

Identification of self-selection biases in field experiments using stated preference experiments

Working Paper [Version: 7 September 2016]

Stefanie Peer^{c,*}, Jasper Knockaert^a, Erik T. Verhoef^{a,b}

^a*Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam*

^b*Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam*

^c*Department of Socioeconomics, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna*

Abstract

If left unidentified and uncorrected, self-selection biases may greatly compromise the external validity of the outcomes of field experiments. We show that self-selection biases in terms of observed and unobserved characteristics can be well identified and corrected by means of a complementary stated preference (SP) experiment conducted among the participants and non-participants of a field experiment. In the SP experiment, respondents are confronted with hypothetical choice situations that closely resemble the choice situations present in the field experiment. The SP experiment does not only allow us to compare participants and non-participants with respect to their behavior and implied preferences in the hypothetical choice situations, but also renders it possible to infer how non-participants would have behaved if they had decided to participate, using an innovative modeling approach to elicit the corresponding preference structures. We apply this approach in the context of a large-scale field experiment in which train commuters received monetary rewards for traveling outside peak hours. We find strong self-selection biases, especially with respect to the marginal utility of income, which is significantly higher among participants of the field experiment.

Keywords: self-selection, representativeness, external validity, departure time decisions, revealed preference data, stated preference data, reward experiment, discrete choice modeling

JEL codes: C9, C25, D01, R4

*Corresponding author: speer@wu.ac.at (Stefanie Peer)

1. Introduction

Field experiments have become increasingly popular over the past years, as they constitute a useful compromise between natural experiments and laboratory experiments (e.g. List, 2007). However, just as traditional laboratory experiments, most of them involve voluntary participation. As the (expected) utility to be gained from participation usually varies across the population, participants may differ from non-participants in terms of observed and unobserved characteristics, potentially hampering the external validity of the results.¹ The resulting sampling biases are usually referred to as self-selection (or participation) biases. They have been studied extensively in the context of laboratory experiments (e.g. Harrison et al., 2009; von Gaudecker et al., 2012; Cleave et al., 2013), whereas they have received less attention in the context of field experiments, although arguably equally relevant (exceptions are the studies by Gautier and Klaauw (2012) and Dohmen and Falk (2010)). In this paper, we show that self-selection biases in field experiments can be well identified by conducting complementary laboratory experiments among participants and non-participants of a field experiment. This method is especially well suited to reveal self-selection biases with respect to underlying behavioral preferences, as – in contrast to the existing approaches aimed at tackling self-selection biases – it makes use of behavioral data from non-participants of the field experiment.

We study self-selection biases in the context of a field experiment that is characterized by voluntary participation and monetary, behavior-dependent rewards. The experiment concerns scheduling decisions of (train) commuters when trips outside the peak period are rewarded financially. The approximately 1000 voluntary participants of the field experiment were recruited among regular train commuters in the Netherlands. Over a period of multiple months, they received financial rewards for off-peak trips ranging between 1.5 and 4.5 Euro per trip. Their travel behavior before, during and after the reward experiment was measured via a smartphone app. Time-differentiated pricing in transportation has been advocated for a long time as a tool to reduce congestion, crowding and capacity requirements during peak hours (see for instance Vickrey, 1969). However, so far, empirical evidence on the effect of time-differentiated prices on the scheduling behavior at the level of the individual traveler is rather limited (e.g. Peer et al., 2015).

We analyze the scheduling preferences of the participants of the field experiment using random-utility-maximization (RUM) models. Persons choose between alternative train connections, each of which is characterized by the following attributes: (monetary) reward, travel time, schedule delays, number of transfers, travel time reliability and crowdedness. From the estimated coefficients, we can derive the marginal rate of substitution between any pair of attributes. Because a monetary attribute (i.e. the reward) is present, the monetary valuations of the attributes can be derived (e.g. the monetary value attached to reducing travel time by one hour). As these valuations are crucial inputs for transport policy appraisals, we are especially interested in identifying to which extent self-selection biases affect these valuations.

The approach we use to identify self-selection biases in the field experiment makes use of complementary choice data collected in a hypothetical, laboratory setting for both participants and non-participants of the field experiment. Such data are usually referred to as *stated preference* (SP) data as opposed to choice data collected in field experiments, which are usually referred to as

¹Note that the (expected) utility gained from participation may not only depend on the expected (monetary) benefits, but may also depend on implicit factors such as social preferences such as the desire to make moral choices (Levitt and List, 2007).

revealed preference (RP) data. We will adopt this conventional terminology in the remainder of the study.

The scheduling choice situations shown in the SP experiment (to both participants and non-participants of the field experiment) closely resemble those present in the the field experiment. Taking into account both the RP data from the field experiment and the SP data from the laboratory experiment, we are then able to estimate different scheduling models, which differ with respect to the classes of respondents for which separate sets of coefficients are estimated. In one variant, class membership in the scheduling model is determined by the observed participation behavior to the field experiment, whereas in the other variant it is determined by predicted participation behavior, which is derived from a model that explains participation behavior. Based on these estimations, we are able to infer the scheduling preferences of non-participants in the field experiment, and subsequently, using simulation techniques, we are able to show to which extent trips by non-participants would have been shifted from peak to off-peak periods if they had participated in the field experiment.

The estimated models provide valuable insights with respect to the differences between participants and non-participants of the field experiment, both in terms of observable characteristics (e.g. socio-economic variables, scheduling restrictions) as well as in terms of (usually unobserved) preference structures. In contrast to existing approaches to tackle selection biases (most importantly, propensity score matching and Heckman-selection-models), our modeling approach enables us to understand how non-participants would have behaved if they had participated in the field experiment. This is due to the inclusion of behavioral (SP) data also for non-participants. Unlike propensity score matching, our approach to identify self-selection biases does not exclusively rely on differences in observable characteristics between participants and non-participants; and unlike Heckman-selection-models, our method does not rely on the presence of strong instruments (i.e. variables that only affect the participation decision, but not the scheduling decisions) either.

The main contribution of this study is the development of a methodology for identifying self-selection biases in field experiments, based on the collection of high-quality complementary SP data among participants and non-participants of the field experiment. While we are not the first to acknowledge the advantage of combining these data sources (see for instance Harrison and List, 2004; List, 2006; Dohmen and Falk, 2010), we are, to our best knowledge, the first ones to use SP experiments conducted among participants and non-participants of a field experiment to identify self-selection biases in the field. Dohmen and Falk (2010) have a similar focus in their study of self-selection into educational professions, however, their identification strategy is different from ours: they use self-selection patterns identified in a laboratory experiment to predict self-selection in the field.

We find strong evidence of self-selection biases in our application. In particular, we are able to identify a striking difference in the marginal utility of income between participants and non-participants: consistent with their decision to participate, participants gain more utility from an additional unit of income. The policy implications of this finding are potentially large. For instance, if the marginal utility of income in field experiments with voluntary participation tends to be higher than for the general population, and these biased estimates of the marginal utility of income are used in policy appraisals, demand shifts due to a change in the price of one or more choice alternatives will be overstated. Policy appraisals may also be severely biased if the monetary valuations of relevant characteristics are based on the biased marginal utility of income. Despite its policy relevance, we are not aware of any previous research that has investigated or even demonstrated the presence of

selection biases with respect to the marginal utility of income.

Evidently, this paper also contributes to the extensive literature on the valuation of travel-related attributes (value of travel time savings, value of comfort, etc.), which generally has not paid much attention to self-selection issues despite their potentially sizable impact on the valuations (among the few exceptions are Mabit (2008), Fosgerau et al. (2010) and Significance et al. (2012)).

The remainder of the paper is structured as follows. Section 2 describes both the field and the laboratory experiment, as well as the corresponding recruitment procedures, and it gives an overview of the available data sources. Section 3 presents the estimation methodology as well as descriptive statistics. Section 4 contains the estimation results, and Section 5 the simulation results. Section 6 concludes.

2. Data

2.1. Overview

The models estimated in this paper try to explain how people who commute by train decide on their departure time for their morning commute, using two primary data sources: revealed preference (RP) data on real-life scheduling choice behavior, and stated preference (SP) data from hypothetical scheduling choice experiments. The RP data are available for participants of the field experiment, during which participants received monetary rewards if they traveled outside the peak. In contrast, SP data are available for participants as well as non-participants of the field experiment, where non-participants are persons who were invited to join the field experiment, but declined. Unlike in standard models aimed at tackling self-selection, behavioral data are thus available also for non-participants (albeit collected in hypothetical choice situations). In addition to the behavioral data, descriptive data on socioeconomic variables, characteristics of the usual travel behavior, scheduling preferences and scheduling restrictions have been collected from participants and non-participants, and will be used in the models. Table 2 gives an overview of the data availability.

Table 1: Data availability

Data	Participants	Non-participants
<i>Behavioral data</i>		
Revealed preference (RP)	×	
Stated preference (SP)	×	×
<i>Descriptive data</i>		
Socioeconomic characteristics	×	×
(Self-reported) scheduling preferences and restrictions	×	×
(Self-reported) characteristics of the usual travel behavior	×	×

2.2. Revealed preference (RP) data

The RP data originate from a peak avoidance experiment conducted in the Netherlands between 2012 and 2013. Participants (typically persons who commute to work by train; the recruitment procedure is described in more detail in Section 2.4) obtained monetary rewards for traveling outside the peak hours, implying that their travel costs were differentiated by time of the day. During the experiment, the travel behavior of the participants was measured using a customized smartphone

app that continuously transmitted the geo-location of the smartphone to a central database. The experiment consisted of 3 phases: 3 weeks of pre-measurement, 15-18 weeks of reward period and 4 weeks of post-measurement. During the pre- and post-measurement periods no rewards were distributed. During the reward period, the participants received a reward if they started their trip from the begin station to the end station of their participation origin-destination (OD)-pair² outside the morning peak (6:30-9:00), and another one if they started their trip from the end station to the begin station of their participation OD-pair outside the evening peak (16:00-18:30).³ Note that in this paper, we focus on the morning commute behavior, as the SP covers only morning commute decisions. The size of the reward was dependent on the length of the participation OD-pair, and on the reward regime (each participant went through a low and a high reward regime in random order), as indicated in Table 2.

Table 2: Rewards per off-peak trip

Distance class of participation OD-pair	Reward regime	
	high	low
≤ 25 km	2.5 €	1.5 €
26–40 km	3.5 €	2.5 €
> 40 km	4.5 €	2.5 €

No rewards were offered for off-peak travel during weekends, nor for a complete suppression of a trip. Moreover, the rewards were only paid if the participants had filled out all required surveys (among which the SP experiment discussed below), as well as various "logbooks" in which they were asked to provide detailed information on the train trips they made during specific weeks. More detailed information on the setup of the RP experiment, as well as the collected data, can be found in Peer et al. (2016).

