Journal of Choice Modelling 31 (2019) 35–50

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

# Context-aware stated preferences with smartphone-based travel surveys

Mazen Danaf[a,*], Bilge Atasoy[b], Carlos Lima de Azevedo[c], Jing Ding-Mastera[a], Maya Abou-Zeid[d], Nathaniel Cox[a], Fang Zhao[e], Moshe Ben-Akiva[f]

[a] *Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA*
[b] *Department of Maritime and Transport Technology, Delft University of Technology, Mekelweg 2, Delft, 2628, CD, the Netherlands*
[c] *Technical University of Denmark, Anker Engelunds Vej 1 Bygning 101A, 2800 Kgs. Lyngby, Denmark*
[d] *American University of Beirut, Bechtel Engineering Building 527, Lebanon*
[e] *Singapore-MIT Alliance for Research and Technology, 1 CREATE Way, #10-01 CREATE Tower, 138602, Singapore*
[f] *Edmund K. Turner Professor of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA*

A R T I C L E I N F O

*Keywords:*
Stated preferences
Smartphone
Mode choice
Context-aware
Random design
Travel behavior

A B S T R A C T

Stated preferences surveys are most commonly used to provide behavioral insights on hypothetical travel scenarios such as new transportation services or attribute ranges beyond those observed in existing conditions. When designing SP surveys, considerable care is needed to balance the statistical objectives with the realism of the experiment. This paper presents an innovative method for smartphone-based stated preferences (SP) surveys leveraging state-of-the-art smartphone-based survey platforms and their revealed preferences sensing capabilities. A random experimental design generates context-aware SP profiles using user specific socio-economic characteristics and past travel data along with relevant web data for scenario generation. The generated choice tasks are automatically validated to reduce the number of dominant or inferior alternatives in real-time, then validated using Monte-Carlo simulations offline. In this paper we focus our attention on mode choice and design an experiment that considers a wide range of possible existing mode alternatives along with a new alternative on-demand mobility service that does not exist in real life. This experiment is then used to collect SP data or a sample of 224 respondents in the Greater Boston Area. A discrete mode choice model is estimated to illustrate the benefit of the proposed method in capturing current context-specific preferences in response to the new scenario.

## 1. Introduction

The pervasiveness of smartphones coupled with developments in information and communication technologies (ICT) and enhanced computing performance has paved the way for shifting towards "smart mobility". Smart mobility is one of the main pillars of smart cities (Giffinger et al., 2007). It is defined as a combination of improved accessibility, availability of ICT, exploration of new data sources and analytics, and modern sustainable and efficient transportation systems. Indeed, technology has not only been

* Corresponding author.
*E-mail addresses:* mdanaf@mit.edu (M. Danaf), b.atasoy@tudelft.nl (B. Atasoy), climaz@dtu.dk (C.L. de Azevedo), jingding@mit.edu (J. Ding-Mastera), ma202@aub.edu.lb (M. Abou-Zeid), nathanael.cox@gmail.com (N. Cox), fang.zhao@smart.mit.edu (F. Zhao), mba@mit.edu (M. Ben-Akiva).

https://doi.org/10.1016/j.jocm.2019.03.001
Received 15 April 2018; Received in revised form 28 February 2019; Accepted 6 March 2019
Available online 11 March 2019
1755-5345/ 

changing the transportation systems in our cities, but also influencing the way transportation surveys are conducted. The popularity of location-enabled devices has greatly expanded transportation data collection options (Susilo et al., 2016). Along with location-based big-data collection for aggregate mobility patterns (Cottrill et al., 2015; Jiang et al., 2017), smartphones have also been used to collect detailed travel diaries (Cottrill et al., 2013; Susilo et al., 2016; Zhao et al., 2015b) in a cheap and non-intrusive manner (Prelipcean et al., 2015). While these technologies have been well established in collecting trip diaries - or revealed preferences (RP) data, they have not yet been utilized in the collection of mobility related stated preferences (SP) data.

SP data are hypothetically created choice situations in which the researcher has the freedom to define the tradeoffs faced by the respondent (Ben-Akiva et al., 2019; Walker et al., 2015). SP methods were first introduced by Thurstone (1931), who proposed experiments of the form “eight hats and eight pairs of shoes versus six hats and __ pairs of shoes”. In the context of transportation, SP surveys are used to analyze hypothetical scenarios, such as testing attribute ranges beyond those observed in RP data or to infer preferences towards new modes and services.

When designing SP surveys, considerable care is needed to balance the statistical objectives with the realism of the experiment (Ben-Akiva et al., 2019). The statistical objectives involve identifying parameter estimates consistently and with low standard errors. The realism of experiments involves accounting for market, personal, or contextual constraints, and presenting alternatives in the same way as their market framing. These objectives can be met by designing context-aware SP surveys, which pertain to a specific context already faced by the respondent. For example, a transportation mode SP survey would refer to a trip performed by the respondent, but present different alternatives and attributes from those originally experienced by this respondent.

This paper presents a generic method for context-aware SP surveys leveraging state-of-the-art smartphone-based RP methods, and presents its application to mode choice of future smart mobility solutions. The context is coming from the observed RP data, e.g., weekly activity pattern or a selected trip for a given day, together with user specific information, e.g., vehicle ownership, usage of car/bike sharing services, etc. In addition to the direct information obtained from the user, we collect external contextual data such as the available activity or transportation alternatives for the user through online sources. The experimental design uses this context in order to generate SP choice experiments with reasonable alternatives and attributes. By accounting for the trip context (e.g. user-specific considerations, and trip-specific constraints) and using smartphone data, this method overcomes several limitations associated with traditional SP surveys related to data quality, realism, and user experience. This methodology can also be used to estimate preferences towards new mobility solutions which then can be used for the design and operation of these solutions.

The remainder of this paper is organized as follows: section 2 presents a brief literature review on prompted recall surveys, context-aware SP surveys, and survey design. Section 3 provides an overview of our proposed system architecture and outlines the methodology of SP data collection. Section 4 presents its application to a case study on FMOD (Future Mobility on Demand). Section 5 presents a discussion of the contributions and limitations of the proposed method. Finally, Section 6 presents future research directions and concludes the paper.

## 2. Literature review

Our proposed SP method subsumes the advantages of prompted recall SP surveys and smartphone-based, context-aware surveys in order to generate more realistic SP choice tasks. The following sections provide a brief background of prompted recall surveys, context-aware surveys, and SP design methods.

### *2.1. Prompted recall SP surveys*

The main limitation of SP surveys is that they only record choices made in hypothetical scenarios (Fifer et al., 2014). This can result in different types of biases such as inattentiveness, attribute nonattendance, and incongruity with actual (revealed preferences) behavior. For example, different studies have indicated that respondents' willingness to pay tends to be higher in SP surveys than in RP (Fifer et al., 2014; Murphy et al., 2005).

A diverse body of literature suggests different approaches for addressing these biases. “Pivoting” is one such example whereby the attributes in the SP experiment are created by changing the attributes of the chosen RP alternative (Hensher, 2004, 2006; 2008; Train and Wilson, 2008). Rose et al. (2008) analyzed different strategies for designing statistically efficient SP experiments where the attributes of alternatives are pivoted off reference alternatives. Hess and Rose (2009) argue that theories derived in behavioral and cognitive psychology and economics support approaches that relate SP experiments to individual specific experiences and perceptions, such as those obtained from a reference alternative.

When respondents are asked about their previous trips, considerable attention is needed for choosing the reference trip for the SP. Fifer et al. (2011) argue that using a 'typical' trip as a reference is problematic because people recall their trip details very poorly. In addition, this might also result in under-sampling some activities (e.g. recreational activities) as most people tend to choose work or education trips. For example, Zhao et al. (2015a) found that survey respondents tend to under-report short trips. One approach to deal with this issue involves asking people about their “last trip” (Train and Wilson, 2008). Another approach is to use user-specific GPS data, and then choose a trip randomly from the user's observed trips. GPS data has been used by Fifer et al. (2011) in order to study how motorists would react to a distance-based charging system that incorporates the risk of driving (which is a function of the distance travelled, night-time travel, and speeding), and by Matyas and Kamargianni (2017) who conducted an SP experiment to analyze the demand for mobility-as-a-service (MaaS) plans.

### 2.2. Web and mobile based surveys

Web- and mobile-based SP surveys have become increasingly popular as they allow researchers to better describe the setting or context to respondents. Unlike the traditional paper-and-pencil questionnaires, web-based surveys can easily integrate visual and auditory effects which can provide more accurate descriptions of attributes. Smartphone based survey platforms have been developed to collect trip and activity diaries such as Future Mobility Sensing (FMS) (Cotrill et al., 2013; Zhao et al., 2015), the SITSS (Smartphone-based Individual Travel Survey System) (Safi et al., 2015), MEILI (Prelipcean et al., 2015; Susilo et al., 2016), UbiActive (Fan et al., 2013), rMOVE (Resource Systems Group, 2015), and others. These platforms have been mainly used to collect RP data.

On the other hand, several web-based surveys have been recently used to collect SP data for new modes and services. For example, Choudhury et al. (2018) investigated the acceptability of three new and emerging smart mobility options (one way car rental, shared taxi, and park and ride with school bus service) and quantified the associated willingness-to-pay values in Lisbon. Similarly, Correia and Viegas (2011) used a web-based survey in Lisbon to investigate a new carpooling structure in the form of clubs for improving trust among carpoolers. Chebli and Mahmassani (2003) used a web-based SP survey to study air travelers' willingness to adopt new services that could help reduce ground access congestion around airports such as transit, rail, and off-airport terminals.

