SHC07 dataset to SHC14 dataset, and then plot the prediction results of commute choice and lunch choice. Fig. 8 presents change of commute and lunch choice between two different weekdays. A relatively clear diagonal only exists in Fig.8 (a), indicating that commute choice is more stable than lunch choice among different weekdays. In general, 51.51% individuals changed their commute choice while 77.87% individuals changed their lunch choice, where a majority of them set their activity schedule half an hour earlier/later. Considering traffic conditions and restaurants are unlikely to change greatly between two adjacent Tuesdays, the changed choice indicates the variability of individual preferences among different days, which brings high overfitting risk to models fitting one observation per individual. Fig.9 and Fig.10 show more details of commute and lunch choice prediction, from which we can find a more obvious overfitting in lunch choice than commute choice, probably because individuals' preference for lunch activities are more flexible.



(a) Comparison of commute choice on two weekdays

(b) Comparison of lunch choice on two weekdays

Fig. 8. Comparison of commute and lunch choice on two different weekdays. x-axis is the alternative on 2019 May 7th, y-axis is the alternative on 2019 May 14th, the value in each cell is normalized by row.

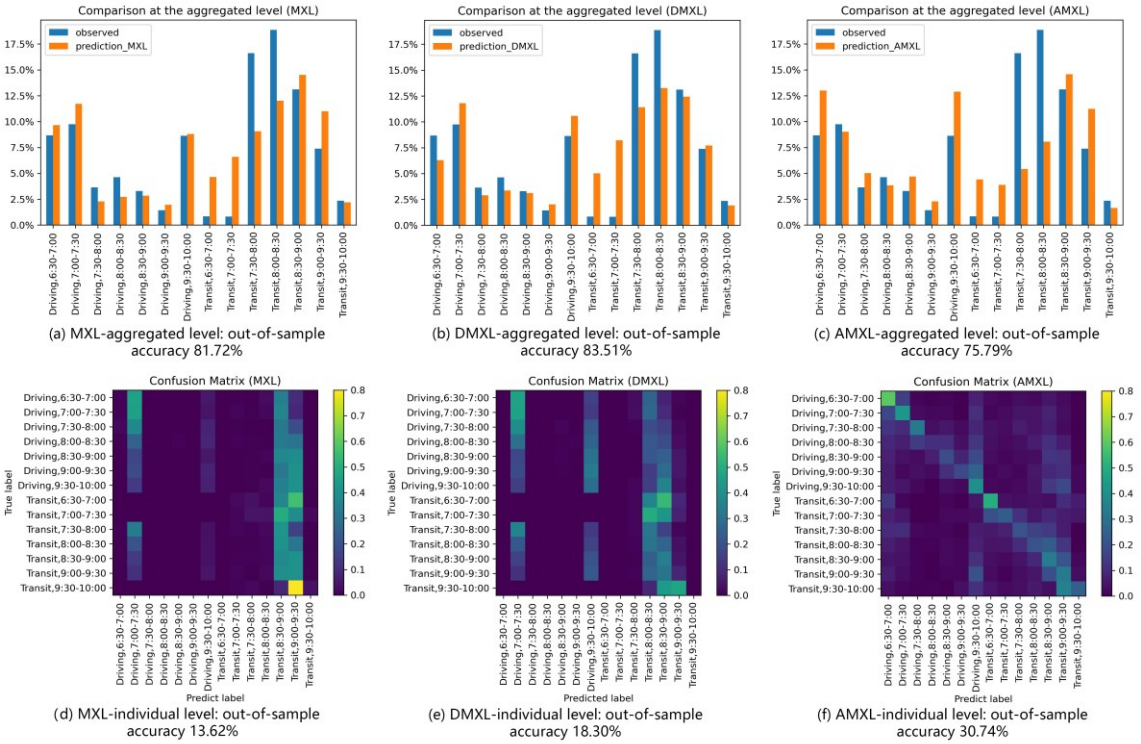


Fig. 9. Commute choice predicted by MNL, DMXL, and AMXL. In (a)-(c), x-axis is the alternative, y-axis is the percentage of individuals choosing the alternative. (d)-(f) are confusion matrixes.