2.3. Stated preference (SP) data

Both participants and non-participants completed an SP experiment consisting of 10 choice situations containing two departure time alternatives for the morning train commute. The choice alternatives differed in terms of departure and arrival times, rewards, expected levels of crowding and travel time variability. The respondents had to indicate their preferred alternative in all 10 choice situations. The choice situations presented in the SP experiment were designed such that they resembled real-life choice situations. This was done to improve the comparability of choice behavior in the SP and RP setting, and also to reduce biases due to the hypothetical character of the SP (e.g. Hensher, 2010). The latter is predominantly achieved by using a realistic framing

²All participants were assigned a so-called ‘participation OD-pair’ when they registered for the experiment, which generally corresponds to the rail link along which they (usually) commute.

³It is important to realize that participants were rewarded *for undertaking an app-registered train trip along their participation OD-pair outside the peak hours* in order to avoid giving them incentives for leaving trips unregistered. This implies that they were not rewarded for working from home, or for choosing alternative modes of transport; and moreover, their potential to earn rewards was independent from their travel behavior (including weekly frequency) during pre-measurement.

of the choice situation⁴ as well as realistic attribute levels, including personalized departure and arrival times, subject to the respondent's preferred arrival time (PAT) and usual travel time on the morning (train) commute.

10 different versions of the SP were designed, which were distributed randomly among participants. They were designed employing the so-called D-efficiency criterion, which seeks to minimize the determinant of the variance-covariance matrix subject to prior parameter estimates (the so-called "D-error"). The D-efficiency criterion is the most popular optimality criterion in the design of choice experiments (Kessels et al., 2006), and is especially adequate in the presence of respondent-specific reference points (as travel time and the PAT are in this paper) (Rose et al., 2008). We selected the optimal design (consisting of 10 versions) among 5000 randomly generated candidate designs subject to a standard additive specification of the utility function, to various constraints⁵, and to draws from prior parameter distributions.⁶ We compared the candidate designs by computing the average D-error across a given set of draws from the prior parameter distributions for each candidate design. More detailed information on the design of the SP can be found in the ancillary technical note by Peer and Knockaert (2016), which is available upon request.

2.4. Recruitment

The participants of the RP experiment were recruited from existing clients of all Dutch passenger railway companies during two recruitment rounds. Only persons who already owned an annual rail pass for a specific OD-pair or the entire Dutch network, were entitled to participate. As a result, 86065 eligible persons received a personal invitation via email during one of the two recruitment rounds. During the second recruitment round, in addition to the emails, also posters at the most frequented stations of the Netherlands were used to approach potential participants. The two recruitment rounds resulted in 1009 participants who downloaded and used the customized app successfully. This implies a response rate of 1.2%, if only those approached via email are considered. And the response rate is even lower if one takes into consideration that some train travelers may have registered as a consequence of having seen the posters at the stations rather than having received an email. A possible explanation for the low response rate is the recruitment method. In contrast to previous peak avoidance experiments (which typically yielded response rates between 5 and 20%⁷), potential participants were approached by email rather than by postal mail. A large

⁴It was randomly assigned whether the departure time decisions for a specific-respondent were framed in a short-run or long-run context, where the short-run context refers to departure time choices on the day of travel (if the usual connections is unexpectedly not available), and the long-run context refers to departure time decisions in a month time, if one would be notified that the timetable will change permanently in a month and the usual connection will not be available any longer. As the different versions of framing the SP did not seem to affect the choice behavior much, we pool the data here.

⁵We first defined the set of values that the attributes could assume, but additionally, we also specified how often specific levels should occur in one version of the SP. Moreover, in order to avoid extremely short or long travel times, as well as extremely early or late arrival times, we chose to use different reference travel times and PATs than the ones provided by the respondents in some cases. Finally, we also fixed that the left alternative always had an arrival time earlier or equal to the respondent-specific PAT, and the right alternative later or equal to the respondent-specific PAT.

⁶Designs optimized subject to prior parameter distributions are often referred to as Bayesian efficient designs, and are generally considered more stable as they will be quite efficient over a larger range of parameter values (e.g. Bliemer et al., 2008). The priors were based on the estimation results obtained during previous SP-based scheduling experiments in the Netherlands.

⁷E.g. Samenwerkingsverband Spitsmijden (2009); Knockaert et al. (2012a,b)

number of recipients may have ignored the emails, expecting that they contain advertisements for train journeys. Other reasons for declining the invitation can be derived from the outcomes of the final survey. In this final survey, 38% of those who ultimately decided to participate stated that they had doubts concerning their participation.⁸

Persons who were invited to participate in the RP experiment but declined (the so-called non-participants) were approached again a few months later (again via email) with the request to fill in a survey focusing on their reasons for declining participation in the RP experiment. In return they would receive a voucher worth 20€ (or alternatively, a donation of the same amount to a charity organization of their choice). The survey covered various questions on socioeconomic characteristics, travel behavior and scheduling restriction, but most importantly it contained the same SP experiment that also had to be filled in by the participants. Among the 3697 recipients of the email, 489 filled in the questionnaire (implying a response rate of ca. 13%).⁹

3. Modeling approach

3.1. Introduction

This section introduces our modeling approach, which is based on conventional (random-utility-maximizing) discrete choice models,¹⁰ and emphasizes the joint use of RP and SP data. The RP data describe choices between different existing train connections in the field experiment (the number of available alternatives differs across OD-pairs, and ranges from 8 to 89), whereas the SP data describe choices between two hypothetical departure time alternatives. Our modeling approach comprises two models:

1. The *scheduling model* explains the scheduling behavior of participants (in the RP as well as the SP experiment) and of non-participants (in the SP experiment) (Section 3.2).
2. The *participation model* explains why persons with specific characteristics choose to participate in the field experiment (Section 3.3).

We present three estimations of the scheduling model with multiple classes each (i.e. one set of coefficients is estimated for each class of respondents). The estimations differ with respect to the specification of the class membership function in the scheduling model (see Section 3.4). In one variant ('sequential estimation') class membership in the scheduling model is determined by the participation model, whereas in the other variant it is determined by observed participation ('independent estimation').

As for the participation model we will present only one estimation; for practical reasons its discussion is featured as the first stage of the 'sequential estimation' because of its role as class membership model in the subsequent estimation of the scheduling model.

⁸Reasons included the technical complexity of the experiment (access to a suitable smartphone: iOS or Android platform, preferably with unlimited internet access; online registration; app download), a perceived lack of information on the setup of the experiments and the reward scheme, the time requirements to fill in the surveys and logbooks as well as privacy concerns.

⁹Due to the rather generous reward of 20€ for filling in the survey, we are quite confident that the non-participants who filled in the survey are representative for all non-participants. Moreover, earlier research has shown that selection biases in laboratory experiments tend to be rather limited (e.g. Cleave et al., 2013; Abeler and Nosenzo, 2014).

¹⁰An in-depth discussion on discrete choice theory can be found in Ben-Akiva and Lerman (1985), Train (1986), Anderson et al. (1992) and Train (2003).

It can be shown that the independent and sequential estimation correspond to different ways to estimate the same behavioural model (see Appendix A). Both variants deliver consistent estimations, with the independent estimation being efficient whereas the sequential estimation is not. Our dataset however does allow to estimate only a subset of choice coefficients using the independent estimation, whereas the sequential estimation allows to estimate all coefficients of the scheduling model.

The descriptive characteristics of the dataset are discussed in Section 3.5.

3.2. Scheduling model

We assume that participants and non-participants make trade-offs between the reward, travel time, schedule delays, the number of transfers, travel time reliability and crowdedness when deciding on the departure time for their commute trips. Table 3 gives an overview of the alternative-specific attribute vector Z and the corresponding data-source-specific coefficient vectors β (for SP) and γ (for RP) that will be estimated.¹¹ The attributes have been defined in consistent ways between SP and RP for most variables. An exception is the number of transfers, which is present among the RP attributes, but has not been included in the SP.

Table 3: Overview of the scheduling choice variables

Variable	Coefficient		Definition
	SP	RP	
Z_R	β_R	γ_R	Reward [€]
Z_{TT}	β_{TT}	γ_{TT}	Travel time [hour]
Z_{SDE}	β_{SDE}	γ_{SDE}	Schedule delay early [hour]
Z_{SDL}	β_{SDL}	γ_{SDL}	Schedule delay late [hour]
Z_{TR}	β_{TR}	γ_{TR}	Number of transfers [#]
Z_{C3}	β_{C3}		Crowded train: seat occupancy is close to capacity
Z_{C4}	β_{C4}		Very crowded train: passengers need to stand
Z_C		γ_C	1 if expected train occupation rate > 90%, 0 otherwise
Z_{CU}		γ_{CU}	1 if no expectation available for occupation rate, 0 otherwise
Z_{REL}	β_{REL}		1 if $P(\text{more than 10 minutes delay at arrival}) > 10\%$, 0 otherwise
Z_{REL}		γ_{REL}	1 if $P(\text{more than 10 minutes delay at arrival}) > 5\%$, 0 otherwise

In the SP domain, the attributes, set of choice alternatives and the chosen alternatives are unambiguous, whereas for the RP they first had to be defined. Peer et al. (2016) describe the exact definitions in detail. Here, we only summarize this by noting that logbook data¹² and data from an app provided by the Dutch national railway company¹³ are used for the purpose of defining the

¹¹For readers familiar with the literature on scheduling models: note that we deviate here from the conventional notation of β (γ) being the shadow price associated with schedule delay early (late).

¹²Logbook data are only available for participants during 6 weeks of their participation. In each of these weeks participants had to provide detailed information on their travel behavior, including delays they encountered. We use the logbook data mostly for defining the expected reliability of each RP departure time alternative.

¹³These app data are used to define the reward, the travel time, the schedule delay, the transfer and the crowding attribute for each RP departure time alternative. Moreover, it is also used to define the available departure time alternatives for each RP departure time alternative.

(person-specific) attributes and the choice set.¹⁴ The chosen alternative is defined by matching the GPS-observed arrival and departure times to the corresponding choice set.