The “vividness” or “representational richness” in web-based environments has been shown to enhance user attentiveness (Hoffman and Novak, 1996). In terms of SP surveys, Sethurman et al. (2005) showed that web-based surveys are superior to the traditional pencil-and-paper surveys in terms of consistency and face validity, even though the authors tried to make the two experiments (web-based and mail-based) as close as possible. In order to better describe the context, more recent studies have integrated video enhancements in SP surveys (Hoffmann et al., 2014; Jensen et al., 2017). The effects of videos have been shown to enhance attentiveness (as Jensen et al. (2017) reported lower scale parameters when using video effects) and result in more plausible results (Hoffmann et al., 2014) as they provide users with a more vivid description of the choice setting.

Recently, Cox (2015) developed an SP survey which uses the Future Mobility Sensing (FMS) platform (Cottrill et al., 2013; Zhao et al., 2015b) to estimate the demand for new transportation modes and services. This web-based survey is context-aware since it refers to a trip that has already been performed by the respondent. Cox used data from GPS and external sources to generate hypothetical scenarios for a large number of modes using the random design approach.

Similarly, Matyas and Kamargianni (2017) designed an SP experiment in order to capture the decision making process of purchasing mobility-as-a-service (MaaS) products. This experiment allows for inferring people's preferences and willingness to pay for flexibility. Individual-specific data is collected using the FMS platform, and users are presented with their current travel patterns and the frequency of usage of each transport mode after validating their travel diaries for multiple days. The SP survey presents users with four hypothetical MaaS plans, three of which are fixed while the fourth allows the user to customize the product. The attributes of the plans include price, transport modes and usage amounts, mode specific features, transferability of the features to the next month, and prizes associated with the plans. This SP survey is long term, as it uses the respondents' activity patterns collected over a long period of time.

### 2.3. SP design methods

Experimental design is a major consideration in SP surveys. The goal is to achieve good statistical identification of the model parameters and reduce their standard errors. To achieve this, the design needs to allow for a considerable span and linearly independent variation in the attribute levels (Ben-Akiva et al., 2019).

Different design methods have been proposed for SP surveys. The simplest is the full factorial design, which generates choice tasks that span all the possible attribute level combinations. These designs are orthogonal, as the attributes in different choice tasks are uncorrelated. However, the number of choice tasks grows exponentially with the number of attributes which makes these designs difficult to implement in practical applications (Rose and Bliemer, 2009).

The fractional factorial design uses a subset of the full factorial design in order to reduce the number of choice situations while maintaining some desired properties of the full factorial design such as orthogonality. Another approach is the random design, whereby choice tasks are randomly chosen from a full factorial design. This design is easy to implement, and Walker et al. (2015) showed that it performs as well as any design especially when it is modified by eliminating dominant and inferior alternatives.

Orthogonal designs are optimal for linear models (e.g. linear regression) as they produce unbiased estimates with the smallest standard errors. This is because the asymptotic variance-covariance (AVC) matrix in these models is independent of the values of the parameters. However, they are not necessarily efficient for non-linear models, such as discrete choice models (Kuhfeld et al., 1994). For example, Bliemer et al. (2017) showed that the presence of dominant or inferior alternatives in the design (which are common in orthogonal designs) might result in biased parameter estimates.

Efficient designs can reduce the standard errors of the parameter estimates (and thus the required sample size) (Rose and Bliemer, 2009; Walker et al., 2015). These designs aim to maximize the information obtained from the SP data and thus reduce the sample size requirements (Rose and Bliemer, 2009; Kessels et al., 2011). The goal is to minimize the standard errors of the estimated parameters via optimization based on the asymptotic variance–covariance (AVC) matrix (Walker et al., 2015), which is a function of the model parameters in discrete choice models. Therefore, these designs require priors for the parameter values. Several studies showed that these models might not be robust if the priors are misspecified, particularly in the mixed MNL model (Walker et al., 2015; Zhu et al., 2017).

Efficient designs were initially developed for Multinomial Logit (MNL) models (Bunch et al., 1996; Huber and Zwerina, 1996). Extensions were proposed for Nested Logit (NL) models (Bliemer et al., 2009; Goos et al., 2010), cross-sectional Mixed Logit (ML) models, (Sándor and Wedel, 2002; Yu et al., 2009), and panel ML models (Bliemer and Rose, 2010; Yu et al., 2011). Rose et al. (2008) also introduced an efficient method for pivoting attribute levels around a reference alternative. More recently, van Cranenburgh et al. (2018) proposed efficient designs that are robust for decision rule uncertainty (such as random regret minimization).

### 2.4. Literature takeaways

The aforementioned studies indicate that prompted recall surveys are widely applied in transportation along with pivoting in order to generate context-aware SP choice tasks. In addition, context-awareness can be further enhanced using smartphones due to their representational richness and sensing capabilities. In this paper, we propose a context-aware SP method that subsumes the advantages of prompted recall and smartphone-based SP surveys. Following the recommendation of Walker et al. (2015), we utilize a variation of the random design in order to construct the SP choice tasks based on RP data, and modify the design to remove any dominant or inferior alternatives.

## 3. Methodology

This paper presents a smartphone-based method for trip-based SP surveys which makes it possible to accurately analyze the demand for new transportation modes and services and analyze hypothetical scenarios. The proposed method can be applied to any activity/trip sensing platform. We leverage on the initial concept proposed by Cox (2015) and detail a new formulation along with its first deployment, data collection and model estimation. In this section the system architecture and the process of generating and validating SP choice tasks are presented.

### 3.1. System architecture

The proposed SP method leverages RP data in generating profiles. A profile is defined as a combination of attribute values and levels for a particular alternative (thus each choice task includes multiple profiles). The individual specific context is coming from (1) the observed RP data, e.g. sensing data, validated trips/activities for a given day (or even the weekly activity pattern); and (2) external contextual data such as the transportation alternatives available to the user through online sources. The RP data includes the departure and arrival times, origin and destination, trip mode and purpose, and activity duration. These are obtained using a smartphone-based sensing app. In our application, we use the Future Mobility Sensing (FMS) platform, explained in Section 3.2. External contextual is data obtained from a trip planner. Our platform uses Google Maps, however, any trip planner can be used in order to obtain the trip data.

These are used together with the user's personal characteristics obtained from the pre-survey, e.g., car ownership, usage of car/bike sharing services, etc. The experimental design uses those data in order to generate SP choice experiments with reasonable alternatives and attributes. The preferences can then be estimated based on both RP and SP data. The proposed concept is presented in Fig. 1, and the different components are explained in the following sections.

From the user's perspective, the action flow and choice task generation process are explained below and presented in Fig. 2:

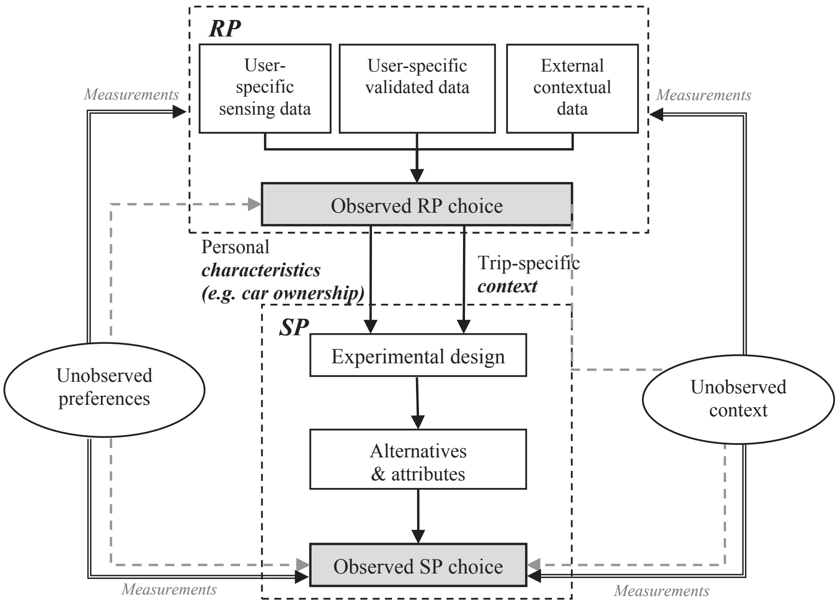


**Fig. 1.** Smartphone-based SP methodology.

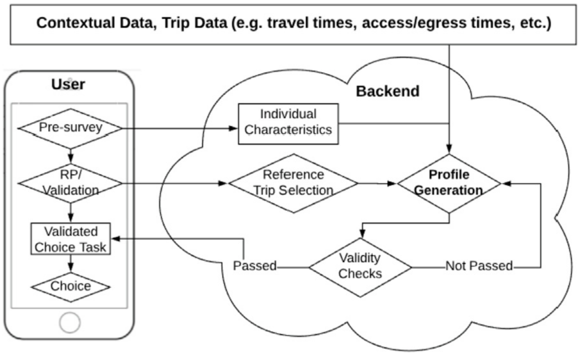


**Fig. 2.** SP data collection process.

1. Upon registering to the platform, a user's socio-demographic data as well as attitudes and perceptions are collected using the pre-survey.
2. The app starts to track the user, who then has to validate the activity diaries with trip and activity information every day directly on the smartphone app. The sensing platform and trip validation are presented in Section 3.2.
3. After a full day is validated, a trip context (activity, origin/destination, and departure time) is selected from the trips that the user has performed as a reference trip. Trip selection is explained in Section 3.3.
4. The alternatives and attributes (e.g., travel times) corresponding to the selected trip context are retrieved from the trip planner.
5. The experimental design presented in Section 3.4 is used to generate SP profiles.
6. The profiles are checked afterwards for realism and internal validity. Validity checks are presented in Section 3.6.
7. A choice task is presented based on a selected trip directly in the smartphone platform. The user is asked about his/her choice if he/she were to repeat the trip under different hypothetical scenarios. Each user is presented with a choice task including multiple alternatives with different modes, travel times, costs, waiting times, and other attributes. The user is required to choose one of the suggested alternatives (which can also include an option for not making the trip at all). This choice task becomes available to the user as soon as validation for that day is completed, however, the user can complete it at any time.