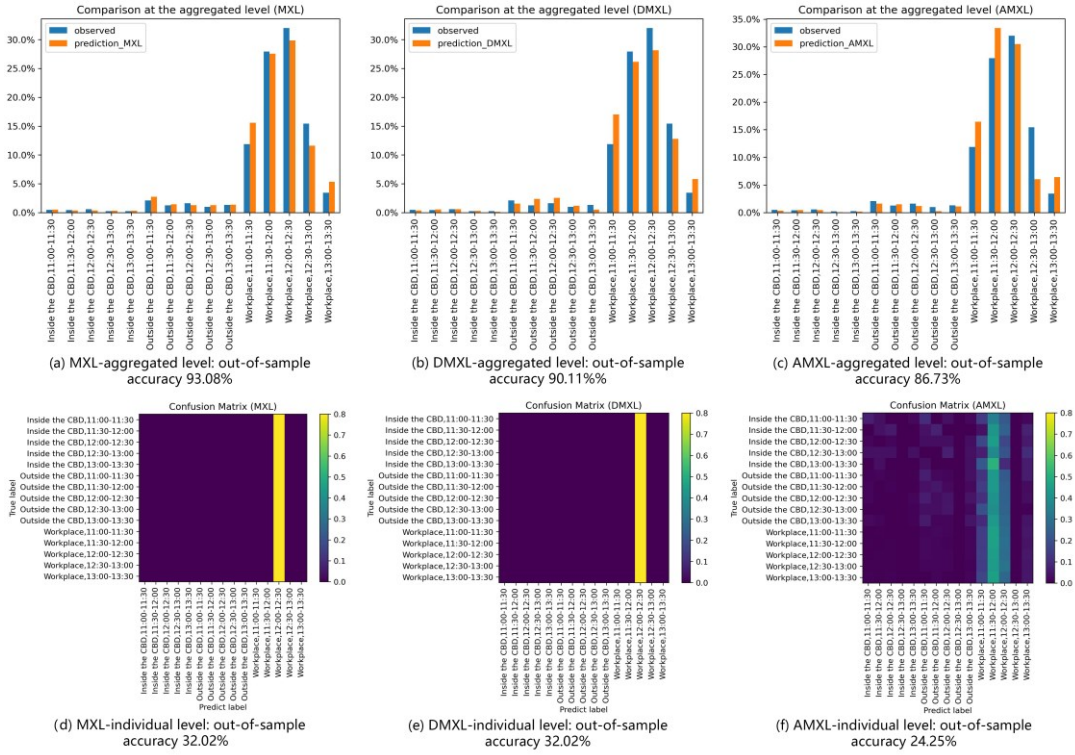


Fig. 10. Lunch choice predicted by MNL, DMXL, and AMXL. In (a)-(c), x-axis is the alternative, y-axis is the percentage of individuals choosing the alternative. (d)-(f) are confusion matrixes.

Notwithstanding overfitting issues at the individual level, AMXL also presents several advantages. (1) AMXL brings a considerable improvement on predicting choice dimensions with stable preferences such as mode choice. For instance, if an individual chose to drive and depart during 7:00 a.m.-7:30 a.m., the most common mistake made by AMXL is to predict "Driving, 6:30-7:00". From Fig.9 (d)-(f) we can find a vague diagonal in the confusion matrix of AMXL, while confusion matrixes of MXL and DMXL are chaotic. (2) AMXL provides more information than MXL and DMXL in choice dimensions with flexible preferences. From Fig.10 (d)-(f) we can find MXL and DMXL are similar to constant models, predicting that all individuals will choose "Workplace, 12:00-12:30", while AMXL gives a time period from 11:00 to 12:30. Though additional information do not bring higher performance in our case, it might help given a dataset with balanced label and personal attributes. (3) AMXL is meaningful in the whole-day activity schedule prediction at least at the aggregated-level, considering its high accuracy of 61.68% compared with 8.66% in DMXL. To this end, we should be careful about the practical applications of AMXL: (1) It can be applied to predicting whole-day activity schedule at the aggregated level; (2) At the individual level, its current form is only valid in predicting choice dimensions with stable preferences; and (3) To predict individual-level choice dimensions with flexible preferences, AMXL should be extended to fit multi-day observations per individual. We share our thoughts of the extended form in the final section.

5.2 Scenario applications

We design two scenarios in this section. The first scenario is to compare the predictions of AMXL and benchmark models. The second scenario is to showcase the capability of AMXL in system design and revenue management.