The topology chosen for the scheduling model is that of a classical multinomial logit model. The deterministic utility V_{jmn} related to person n , scheduling choice alternative j and choice situation m is linear in parameters, and defined as follows:

$$V_{jmn} = \begin{cases} \beta Z_{jmn} & \text{for SP choice alternatives} \\ \gamma Z_{jmn} & \text{for RP choice alternatives} \end{cases} \quad (1)$$

The random utility is defined by adding a random disturbance ϵ which is Gumbel distributed:

$$U_{jmn} = V_{jmn} + \epsilon_{jmn} \quad (2)$$

The choice observation y_{mn} is then defined as the alternative j with the highest utility U_{jmn} :

$$y_{mn} = \operatorname{argmax}_j U_{jmn} \quad (3)$$

Finally, the probability mass function P_s of alternative i in choice situation m by person n is:¹⁵

$$P_s(i|Z_{mn}; \beta, \gamma) = \frac{e^{V_{imn}}}{\sum_j e^{V_{jmn}}} \quad (4)$$

Extending the scheduling model to a class specification involves defining a class-specific vector of coefficients β_c and γ_c (one set of coefficients for each class $c = 0, \dots, C-1$ with C the number of classes), and a class-membership (participation) model with probability mass function $P_m(c|X_n; \alpha)$, which will be defined in Section 3.3. The probability mass function P_1 of the scheduling decisions $y_{1n}, \dots, y_{Mn}$ by individual n is then:

$$P_1(y_{1n}, \dots, y_{Mn}|Z_n, X_n; \alpha, \beta_0, \beta_1, \gamma_0, \gamma_1) = \sum_{c=0,1} \left[P_m(c|X_n; \alpha) \prod_m P_s(y_{mn}|Z_{mn}; \beta_c, \gamma_c) \right] \quad (5)$$

which accounts for the panel specification of the observations in the dataset. We will discuss different specifications of P_m in Section 3.4.

3.3. Participation model

We expect that participation is mainly driven by the expected aggregate reward and the effort associated with collecting the reward. In the context of this paper this may depend on socioeconomic variables, characteristics of the usual travel behavior, scheduling preferences and scheduling restrictions. Data on these variables have been collected from surveys among participants and non-participants of the RP experiment. Overall, 36 different variables (all specified as binary variables)

¹⁴The choice set has been defined based on timetable information from a popular smartphone app provided by the national Dutch railway company, which shows the relevant connections along a specific OD-pair. We take into account all connections between 5.30am and 10am, resulting in choice sets that vary in size (from 8 to 89 alternatives) across different OD-pairs.

¹⁵To simplify our notations we will note as Z_{mn} the set of the attribute vectors $\{Z_{1mn}, \dots, Z_{Jmn}\}$ that correspond to the choice set m evaluated by person n . Similarly we will note as Z_n the set of all attribute vectors Z_{jmn} faced by person n across all choices made.

enter the participation model as control variables. A full overview of the variables is presented in Table A.11. We note the vector of variables for person n as X_n , and the vector of corresponding coefficients as λ . The typology chosen for this model is that of an ordered probit. We define X_n^* as a distributed variable with disturbance term ω_n :

$$X_n^* = \lambda X_n + \omega_n \quad \text{and} \quad \omega_n \sim \mathcal{N}(0, 1) \quad (6)$$

The probability density function of X_n^* is $\phi(\omega_n - \lambda X_n)$ (with ϕ the standard normal probability density function). The observation I_n (0 for non-participation and 1 for participation) is then defined using an (estimated) threshold value α as:

$$I_n = \begin{cases} 0 & \text{if } X_n^* < \alpha \\ 1 & \text{if } X_n^* > \alpha, \end{cases} \quad (7)$$

and the probability mass function P_2 of I_n by:

$$P_2(0|X_n^*; \alpha) = \begin{cases} 0 & \text{if } X_n^* > \alpha \\ 1 & \text{if } X_n^* < \alpha \end{cases} \quad \text{and} \quad P_2(1|X_n^*; \alpha) = \begin{cases} 0 & \text{if } X_n^* < \alpha \\ 1 & \text{if } X_n^* > \alpha. \end{cases} \quad (8)$$

Note that the bivariate ordered probit applied here is mathematically equivalent to a binomial probit choice model.

3.4. Estimation

We discuss three model estimations corresponding to three different specifications of the class membership probability density function in the scheduling model, P_m :

1. *Independent estimation* (scheduling model with 2 classes): deterministic class membership depending on participation in the field experiment (Section 3.4.1)
2. *Independent estimation with latent classes* (scheduling model with 4 classes): within the classes of participants and non-participants of the field experiment, respectively, a pure latent class model determines class membership (Section 3.4.2)
3. *Sequential estimation* (2 classes): the participation model determines class membership in the scheduling model (Section 3.4.3)

3.4.1. Independent estimation

Here we estimate separate (and independent) scheduling models for each group: one model for participants, and a second model for non-participants. To estimate each model, we optimize the log-likelihood of the observations provided by the corresponding group. It will deliver us with two (SP-related) coefficient vectors β_c , one for participants (β_1) and one for non-participants (β_0), as well as one (RP-related) coefficient vector γ_c for participants (γ_1). The coefficient vector γ_0 for non-participants cannot be estimated, as for this group we do not have RP observations. The estimation defined here is equivalent to defining the class membership probability $P_m(c|I_n) = |c + I_n - 1|$. The log likelihood LL_n of the observations by participant n becomes then:

$$\begin{aligned} LL_n &= \ln \sum_{c=0,1} |c + I_n - 1| \prod_m P_s(y_{mn}|Z_{mn}; \beta_c, \gamma_c) \\ &= \begin{cases} \sum_m \ln P_s(y_{mn}|Z_{mn}; \beta_0) & \text{if } I_n=0 \text{ (non-participant)} \\ \sum_m \ln P_s(y_{mn}|Z_{mn}; \beta_1, \gamma_1) & \text{if } I_n=1 \text{ (participant)} \end{cases} \quad (9) \end{aligned}$$

3.4.2. Independent estimation with latent classes

In this estimation, we extend the independent estimation with two latent classes for each group. The latent class model is a specific case of a mixed logit specification and allows to capture unobserved heterogeneity in preferences across participants. The estimation of the models for the two groups (participants and non-participants) is still independent and done exclusively on observations provided by participants of the corresponding group. The log likelihood LL_n of the observations by participant n becomes then:

$$LL_n = \begin{cases} \ln \sum_{c=0,1} \alpha_c \prod_m P_s(y_{mn}|Z_{mn}; \beta_c) & \text{if } I_n=0 \text{ (non-participant)} \\ \ln \sum_{c=2,3} \alpha_c \prod_m P_s(y_{mn}|Z_{mn}; \beta_c, \gamma_c) & \text{if } I_n=1 \text{ (participant)} \end{cases} \quad (10)$$

with α_1, α_3 being estimated constant probabilities, and $\alpha_0 = 1 - \alpha_1$ and $\alpha_2 = 1 - \alpha_3$.

3.4.3. Sequential estimation

In the sequential estimation, we first estimate the participation model and subsequently use the simulated class membership probabilities P_2 as a class membership model P_m when estimating the scheduling models for both classes. Because each participant has a strictly positive probability to be member of any of both classes ($\forall c : P_2(c|X_n^*; \alpha) > 0$), this estimation provides an RP coefficient vector γ also for the class of non-participants. Note that the methodology applied here is mathematically equivalent to that of a (distributed) propensity score, where X_n^* is the propensity score. Estimating the participation model implies identifying values for the coefficients α and λ that maximise the log likelihood LL_p :

$$LL_p = \sum_n \ln \underbrace{\int \phi(\omega_n - \lambda X_n) \cdot P_2(I_n|X_n^*; \alpha) d\omega}_{L_{p,n}} \quad (11)$$

As $P_2(I_n|X_n^*; \alpha)$ is binary, we can solve the integral (as a definite integral over the interval where $P_2(I_n|X_n^*; \alpha) = 1$) and express the likelihood $L_{p,n}$ of the observation I_n (0 for non-participation and 1 for participation) by person n using the standard normal cumulative distribution function Φ :

$$L_{p,n} = \begin{cases} \Phi(\alpha - \lambda X_n) & \text{if } I_n=0 \\ 1 - \Phi(\alpha - \lambda X_n) & \text{if } I_n=1 \end{cases} \quad (12)$$

In the second step, we use the identified values $\hat{\alpha}$ and $\hat{\lambda}$ (so in this step $\hat{X}_n^* = \hat{\lambda}X_n + \omega_n$) to estimate the scheduling model, which implies identifying values for each β_c and γ_c that maximize the log likelihood LL_c :

$$LL_c = \sum_n \ln \int \phi(\omega - \hat{\lambda}X_n) \cdot P_1(y_{1n}, \dots, y_{Mn}|Z_n, \hat{X}_n^*; \hat{\alpha}, \beta_0, \beta_1, \gamma_0, \gamma_1) d\omega \quad (13)$$

$$= \sum_n \ln \underbrace{\int \phi(\omega - \hat{\lambda}X_n) \sum_{c=0,1} P_2(c|\hat{X}_n^*; \hat{\alpha}) \prod_m P_s(y_{mn}|Z_{mn}; \beta_c, \gamma_c) d\omega}_{L_{c,n}} \quad (14)$$

As the conditional choice probabilities $P_s(y_{mn}|Z_{mn}; \beta, \gamma)$ are not dependent on $\hat{X}_n^*$ and $P_2(c|\hat{X}_n^*; \hat{\alpha})$ is binary, we can again solve the integral:

$$L_{c,n} = \Phi(\hat{\alpha} - \hat{\lambda}X_n) \prod P_s(y_{mn}|Z_{mn}; \beta_0, \gamma_0) + (1 - \Phi(\hat{\alpha} - \hat{\lambda}X_n)) \prod P_s(y_{mn}|Z_{mn}; \beta_1, \gamma_1) \quad (15)$$

As noted in Appendix A, the sequential estimation of the model, with integration of the choice model over the distribution of the membership model variable $\hat{X}_n^*$ as is done here, is consistent, but not efficient.

3.5. Descriptives

The observational dataset is a combination of SP and RP observations. As the SP survey is focused on the morning peak, we will only take into account RP observations collected during the morning peak. For each participant to the SP experiment, all 10 SP observations are included. For participants to the RP experiment we sampled from the set of available RP observations:¹⁶ for participants with more than 11 RP observations we randomly sampled 11 observations; for participants with less than 11 RP observations we included them all.¹⁷ The resulting RP-dataset contains on average 10 observations per participant. Participants for whom no RP observations were included in the dataset are excluded from further analyses.

Table 4 provides the relative share of peak trips during the three phases of the RP experiment. We find that the relative number of peak trips is significantly lower during the reward period than during pre- and post-measurement, implying that the financial reward for off-peak trips indeed led to changes in scheduling behavior. More precisely, the relative share of peak trips decreased by 33% during the reward period compared to the pre-measurement, and by 13% during post-measurement (again compared to the pre-measurement). Interestingly, during the post-measurement the relative share of peak trips did thus not return to the pre-measurement level, but remained lower, possibly because participants got used to traveling outside the peak hours.