### 3.2. Sensing platform and trip validation

The sensing platform is needed to collect RP data (activity and trip diaries) for selecting the reference trip and obtaining contextual data (e.g. trip purpose, accompanying persons, etc.). In our application, we leverage the Future Mobility Sensing (FMS) platform (Cottrill et al., 2013; Zhao et al., 2015b). FMS is a smartphone and prompted-recall-based integrated activity-travel survey system. It is mainly used to collect RP data in the form of activity diaries which include stop locations, durations (start and end time), activities performed, travel modes taken between stops, and other information such as travel costs, accompanying persons, etc. As an outcome of this study, the FMS platform has been extended to include the SP feature.

Once installed on the smartphone, the app collects location (GPS, WiFi, GSM), accelerometer data and other information (e.g. battery level) on a continuous basis. The backend system processes the collected data in-real time in order to detect stops and infer the trip mode, travel times and activity type using machine learning algorithms. An app interface presents the users with maps showing their stop locations and trip trajectories in addition to partially filled activity diaries where they can verify and validate their trips and activities periodically. This overcomes the main limitations associated with traditional travel surveys such as under-reporting of trips and inaccurate or incomplete time, location and route information (Zhao et al., 2015b).

Trip diaries are complemented with a pre-survey which provides socio-demographic data about users, such as their age, gender, working status, income, car ownership, bike ownership, and others, and is filled out before the smartphone data collection.

Fig. 3 presents screenshots of a sample day trajectory and activity diary validated through the sensing platform (these screens were already existing in the FMS platform) and the reference trip and SP choice task (these screens were added to this platform as part of this study). In the SP, the user can browse through different tabs showing different categories of alternatives (e.g. transit, non-motorized, car, taxi/on-demand, and FMOD tabs). The user can also scroll down to view different alternatives listed under a tab. For example, under the transit tab, the user can see “walk to transit”, “drive to transit”, and “bike to transit”, and under the FMOD tab, the user can see “taxi”, “shared taxi”, and “minibus” alternatives (which are explained in Section 4). Fig. 3d, e, and 3f show examples of alternatives under the transit, on-demand, and non-motorized tabs.

This interface was used in the data collection presented in Section 4. It was tested on a smaller sample (38 users) in a pilot phase and improved based on this pilot. The SP attributes are presented in a way that is similar to how users would see them in real life. For example, travel times are presented in minutes, fares are presented in Dollars with two decimal places (which is similar to apps such as Uber and Lyft and to transit fares in Boston), and subscription fees are presented as whole numbers.

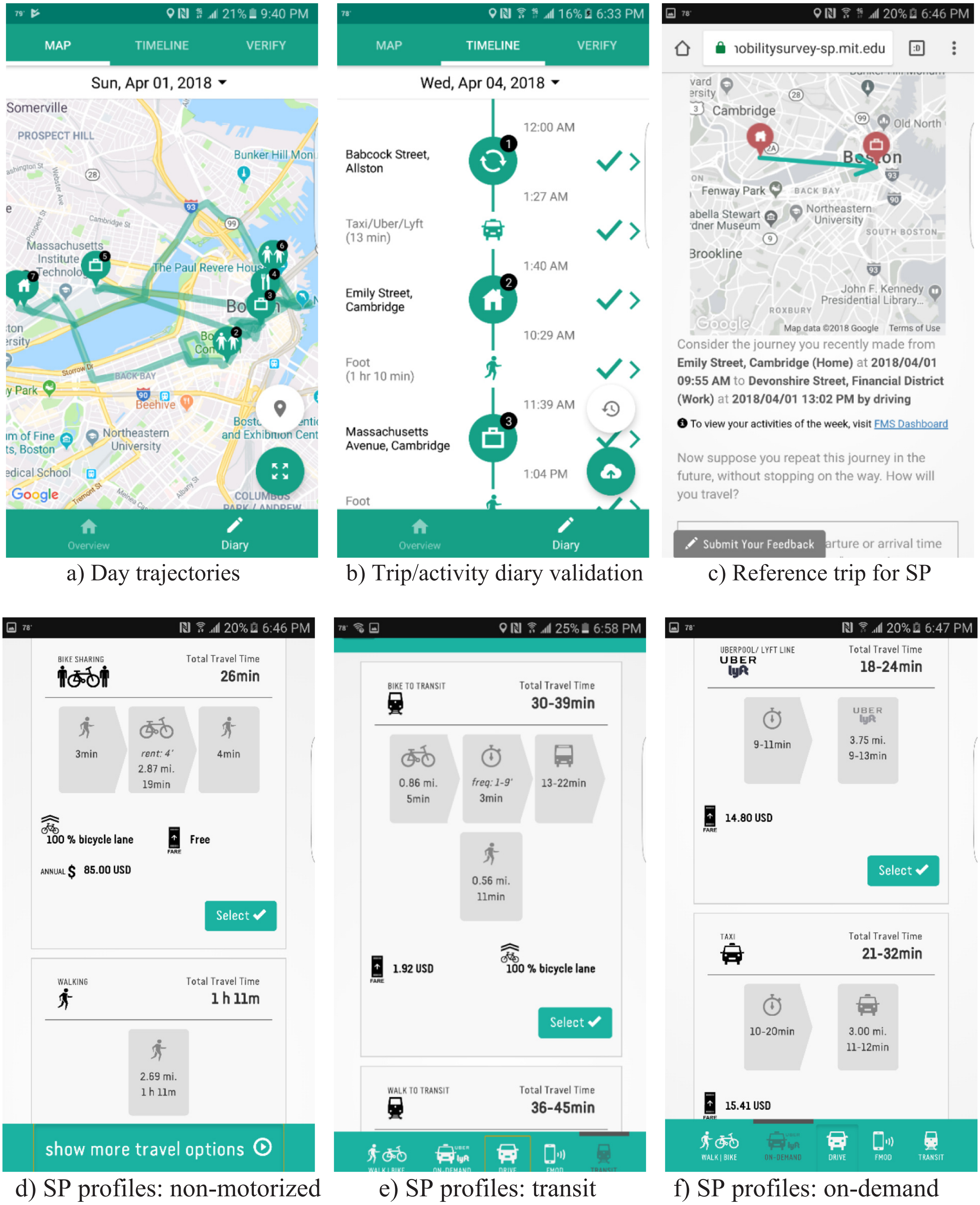


**Fig. 3.** Screenshots from the FMS platform in the same order in which the user sees them (except for SP tabs where the order of tabs is shuffled).

### 3.3. Reference trip selection

For a given validated day, a realized trip recorded in the trip diary is selected by our proposed framework for the daily SP experiment. Any validated trip can be selected as long as it satisfies a set of pre-defined conditions. These conditions are specific to the research question at hand. Currently, a single trip can be selected, or intermediate stops can be dropped in order to form a half-tour; the trip or half-tour can be either home-based or non-home based; a minimum trip length for a trip to be selected can also be specified in order to avoid long non-commuting trips. In our applications, users are presented with one SP choice experiment per day (however, this can also be modified).

### 3.4. Experimental design

Profile generation is based on a variation of the random design (whereby constraints are used in order to enhance realism and remove dominant and inferior alternatives). This design is found to be the most convenient for our application. First, there is a one-to-one correspondence between the realized trips and the choice tasks. The attributes of the realized trip can be easily pivoted using the random design, and checked for external and internal validity. In addition, this design can handle large number of alternatives and attributes (unlike some other designs such as orthogonal designs). Finally, previous research has shown that the random design performs relatively well, and in the absence of good priors, it might be preferred to efficient designs (Walker et al., 2015; Street and Burgess, 2007). Nevertheless, implementing efficient designs on smartphone-based platforms is a promising future research direction, as it can build upon the results obtained from this paper to produce good priors.

SP profiles are generated on the backend server in real-time. Despite the large number of alternatives and attributes, most of the time a choice task is generated and validation checks are completed in less than 1 s.

#### 3.4.1. Pivoted alternative attributes

In our application, profile attributes are generated using random “design parameters” in addition to information from Google Maps or the activity diary. After selecting a reference trip from the activity diary, the origin, destination, and departure time of that trip are fed into the Google Maps API in order to obtain some baseline attributes, such as travel times using different modes and transit access and egress times. These attributes are then “pivoted” using random “design parameters” in order to construct the attributes shown in the profiles.

Design parameters are chosen from a pre-defined set of levels. For example, the travel time attribute for car is calculated by multiplying the travel time obtained from Google Maps by a design parameter, the *travel time ratio*, with levels of 0.7, 0.85, 0.95, 1.00, 1.05, 1.15, 1.20, or 1.50. This means that the pivoted travel time can range between 70% and 150% of the travel time obtained from Google Maps for the same trip. A uniform distribution is used such that each level of the design parameter has an equal probability of being selected (i.e., the travel time ratio is equally likely to be chosen from any of the values above).

Other design parameters represent travel time variability (as a percentage of the total trip time). These are used along with the pivoted travel times in order to present users with ranges that reflect uncertainty in travel time. For example, the drive alone travel time in a given choice task can range between 30 and 35 min. Similar design parameters are used to represent uncertainty in taxi waiting times and transit headways.

