Commuters' Preferences for Fast and Reliable Travel

Paul Koster
Hans Koster

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute.

Commuters' preferences for Fast and Reliable travel: A Semi-parametric Estimation Approach[☆]

Paul R. Koster[a,b,*], Hans R.A. Koster[a,b]

[a]*Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b]*Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands*

Abstract

We employ a semi-parametric estimation approach to analyse observed and unobserved heterogeneity in the value of savings in travel time and schedule delay. Our econometric approach allows for the estimation of unobserved and observed heterogeneity in preferences in a flexible way, meaning that we do not put any structure on how individual characteristics (such as income and age) relate to the value of savings in travel time and schedule delay. Using data from a stated choice experiment, we illustrate the estimation approach and find that there is substantial heterogeneity in the value of savings in travel time and schedule delay. For our data, we find that unobserved heterogeneity is more important than heterogeneity related to individual characteristics.

Keywords: local maximum likelihood, latent class, semiparametric Logit, normalised entropy criterion, value of time, value of reliability

1. Introduction

Preference heterogeneity is of key importance when evaluating the effect of transport economic policies. For example, the success of off-peak rewarding policies is strongly related to the potential of travelling outside the peak for a group of travellers. Some studies have shown that the distributional effects of congestion pricing policies strongly depend on the heterogeneity in individuals preferences (Arnott et al., 1992, 1994; Van den Berg and Verhoef, 2011a,b; Börjesson and Kristoffersson, 2014). Furthermore, heterogeneity is important when evaluating the benefits of private provision of highways, because operators' profits depend on the marginal willingness to pay (WTP) for reductions in travel time and travel time variability of travellers (Mills, 1981; Winston and Yan, 2011; Tan and Yang, 2012). Heterogeneity is also important when studying the effects of congestion pricing in the presence of alternative privately or publicly operated transport modes (Huang, 2000; Van den Berg and Verhoef, 2013). In all cases, using average WTP values may lead to biased welfare estimates and imprecise policy recommendations.

In this paper, we use a semi-parametric econometric framework to analyse heterogeneity related to observed individual characteristics, as well as unobserved heterogeneity, using panel data from a stated choice experiment. We estimate a semi-parametric panel latent class model that allows for a flexible specification of both unobserved and observed heterogeneity. Our method is applied to the estimation of value of travel time savings (VOT) and the values of savings in schedule delay early (VSDE) and late (VSDL). We obtain distributions of preferences, by employing local-likelihood estimation methods introduced by Tibshirani and Hastie (1987) and Fan et al. (1995, 1998), and by assuming that individuals who are similar in terms of characteristics (e.g. income, age) will have more similar values of savings in travel time and schedule delay.

Our semi-parametric panel latent class model has several features. First, our method does not make any assumptions on the shape of the distribution of preferences that is estimated. Second, our econometric procedure allows for the estimation of both unobserved and observed heterogeneity. It extends the local logit model used by Fosgerau (2007), Börjesson et al. (2012), Hjorth and Fosgerau (2012) and the local multinomial and local cross-sectional latent class model of Dekker et al. (2014) to allow for unobserved heterogeneity and panel data. Third, kernel smoothing techniques reduce the well-known curse of dimensionality, meaning that the inclusion of a large number of covariates will not necessarily lead to imprecise estimates (but it may increase the bias compared to a fully saturated model). The use of kernel smoothing allows for the incorporation of all interactions between individual characteristics and preferences without the need to specify these interactions explicitly. Hence, our approach is particularly useful for datasets with a large number of individual characteristics. Fourth, our estimation method takes into account the repeated nature of the choices and therefore correctly deals with the panel dimension of the data. It results in a unique semi-parametric distribution of preferences for each unique combination of individual characteristics in our dataset.

The data to estimate commuters' VOT and value of arriving at the preferred arrival time at work are obtained from a stated choice survey held among participants of a real-world

rewarding experiment to combat congestion. We estimate the VOT, VSDE and VSDL using the classical linear specification of schedule delay (Vickrey, 1969).

We find that the average values are somewhat higher than in the previous literature, most likely because respondents participate in a reward experiment and people are arguably more sensitive to costs than to rewards. We also show that there is substantial heterogeneity in the estimated VOT, VSDE and VSDL. The VOT, for example, ranges from about €20 to €80. The results suggest that unobserved heterogeneity seems to be more important than heterogeneity related to observable individual characteristics. Nevertheless, we find statistically significant impacts of income, education and household composition on the VOT, the VSDE and the VSDL.

The paper continues as follows. Section II discusses the design and set-up of the stated choice experiment and specification of the utility function, which is followed by the econometric framework in Section III. Section IV presents the main results and Section V concludes and discusses the limitations and directions for further research.