Table 4: Peak shares

Period	Number of observations	Relative share of peak trips
pre-measurement	568	68.8%
experiment	3410	46.3%
post-measurement	499	59.9%

Figure 1 provides the graphical illustration of the distribution of departure times during the three phases of the experiment (pre-measurement, reward period and post-measurement) of all RP observations taken into account in the analysis. The figure provides a strong indication of self-selection with respect to the usual travel behavior. It shows that already before the start of the experiment, participants tended to travel outside peak hours, in particular before the start of the morning peak period. Clearly, this pattern is not representative for the overall demand for train

¹⁶We start from the selection of observations used by Peer et al. (2016) for the estimation of RP models. In this study, we however include RP and SP observations by participants who have no trip registered in either pre-measurement or reward period, as well as observations from participants with a large difference in travel frequency between pre-measurement and reward period.

¹⁷We repeated the estimations presented in Section 4 for multiple random samples, and found that the estimation results barely differed.

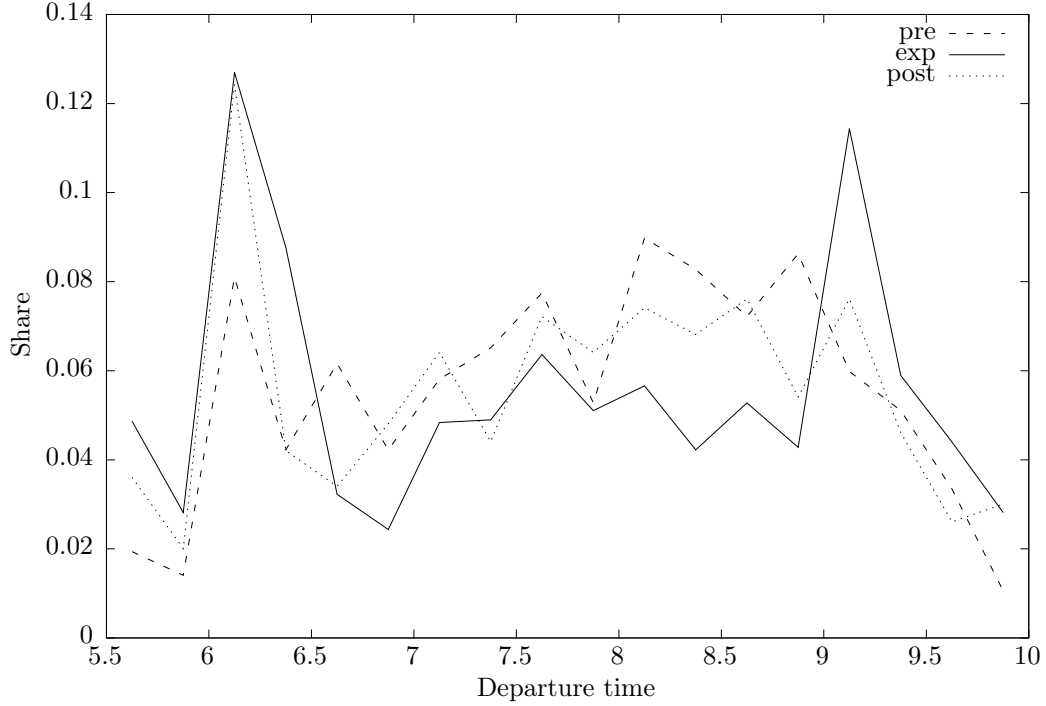


Figure 1: Distribution of selected RP-observations across the morning peak

travel among Dutch train commuters. Instead, it is an indication that those who can earn the rewards most easily (i.e. those who do not have to change their scheduling behavior in order to receive the rewards) are relatively more likely to participate. Further descriptive evidence concerning self-selection biases will be provided in Section 4.3.1, where the results of the participation model are discussed.

4. Model estimations

4.1. Introduction

In this section, we first present the results of the independent estimation of two scheduling model variants as introduced in Sections 3.4.1 and 3.4.2, respectively. The results of the scheduling model with deterministic membership is displayed in Table A.9, whereas the results of the latent class model are presented in Table A.10. We then proceed with presenting the results of the sequential estimation as introduced in Section 3.4.3, consisting of the participation model and the scheduling model. The results of the participation model are shown in Table A.11, and those of the scheduling model in Table A.9.

All models are estimated using a pre-release version of Pythonbiogeme 2.4 (Bierlaire, 2006). To find the global optimum, we run a series of randomly initialized estimations. Based on the identified coefficient values, the estimations are clustered in order to identify (local) optima. The number of estimations as well as the number of identified (local) optima are used to evaluate a stopping criterion after which the optimum with the best log likelihood value is assumed to represent the

global optimum. Both the clustering and the stopping criteria use an ad hoc heuristic. Robust clustered standard errors are used. The estimation statistics for all models discussed below can be found in Table 5.

4.2. Independent estimations

First, we estimate the independent model in which class membership is deterministic and determined by observed participation behavior (Section 4.2.1). This means, separate coefficient vectors are estimated for participants and non-participants of the field experiment. Second, we estimate a latent class version of this model, where the existing scheduling model is extended with two latent classes for participants and two latent classes for non-participants. This model is presented in Section 4.2.2.

4.2.1. Independent estimation (with deterministic class membership)

In this first model, we estimate RP-based coefficients for the scheduling model for the participants of the field experiment, and SP-based coefficients for both participants and non-participants (the results are displayed in Table A.9). Across these three sets of estimated coefficients, we consistently find that travel time and schedule delays (i.e. the earliness/lateness with respect to the preferred arrival time) are the largest determinant of the scheduling decisions. In particular, we find that for all three sets of coefficients it holds that the disutility associated with 1 hour of travel time is larger than the disutility associated with 1 hour of schedule delay early, which in turn is larger than the disutility associated with 1 hour of schedule delay late. Note that the same relationship has been found in a similar field experiment targeted at car drivers (e.g. Peer et al., 2013, 2015). The remaining coefficients have the expected sign, although not all of them are significant. We find that the participants exhibit significant disutility from transfers in the field experiment (transfers have not been included as attribute in the SP). The reliability and crowding coefficients are insignificant in the RP domain¹⁸, but significant for the two sets of SP-based coefficients (except for the coefficient for "slight crowding", β_{C3}).

¹⁸See Peer et al. (2016) for a detailed overview of the possible reasons, which include low levels of variation in the reliability and crowding attributes across the choice alternatives, travelers' lack of awareness of the variation, correlations between attribute values, endogeneity, and issues related to data quality.

Table 5: Estimation statistics

Statistic	Independent	Independent (LC)	Sequential	
			participation	scheduling
Model null log likelihood	-19384.4	-19384.4	-655.0	-19384.4
Model final log likelihood	-15956.0	-14904.6	-482.1	-15328.5
Number of model coefficients	22	46	37	30
Number of RP observations	4477	4477		4477
Number of SP observations	9450	9450		9450
Number of individuals	945	945	945	945
Number of converged estimations	7	21	7	39
Number of (sub)optima	1	3	1	6
Number of unconverged estimations	0	8	0	0

Next, we compare the SP-based coefficient estimates for participants and non-participants. Here, also a comparison in terms of the coefficient estimates, not only in terms of the valuations, seems justifiable, as the experiment was the same for both groups, and hence we assume that the scale (of utility) is not different. However, despite the SP-design being identical for participants and non-participants of the field experiment, we find substantial differences in the coefficient estimates across the two groups. The most striking difference appears in the reward coefficient (which is usually interpreted as the marginal utility of income), which is more than four times higher for the participants than the non-participants. Since the other coefficients differ by a much smaller order of magnitude, also the SP-based valuations of the trip attributes are about four times as high as the corresponding RP-based valuations. The lower marginal utility of income found for the non-participants is clearly consistent with their decision not to participate in the field experiment: since they gain less utility from an additional Euro of income, they are less inclined to participate. We would have expected that the disutility from schedule delays is higher for non-participants, which would also be a reason for not participating. However, that does not seem to be the case in general: the coefficient for schedule delay early is higher among non-participants (in absolute terms), but for schedule delay late actually the reverse is found. We find small differences in the disutility from (heavy) crowding and unreliability, both being slightly higher among participants (again in absolute terms). Also this is a plausible result, as commuters with a higher disutility from crowding and unreliability will be more inclined to participate in a peak avoidance experiment.

Generally, we can conclude that the differences in SP coefficient estimates (specifically, the difference in the reward coefficient) between participants and non-participants imply that self-selection is present in terms of behavioral preferences. However, based on the independent estimation we are not able to tell to which extent the selection biases in terms of behavioral preferences are caused by differences in observed characteristics. The sequential model will shed more light on this question.

We proceed with comparing the RP and SP estimates for the participants in more detail, in order to identify possible differences between the scheduling behavior of the participants in real-life and in the hypothetical setting. Note that the coefficients cannot be compared directly in terms size, as the scale is expected to differ between the SP- and the RP-based coefficients. We find that in general the monetary valuations are higher for the SP-based coefficients. We speculate that there may be a hypothetical bias in the reward coefficient, which renders the reward coefficient higher in the RP than in the SP setting. A plausible explanation is that participants react more strongly to a real reward than a hypothetical one. The same result has for instance been found for the peak avoidance experiment targeted at car commuters (Peer et al., 2013). Moreover, we find that the ratio of the value attached to travel time reductions and the value attached to schedule delay reductions is substantially higher in the RP than in the SP domain (2.9 vs. 1.7 for schedule delay early, and 3.7 vs. 2.2 for schedule delay late, respectively). Also that result is consistent with the findings of Peer et al. (2013). A possible reason is that the scheduling restrictions may appear more binding (and hence more costly) in the hypothetical setting.

While the SP and RP valuations for the participants are not identical (which is consistent with most studies that compare SP and RP estimates), we can still conclude that the elicited SP-preferences are a good representation of the real-life (RP) preferences of the participants. First, the differences are fairly small in absolute size (in the range of a few Euro). Second, the relative importance of the trip attributes is very consistent between SP and RP: in both data sources, travel time is valued highest, then followed by schedule delay early and schedule delay late. The valuations of the remaining attributes are of relatively minor size. The latent class model presented in the

following section will allow for further insights into the comparability of SP and RP estimates for the participants.

4.2.2. Independent estimation with latent classes

The model discussed in this section (see Table A.10) is identical to the one discussed in the previous section, with the exception that for both participants and non-participants two latent classes (instead of one class each) are introduced. The latent classes capture unobserved heterogeneity within each of the two groups.¹⁹ The class membership equation is simulated using a constant probability density, and hence is not determined by any socio-economic variables or similar.²⁰

We find that main results of the model discussed in the previous section remain valid also when heterogeneity is explicitly taken into account. Specifically, the SP-based reward coefficient is again significantly smaller for both latent classes of non-participants than for both latent classes of participants. Nevertheless, the results also hint at significant heterogeneity within the group of participants as well as the non-participants.

For the non-participants, we find a large degree of heterogeneity with respect to the travel time and schedule delay coefficients: the first class exhibits a high degree of disutility from schedule delay late and a limited disutility from travel time, whereas the second class has a relatively high disutility from earliness and from travel time. Since the reward coefficients for the two classes of non-participants are quite similar, this finding translates also into heterogeneity in terms of the valuations attached to reducing travel time and schedule delays.