The procedure mentioned above results in profiles that carry the same information as those generated by the standard random design. However, the attributes are presented to users in a way that is easier for them to process. For example, using the random design, the user might be asked to recall a reference trip. Afterwards, he/she will be asked about his/her choice if the travel time was 15% greater than that of the reference trip and the travel time variability was $\pm$ 10% of the total travel time. On the other hand, this method presents the user with a travel time range in minutes that is calculated using the same information. Following the same example, and assuming that the reference trip is 30 min long, the corresponding range is 31–38 min, and is calculated as $(1.15 \times 30) \mp (0.1 \times 1.15 \times 30)$. Therefore, the presented attributes are consistent with the way users perceive them in real life. In addition, this saves mental effort for users in estimating the total travel time and travel time variability from the attributes presented in the profiles.

#### 3.4.2. Random alternative attributes

It may not be possible for the researcher to obtain all the alternative attributes of the reference trip and pivot around these attributes. Therefore, some attributes are constructed using random design parameters that directly correspond to attribute levels. For example, the taxi waiting time is equally likely to be 2, 5, 8, 10, or 15 min, and hence, this attribute is not calculated based on information from Google Maps or the activity diary.

Another design parameter might represent the parking cost per hour. For example, the possible levels for the hourly parking cost are specified as 0.5, 1.0, 1.5, 3.0, 4.0, or 8.0 US Dollars. In order to calculate the corresponding attribute (total parking cost), the hourly parking rate is multiplied by the activity duration (which is obtained from the travel diary). The researcher might choose to present the user with the total parking cost or the hourly parking cost, to be consistent with the actual market framing of parking costs. Similarly, the user is presented with the fuel cost (gas price per gallon), parking cost (per hour or per day), and toll cost separately, and not the total car cost, which is also consistent with the way users see these attributes in real life.

#### 3.4.3. Levels of the design variables

The levels of design parameters are chosen to guarantee sufficient variability in the data, while ensuring realistic attribute ranges.

**Table 1**
Considered attributes across alternatives.

| | Non-motorized | | | Motorized | | On-demand | | | Transit |
|---|---|---|---|---|---|---|---|---|---|
| | Walk | Bike | Bike-sharing | Drive alone | Carpool | Taxi | Uber/Lyft | Car-sharing | |
| Walking time | x | | | | | | | | |
| Biking time | | x | x | | | | | | |
| Waiting time | | | | | | x | x | | |
| Schedule delay | | | | | | | | | |
| Access/egress time | | | x | x | x | | | x | x |
| In-vehicle travel time | | | | x | x | x | x | x | x |
| Parking time | | | | x | x | | | | |
| % Bike lane | | x | x | | | | | | |
| Annual subscription cost | | | x | | | | | x | |
| Distance-/time- based variable cost/fare | | | x | | | x | x | x | x |
| Fuel cost | | | | x | x | | | | |
| Toll cost | | | | x | x | | | x | |
| Parking cost | | | | x | x | | | x | |
| Transfers | | | | | | | | | x |
| Headway | | | | | | | | | x |

According to Sanko (2001), the researcher's domain knowledge is essential is deciding upon these levels. These levels should have a sufficient range that will likely cover the users' "boundary values", defined as the points as at which users will trade-off. Furthermore, these attribute levels should be close enough to each other, and unequally spaced in order to accurately estimate these boundary values (Sanko, 2001).

### 3.5. Basic alternatives

This section introduces the basic alternatives presented in SP (walking, biking, bike-sharing, drive alone, carpooling, car-sharing, taxi, on-demand services, and transit). However, it is important to note that the generic SP framework provides flexibility for introducing new alternatives and attributes. Section 4 presents an application with a new service that does not exist in real life, Future Mobility on Demand (FMOD) (Atasoy et al., 2015). Other potential applications include travel advisers (e.g. Xie et al., 2019), automated vehicles (Seshadri et al., 2019), mobility-as-a-service packages, electric mobility, or personal and active mobility modes.

Considered attributes across all modes are listed in Table 1. Under non-motorized modes, bike-sharing has additional attributes of access and egress times (as users pick up/drop off bikes at the stations), annual subscription, and time-based rental costs. The availability of these modes is determined by pre-specified maximum walking and biking distances. Bike availability is contingent on bike ownership; however, bike-sharing is always displayed given that maximum biking distance is not violated.

In our SP platform the experimental design explained in Section 3.4 is applied as follows. For drive alone and carpool alternatives, the existence of toll and parking costs is randomly decided using Boolean design variables. If there is a non-zero parking cost, it is calculated based on the activity duration. The experimental design ensures that the drive alone travel time is not greater than carpooling travel time. Drive alone availability is determined by car ownership and having a valid driving license. If these two conditions are satisfied, the user might also choose carpooling "as a driver" or "as a passenger". Otherwise, only carpooling "as a passenger" is available.

For on-demand services (e.g. Uber, Lyft, etc.), the design ensures that travel time for shared alternatives (e.g., UberPool) is never less than that of the private ones (e.g. UberX). They are assumed to be always available except car-sharing which necessitates a valid driving license. Note that the new mobility service studied in Section 4 (FMOD) is also an on-demand service. For subscription-based services (e.g., car-sharing or bike-sharing), users who are not subscribed to these services are presented with annual subscription fees.

Finally, transit modes include bus and train, with walk, bike, or car access (i.e., park-and-ride). The number of transfers is randomly determined in a time-based manner, e.g., trips shorter than 10 min have no transfers, those between 10 and 20 min may have up to one transfer, and trips longer than 20 min may have up to two transfers. The fares include a fixed and a distance-based component for flexibility (e.g., setting the distance-based component to zero results in a flat fare). A transit alternative may include bus, train, or a combination of the two. The availability is either based on the existing conditions (referring to Google Maps), or defined by the researcher in order to test hypothetical scenarios.

For new/non-existing modes (such as the mobility-on-demand application presented in Section 4), alternative attributes can be specified as functions of the attributes of existing modes (e.g. car, taxi, etc.).

### 3.6. Validity checks

Since SP profiles are generated based on the random experimental design, some combinations of attributes might not be realistic to users. For example, walking time in a particular choice task may be shorter than biking time. These two attributes are calculated by multiplying the walking/biking times obtained from Google Maps by design parameters defined as "travel time ratios". The latter case might occur if the worst biking time ratio and the best walking time ratio are both selected.

In addition, a choice task may include a dominant alternative (i.e. having both its total travel time and travel cost better than those of any other alternative) or an inferior alternative (having its travel time and cost worse than any other alternative).

In order to avoid such cases, validity checks are used by defining a set of "flags", which correspond to unrealistic combinations of attributes and dominant/inferior alternatives. Below are some examples:

- Drive alone is cheaper than transit.
- Drive alone is cheaper and faster than any other mode.
- Walking is faster than transit, biking, taxi, etc. (except for short trips, e.g., shorter than 1 km).

Although such cases might still occur in reality, Bliemer et al. (2017) argue that strictly dominant alternatives should be avoided in SP surveys because they bias the estimates in logit models.

Furthermore, some flags correspond to the difference between the maximum and minimum costs (or travel times) among all alternatives exceeding a certain threshold.

Different choice tasks are generated until a valid one is found. Ideally, users should be presented with choice tasks without any flags. However, this might be difficult to achieve depending on the complexity of the experiment. Therefore, a choice task is accepted if the number of flags does not exceed a pre-defined threshold. For example, the maximum allowable number of flags cannot exceed 2 or 3. Increasing this threshold can result in unrealistic choice tasks (having many unrealistic combinations). On the other hand, decreasing this threshold can result in excessive running times as the number of generated choice tasks will be very high until a valid one is found. Another possible approach is generating several choice tasks (e.g. 1 000), and choosing the one with the least flags.

Using this design, and unlike the standard random design, not all attribute combinations have an equal probability of being chosen. As indicated above, not all of the generated choice tasks are accepted; they are evaluated based on pre-specified constraints that increase realism and remove dominant/inferior alternatives. In addition, some attribute levels are dependent on others. For example, the toll amount is contingent on a Boolean variable indicating whether an alternative includes a toll cost or not (if not, then the toll amount is zero 50% of the time).

### *3.7. Design validation*

The profile generation algorithm is validated using Monte-Carlo simulations to ensure that the data can be used in behavioral models. This step simulates data that could be collected using the survey, and is done offline prior to deployment. It is needed to prove that the behavioral models estimated using the SP choice tasks result in consistent estimates and acceptable confidence intervals. The design efficiency can also be assessed analytically using the covariance matrix of the attributes (or the D-error), however, the Monte Carlo experiment also allows us to identify other issues such as inconsistency (differences between the estimates and the true values of the parameters).

For a sample of trips, choices are simulated using a logit model (or a nested logit model). Different attributes are assigned predetermined coefficients (weights) which are known to the researcher. The systematic utilities of all alternatives are then calculated using the assumed coefficients and the profile attributes. An extreme value error term is added in order to simulate the total utilities. The simulated choices correspond to those with the highest utility among all alternatives.

Afterwards, a new logit model (or nested logit model) is estimated using the simulated choices in order to check whether the true values of the coefficients can be recovered. This procedure enables the identification of specific issues in the design such as bias, multicollinearity or insufficient variations in attributes which can be fixed accordingly. Model structures other than logit can also be used to account for correlations among individuals and alternatives (e.g. probit, mixed logit, etc.).