5.2.1 Scenario1: Peak-hour congestion

In this scenario, we decrease the driving time within 7:30 a.m.-9:30 a.m. by 10% and 20%, simulating a relief of the peak-hour congestion. We compare the prediction of MXL (the worst model in benchmarks), DMXL (the best model in benchmarks), and AMXL. Fig. 9 shows the results. Generally, individuals changed their mode choice from transit to driving during peak hour. The proportion of driving in peak-hour increased by 3% and 7% after the driving time decreased by 10% and 20%.

Moreover, it is interesting to compare the prediction results of benchmark models and AMXL: time-schedule shift accounts for a larger proportion in the choice shift predicted by AMXL, mode shift is larger in MNL, DMXL is in the middle of MNL and AMXL. To be specific, the choice shift predicted by AMXL comes from individuals who previously departed slightly earlier or later than the peak hour to avoid congestion. On the other hand, individuals choosing other alternatives generally decreased in the prediction of MNL and DMXL. To this end, the prediction of AMXL is slightly different from benchmark models, but is reasonable from a behavioral perspective.

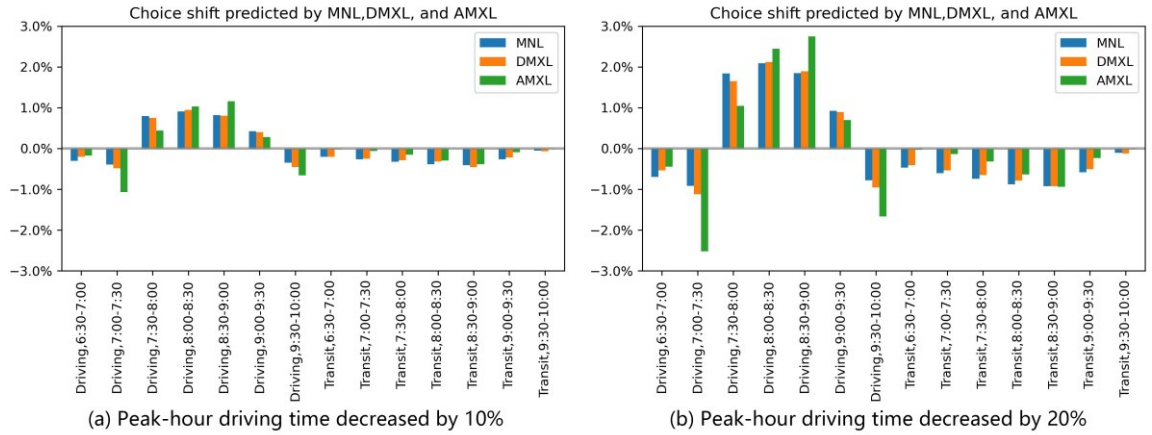


Fig. 11. Percentage of choice shift after the peak-hour driving time decreased by 10% and 20%. In (a)-(b), x-axis is the alternative, y-axis is the percentage change of commuters choosing the alternative.

5.2.2 Scenario2: Sending lunch coupons to attract commuters

This scenario is to check the compatibility of AMXL with optimization models applicable to system design and revenue management. Let us assume that a new restaurant will be built in a specific street-block in the CBD, whose location is shown in Fig. 10. To attract commuters, the restaurant manager tends to send lunch coupons to commuters in the CBD. The coupon is a kind of limited resource, and they can only be sent to a finite number of street blocks in the CBD considering the labor cost. The manager needs to decide which street blocks to send coupons to maximize profits.