Also for the participants the reward coefficients of both classes are very close to each other. Moreover, participants appear to be rather uniform with respect to the value attached to reducing travel time and schedule delays in the RP domain. Heterogeneity can mainly be observed in the SP domain concerning the scheduling valuations: the first class of participants values schedule delay late significantly higher than schedule delay early, while the reverse is true for the second class. For the participants, we can moreover observe that on average the SP and RP valuations converge, compared to the independent estimation without latent classes, implying that heterogeneity may be one of the reasons why SP and RP valuations differ.

4.3. Sequential estimation

In the sequential model estimation, the (probabilistic) membership of the two classes (participants and non-participants) is simulated by the participation model.²¹ Class-specific coefficient vectors β and γ are identified. Note that this estimation methodology allows us to derive RP coefficients also for the class of non-participants based on the RP behavior of participants, as is in the membership simulation for every person a strictly positive probability is associated with both classes. The results of the participation model are discussed in Section 4.3.1 and those of the scheduling model in Section 4.3.2.

¹⁹Also a higher number of latent classes than two has been tested, however, the size of one latent class becomes quite small then, leading to insignificant coefficients and unstable optima (i.e. a large number of sub-optima with similar log likelihoods, but fairly different coefficient values).

²⁰The class membership probability is computed at the person-level and hence equal across all (SP and RP) observations available for one person. The constant has been chosen after it has been verified that heterogeneity is mainly driven by unobserved factors rather than the variables considered in the participation model.

²¹We tried to estimate also a model with a latent class structure (similar to the one discussed in Section 4.2.2), however, it turned out to be too demanding on the data.

4.3.1. Participation model

Table A.11 presents the results of the participation model, which aims at explaining the participation decision using socioeconomic variables, characteristics of the usual travel behavior, scheduling preferences and scheduling restrictions. All 37 explanatory variables have been defined as dummies. Besides the estimation results, the table also presents the sample averages for these dummies.

The model results indicate that significant self-selection biases with respect to some of the 37 included variables are present. When looking at both the sample averages and the estimation results, one can observe that some variables exhibit fairly large differences in terms of the sample averages, but turn out insignificant in the (multivariate) participation model. In our elaborations below, we put the focus on those variable groups within which at least one or more coefficients are significant at the 5% level in the participation model, as due to the large number of estimated coefficients not each coefficient can be discussed separately.

As expected, we find that those persons who can earn rewards more easily are more likely to participate. First, participants are more likely to face less stringent (work-related and private) scheduling restrictions than non-participants ($x_{21} - x_{27}$), which renders it easier to shift their departure times to off-peak hours in order to collect the rewards.²² Second, participants have a higher probability of having their usual scheduled departure time outside the peak hours ($x_{13} - x_{14}$) than non-participants.²³ In that case, little or no adaptation of their schedules is necessary for receiving the rewards. Third, also the preferred departure time of the participants is more likely to be located outside peak hours ($x_{17} - x_{18}$) than among non-participants, which again decreases the costs of participation as it is attractive to the participants to travel outside peak hours.²⁴ Note that these first three categories of variables may be endogenous to some degree. Nevertheless, we find that dropping these variables barely affects the results of the remaining coefficients of the participation model, and neither the coefficients of the scheduling model (see Section 4.3.2).

Moreover, participation is positively correlated to travel frequency (specifically, participants are more likely to commute 5 days/week than non-participants), as a higher number of weekly trips also implies a higher number of opportunities to gain the rewards for off-peak travel (x_{31}). Finally, participants are also less likely to have administrative jobs or to be active in the service sector than non-participants ($x_9 - x_{12}$).²⁵ A plausible explanation is that these jobs are typically associated with fairly inflexible working (start/end) hours.

Besides the above mentioned variables, we also controlled for various other socio-economic variables, such as income, education, household structure, most of which are not significant in the participation model. One exception is that participants are more likely to be member of the lowest income class (x_1), possibly rendering the rewards more attractive to them due to a higher marginal utility of income. The other exception is that participants are more likely to have a higher education level than non-participants (x_5). A plausible explanation may be that highly educated individuals are more interested in participating in scientific experiments, but also that they tend to have more flexible jobs (which may not fully be captured by the questions concerning scheduling restrictions

²²Although for all 7 variables it is true that participants are more flexible than non-participants, only two are significant at the 5% level in the participation model.

²³While this is true for the morning and evening off-peak hours, only the evening off-peak hour is significant at the 5% level.

²⁴Again, the coefficient is only significant at the 5% level for the preferred departure time during the evening commute.

²⁵Note that the reference category corresponds to "management" professions.

in contained in $x_{21} - x_{27}$). Finally, we find that ownership of a 1st class travel pass is positively correlated to participation (x_{34}).

We can conclude that the participation model yields plausible results, which indicate that the decision to participate in the field experiment has generally been taken on rational grounds. It implies that participants and non-participants are significantly different from each other along various (observable) dimensions. Clearly, the predictive power of the model is not perfect. However, it provides a valid indicator for the participation likelihood of a specific person. Moreover, it is interesting to note that most variables that turned out significant are variables that can typically only be gathered via dedicated surveys (e.g. scheduling restrictions, usual travel behavior) rather than from pre-existing databases that typically only include socio-economic variables.

4.3.2. Scheduling model

Table A.9 displays the results of the scheduling model that is part of the sequential estimation. As the participation decision is not predicted perfectly, the differences in the estimated coefficients across the two classes are not as distinct as in the independent models. However, since we still find significant differences between the coefficients associated with the two classes, we can conclude that selection biases with respect to observable characteristics and selection biases with respect to the underlying preference structures are correlated with each other. More specifically, we can confirm that the reward coefficient shows a similar pattern in the RP domain as in the SP domain: we find that the reward coefficient is higher for the class of participants than for the class of non-participants both in the SP as well as in the RP domain. Also the estimates for the other coefficients are largely consistent with the findings of the previous model. We find that the selection biases in the observed and unobserved characteristics (preferences) are correlated with the participation decision; otherwise we would not find significant differences between participants and non-participants.

Comparing the SP results obtained in this sequential estimation to the ones in obtained in the independent estimation, we observe that the behavioral difference (captured by the scheduling model) between the classes (participants and non-participants) is understated. As noted before, these are two consistent estimates of the same behavioural model (see Appendix A), with only the independent estimation approach delivering efficient estimates. Hence we expect the differences obtained for the RP results to be understated to a similar extent when using the (same) inefficient estimation approach.

As noted before, the sequential estimation can be interpreted as an application of propensity score matching, hence this methodology suffers from the same limitation of delivering inefficient estimates and as a result underestimating behavioral differences between the classes.

5. Simulations

In this section, we will use the models estimated in the previous section to simulate behavior of participants and non-participants as a function of reward levels. The aim is to gauge the behavioral adaptations made by non-participants and to assess how this is different from how participants behave.

5.1. Behavioural indicators

In our simulations, we will use the probability of choosing a departure time in the peak hour interval as the relevant behavioral indicator, as it corresponds to the (expected) share of peak trips.

We will calculate this probability in two ways. A first indicator P_{part} exclusively uses the choice coefficients β_1, γ_1 of the class of participants:

$$P_{part} = \frac{\sum_n \sum_m \sum_j p_{jmn} P_s(j|Z_{mn}; \beta_1, \gamma_1)}{\sum_n \sum_m 1} \quad (16)$$

with $p_{jmn} = 1$ if alternative j in choice situation m by person n has an in-peak scheduled departure time, and $p_{jmn} = 0$ otherwise. The change in P_{part} as a function of the reward level can be considered as an upper limit to behavioral adaptation. A second indicator $P_{nonpart}$ similarly exclusively uses the choice coefficients β_0, γ_0 of the class of non-participants:

$$P_{nonpart} = \frac{\sum_n \sum_m \sum_j p_{jmn} P_s(j|Z_{mn}; \beta_0, \gamma_0)}{\sum_n \sum_m 1} \quad (17)$$

The change in $P_{nonpart}$ as a function of the reward level can be considered as a lower limit to behavioral adaptation.

5.2. Reward Scenarios

For a series of reward scenarios both behavioral indicators P_{part} and $P_{nonpart}$ are calculated. For our simulations we use the levels of the independent variables X_n and Z_{mn} in the observational dataset used for model estimation, and set the reward variable arbitrarily, to define the different simulation scenarios.

In Table 6 we present the value of both behavioral indicators for different levels of the reward variable for a series of scenarios defined by (overlapping) selections of observations:

- observations by *participants* (observed participation: $I_n = 1$)
- observations by *non-participants* (observed non-participation: $I_n = 0$)
- observations by *probably participants* (observed participants with a high probability of participation: $I_n = 1$ and $P_2(1|X_n^*; \alpha) > 0.5$)
- observations by *maybe participants* (observed participants with a low probability of participation: $I_n = 1$ and $P_2(0|X_n^*; \alpha) \geq 0.5$)

For each selection of observations we simulate the behavioural indicators for three levels of the reward variable. In the reference scenario we set the reward variable Z_R to zero for all observations. Subsequently we simulate behavior with the reward variable Z_R set at 2€ and 4€ for all observations. The combination of the three off-peak reward levels and four selections of (independent variable levels of) observations delivers twelve simulation scenarios.

For each of the twelve scenarios both behavioral indicators will be calculated using the behavioral coefficients (β_c, γ_c) of the independent model (without latent classes) as well as of the sequential model.

We will present separately simulations based on RP-observations (used to estimate the coefficient vectors γ_c) vs SP-observations (used to estimate the coefficient vectors β_c).