## 4. Case study: Future Mobility on Demand

This section presents an application of the SP methodology described above to FMOD, which is an app-based service that provides a personalized menu of travel options in real-time from the available modes operated by FMOD: taxi, shared-taxi, and minibus (Atasoy et al., 2015). The traveler can select one option in the menu for their trip or reject all options. FMOD taxi provides door-to-door service in a private vehicle, which is typically the highest priced service. FMOD shared-taxi serves multiple passengers in the same vicinity, but travel time may increase due to the pick-up and drop-off of other passengers. FMOD minibus runs along fixed routes with fixed stops but adapts to passengers' schedule and typically has the lowest fare. The FMOD operator manages the requests and the flexible fleet of vehicles in real-time (in terms of dispatching, passenger-vehicle matching, routing, etc.), thus conditioning the attributes of the menu offers to its operational performance. FMOD is a demand responsive system and provides flexibility to both passengers and operators; passengers can choose from a menu that can be optimized in an assortment optimization framework to tailor their needs and preferences, while operators can allocate their homogeneous vehicle fleet among all three types of services.

Since FMOD is a schedule-based service, users are asked about their preferred arrival or departure time window. The user request will preferably be assigned to a vehicle that matches their preferred time window. However, this might not be always possible due to the limited fleet size or operational costs. This introduces schedule delay, which is defined as the difference between the actual pick-up (drop-off) time and the boundary of the preferred departure (arrival) time window.

In the SP design, FMOD taxi and shared-taxi attributes include waiting time, travel time, fare, and schedule delay. Schedule delay is not directly provided to the user and only the estimated pick-up and drop-off times are displayed. These may or may not fall within the specified time window. Moreover, FMOD minibus has additional attributes of access and egress time. Therefore, the user is

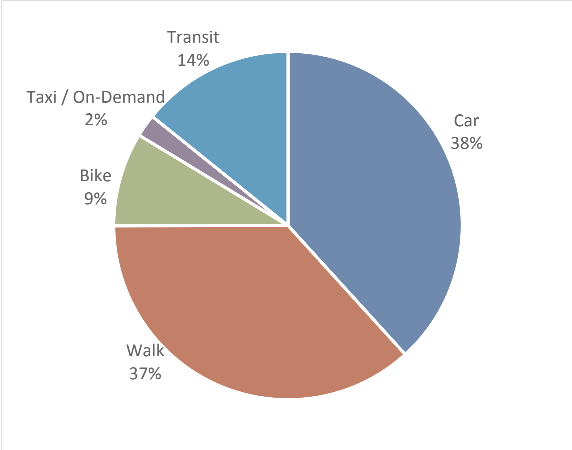


**Fig. 4.** Mode shares in the revealed preferences data.

presented with a departure time from home, arrival time at station, pick-up time, drop-off time, and arrival time at the final destination. FMOD is assumed to be always available to users (except for very short trips, in which FMOD minibus is not available). It is presented as a separate tab in the SP user interface which includes all three services. The FMOD-SP design was validated using the Monte Carlo procedure presented in section 3.7.

The SP survey selects home-based half-tours, since the goal is to deploy the estimated mode choice model in an activity-based travel demand model (namely the one presented in Lu et al., 2015).

### 4.1. Sample demonstration

An FMOD case study was conducted with a sample of 224 respondents in the Greater Boston Area, who completed 908 SP choice tasks. The number of choice tasks per respondent ranged between 1 and 11, with an average of 4. The sample was fairly split between males (46%) and females (49%), while the remaining 3% refused to reply. Respondents' ages ranged between 11 and 69 years, with a mean of 30.7 and a standard deviation of 9.6. Full-time employees accounted for 56% of the sample, and students accounted for 30%. The remaining users were either unemployed, part-time employees, retired, or self-employed. The mode shares in the RP data are presented in Fig. 4.

The SP alternatives included driving alone, carpooling, car-sharing, motorcycle, walking, biking, bike sharing, transit (with walk, bike, or car access), and on-demand (taxi, private ride hailing, and shared ride hailing), in addition to the three FMOD modes. The current SP design resulted in FMOD modes being chosen 24% of the time. Sufficient observations were collected for each alternative except for car-sharing and motorcycle (3 and 4 observations respectively).

In this application, the realism of the experiment was ensured using the experimental design as well as the validity checks (see section 3.6.). The maximum allowable number of flags was specified to be 2. Nevertheless, we still observed some dominant and inferior alternatives with a frequency of 4% and 7%, respectively. By observing the data, we noticed that these correspond to special cases (such as short trips, where walking is dominant or driving to transit is inferior). Other flags were less likely to occur; for example, car was cheaper than transit in less than 2% of the cases, walking was faster than car in less than 2% of the cases, and faster than transit in less than 5% of the cases. Other constraints were forced in the experimental design; for example, carpooling options were always specified to be slower and less expensive than driving alone, and private on-demand and FMOD alternatives were always specified to be faster and more expensive than shared alternatives.

### 4.2. Model estimation

The SP data are used in order to estimate a mode choice model among the existing modes and FMOD modes. Since some choice tasks were completed within a very short time (e.g., less than 10 s), some of these might have been filled out at random. Therefore, a latent class model is used in order to model two choice protocols: utility maximization and random choice. A binary logit model is used to model the (latent) choice protocol. In the case of utility maximization, a mixed logit model with random parameters and error terms following the normal or the log-normal distribution is used. In the case of random choice, equally likely probabilities are assumed to all the available alternatives.

The attributes used in the utility maximization model are cost, travel time (in-vehicle, out-of-vehicle, and non-motorized travel time), and travel time variability. A dummy variable is included in the SP utility equations representing whether the SP mode was chosen in the reference trip or not. This variable accounts for the context specific to each trip.

#### 4.2.1. Model specification

The choice protocol model is specified as a binary logit that includes socio-demographic variables, the choice task number, and the time for completing the choice task divided by the number of available alternatives. Age is the only socio-demographic variable that was found to be significant. The utility equations are presented in equation (1) below:

$$U_{UM} = ASC_{UM} + \beta_{Age} * Age + \beta_{ChoiceNum} * ChoiceNum + \beta_{Dur} * \frac{Duration}{NumAlt} + \varepsilon_{UM}$$
$$U_{Random} = 0 + \varepsilon_{Random} \tag{1}$$

Where $Age$ is the respondent's age (divided by 100), $ChoiceNum$ is the choice number, Duration is the completion time measured in seconds, and $NumAlt$ is the number of available alternatives in the choice task.

The utility functions of the utility maximization model are presented in Equation (2). A scale parameter is estimated while the cost coefficient is fixed to $-1$. Therefore, all other parameters are in the willingness to pay space (money-metric utility). A higher scale parameter indicates a higher explanatory power of the model variables compared to the error component and vice versa.

The scale parameter and the parameters of all attributes are modeled with inter-personal heterogeneity in order to account for the panel nature of the data (since multiple observations are available from the same user). Travel time and travel time variability coefficients follow the log-normal distribution through the exponential form which ensures that all of these parameters are negative. Exponentiation is also used for the scale parameter in order to ensure that this parameter is positive. The estimated coefficients are the mean and standard deviation of the underlying normal distributions (thus, these can take any sign, while the actual parameters maintain their correct sign). A full set of alternative specific constants (ASC) is used, $ASC_{Transit}$ is arbitrarily normalized to 0. Normally distributed error terms are used to account for correlations among alternatives (namely correlations among car modes, transit modes, on-demand modes, FMOD, and non-motorized modes). A full set of alternative specific constants (ASC) is used, $ASC_{Transit}$ is arbitrarily normalized to 0.

For car, in-vehicle travel time includes both average travel time and parking time, while for other modes it only includes average travel time. Two different in-vehicle travel time coefficients are estimated, for drive alone and for other motorized modes. For transit with car access, this includes the total in-vehicle travel time in car and in transit. Out-of-vehicle travel time includes access time (car modes, transit, and FMOD minibus), egress time (car modes, transit, and FMOD minibus), and waiting time (transit, on-demand modes, and FMOD minibus). Non-motorized travel time is the travel time for walking or biking modes. Cost is the fare for all modes except for car, which is the sum of fuel cost, parking cost and toll cost. In order to obtain similar magnitudes of all coefficients, cost is measured in tens of USD. Finally, travel time variability is defined as the difference between the maximum and the minimum travel time, divided by their mean.

$$U_i = \exp(\text{Scale})(-Cost_i + ASC_i + \exp(B_{IVTT})IVTT_i + \exp(B_{OVTT})OVTT_i + \exp(B_{NMTT})NMTT_i$$
$$+\exp(B_{TTVAR})TTVAR_i + B_{Inertia}RP_i + B_Z Z + \eta_i) + \varepsilon_i \tag{2}$$

Where:

$B_{IVTTCar} \sim N(B_{IVTTCar\_mean}, B_{IVTTCar\_std})$: coefficient for in-vehicle travel time by car
$B_{IVTT} \sim N(B_{IVTT\_mean}, B_{IVTT\_std})$: coefficient for in-vehicle travel time by other modes
$B_{OVTT} \sim N(B_{OVTT\_mean}, B_{OVTT\_std})$: coefficient for out-of-vehicle travel time
$B_{NMTT} \sim N(B_{NMTT\_mean}, B_{NMTT\_std})$: coefficient for non-motorized travel time
$B_{TTVAR} \sim N(B_{TTVAR\_mean}, B_{TTVAR\_std})$: coefficient for travel time variability
$B_{intertia} \sim N(B_{intertia\_mean}, B_{intertia\_std})$: coefficient for inertia
Scale $\sim N(Scale_{mean}, Scale_{std})$: scale parameter
IVTT, OVTT, and NMTT: in-vehicle, out-of-vehicle, and non-motorized travel time respectively (min)
TTVAR: travel time variability (unitless)
COST: travel cost (tens of USD)
RP: a dummy variable indicating whether the SP mode was chosen in the reference trip or not
$Z$: a vector of individual characteristics and socio-demographics (age, gender, student status, income, and car ownership) with coefficients $B_Z$
$\eta_i \sim N(0, \sigma_i)$: a normally distributed error term to capture correlations among modes.
$\varepsilon_i$: an iid error component following the extreme value distribution.