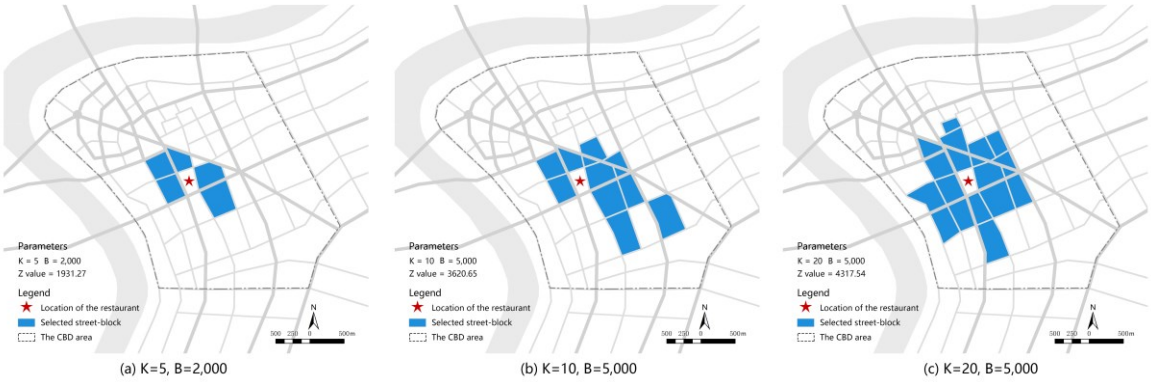


Fig. 12. A summary of optimal solutions given different *K* and *B*.

This is a typical binary programming (BP) problem in which the "profits" can be calculated from our proposed AMXL model if the manager had access to this data. The BP problem can be formulated as Eqs. (35) – (39).

$$\max Z = c_1 x_1 + c_2 x_2 + \cdots + c_{80} x_{80} \tag{35}$$

subject to:

$$c_b = \sum_{i \in P_b} \left(\theta^{t}_{work-lunch,i} \Delta t_{work-lunch,i} + \theta^{des,1}_{lunch,i} \Delta des^{1}_{lunch,i}\right), \qquad b = 1,\ldots,80 \tag{36}$$

$$|P_1| x_1 + |P_2| x_2 + \cdots + |P_3| x_{80} \le B \tag{37}$$

$$x_1 + x_2 + \cdots + x_{80} \leq K \quad (38)$$

$$x_1, x_2, \ldots, x_{80} \in \{0,1\} \quad (39)$$

where $x_1, x_2, \ldots, x_{80}$ are binary decision variables indicating whether each of the 80 street-blocks in total is selected to send coupons. The coefficients of these decision variables in the objective function are directly calculated by an estimated AMXL. Specifically, Eq. (36) defines the profit obtained by sending coupons to street block b, c_b, which can be reflected by the summation of increased utilities derived from commuters having lunch at the restaurant with coupons. For each commuter i working in street-block b ($i \in P_b$), we assume having lunch with coupons increases the utility from negative to 0 (the same as the reference group, having lunch in the workplace), by setting $\Delta des^1_{lunch,i}$ to -1 (hence $\theta^{des,1}_{lunch,i}\Delta des^1_{lunch,i}$ is positive). In addition, having lunch outside the workplace will cost a travel time distance between the workplace and the lunch spot, whose negative utility is denoted as $\theta^t_{work-lunch,i}\Delta t_{work-lunch,i}$. Eq. (37) ensures that the total coupons sent out is no more than a budget B, and Eq. (38) ensures that the total selected street-blocks is no more than K.

Fig. 10 summarizes the optimal solutions of this BP problem given different K and B. When $K = 5$ and $B = 2{,}000$, five street blocks surround the restaurant are selected, and the total increased utility equals 1931.27. When $K = 10$ and $B = 5{,}000$, the optimal solution suggested to select larger street blocks in the south of the restaurant (limited by K), and the total increased utility equals 3620.65. When $K = 20$ and $B = 5{,}000$, the model chooses relatively small street blocks (limited by B) and increases the objective value to 4317.54, which is a 19.25% increase on the base of the second strategy.

6. Conclusion

Demand for agent-based modeling is on the rise, and understanding the joint activity scheduling choice requires innovative methods to deal with big datasets and large choice sets. We propose an agent-based mixed-logit model (AMXL) combined with an inverse optimization (IO) estimation method, an agent-level machine learning method that is theoretically constituent with a utility-maximizing mixed logit model framework. This method is designed for a ubiquitous dataset representing a whole population which is possible with big data. It contributes to the study field by overcoming three limitations of conventional DCMs given ubiquitous datasets.

The methodological contributions of this study are as follows. First, we decompose the whole-day scheduling choice into a series of inter-related sub-choices and jointly estimated by the AMXL model through fixing shared coefficients. By doing this, we reduce the size of the choice set from a product of alternatives in each choice dimension to a summation of them. Second, AMXL provides individual-specific estimation, allowing modelers to obtain empirical distributions of coefficients given observation datasets. The experimental results based on 26,149 samples show that empirical distributions of coefficients are neither fixed-point nor Gaussian. Instead, it seems to be a combination of them. These two improvements increase the prediction accuracy from 8.66% (DMXL) to 61.68% (AMXL). Third, the deterministic estimation in AMXL provides linear demand functions that can be integrated into optimization models. Estimated coefficients in AMXL can be directly used to calculate coefficients in an optimization problem, which is more efficient than relying on simulation.