Table 6: Simulations (results for 2€ reward levels not shown)

model	indicator	independent variable scenario (selection of observations and reward level)							
		participants		non-part.		probably part.		maybe part.	
		0€	4€	0€	4€	0€	4€	0€	4€
<i>stated preference</i>									
independent	$P_{nonpart}$	53,5%	51,0%	53,2%	50,7%	53,5%	51,0%	53,5%	50,9%
	P_{part}	53,5%	42,6%	53,2%	42,3%	53,5%	42,6%	53,3%	42,4%
sequential	$P_{nonpart}$	54,0%	48,1%	53,7%	47,7%	54,0%	48,1%	54,1%	48,2%
	P_{part}	52,7%	44,8%	52,6%	44,6%	52,8%	44,9%	52,4%	44,5%
<i>revealed preference</i>									
independent	$P_{nonpart}$								
	P_{part}	66,9%	43,8%			65,2%	41,8%	71,9%	49,4%
sequential	$P_{nonpart}$	62,9%	47,6%			60,6%	45,1%	69,6%	54,6%
	P_{part}	64,3%	44,3%			62,6%	42,7%	69,2%	48,9%

Table 7: Simulations (change vs corresponding 0€ reward scenario expressed in percentage point)

model	indicator	independent variable scenario (selection of observations and reward level)							
		participants		non-part.		probably part.		maybe part.	
		2€	4€	2€	4€	2€	4€	2€	4€
<i>stated preference</i>									
independent	$\Delta P_{nonpart}$	-1.3	-2.5	-1.3	-2.5	-1.3	-2.5	-1.3	-2.5
	ΔP_{part}	-5.5	-10.9	-5.5	-10.9	-5.5	-10.9	-5.5	-10.9
sequential	$\Delta P_{nonpart}$	-2.9	-5.9	-3.0	-5.9	-2.9	-5.9	-3.0	-5.9
	ΔP_{part}	-4.0	-7.9	-4.0	-7.9	-4.0	-7.9	-3.8	-7.9
<i>revealed preference</i>									
independent	$\Delta P_{nonpart}$								
	ΔP_{part}	-11.2	-23.2			-11.5	-23.4	-10.6	-22.5
sequential	$\Delta P_{nonpart}$	-7.4	-15.4			-7.5	-15.5	-7.1	-14.9
	ΔP_{part}	-9.7	-20.0			-9.7	-19.9	-9.5	-20.2

Table 8: Subselection of participants in simulation

Selection	Definition	Individuals	RP		SP	
			Indiv	Obs	Indiv	Obs
participants	$I_n = 1$	447	447	4477	447	4470
probably participants	$(P_2(1 X_n^*; \alpha) \geq 0.5) \& (I_n = 1)$	329	329	3308	329	3290
maybe participants	$(P_2(0 X_n^*; \alpha) > 0.5) \& (I_n = 1)$	118	118	1169	118	1180
non-participants	$I_n = 0$	498	0	0	498	4980

5.3. Simulation results

The results of the simulations are presented in Tables 6 (peak trip shares) and 7 (change in peak trip share resulting from the introduction of a reward). The columns each represent a simulation scenario (choice situations defined by values of the independent variables), whereas each row represents a set of consumer preferences (defined by a coefficient vector estimated in Section 4).

For stated preference the initial share of peak trips (reward level 0€) is very similar across the different scenarios (between 52.4% and 54.1%; see Table 6). This is caused by the design of the SP survey. For any scenario the difference in initial share of peak trips between the two indicators P_{part} and $P_{nonpart}$ is small (between 0.0pp and 1.7pp). When the reward is set to 4€, a reduction in peak share (compared to the corresponding 0€ reference scenario) results for all indicators and scenarios (−10.9pp for the independent model and −7.9pp for the sequential model; see Table 7).

For revealed preference based simulations not all scenarios can be defined: the $P_{nonpart}$ indicator cannot be determined for the independent model as we do not have an estimation for the corresponding model coefficient vector γ_0 . Similarly the scenarios based on (the levels of independent variables of) observations by non-participants cannot be defined, as we lack any such observations. We do observe that the initial share of peak trips is larger for participants with lower values for $P_2(1|X_n^*; \alpha)$ (maybe participants; initial peak share between 69.2% and 71.9%; see Table 6) compared to those with larger probabilities of participation (probably participants; initial peak share between 60.6% and 65.2%) as simulated by the estimated participation model. This is in line with expectations. When the reward is set at 4€, a reduction in peak share results for all indicators and scenarios (between −14.9pp and −23.4pp; see Table 7). This reduction is bigger than for the corresponding SP-scenario.

The use of the different scenarios based on different selections of observations is to study the impact of the initial peak shares (at 0€ reward) on the reduction in peak travel when a reward is introduced (expressed by ΔP_{part} and $\Delta P_{nonpart}$). From Table 7 it becomes obvious that even when these initial levels are different, the impact of a reward on the peak shares is (almost) identical across scenarios based on different selections of the observations (i.e. in Table 7 the columns for probably participants are almost identical to the columns for participants, and so on). In the subsequent discussion of variation in ΔP_{part} and $\Delta P_{nonpart}$ across different scenarios, models, and SP vs RP, we will limit to the scenarios defined by the (levels of the independent variables in) observations from participants (the two leftmost columns of Tables 6 and 7).

Focusing on the indicators based on the sequential model, we observe that ratio of the reduction in peak share (expressed in percentage point and presented in Table 7) between $\Delta P_{nonpart}$ and ΔP_{part} is very similar for the stated preference and revealed preference simulations: 74.4% (−5.9/−7.9) vs 76.8% (−15.4/−20.0) (at 4€ reward). This result is not implied by the model specification, and is hence a genuine behavioural observation.²⁶ For the independent model we can only calculate this ratio for the stated preference simulations, and we find a lower value 23.4% (−2.5/−10.9). If we assume that also for this model the ratio is similar for revealed preference, we may expect a reduction in peak share of 5.4pp for $\Delta P_{nonpart}$ compared to the simulated 23.2pp for ΔP_{part} (at 4€ reward). The lower value identified by the independent model could be considered as a lower boundary whereas the higher value from the sequential model the represents an upper boundary

²⁶This finding means that the reaction to a reward is to a similar extent stronger in RP situations compared to SP situations for travellers with a lower simulated participation probability (a factor of $-15.4/-5.9 = 2.6$ stronger) compared to travellers with a higher simulated participation probability ($-20.0/-7.9 = 2.5$).

for the reduction in peak trips among non-participants, which is hence expected to be between 25% and 75% of the reduction among participants.

From the simulations it becomes clear that the sequential model (which is equivalent to a distributed propensity score approach) underestimates (compared to the independent model) the differences in behavior between participants (ΔP_{part}) and non-participants ($\Delta P_{nonpart}$) in a stated preference setting. As discussed in Appendix A both models correspond to different estimations of the same integrated behavioural model, and both deliver consistent estimates but only the sequential model is an efficient estimate of the underlying integrated model. The sequential model is not efficient in estimating the integrated behavioural model because of the limited predictive power of the participation model identified in this study, which results in a different estimate for the class-specific scheduling models (participants vs non-participants).

Another cause of the differences identified may be found in the stochastic nature of the participation decision in the sequentially estimated model that allows for some other heterogeneity (unrelated to the participation decision) to be picked up by the scheduling model (which would point to the independent model without latent classes being a simplification and hence a misspecification). This explanation seems to be supported by the better scheduling decision likelihood of the sequentially estimated model (-15328.5 vs -15956.0 for the independent model; see Table 5), and may also underpin the observation that for the revealed preference observations the sequentially estimated model obtains weaker responses as a result of the reward for both indicators compared to the independent model.

6. Conclusions

This paper shows that self-selection biases in field experiments can be well identified by means of collecting additional, complementary data from stated preference (SP) experiments from participants and non-participants of a field experiment. We estimate multiple models, in which the first stage models the decision to participate in the field experiment, and the second stage models the behavior of participants in the revealed preference (RP) and the SP experiment, and of non-participants in the SP experiment. In contrast to existing methods (e.g. propensity score matching, Heckman selection model), this setup is especially well suited for revealing selection biases with respect to the underlying preference structures.

We applied this approach in the context of a peak avoidance experiment, in which train commuters were able to receive rewards for traveling outside peak hours. Such time-differentiated prices are often a more cost-efficient alternative to the typically very costly expansions of train and network capacity. The reward experiment was successful in shifting trips from peak- to off-peak periods: the relative share of morning peak trips among participants dropped by 33% as a result. However, this result should not be generalized to the wider population of Dutch train commuters, as we find that participants of the field experiment differ significantly from non-participants, due to the presence of substantial self-selection biases: persons who could earn the rewards for off-peak travel more easily (e.g. due to less stringent scheduling restrictions, a higher commuting frequency) are found to be more likely to participate. Moreover, we find that participants have a significantly higher marginal utility of income than non-participants, implying that the monetary valuations of the trip attributes are significantly lower among participants. Simulation results then indicate that the reduction in peak trips among non-participants is expected to be between 25–75% of the reduction among participants. We found that the result based on estimations equivalent to a propensity score matching approach corresponds to the upper bound of this interval. It implies that the propensity

score matching approach obtains a low value for differences in behavior between participants and non-participants.

The results have strong policy implications that clearly exceed the scope of pricing experiments in transportation. Especially the selection bias with respect to the marginal utility of income is highly policy-relevant, and we are not aware of any earlier research focusing on this specific aspect of selection bias. The lack of research probably results from the fact that the marginal utility of income cannot be observed directly from standard surveys (unlike socio-economic characteristics, but also unlike "softer" characteristics like attitudes that can in most cases be derived from questionnaires). Biases in the marginal utility of income can have severe consequences when the marginal utility of income is used as an input to policy appraisals, for instance, in forecasts of the effects of pricing policies, welfare aggregations, or when used to derive monetary valuations of decision attributes.

Our results also have implications for the future design of field experiments. They provide a clear indication that self-selection biases in field experiments must be identified and corrected in order to ensure the generalizability of the results. Clearly, it is favorable if the avoidance or reduction of self-selection biases already happens *ex ante* (i.e. before the conduction of the experiment), rather than *ex-post* as in this study. In some cases it might be feasible to make participation obligatory, or to make it so attractive that (almost) all invited persons participate (e.g. by offering generous (behavior-independent) rewards for participation). If such *ex ante* measures are not feasible, our study shows that the conduction of complementary SP experiments is very insightful as underlying selection biases with respect to behavioral preferences can be well identified. If our approach is followed, well-designed SPs that are highly consistent with the RP setting are crucial for limiting hypothetical biases that drive SP and RP estimates apart.

More generally, our model provides evidence on how participation behavior and behavior in a reward experiment (both in the RP as well as in the SP setting) are related to each other. Here, we use the knowledge about this relation mainly to predict how non-participants would have behaved if they had participated in the field experiment. However, it would also be an interesting exercise to try inferring participation behavior in the field experiment from the participants' and non-participants' behavior in the SP.

Appendix A. Integrated behavioural model

In this appendix we show how the independent and the sequential estimation discussed in Section 3.4 derive mathematically from a single behavioral model. This model is defined by integrating the scheduling model and the participation model presented in sections 3.2 and 3.3 (see figure A.2).

In the integrated model the participation model is defined by the distributed latent variable X_n^* (equation (6)) and the probability mass function $P_2(I_n|X_n^*; \alpha)$ (equation (8)) of participation observation I_n . The scheduling model is defined by the alternative-specific utilities U_{jmn} (equation (2)), with the coefficient vectors β (for SP) and γ (for RP) binomially distributed with a probability mass function $P_m(c|X_n; \alpha)$, leading to the class specification defined by expression (5) for the probability mass function $P_1(y_{1n}, \dots, y_{Mn}|Z_n, X_n; \alpha, \beta_0, \beta_1, \gamma_0, \gamma_1)$ of the scheduling decisions $y_{1n}, \dots, y_{Mn}$ by individual n . In our application we use for P_m the same probability mass function as for P_2 , implying that the classes in the scheduling coincide with these of the participation model (one class for participants and one for non-participants).