**Table 2**
Estimation results for the choice protocol model.

| Parameter | Estimate | Robust Std err | Robust t-test | p-value |
|---|---|---|---|---|
| Constant (UM) | 6.84 | 2.40 | 2.84 | < 0.01 |
| Age | −0.133 | 0.0437 | −3.04 | < 0.01 |
| Choice Task | −0.421 | 0.253 | −1.67 | 0.100 |
| Duration/NumAlt | 0.307 | 0.162 | 1.89 | 0.06 |

**Table 3**
Estimation results for the utility maximization model.

| Parameter | Estimate | Robust Std err | Robust t-test | p-value |
|---|---|---|---|---|
| Scale - mean | 0.516 | 0.265 | 1.94 | 0.05 |
| Scale - std dev. | 1.75 | 0.400 | 4.38 | < 0.01 |
| IVTT by car - mean | −2.51 | 0.251 | −10.01 | < 0.01 |
| IVTT by car - std dev. | 0.278 | 0.0826 | 3.37 | < 0.01 |
| IVTT by other modes - mean | −3.47 | 0.213 | −16.3 | < 0.01 |
| IVTT by other modes - std dev. | 0.529 | 0.0906 | 5.84 | < 0.01 |
| NMTT - mean | −2.83 | 0.176 | −16.06 | < 0.01 |
| NMTT - std dev. | 0.602 | 0.0635 | 9.49 | < 0.01 |
| OVTT - mean | −4.29 | 0.518 | −8.27 | < 0.01 |
| OVTT - std dev. | 1.67 | 0.232 | 7.23 | < 0.01 |
| Travel Time Var.- mean | −0.251 | 0.332 | −0.760 | 0.45 |
| Travel Time Var.- std dev. | 0.118 | 0.065 | 1.82 | 0.07 |
| Inertia - mean | 0.884 | 0.189 | 4.67 | < 0.01 |
| Inertia - std dev. | 0.650 | 0.317 | 2.05 | 0.04 |
| Drive alone constant | 0.877 | 0.446 | 1.97 | 0.05 |
| Carpool constant | −0.467 | 0.224 | −2.09 | 0.04 |
| Walk constant | 0.876 | 0.340 | 2.58 | 0.01 |
| Bike constant | 0.995 | 0.335 | 2.97 | < 0.01 |
| Bikeshare constant | −0.548 | 0.249 | −2.20 | 0.03 |
| FMOD minibus constant | −1.66 | 0.422 | −3.95 | < 0.01 |
| FMOD shared taxi constant | −1.95 | 0.529 | −3.70 | < 0.01 |
| FMOD taxi constant | −2.22 | 0.633 | −3.50 | < 0.01 |
| Taxi constant | −3.23 | 0.846 | −3.82 | < 0.01 |
| Uber/Lyft constant | −1.90 | 0.504 | −3.76 | < 0.01 |
| Uberpool/LyftLine constant | −1.75 | 0.438 | −4.00 | < 0.01 |
| Driving modes - std dev. | 0.0516 | 0.0768 | 0.67 | 0.5 |
| On Demand modes - std dev. | 2.16 | 0.364 | 5.93 | < 0.01 |
| Non-motorized modes - std dev. | 0.317 | 0.0564 | 5.61 | < 0.01 |
| Transit modes - std dev. | 0.0623 | 0.114 | 0.550 | 0.58 |
| FMOD modes - std dev. | 0.0549 | 0.0561 | 0.98 | 0.33 |
| Age (FMOD) | −2.29 | 0.594 | −3.85 | < 0.01 |
| Age (Non-motorized) | −0.730 | 0.293 | −2.49 | 0.01 |
| Student (Non-motorized) | 0.673 | 0.352 | 1.91 | 0.06 |
| Student (Transit) | −0.753 | 0.427 | −1.76 | 0.08 |
| Income (Driving) | 0.705 | 0.258 | 2.73 | 0.01 |
| Income (FMOD) | 2.34 | 0.370 | 6.31 | < 0.01 |

### *4.2.2. Estimation results*

The model was estimated using Maximum Simulated Likelihood (MSL) in Python Biogeme (Bierlaire, 2016). Trips exceeding 30 miles were excluded. The estimation results are presented in Tables 2 and 3. The model converges with 10,000 random Halton draws of the random parameters. Using the same starting values, the model coefficients do not vary significantly when the number of draws is increased beyond 10,000 (however, using different starting values sometimes results in a different local optimum, with significantly different estimates and a lower log-likelihood).

The estimation results of the Choice Protocol model are presented in Table 2. The constant is positive, indicating that respondents are likely to make a choice according to utility maximization rather than at random. However, the coefficient of age is negative, indicating that older respondents are more likely to choose at random. The choice task coefficient is also negative, indicating that in the later choice tasks, respondents are more likely to choose at random. Finally, duration divided by the number of alternatives is positive, indicating that the more time respondents spend per alternative, the less likely they are to choose at random.

The estimation results of the utility maximization model are presented in Table 3. Significant heterogeneity is present in the scale parameter and in all travel time, travel time variability, and inertia coefficients. The mean of the inertia coefficient is positive as expected, indicating that respondents are more likely to choose the RP chosen mode in the SP choice task. This can be due to unobserved constraints associated with that specific trip favoring that mode, but can also capture other effects such as justification bias. The standard deviations of on-demand and non-motorized modes are significant, indicating strong correlations among the

**Table 4**
Estimated values of time.

| | Mean VOT (USD/hr) |
|---|---|
| In-vehicle travel time (drive alone) | 50.68 |
| In-vehicle travel time (Other modes) | 21.47 |
| Non-motorized travel time | 42.44 |
| Out-of-vehicle travel time | 33.16 |

**Table 5**
Summary statistics and comparison to MNL.

| | Estimated Model | Multinomial Logit (MNL) |
|---|---|---|
| Number of estimated parameters | 40 | 24 |
| Final log likelihood | −636.12 | −662.957 |
| Final gradient norm | 7.15E-05 | 4.53E-05 |
| Run time | 02 h 15:00 min | 0:02 s |
| Number of draws | 10,000 | – |
| Sample size | 640 | |

alternatives belonging to these groups. The correlations between car modes, FMOD modes, and transit are not significant.

The coefficients of travel time, travel time variability, and the scale parameter are log-normally distributed. The estimated mean and standard deviation are for the underlying Normal distribution. The mean of a log-normally distributed variable can be calculated as $\exp(\mu + \sigma^2/2)$ where $\mu$ and $\sigma^2$ are the mean and variance of the corresponding normal distribution, respectively. Therefore, we can obtain the mean of the IVTT, OVTT, and NMTT coefficients. As everything is in willingness to pay space, the values of IVTT by car, IVTT by all other modes, OVTT, and NMTT (in USD/hour) are presented in Table 4. The signs to the mean values of travel times are negative because the corresponding coefficients are given negative signs in the utility function. The values of travel time are close to those found by Xie et al. (2019), who used the same survey platform to collect data from a similar sample of respondents in the Greater Boston Area. In addition, they are close to the values obtained by Atasoy et al. (2018) who used a convenience sample of 38 users in a pilot of this survey.

The value of time of non-motorized modes is very high, indicating that users are very sensitive to travel time by these modes. This is expected, especially because of the cold weather in winter in the Greater Boston Area. Except for car, the value of IVTT is the lowest indicating participants are more tolerant to IVTT in comparison to NMTT and OVTT, which is reasonable due to the comfort of staying in vehicles and the possibility of multi-tasking. On the other hand, the IVTT by car is the highest. This could be potentially due to the discomfort associated with driving and the inability to multitask. However, it can be due to the fact that people with a higher value of time are more likely to drive alone.

The constants of FMOD modes are comparable to those of Uber/Lyft and UberPool/LyftLine, and substantially higher than the taxi constant, which might indicate that respondents are willing to adopt these services in comparable rates. In addition, the results indicate that respondents with high income and young respondents are more likely to use FMOD compared to other market segments.

The estimated model accounts for heterogeneity as well as the random response bias. In order to justify its complexity, we compare it to an MNL model with the same specification, but without heterogeneity in any of the parameters and without the choice protocol model. The results are presented in Table 5. Both models achieve convergence as indicated by the low values of the final gradient norm. However, the complex model shows an improvement of 26.8 points in the log-likelihood over MNL. The likelihood ratio test results in a test statistic of 53.67, which is higher than the critical value at the 95% level of confidence with 16 degrees of freedom (26.30). This indicates that the estimated model is superior to MNL. It is also important to note that the MNL estimates are not realistic; for example, the estimated parameter for walking or biking time is positive and very close to zero. This reinforces the importance of accounting for heterogeneity as otherwise even biased estimates might be obtained.

#### *4.2.3. Joint SP/RP estimation*

In this work, the model was estimated using SP data only. However, the RP data might be used in order to estimate a joint SP/RP model. An example is presented in Atasoy et al. (2018) whereby a joint model is estimated for a pilot sample of 38 respondents. The combined SP/RP model relies on the construction of realistic choice sets for each RP observation from such data sources, contrarily to the SP experiment where the full choice set is provided by design. Such choice set reconstruction requires careful modeling and data (attributes) estimation which can be error prone and significantly affect estimation results. In this application, choice sets were obtained from Google Maps and from the pre-survey (car and bike ownership). Attributes such as travel times and access/egress times were also obtained from Google Maps and parking costs, taxi rates and public transportation fares were estimated from public online sources for the Boston/Cambridge area. However, some assumptions had to be made in this process both for the choice set (e.g.: limited by Google Maps API outputs) and its attributes (such as average transit and taxi waiting times) resulting in an unstable estimation procedure. Besides the need for testing such method to larger samples, such as the one in this current study, further research is needed to relax the limitations found in the choice set generation method proposed in Atasoy et al. (2018).