The overfitting issue is essential to the value of AMXL model. Our experimental results indeed show evidence of overfitting, which is more obvious in choice dimensions with flexible preferences (e.g., time to have lunch and lunch location). This is because AMXL cannot capture the intra-individual heterogeneity on multiple days. Nevertheless, AMXL is still useful in predicting the whole-day schedule at the aggregated level and predicting choice dimensions with stable preferences at the individual level. Actually, the overfitting issue can be reduced by formulating multiday inverse utility maximization problems, in which coefficients are no longer perturbed to fit one observation

per individual. Instead, multiday observations per individual can be grouped together to estimate individual-specific coefficients. This can be fulfilled by modifying the formulation of IO at the beginning, such as relaxing the constraints of utility ranking and include a log-likelihood-formed item into the objective function. Our future study will focus on the extended AMXL model that considers both intra- and inter-individual heterogeneity.

Besides the overfitting issue, there are many new research opportunities and questions to be addressed. While the study looks only at commuters' work, lunch, and afterwork activities, the model can be customized to examine more diverse activity schedules specific to population segments. Also, information for economic interpretation obtained by AMXL and DCMs, such as elasticity, marginal rate of substitution, and change of social welfare, should be compared in detail. Last but not least, AMXL takes a much longer time to converge compared with MNL and MXL. Though paralleling computation and pre-trained priors can help to reduce the time cost, a more efficient way might be first categorizing individuals into groups and estimating a set of coefficients for each group after that. As discussed in Section 3.2.1, agents can also represent segments of the population, in which multiple observations from the same segment would share the same set of coefficients. This can make the proposed method applicable to a much wider set of choice modeling scenarios, such as simultaneously estimating a large set of segment models to represent a whole population. Estimating coefficients at a segment level by constraining the agents sharing the same segment will be another direction of our future study.

Acknowledgments

This research is partially supported by NSF CMMI-1652735 and the C2SMART Center (USDOT #69A3551747124). The authors wish to express their thanks to Professor De Wang's team in Tongji University for providing the dataset for experiments.

References

Adler, T., & Ben-Akiva, M. (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, 13(3), 243–257.

Aguiléra, A. (2018). Smartphone and individual travel behavior. In *Urban Mobility and the Smartphone: Transportation, Travel Behavior and Public Policy*. Elsevier Inc.

Aguirregabiria, V., & Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1), 38–67.

Ahas, R., Silm, S., Saluveer, E., & Järv, O. (2009). Modelling Home and Work Locations of Populations Using Passive Mobile Positioning Data. *Lecture Notes in Geoinformation and Cartography*, 0(199079), 301–315.

Ahuja, R. K., & Orlin, J. B. (2001). Inverse Optimization. *Operations Research*, 49(5), 771–783.

Axhausen, K., Horni, A., & Nagel, K. (2016). *The multi-agent transport simulation MATSim*. Ubiquity Press.

Becker, F., Danaf, M., Song, X., Atasoy, B., & Ben-Akiva, M. (2018). Bayesian estimator for logit mixtures with inter-and intra-consumer heterogeneity. *Transportation Research Part B: Methodological*, 117, 1–17.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9). MIT press.

Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35(1), 1–28.

Budziński, W., Campbell, D., Czajkowski, M., Demšar, U., & Hanley, N. (2018). Using Geographically Weighted Choice Models to Account for the Spatial Heterogeneity of Preferences. *Journal of Agricultural Economics*, 69(3), 606–626.

Burton, D., & Toint, P. L. (1992). On an instance of the inverse shortest paths problem. *Mathematical Programming* 1992 53:1, 53(1), 45–61.

Chan, T. C., & Kaw, N. (2020). Inverse optimization for the recovery of constraint parameters. *European Journal of Operational Research*, 282(2), 415-427.

Charypar, D., & Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation* 2005 32:4, 32(4), 369–397.

Chen, Y.-C., DUANN, L.-S., & HU, W.-P. (2005). The estimation of Discrete choice models with large choice set. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 1724–1739.