For the further discussion of the estimation of the integrated model it is of use to note that the integrated model defined here is mathematically equivalent to an integrated choice and latent

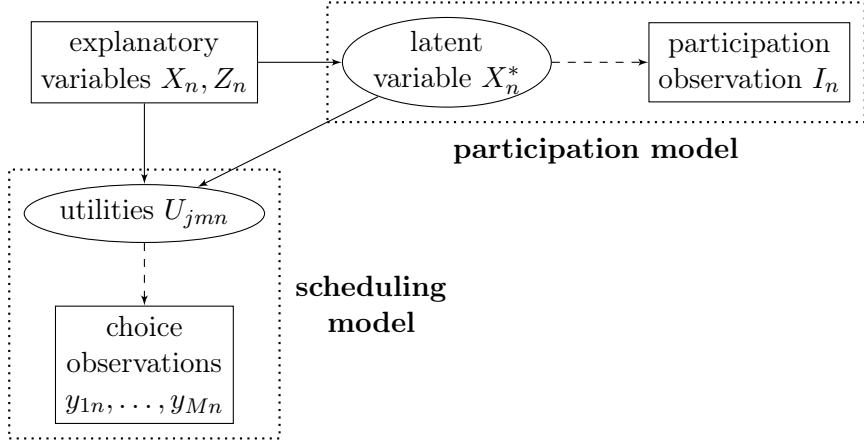


Figure A.2: Integrated scheduling and participation model

variable model (ICLV), where the participation model is the latent variable model, and the scheduling model is the choice model. In an ICLV model the dashed lines in figure A.2 would represent the measurement equations, and the solid lines the structural equations. Walker (2001) discusses a number of ways to estimate integrated choice and latent variable models, including the simultaneous and the sequential estimation.

A *simultaneous* estimation of the integrated scheduling and participation model implies simultaneously identifying values for the all coefficients α , β , γ and λ that maximise the joint log likelihood LL :

$$LL = \sum_n \ln \int \underbrace{\phi(\omega - \lambda X_n)}_{\text{pdf of } \omega} \cdot \underbrace{P_1(I_n | X_n^*; \alpha)}_{\text{participation}} \cdot \underbrace{P_2(y_{1n}, \dots, y_{Mn} | Z_n, X_n^*; \alpha, \beta_0, \beta_1, \gamma_0, \gamma_1)}_{\text{scheduling}} d\omega \quad (\text{A.1})$$

$$= \sum_n \ln \underbrace{\int \phi(\omega - \lambda X_n) \cdot P_1(I_n | X_n^*; \alpha) \sum_{c=0,1} P_1(c | X_n^*; \alpha) \prod_M P_s(y_{mn} | Z_{mn}; \beta_c, \gamma_c) d\omega}_{LL_n} \quad (\text{A.2})$$

As $P_1(I_n | X_n^*; \alpha)$ is a binary function of ω (see equation (8)), we can solve the integral (as a definite integral over the interval where $P_1(I_n | X_n^*; \alpha) = 1$) and express the joint log likelihood LL_n of the observations I_n (participation) and $y_{1n}, \dots, y_{Mn}$ (scheduling) by person n using the standard normal cumulative distribution function Φ :

$$LL_n = \begin{cases} \ln [\Phi(\alpha - \lambda X_n)] + \sum_m \ln P_s(y_{mn} | Z_{mn}; \beta_0, \gamma_0) & \text{if } I_n=0 \text{ (non-participant)} \\ \ln [1 - \Phi(\alpha - \lambda X_n)] + \sum_m \ln P_s(y_{mn} | Z_{mn}; \beta_1, \gamma_1) & \text{if } I_n=1 \text{ (participant)} \end{cases} \quad (\text{A.3})$$

The first term is the log likelihood of the participation model, whereas the second term is only a function of the coefficient vectors β_c and γ_c , with each β_c and γ_c estimated on the observations provided by the corresponding class of participants c . As a result it is possible to estimate the participation model and both scheduling models (participant vs non-participant) independently, hence our reference to these estimations of the scheduling models as the "independent" ones.

The simultaneous estimation is both efficient and consistent, as pointed out by Walker (2001) for the integrated choice and latent variable model. Our dataset however does not allow the (simultaneous) estimation of a coefficient vector γ_0 for non-participants due to a lack of RP observations by respondents with revealed non-participation behaviour.

In a *sequential* estimation we first estimate the participation model and next the scheduling model as discussed in Section 3.4.3. The estimation of the participation should deliver the same results for both approaches, as is easily observed by comparing equation (A.3) to equation (12).

As noted by Walker (2001) for the integrated choice and latent variable model, the sequential estimation (with integration of the choice model over the distribution of the latent variable) is consistent but not efficient. An advantage of this estimation methodology is however that we obtain two vectors of coefficients γ_c even if we have RP scheduling observations only for one class (conditional upon having participation observations for both classes): in the second step each respondent has a strictly positive probability of membership of both classes.

Finally note that the sequential estimation is mathematically equivalent to that of a (distributed) propensity score, where X_n^* is the propensity score.

References

- Abeler, J., Nosenzo, D., 2014. Self-selection into laboratory experiments: pro-social motives versus monetary incentives. *Experimental Economics* 18 (2), 195–214.
- Anderson, S. P., De Palma, A., Thisse, J. F., 1992. *Discrete choice theory of product differentiation*. MIT Press, Cambridge, MA.
- Ben-Akiva, M. E., Lerman, S. R., 1985. *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge, MA.
- Bierlaire, M., 2006. BIOGEME: a free package for the estimation of discrete choice models. In: *Swiss Transport Research Conference*.
- Bliemer, M. C., Rose, J. M., Hess, S., 2008. Approximation of bayesian efficiency in experimental choice designs. *Journal of Choice Modelling* 1 (1), 98–126.
- Cleave, B. L., Nikiforakis, N., Slonim, R., 2013. Is there selection bias in laboratory experiments? the case of social and risk preferences. *Experimental Economics* 16 (3), 372–382.
- Dohmen, T., Falk, A., 2010. You get what you pay for: Incentives and selection in the education system. *The Economic Journal* 120, 256–271.
- Fosgerau, M., Hjorth, K., Lyk-Jensen, S. V., 2010. Between-mode-differences in the value of travel time: Self-selection or strategic behaviour? *Transportation research part D: transport and environment* 15 (7), 370–381.
- Gautier, P. A., Klaauw, B. v. d., 2012. Selection in a field experiment with voluntary participation. *Journal of Applied Econometrics* 27 (1), 63–84.
- Harrison, G. W., Lau, M. I., Rutström, E. E., 2009. Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization* 70 (3), 498 – 507, field Experiments in Economics. URL <http://www.sciencedirect.com/science/article/pii/S0167268109000134>
- Harrison, G. W., List, J. A., 2004. Field experiments. *Journal of Economic Literature* 42 (4), pp. 1009–1055. URL <http://www.jstor.org/stable/3594915>
- Hensher, D. A., 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological* 44 (6), 735 – 752, methodological Advancements in Constructing Designs and Understanding Respondent Behaviour Related to Stated Preference Experiments. URL <http://www.sciencedirect.com/science/article/pii/S0191261509001477>
- Kessels, R., Goos, P., Vandebroek, M., 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 409–419.
- Knockaert, J., Ettema, D., Verhoef, E., Koster, P., Peer, S., Tseng, Y., 2012a. Spitsmijden Gouda-Zoetermeer [scientific report]. URL http://www.spitsmijden.nl/downloads/SM2D_scientific_report_final.pdf
- Knockaert, J., Tseng, Y.-Y., Verhoef, E. T., Rouwendal, J., 2012b. The Spitsmijden experiment: A reward to battle congestion. *Transport Policy* 24, 260–272. URL <http://dx.doi.org/10.1016/j.tranpol.2012.07.007>

- Levitt, S. D., List, J. A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives* 21 (2), pp. 153–174.
URL <http://www.jstor.org/stable/30033722>
- List, J. A., 2006. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy* 114 (1).
- List, J. A., 2007. Field experiments: a bridge between lab and naturally occurring data. *The BE Journal of Economic Analysis & Policy* 5 (2).
- Mabit, S. L., 2008. Sample selection and taste correlation in discrete choice transport modelling. Ph.D. thesis, Technical University of Denmark, Department of Transport.
- Peer, S., Knockaert, J., 2016. Design of a stated preference (SP) experiment for *Spitsmijden in de Trein*. Tech. rep., Vrije Universiteit Amsterdam.
- Peer, S., Knockaert, J., Koster, P., Verhoef, E., 2013. Overreporting vs. overreacting: Commuters' perceptions of travel times. Tech. rep., Tinbergen Institute Discussion Paper.
- Peer, S., Knockaert, J., Verhoef, E. T., 2016. Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. *Transportation Research Part B: Methodological* 83, 314 – 333.
URL <http://www.sciencedirect.com/science/article/pii/S0191261515002611>
- Peer, S., Verhoef, E., Knockaert, J., Koster, P., Tseng, Y.-Y., 2015. Long-run vs. short-run perspectives on consumer scheduling: Evidence from a revealed-preference experiment among peak-hour road commuters. *International Economic Review* 56 (1), 303–323.
URL <http://dx.doi.org/10.1111/iere.12103>
- Rose, J. M., Bliemer, M. C., Hensher, D. A., Collins, A. T., 2008. Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological* 42 (4), 395–406.
- Samenwerkingsverband Spitsmijden, 2009. Effecten van belonen in Spitsmijden in het OV: Hoe verleid je OV-reizigers?
URL <http://www.spitsmijden.nl/resultaten/rresultaten2c/Rapport%20Spitsmijden%20in%20het%20OV%20in%20Spitsmijden%202%20aug2009.pdf>
- Significance, VU University Amsterdam, John Bates Services, 2012. Values of time and reliability in passenger and freight transport in The netherlands. Tech. rep., Report for the Ministry of Infrastructure and the Environment.
- Train, K., 1986. *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. MIT Press, Cambridge, MA.
- Train, K., 2003. *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, UK.
- Vickrey, W. S., 1969. Congestion theory and transport investment. *American Economic Review* 59 (2), 251–60.
URL <http://www.jstor.org/stable/1823678>
- von Gaudecker, H.-M., van Soest, A., Wengström, E., 2012. Experts in experiments. *Journal of Risk and Uncertainty* 45 (2), 159–190.
URL <http://dx.doi.org/10.1007/s11166-012-9151-7>
- Walker, J. L., 2001. Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Ph.D. thesis, Massachusetts Institute of Technology.