## 5. Discussion

This section presents the main contributions and limitations of the proposed SP method.

### *5.1. Methodological contributions*

By accounting for the trip-specific context and using smartphone sensing capabilities, this method overcomes several limitations associated with traditional pencil-and-paper surveys in terms of data quality and user experience.

#### 5.1.1. Data quality

This methodology is context-aware, as the user can easily relate to the RP trip. The SP attributes are generated based on the RP attributes together with the user specific and trip specific context, which ensures that the values presented to the user are realistic. The random design and the validation checks are easy to implement in real-time, and ensure that dominant and inferior alternatives are avoided automatically. The user is presented with the exact trip details (such as the arrival/departure times, trajectory, and origin/destination) and asked to validate these details before completing the choice task. Therefore the context described to the user is very similar to his/her real experience for the recalled trip. In terms of modeling, the unobserved context can be partially captured by an inertia parameter (similar to the one that was estimated in the application presented in Section 4). In addition, other contextual variables specific to the trip can be included such as the trip time of day, weather conditions, etc.

In traditional SP surveys, choice tasks are generated based on the same reference trip. On the other hand, this methodology allows the researcher to obtain different SP observations from the same individual corresponding to different trips. This allows for modeling more complex phenomena such as intra-personal heterogeneity (Becker et al., 2018), which is difficult to identify using choice tasks filled out in a short time period and corresponding to the same reference trip.

Compared to traditional SP surveys, smartphones allow for better and more flexible sampling of trips. The researcher is able to impose constraints on trip selection based on the research needs. For example, he/she might be interested in short (or long trips), work trips only, or trips involving multiple household members in order to study intra-household interactions.

Since users are presented with one choice task per day, some biases associated with traditional SP surveys are minimized. For example, the user's response in one day is less likely to affect his/her responses in the following days. In addition, the “fatigue” bias is reduced as users spend less time each day on the single choice task. Finally, since the time required for completing the choice task is recorded, it can be used to filter outliers (either directly or using a modeling approach similar to the one in Section 4).

#### 5.1.2. User experience

From the users’ perspective, this methodology reduces the mental burden which is typical in traditional SP surveys. Users are presented with profile attributes that are easy for them to understand and process as they are based on realized trips and mimic the presentation of these attributes in the market (e.g. in the case of app-based services). Uncertainty is also captured by presenting users with travel time ranges (in minutes) calculated as a percentage of the total travel time rather than showing percentages or fixed ranges.

Finally, users are not constrained by the time at which they have to complete the choice task, as the smartphone-based survey is easily accessible at any time. While this might be advantageous to users, it might result in inconsistencies depending on the time interval between the reference trip and completing the choice task. In our application, half of the respondents completed the choice task within the two days of the recalled trip, and 76% completed it within one week. Further research can investigate the effect of time between completing the reference trip and filling out prompted recall surveys.

### 5.2. Limitations

Although this method overcomes several limitations associated with SP surveys, there are still several limitations, most of which are common in SP survey design.

First, this method relies on design variables which have to be specified by the researcher. As in most SP designs, the researcher has to decide on the levels and ranges of these variables. Thus, the design needs to be checked for face validity and internal validity. Therefore, the logical checks presented in Sections 3.6 and 3.7 are necessary to ensure that the presented profiles are realistic, and that the data can be used in estimating behavioral models. The interface design can also be tested and further enhanced in order to ensure that respondents are not attentive to some attributes more than others.

Furthermore, smartphone data might have measurement errors (e.g., in distances and travel times), however, these are not problematic for the SP data. First, users are asked to validate their travel times beforehand and correct any errors recorded by the phone app. Second, by design, the SP attributes are pivoted off the RP attributes which are obtained from the trip planner (e.g., Google Maps). In model estimation, the researcher only uses the pivoted SP attributes (and not the initial attributes obtained from the trip planner). Nevertheless, these errors might become problematic if the researcher estimates a joint SP/RP model using the attributes obtained from travel advisors as these might have some measurement errors.

Another limitation associated with the joint estimation of SP/RP data is that the sensing platform does not collect data about the choice and consideration sets of users. Even though choice sets are obtained from trip planners such as Google Maps, the user might not consider the same available alternatives while making a choice. This issue can be overcome by modeling the consideration sets explicitly (e.g., Shocker et al., 1991; Swait and Ben-Akiva, 1987).

As in most RP-based SP surveys, justification bias might be expected as users might be more likely to select the alternative they chose in the RP in order to justify their previous choice. While this might introduce bias to the estimation results, it can be mitigated by including an “inertia” variable in the estimation as in the previous case study.

While this smartphone-based data collection method can be easily scalable, such platforms might be likely to attract young educated users. Therefore, considerable care is needed for sampling, or behavioral models should account for sampling bias.

### 5.3. FMOD application

This paper presented a novel application to a non-existing transportation service, FMOD, which has been studied before from the supply/operational side, but not from the demand angle (Atasoy et al., 2015). Understanding the factors affecting the demand for FMOD modes is critical for operators. First, it allows them to assess the feasibility/profitability of the service. It also allows for optimizing the operational aspects (such as fleet size, pricing of different types of services, etc.). During operation, it also enables operators to test different policies, such as changes in pricing, constraints on maximum waiting times, and allowable route deviations due to pooling passengers. Finally, it allows operators to identify potential market segments that are more likely to use the service; in our application, the results showed that younger respondents and respondents with higher income are more likely to use the service compared to others.

## 6. Conclusion

This paper presented a generic method for context-aware SP surveys leveraging state-of-the-art smartphone-based RP methods, in addition to an application to mode choice of future smart mobility solutions. This method utilizes a random experimental design and is flexible in terms of adding new modes or services. The current version includes a variety of modes including non-motorized modes, on-demand modes, private motorized modes, and transit. In the FMOD case study, the estimation of the mode choice model resulted in reasonable values of the model parameters and the estimated value of time.

A number of extensions can be considered in future research. Although the context-aware nature of our method overcomes common issues with SP surveys such as framing and realism, we might still expect some biases such as attribute non-attendance and justification bias. Future research will focus on making the SP survey incentive compatible in order to eliminate such biases (Ben-Akiva et al., 2019). In addition, smartphone-based SP surveys can be compared to traditional paper-based surveys and web-based surveys, and the effects on behavior can be investigated. Future work will also focus on parallelization and distributed processing in order to overcome the tradeoff between design realism (number of allowable flags) and the time needed to generate a valid choice task.

While this paper presented an application to FMOD, similar context-based SP surveys were deployed in Singapore and in the Greater Boston Area (GBA). In Singapore, data was collected on automated mobility on demand (AMOD) from a sample of 350 respondents who completed more than 2 500 choice tasks. A similar SP survey was implemented in GBA where 1 155 choice tasks were completed by 183 respondents to collect data for an app (*Tripod*) that incentivizes users to switch towards energy efficient travel choices.

The information obtained from the pre-survey allows for developing behaviorally rich demand models that do not only model choices from the SP menu, but also other choices such as subscriptions to app-based services and using these services for a particular trip. Xie et al. (2019) propose a framework for estimating demand for new transportation surveys using pre-survey, SP, and post-survey data whereby a nested structure is proposed for different models including (1) an app subscription model which predicts whether an individual will subscribe to the new service or not, (2) an app usage model, which predicts whether the individual will access the service (conditional on subscription), and (3) a model predicting the user's mode choice from the menu presented in the app.

Finally, the current version of the SP survey is trip-based; it considers half-tours only. Ongoing research is focusing on activity-based SP where travelers can be presented with alternatives for their entire travel and activity pattern. One of the initial attempts focusing on travel patterns was in the context of mobility-as-a-service packages (Matyas and Kamargianni, 2017) described in Section 2.2. Such research stream can further shed light into higher level individual context in fundamental travel decisions, at the reach of a smartphone.

## Acknowledgment

Part of the SP research was conducted under the “Urban Mobility Behaviors and Preferences Test Bed” project in collaboration with Ford and under the Tripod project funded by the U.S. Department of Energy, Advanced Research Projects Agency-Energy (ARPA-E), grant number: DE-AR0000611. We would like to thank the support of our colleagues Christina Lui, Dao Trung Hieu, and William Ko from SMART in the development of all app extensions and in the data collection process.

## References

Atasoy, B., Ikeda, T., Song, X., Ben-Akiva, M.E., 2015. The concept and impact analysis of a flexible mobility on demand system. Transport. Res. C Emerg. Technol. 56, 373–392.

Atasoy, B., Azevedo, C., Danaf, M., Mastera, J., Abou-Zeid, M., Cox, N., Zhao, F., Ben-Akiva, M., 2018. Context-aware stated preferences surveys for smart mobility. In: Presenteed at the 15th International Conference on Travel Behavior Research, Santa Barbara, California.

Becker, F., Danaf, M., Song, X., Atasoy, B., Ben-Akiva, M., 2018. Bayesian estimator for Logit Mixtures with inter- and intra-consumer heterogeneity. Transp. Res. Part B: Methodol. 117, 1–17.

Ben-Akiva, M., McFadden, D., Train, K., 2019. Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis. Foundations and Trends in Econometrics 10 (1–2), 1–144.

Bierlaire, M., 2016. PythonBiogeme: a Short Introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory. School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Bliemer, M.C., Rose, J.M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transp. Res. Part B Methodol. 44 (6), 720–734.