Chow, J. Y. J., & Recker, W. W. (2012). Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. *Transportation Research Part B: Methodological*, 46(3), 463–479.

Chow, J. Y. J., & Djavadian, S. (2015). Activity-based market equilibrium for capacitated multimodal transport systems. *Transportation Research Part C: Emerging Technologies*, *59*, 2-18.

Chow, J. Y. J. (2018). *Informed Urban transport systems: Classic and emerging mobility methods toward smart cities*. Elsevier.

Cordone, R., & Redaelli, F. (2011). Optimizing the demand captured by a railway system with a regular timetable. *Transportation Research Part B: Methodological*, 45(2), 430–446.

Danalet, A., Bierlaire, M., Verità, M., & Ascona. (2015). Importance sampling for activity path choice. In 15th Swiss Transport Research Conference.

Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, 100, 65–80.

Dong, X., Chow, J. Y. J., Waller, S. T., & Rey, D. (2022). A chance-constrained dial-a-ride problem with utility-maximising demand and multiple pricing structures. *Transportation Research Part E: Logistics and Transportation Review*, 158, 102601.

Dumont, J., Giergiczny, M., & Hess, S. (2015). Individual level models vs. sample level models: contrasts and mutual benefits. *Transportmetrica A: Transport Science*, 11(6), 465–483.

Ettema, D., Bastin, F., Polak, J., & Ashiru, O. (2007). Modelling the joint choice of activity timing and duration. *Transportation Research Part A: Policy and Practice*, 41(9), 827–841.

Ghobadi, K., & Mahmoudzadeh, H. (2021). Inferring linear feasible regions using inverse optimization. *European Journal of Operational Research*, 290(3), 829-843.

Gilbert, F., Marcotte, P., & Savard, G. (2015). A Numerical Study of the Logit Network Pricing Problem. *Transportation Science*, 49(3), 706–719.

Habib, K. M. N., & Miller, E. J. (2006). Modelling workers‟ skeleton travel-activity schedules. *Transportation Research Record*, 1985, 88–97.

Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, 273–282.

Hasnine, M. S., & Habib, K. N. (2018). What about the dynamics in daily travel mode choices? A dynamic discrete choice approach for tour-based mode choice modelling. *Transport Policy*, 71, 70–80.

He, B. Y., Zhou, J., Ma, Z., Chow, J. Y. J., & Ozbay, K. (2020). Evaluation of city-scale built environment policies in New York City with an emerging-mobility-accessible synthetic population. *Transportation Research Part A: Policy and Practice*, 141, 444–467.

Hess, S. (2010). Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. *Journal of Choice Modelling*, 3(2), 134–152.

Hess, S., & Hensher, D. A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological*, 44(6), 781–790.

Iraj, E. H., & Terekhov, D. (2021). Comparing Inverse Optimization and Machine Learning Methods for Imputing a Convex Objective Function. *arXiv preprint arXiv:2102.10742*.

Iskhakov, F., Rust, J., & Schjerning, B. (2020). Machine learning and structural econometrics: contrasts and synergies. *The Econometrics Journal*, 23(3), S81–S124.

Kancharla, S. R., & Ramadurai, G. (2020). Electric vehicle routing problem with non-linear charging and load-dependent discharging. *Expert Systems with Applications*, 160, 113714.

Kline, D. M., & Berardi, V. L. (2005). Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Computing and Applications*, 14(4), 310–318.

Krueger, R., Bierlaire, M., Daziano, R. A., Rashidi, T. H., & Bansal, P. (2021). Evaluating the predictive abilities of mixed logit models with unobserved inter-and intra-individual heterogeneity. *Journal of Choice Modelling*, 41, 100323.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature* 2015 521:7553, 521(7553), 436–444.

Lemp, J. D., & Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice*, 46(3), 602–613.

Liao, Q., & Poggio, T. (2018). When Is Handcrafting Not a Curse? Center for Brains, Minds, and Machines, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA.

Liu, H. X., He, X., He, B., Liu, H. X., He, X., & He, B. (2007). Method of Successive Weighted Averages (MSWA) and Self-Regulated Averaging Schemes for Solving Stochastic User Equilibrium Problem. *Networks and Spatial Economics* 9(4), 485–503.

Lizana, P., Ortúzar, J. de D., Arellana, J., & Rizzi, L. I. (2021). Forecasting with a joint mode/time-of-day choice model based on combined RP and SC data. *Transportation Research Part A: Policy and Practice*, 150, 302–316.