Table A.9: Scheduling model

	unit	independent estimation				sequential estimation			
		non-participants		participants		non-participants		participants	
		val	std err	val	std err	val	std err	val	std err
<i>model coefficients*</i>									
β_R	SP Reward	€	0.0323	0.0081	0.138	0.010	0.0847	0.0118	0.0131
β_{TT}	SP Travel time	hour	-2.78	0.17	-3.00	0.19	-3.56	0.27	-2.45
β_{SDE}	SP Schedule delay early	hour	-2.04	0.10	-1.72	0.10	-2.45	0.16	-1.46
β_{SDL}	SP Schedule delay late	hour	-1.29	0.09	-1.40	0.09	-0.922	0.115	-1.97
β_{C3}	SP Crowded train		-0.0695	0.0516	-0.0447	0.0564	0.0216	0.0627	-0.155
β_{C4}	SP Very crowded train		-0.327	0.042	-0.443	0.052	-0.387	0.053	-0.424
β_{REL}	SP P(delay>10min)>10%		-0.236	0.039	-0.432	0.049	-0.314	0.053	-0.374
γ_R	RP Reward	€			0.288	0.029	0.206	0.058	0.047
γ_{TT}	RP Travel time	hour			-4.21	0.64	-3.85	1.19	-4.46
γ_{SDE}	RP Schedule delay early	hour			-1.44	0.11	-5.13	0.90	-1.07
γ_{SDL}	RP Schedule delay late	hour			-1.13	0.06	-0.734	0.096	-2.23
γ_{TR}	RP Number of transfers	#			-0.687	0.101	-0.591	0.178	-0.756
γ_C	RP Crowded train expected				-0.0759	0.0829	-0.0630	0.1112	-0.0645
γ_{CU}	RP missing occupancy expectation				0.0411	0.1357	-0.156	0.223	0.231
γ_{REL}	RP P(delay>10min)>5%				-0.113	0.169	0.235	0.310	-0.380
<i>valuations†</i>									
W_{TT}	SP Travel time	€/hour	86.2	21.2	21.8	1.9	42.0	7.2	25.1
W_{SDE}	SP Schedule delay early	€/hour	63.2	15.5	12.5	1.0	28.9	4.5	14.9
W_{SDL}	SP Schedule delay late	€/hour	39.9	10.1	10.1	0.9	10.9	1.8	20.2
W_{C3}	SP Crowded train	€	(2.15)	1.70	(0.325)	0.410	(-0.255)	0.749	1.58
W_{C4}	SP Very crowded train	€	10.1	2.8	3.22	0.41	4.57	0.85	4.34
W_{REL}	SP P(delay>10min)>10%	€	7.33	2.16	3.14	0.42	3.71	0.87	3.83
W_{TT}	RP Travel time	€/hour			14.6	2.5	18.7	7.0	15.5
W_{SDE}	RP Schedule delay early	€/hour			5.00	0.55	24.9	9.5	3.72
W_{SDL}	RP Schedule delay late	€/hour			3.92	0.39	3.57	0.92	7.74
W_{TR}	RP Number of transfers	€/h			2.39	0.43	2.87	1.16	2.63
W_C	RP Crowded train expected	€			(0.263)	0.293	(0.307)	0.567	(0.224)
W_{CU}	RP missing occupancy expectation	€			(-0.143)	0.471	(0.759)	1.106	(-0.800)
W_{REL}	RP P(delay>10min)>5%	€			(0.392)	0.592	(-1.14)	1.53	(1.32)
<i>class share in sample</i>			52.7%		47.3%		52.6%		47.4%

*Coefficients that are not significant at the 5% level are in italics.

†Valuations that are not significantly different from 0 at the 5% level are in italics. Valuations for which either the corresponding model coefficient or the reward coefficient is insignificant are put into brackets.

Table A.10: Scheduling model (with latent classes, independent estimation)

	unit	non-participants				participants			
		class 1		class 2		class 1		class 2	
		val	std err	val	std err	val	std err	val	std err
<i>model coefficients*</i>									
β_R	€	0.0620	0.0178	0.0484	0.0116	0.139	0.015	0.138	0.015
β_{TT}	hour	-1.85	0.49	-4.44	0.29	-2.85	0.31	-3.13	0.26
β_{SDE}	hour	-1.23	0.29	-3.19	0.18	-1.88	0.18	-1.61	0.15
β_{SDL}	hour	-3.36	0.37	-0.901	0.159	-1.70	0.16	-1.16	0.14
β_{C3}	SP Crowded train	-0.301	0.119	0.107	0.083	-0.0662	0.0976	-0.0346	0.0817
β_{C4}	SP Very crowded train	-0.519	0.117	-0.351	0.065	-0.467	0.083	-0.431	0.074
β_{REL}	SP P(delay>10min)>10%	-0.0798	0.0953	-0.381	0.064	-0.483	0.074	-0.394	0.072
γ_R	€					0.241	0.048	0.253	0.049
γ_{TT}	hour					-4.45	1.20	-4.14	1.01
γ_{SDE}	hour					-0.960	0.144	-4.12	0.49
γ_{SDL}	hour					-2.95	0.34	-0.835	0.102
γ_{TR}	RP Number of transfers					-0.706	0.197	-0.619	0.167
γ_C	RP Crowded train expected					-0.0752	0.1345	-0.0477	0.1064
γ_{CU}	RP missing occupancy expectation					0.263	0.274	-0.0515	0.1784
γ_{REL}	RP P(delay>10min)>5%					-0.532	0.327	0.0461	0.2310
<i>valuations †</i>									
W_{TT}	€/hour	29.8	10.9	91.6	22.8	20.5	2.8	22.6	2.9
W_{SDE}	€/hour	19.8	6.4	65.8	15.7	13.6	1.6	11.6	1.3
W_{SDL}	€/hour	54.2	16.0	18.6	4.5	12.2	1.6	8.39	1.21
W_{C3}	€	4.86	2.20	(-2.22)	1.88	(0.477)	0.703	(0.250)	0.588
W_{C4}	€	8.37	3.02	7.25	2.15	3.36	0.67	3.12	0.58
W_{REL}	€	(1.29)	1.53	7.87	2.49	3.48	0.63	2.85	0.63
W_{TT}	€/hour					18.4	5.6	16.4	4.8
W_{SDE}	€/hour					3.98	0.97	16.3	4.2
W_{SDL}	€/hour					12.2	3.3	3.31	0.58
W_{TR}	€/h					2.92	1.02	2.45	0.88
W_C	€					(0.311)	0.569	(0.189)	0.422
W_{CU}	€					(-1.09)	1.16	(0.204)	0.705
W_{REL}	€					(2.20)	1.38	(-0.182)	0.904
<i>class share in sample</i>		15.3%		37.4%		21.4%		25.9%	

*Coefficients that are not significant at the 5% level are in italics.

†Valuations that are not significantly different from 0 at the 5% level are in italics. Valuations for which either the corresponding model coefficient or the reward coefficient is insignificant are put into brackets.

Table A.11: Participation model (estimation log likelihood: -482.1)

Variable	Description	Definition	Sample average (%)		Coefficient estimation*	
			part	non-part	value	std err
x_1	low income	net household income < 2000€/month	11.6	8.8	0.547	0.173
x_2	high income	net household income > 3500€/month	31.8	28.7	0.0249	0.1308
x_3	income unknown	net household income unknown	27.5	30.1	0.0684	0.1297
x_4	low education		12.3	21.1	-0.102	0.148
x_5	high education		42.7	25.1	0.225	0.111
x_6	other education	"other" was chosen from the list of alternatives	1.3	1.6	0.0920	0.3346
x_7	single household		19.2	17.5	-0.0162	0.1469
x_8	household with children		43.2	46.0	0.194	0.114
x_9	military/agriculture	including fisheries and forestry	1.3	0.0	3.53	0.26
x_{10}	training/R&D	including specialists and education	53.9	34.5	-0.0107	0.1423
x_{11}	administration/services		29.5	54.6	-0.484	0.153
x_{12}	production	excluding fisheries, forestry and agriculture	0.7	0.6	-0.0645	0.7463
x_{13}	morning commute off-peak	usual scheduled departure before 6.30am or after 9am	24.6	14.9	0.276	0.176
x_{14}	evening commute off-peak	usual scheduled departure before 4pm or after 6.30pm	19.2	6.8	0.534	0.204
x_{15}	morning commute peak shoulder	usual scheduled departure 6.30am-7am or 8.30am-9am	26.8	27.5	0.114	0.118
x_{16}	evening commute peak shoulder	usual scheduled departure 4pm-4.30pm or 6pm-6.30pm	26.2	20.9	0.189	0.122
x_{17}	morning pdt off-peak	preferred departure before 6.30am or after 9am	36.7	23.5	-0.0158	0.1603
x_{18}	evening pdt off-peak	preferred departure before 4pm or after 6.30pm	28.0	8.2	0.812	0.178
x_{19}	morning pdt peak shoulder	preferred departure 6.30am-7am or 8.30am-9am	22.8	24.1	-0.0356	0.1298
x_{20}	evening pdt peak shoulder	preferred departure 4pm-4.30pm or 6pm-6.30pm	27.1	26.5	0.156	0.122
x_{21}	can start work later		85.5	71.3	0.0955	0.1451
x_{22}	can start work earlier		89.0	82.7	-0.147	0.157
x_{23}	can leave home earlier	only if unlimited in time	64.9	47.0	0.0968	0.1071
x_{24}	can leave home later	only if unlimited in time	73.6	42.0	0.455	0.120
x_{25}	can end work later		94.6	87.3	0.301	0.211
x_{26}	can arrive home later	only if unlimited in time	70.9	39.2	0.458	0.114
x_{27}	can arrive home earlier	only if unlimited in time	67.8	38.8	0.150	0.118
x_{28}	medium distance class	participation traject 26km-40km	24.8	22.9	0.168	0.133
x_{29}	long distance class	participation traject > 40km	53.9	44.0	0.177	0.121
x_{30}	medium travel frequency	usual train travel frequency 3-4 days/week	50.6	66.3	0.756	0.418
x_{31}	high travel frequency	usual train travel frequency 5 days/week	48.8	31.1	1.14	0.42
x_{32}	frequent morning peak commute	commute with train dept 6.30am-9am > 3 days/week	66.4	71.1	-0.121	0.152
x_{33}	frequent evening peak commute	commute with train dept 4pm-6.30pm > 3 days/week	65.5	70.3	0.0252	0.1449
x_{34}	train travel in 1st class		14.3	5.8	0.517	0.168
x_{35}	young	born after 1981	25.7	14.3	0.303	0.139
x_{36}	old	born before 1962	17.9	22.5	-0.127	0.125
	modelling constant α				2.37	0.51

*Coefficients that are not significant at the 5% level are in italics.