Bliemer, M.C., Rose, J.M., Chorus, C.G., 2017. Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. Transp. Res. Part B Methodol. 102, 83–104.

Bliemer, M.C., Rose, J.M., Hensher, D.A., 2009. Efficient stated choice experiments for estimating nested logit models. Transp. Res. Part B Methodol. 43 (1), 19–35.

Bunch, D.S., Louviere, J.J., Anderson, D., 1996. A Comparison of Experimental Design Strategies for Multinomial Logit Models: the Case of Generic Attributes.

University of California Davis Graduate School of Management Working Paper, pp. 11–96.
Chebli, H., Mahmassani, H.S., 2003. January). Air travelers' stated preferences towards new airport landside access mode services. In: Annual Meeting of Transportation Research Board, Washington DC.
Choudhury, C.F., Yang, L., e Silva, J.D.A., Ben-Akiva, M., 2018. Modelling preferences for smart modes and services: a case study in Lisbon. Transport. Res. Pol. Pract. 115, 15–31.
Correia, G., Viegas, J.M., 2011. Carpooling and carpool clubs: clarifying concepts and assessing value enhancement possibilities through a Stated Preference web survey in Lisbon, Portugal. Transport. Res. Pol. Pract. 45 (2), 81–90.
Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M., Zegras, P., 2013. Future mobility survey: experience in developing a smartphone-based travel survey in Singapore. Transport. Res. Rec.: Journal of the Transportation Research Board (2354), 59–67.
Cottrill, Caitlin D., Derrible, Sybil, 2015. Leveraging big data for the development of transport sustainability indicators. J. Urban Technol. 22 (1), 45–64.
Cox, N., 2015. Estimating Demand for New Modes of Transportation Using a Context-Aware Stated Preference Survey. MSc. Thesis, Massachusetts Institute of Technology.
Fan, Y., Chen, Q., Liao, C.F., Douma, F., 2013. January). UbiActive: a smartphone-based tool for trip detection and travel-related physical activity assessment. In: TRB 92nd Annual Meeting Compendium of Papers, Transportation Research Board, TRB 2013 Annual Meeting.
Fifer, S., Greaves, S., Rose, J., Ellison, R., 2011. A combined GPS/stated choice experiment to estimate values of crash-risk reduction. Journal of Choice Modelling 4 (1), 44–61.
Fifer, S., Rose, J., Greaves, S., 2014. Hypothetical bias in Stated Choice Experiments: is it a problem? And if so, how do we deal with it? Transport. Res. Pol. Pract. 61, 164–177.
Giffinger, R., Fertner, C., Kramar, H., Meijers, E., 2007. City-ranking of European Medium-Sized.
Goos, P., Vermeulen, B., Vandebroek, M., 2010. D-optimal conjoint choice designs with no-choice options for a nested logit model. J. Stat. Plan. Inference 140 (4), 851–861.
Hensher, D., 2004. Accounting for stated choice design dimensionality in willingness to pay for travel time savings. Transport. Res. B 38, 425–446.
Hensher, D., 2006. How do respondents process stated choice experiments? attribute consideration under varying information load. J. Appl. Econom. 21 (6), 861–878.
Hensher, D., 2008. Joint estimation of process and outcome in choice experiments and implications for willingness to pay. J. Transp. Econ. Policy 42 (2), 297–322.
Hess, S., Rose, J.M., 2009. Should reference alternatives in pivot design SC surveys be treated differently? Environ. Resour. Econ. 42 (3), 297–317.
Hoffman, D.L., Novak, T.P., 1996. Marketing in hypermedia computer-mediated environments: conceptual foundations. J. Mark. 50–68.
Hoffmann, S., Winter, J., Caro, F.G., Gottlieb, A., 2014. Effects of Video Enhancement in a Stated-Choice Experiment on Medical Decision Making.
Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice designs. J. Mark. Res. 307–317.
Jensen, A.K., Uggeldahl, K., Olsen, S.B., 2017. Watch and learn: using videos to provide information in choice experiments”. In: Proceedings of the International Choice Modeling Conference (ICMC).
Jiang, S., Ferreira, J., Gonzalez, M.C., 2017. Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. IEEE Transactions on Big Data 3 (2), 208–219.
Kessels, R., Jones, B., Goos, P., Vandebroek, M., 2011. The usefulness of Bayesian optimal designs for discrete choice experiments. Appl. Stoch Model Bus. Ind. 27 (3), 173–188.
Kuhfeld, W.F., Tobias, R.D., Garratt, M., 1994. Efficient experimental design with marketing research applications. J. Mark. Res. 31 (4), 545–557.
Lu, Y., Adnan, M., Basak, K., Pereira, F.C., Carrion, C., Saber, V.H., Loganathan, H., Ben-Akiva, M.E., 2015. Simmobility mid-term simulator: a state of the art integrated agent based demand and supply model. In: Transportation Research Board 94th Annual Meeting, Washington, DC.
Matyas, M., Kamargianni, M., 2017. A stated preference experiments for mobility-as-a-service plans. In: Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference. IEEE, pp. 738–743.
Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environ. Resour Econ. 30 (3), 313–325.
Prelipcean, A.C., Gidofalvi, G., Susilo, Y.O., 2015, June. Comparative framework for activity-travel diary collection systems. In: Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference on. IEEE, pp. 251–258.
Resource Systems Group (RSG), 2015. rMove, webpage. http://rmove.rsginc.com/index.html.
Rose, J.M., Bliemer, M.C., 2009. Constructing efficient stated choice experimental designs. Transport Rev. 29 (5), 587–617.
Rose, J.M., Bliemer, M.C., Hensher, D.A., Collins, A.T., 2008. Designing efficient stated choice experiments in the presence of reference alternatives. Transp. Res. Part B Methodol. 42 (4), 395–406.
Safi, H., Assemi, B., Mesbah, M., Ferreira, L., Hickman, M., 2015. Design and implementation of a smartphone-based travel survey. Transport. Res. Rec.: Journal of the Transportation Research Board 99–107 2526.
Sándor, Z., Wedel, M., 2002. Profile construction in experimental choice designs for mixed logit models. Market. Sci. 21 (4), 455–475.
Sanko, N., 2001. Guidelines for stated preference experiment design. Ecole Nationale des Ponts et Chaussées MB Thesis.
Sethuraman, R., Kerin, R.A., Cron, W.L., 2005. A field study comparing online and offline data collection methods for identifying product attribute preferences using conjoint analysis. J. Bus. Res. 58 (5), 602–610.
Seshadri, R., Kumarga, L., Atasoy, B., Danaf, M., Xie, Y., Azevedo, C., Zhao, F., Zegras, C., Ben-Akiva, M., 2019. Understanding preferences for automated mobility on demand using a smartphone-based stated preference survey: a case study of Singapore. In: 97th Annual Conference of the Transportation Research Board, Washington D.C.
Shocker, A.D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration set influences on consumer decision-making and choice: issues, models, and suggestions. Market. Lett. 2 (3), 181–197.
Street, D.J., Burgess, L., 2007. The construction of optimal stated choice experiments: Theory and methods Vol. 647 John Wiley & Sons.
Susilo, Y.O., Prelipcean, A.C., Gidofalvi, G., Allström, A., Kristoffersson, I., Widell, J., 2016. Lessons from a trial of MEILI, a smartphone based semi-automatic activity-travel diary collector. In: Stockholm City, Sweden.
Swait, J., Ben-Akiva, M., 1987. Empirical test of a constrained discrete choice model: mode choice in Sao-Paulo. Brazil. Transport. Res. B 21B, 103–115.
Thurstone, L.L., 1931. The measurement of social attitudes. J. Abnorm. Soc. Psychol. 26 (3), 249.
Train, K., Wilson, W.W., 2008. Estimation on stated-preference experiments constructed from revealed-preference choices. Transp. Res. Part B Methodol. 42 (3), 191–203.
van Cranenburgh, S., Rose, J.M., Chorus, C.G., 2018. On the robustness of efficient experimental designs towards the underlying decision rule. Transport. Res. Pol. Pract. 109, 50–64.
Walker, J., Wang, Y., Thorhauge, M., Ben-Akiva, M., 2015. D-efficient or deficient? A robustness analysis of experimental designs in a VOT estimation context. In: Transportation Research Board 94th Annual Meeting, Washington, DC.
Xie, Y., Danaf, M., Azevedo, C.M.L., Prakash, A., Atasoy, B., Jeong, K., et al., 2018. Modeling travel behavior with on-demand systems: the case of real-time sustainable travel incentives. In: hEART 2018: 7th Symposium of the European Association for Research in Transportation.
Yu, J., Goos, P., Vandebroek, M., 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. Market. Sci. 28 (1), 122–135.
Yu, J., Goos, P., Vandebroek, M., 2011. Individually adapted sequential Bayesian conjoint-choice designs in the presence of consumer heterogeneity. Int. J. Res. Mark. 28 (4), 378–388.
Zhao, F., Pereira, F.C., Ball, R., Kim, Y., Han, Y., Zegras, C., Ben-Akiva, M., 2015a. Exploratory analysis of a smartphone-based travel survey in Singapore. Transport. Res. Rec.: Journal of the Transportation Research Board 2 (2494), 45–56.
Zhao, F., Ghorpade, A., Pereira, F.C., Zegras, C., Ben-Akiva, M., 2015b. Quantifying mobility: pervasive technologies for transport modeling. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. ACM, pp. 1039–1044.
Zhu, H., Luo, X., Li, Y., Zhu, Y., Huang, Q., 2017. Comparing the efficiency and robustness of state-of-the-art experimental designs for stated choice modeling: a simulation analysis. Adv. Mech. Eng. 9 (2), 1–11.