2. Stated choice experiment and utility specification

2.1. Stated choice experiment

We use data from a stated choice experiment among Dutch morning-commuters participating in a peak-avoidance reward experiment. In order to reduce congestion, these commuters received a reward if they did not travel between cameras A and B during the morning peak (6:30-9:30). These cameras were placed on a congested highway. An example of a choice scenario is given in Figure 1. Respondents were asked to choose between two departure times. To account for travel time variability, each departure time has two possible travel times with a corresponding probability, arrival time at work and reward. Hence, commuters made a trade-off between earlier or later arrivals, shorter or longer travel times, and receiving a monetary reward. The preferred arrival time of the traveller was given as a reminder. It was based on previous questions in the questionnaire and was defined as the time that a traveller would like to arrive at work if there is no possibility to receive a reward, and there is no travel time delay. The lay-out was pre-tested in a focus-group, and internet pre-tests were carried out to ensure that respondents understand the questions well. Several considerations regarding the screen lay-out were made. First, we made explicit what is important for travellers: departure time, probabilities, travel times, arrival times and rewards. Second, travel times of the separate parts of the trip were given. This is to show the respondents why they (do not) receive the reward and how the travel time is built up from the different components. Fourth, we used bold values for the variables that would have been (potentially) important in the trade-offs of the respondents.

The attribute values for travel times were pivoted around the average travel time of the respondent to enhance realism (Hensher, 2001). Arrivals at Camera A were spread over the whole peak hour to have sufficient variation in arrival times. Several other constraints were put in the design attributes to enhance realism. These are described in detail in Knockaert et al. (2012). The experimental design was pre-tested using extensive simulation, so that we are sure that the design is able to recover a broad range of parameters.

	Alternative A		Alternative B	
Departure time from home	6:05		6:50	
Probability	*80%*	*20%*	*90%*	*10%*
Total travel time	30 min	40 min	20 min	35 min
Travel time from home to camera A	15 min	15 min	10 min	10 min
Travel time from camera A to camera B	5 min	10 min	5 min	15 min
Travel time from camera B to work	10 min	15 min	5 min	10 min
Arrival time at work	6:35	6:45	7:10	7:25
Reward	4 euro	4 euro	0 euro	0 euro

Figure 1: Example of a choice scenario

We exclude respondents who chose randomly or for whom no individual characteristics are available (29% of the data). For the remaining respondents we have information on gross monthly household income, level of education, gender, age, household composition (single, children, etc.) and their residential and working location. The summary statistics of the individual characteristics that are used in the estimation are presented in Table 1. Compared to the Dutch average, we have a large share of high incomes and highly educated travellers. In our sample, about 84 percent has a bachelor degree or higher. We also have a relatively high share of males in our sample (76 percent), whereas single persons are under-represented (17 percent).

2.2. Utility specification

Our empirical framework includes two latent variables: utility and class membership. We assume that individual n, conditional on class membership c, chooses alternative i from a set of J alternatives in choice scenario t, with corresponding deterministic utility (V_{ncti}). V_{ncti} is a function of the preferences β_{nc} and the independent variables. These independent variables are given by the expected reward $E[R]_{nti}$, the expected travel time $E[T]_{nti}$, expected schedule delay early $E[SDE]_{nti}$ and expected schedule delay late $E[SDL]_{nti}$. The expected values of the variables are given by the probability weighted averages of M mass points, so $E[X]_{nti} = \frac{1}{M}\sum_{m=1}^{M} p_{ntim} X_{ntim}$, for attribute $X \in \{R, T, SDE, SDL\}$. As Figure 1 shows, the choice experiment has two possible travel times and therefore $M = 2$. Schedule delay is defined as the deviation of an arrival time from the individuals' preferred arrival time PAT_n, where arrivals different from PAT_n result in a dis-utility. Following Noland and Small (1995), we assume a linear-additive specification of the systematic utility. The random utility U_{ncti} of alternative i is then given by:

$$U_{ncti} = \beta_n^{ER} E[R]_{nti} + \beta_{nc}^{ET} E[T]_{nti} + \beta_{nc}^{ESDE} E[SDE]_{nti} + \beta_{nc}^{ESDL} E[SDL]_{nti} + \epsilon_{ncti}, \qquad (1)$$

Table 1: Descriptives of the sample

	Mean
Income <€2500	0.066
Income €2500-€3500	0.340
Income €3500-€5000	0.401
Income >€5000	0.193
Education Primary or Secondary	0.043
Education Vocational	0.117
Education Bachelor Degree or higher	0.840
Female	0.237
Single	0.165
No Children	0.430
Children <5 years	0.212
Children 5-12 years	0.263
Children at Secondary School (>12 years)	0.095
Age<25	0.014
Age 25-50	0.761
Age>50	0.224
Number of Individuals	487

where the scale of utility is normalised to one.[1] Furthermore, we restrict β_n^{ER} to be equal over all classes to enhance empirical identification. We are interested in the VOT, the VSDE and the VSDL of individual n belonging (with probability) to class c. These are given by the following ratios:

$$VOT_{nc} = -\frac{\beta_{nc}^{ET}}{\beta_n^{ER}}, \tag{2}$$

$$VSDE_{nc} = -\frac{\beta_{nc}^{ESDE}}{\beta_n^{ER}}, \tag{3}$$

$$VSDL_{nc} = -\frac{\beta_{nc}^{ESDL}}{\beta_n^{ER}}, \tag{4}$$

The next section describes the econometric approach we use to estimate these values.