Ljubić, I., & Moreno, E. (2018). Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. *European Journal of Operational Research*, 266(1), 46–56.

Ma, T. Y., Chow, J. Y. J., & Xu, J. (2017). Causal structure learning for travel mode choice using structural restrictions and model averaging algorithm. *Transportmetrica A: Transport Science*, 13(4), 299-325.

Nurul Habib, K. (2018). A comprehensive utility-based system of activity-travel scheduling options modelling (CUSTOM) for worker's daily activity scheduling processes. *Transportmetrica A: Transport Science*, 14(4), 292–315.

Omrani, H. (2015). Predicting Travel Mode of Individuals by Machine Learning. *Transportation Research Procedia*, 10, 840–849.

Pacheco, M. P., Bierlaire, M., Gendron, B., & Sharif Azadeh, S. (2021). Integrating advanced discrete choice models in mixed integer linear optimization. *Transportation Research Part B: Methodological*, 146, 26–49.

Pougala, J., Hillel, T., & Bierlaire, M. (2021). Choice set generation for activity-based models. Swiss Transport Research Conference.

Pulugurta, S., Arun, A., & Errampalli, M. (2013). Use of Artificial Intelligence for Mode Choice Analysis and Comparison with Traditional Multinomial Logit Model. Procedia - Social and Behavioral Sciences, 104, 583–592.

Qiu, F., & Wang, J. (2015). Distributionally Robust Congestion Management with Dynamic Line Ratings. *IEEE Transactions on Power Systems*, 30(4), 2198–2199.

Replica (2022). https://replicahq.com/, last accessed Aug. 8, 2022.

Richter, L. L., & Pollitt, M. G. (2018). Which smart electricity service contracts will consumers accept? The demand for compensation in a platform market. *Energy Economics*, 72, 436–450.

Robenek, T., Azadeh, S. S., Maknoon, Y., de Lapparent, M., & Bierlaire, M. (2018). Train timetable design under elastic passenger demand. *Transportation Research Part B: Methodological,* 111, 19–38.

Sarrias, M. (2020). Individual-specific posterior distributions from Mixed Logit models: Properties, limitations and diagnostic checks. *Journal of Choice Modelling*, 36, 100224.

Shaaban, K., & Pande, A. (2016). Classification tree analysis of factors affecting parking choices in Qatar. *Case Studies on Transport Policy*, 4(2), 88–95.

Sun, J., Guo, J., Wu, X., Zhu, Q., Wu, D., Xian, K., & Zhou, X. (2019). Analyzing the Impact of Traffic Congestion Mitigation: From an Explainable Neural Network Learning Framework to Marginal Effect Analyses. *Sensors* 19(10), 2254.

Tan, Y., Delong, A., & Terekhov, D. (2019). Deep Inverse Optimization. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11494 LNCS, 540–556.

Tanwanichkul, L., Kaewwichian, P., & Pitaksringkarn, J. (2019). Car ownership demand modeling using machine learning: Decision trees and neural networks. *GEOMATE Journal*, 17(62), 219–230.

Tribby, C. P., Miller, H. J., Brown, B. B., Werner, C. M., & Smith, K. R. (2017). Analyzing Walking Route Choice through Built Environments using Random Forests and Discrete Choice Techniques. *Environment and Planning. B, Urban Analytics and City Science*, 44(6), 1145–1167.

Västberg, O. B., Karlström, A., Jonsson, D., & Sundberg, M. (2020). A dynamic discrete choice activity-based travel demand model. *Transportation Science*, 54(1), 21–41.

Wang, S., Mo, B., & Zhao, J. (2020). Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. *Transportation Research Part C: Emerging Technologies*, 112, 234–251.

Wang, S., Wang, Q., & Zhao, J. (2020). Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C: Emerging Technologies*, 118, 102701.

Xu, S. J., Nourinejad, M., Lai, X., & Chow, J. Y. J. (2018). Network learning via multiagent inverse transportation problems. *Transportation Science*, 52(6), 1347–1364.

Zhao, Y., Pawlak, J., & Polak, J. W. (2018). Inverse discrete choice modelling: theoretical and practical considerations for imputing respondent attributes from the patterns of observed choices. *Transportation Planning and Technology*, 41(1), 58–79.