3. Econometric setup

3.1. Local estimation

This section introduces the 'local' estimation approach. In standard latent class models one estimates the semi-parametric distribution of preferences by estimating class membership

---

[1]We allow for scale heterogeneity for observed characteristics because the marginal utilities depend on n-specific covariates.

probabilities jointly with the marginal (dis)utilities as specified in Equation (1), where one takes into account that a respondent makes repeated choices. Observed heterogeneity can then be included by explaining the class probabilities by covariates, or by including covariates in the utility function Equation (1).[2] With a low number of covariates these interactions can easily be included in a standard latent class model. However, when the number of covariates increases, the number of potential parameters to be estimated rapidly increases as also higher order interaction effects are included. The contribution of our econometric approach is that we allow for maximum flexibility by employing local kernel estimation in order to avoid a priori assumptions on which interactions are important.

Tibshirani and Hastie (1987) and Fan et al. (1995, 1998) introduce local likelihood estimation. The term 'local' implies that each local point (in our case each individual) is treated as a reference point. Conditional on the local point, a vector of kernel weights is determined, reflecting the multidimensional distance between the local point and the other points in the dataset. We let the kernel weights depend on the $N \times Q$ matrix of individual characteristics $\mathbf{z}$, where N is the number of individuals and Q is the number of characteristics that are included in the model. The kernel weights account for the fact that more similar people in observable characteristics have more similar preferences. Likewise, individuals with exactly the same characteristics have the same weights in the likelihood function and therefore the estimated semi-parametric distribution of preferences is the same for these individuals.

We have a balanced panel with N individuals making T choices each, assuming that there are C latent classes. Let $\beta_{nc} \equiv \beta_c(z_n)$ be the vector of marginal utilities of class c, which are a function of the $1 \times Q$ vector of personal characteristic of individual n: z_n. Let y_{ntj} be the choice of individual n for choice occasion t. The probability that individual n chooses alternative i from the set of J alternatives conditional on membership of class c is then defined as the logit function:

$$\breve{P}_{nti|c} = \frac{\exp(X_{nti}\beta_c(z_n))}{\sum_{j=1}^{J} \exp(X_{ntj}\beta_c(z_n))}. \tag{5}$$

Because individuals make repeated choices we will consider the probability of a choice sequence rather than analyse the choices in isolation. Let y_{nti} be the choice of individual n for choice occasion t, where $y_{nti} = 1$, when individual n chooses alternative i, and $y_{nti} = 0$ otherwise. The joint probability of a sequence of choices, conditional on the class assignment, is given by:

$$P_{n|c} = \prod_{i=1}^{J} \prod_{t=1}^{T} \breve{P}_{nti|c}^{y_{nti}} \tag{6}$$

Because $\breve{P}_{nti|c}^{y_{nti}} = 1$ for $y_{nti} = 0$, and $\breve{P}_{nti|c}^{y_{nti}} = P_{nti|c}$ for $y_{nti} = 1$, $P_{n|c}$ is the multiplication of the probabilities of the chosen alternatives (conditional on the class assignment). The

---

[2]This may be clarified by an example. Consider a covariate gender that we include in the utility function Equation (1). This means that the estimated parameters of Equation (1) will be different for males and females. If we also interact gender with the latent variable class membership we obtain gender-specific *distributions* of the VOT, VSDE and VSDL.

probability that individual n belongs to a latent class c is given by $p_{nc} \equiv p_c(z_n)$, meaning that class membership probabilities may depend on the individuals personal characteristics. We ensure $\sum_{c=1}^{C} p_{nc} = 1$, so that the model is identified. The likelihood that individual n makes the observed sequence of choices is given by the weighted average of the probabilities of Equation (6), with weights given by the class membership probabilities:

$$LL_n = \sum_{c=1}^{C} p_{nc} P_{n|c}. \tag{7}$$

The class probabilities p_{nc} and the preference parameters β_{nc} will be estimated using local likelihood estimation. This means that the parameter estimates for a reference individual $\bar{n}$ are obtained by maximising the following (weighted) likelihood function:

$$WLL_{\bar{n}} = \max_{\beta_{\bar{n}}, p_{\bar{n}}} \sum_{n=1}^{N} w_{\bar{n}}[\mathbf{z}; \lambda] \log(LL_n), \tag{8}$$

where $p_{\bar{n}}$ is the $C \times 1$ vector with the class probabilities and $\beta_{\bar{n}}$ the $C \times 1$ vector with preference parameters. The local log-likelihood is the log of Equation (7), multiplied by an $N \times 1$ vector of kernel weights, $w_{\bar{n}}[\mathbf{z}; \lambda]$. These kernel weights depend on the socio-economic distance of an individual $\bar{n}$ to all other individuals in the dataset, where this distance is governed by the vector of bandwidths λ. More details on the functional form of the kernel weights will be given in the next subsection. The estimation procedure results in the $C \times 1$ vectors $p_{\bar{n}}$ and $\beta_{\bar{n}}$, and is repeated for every unique individual in the dataset. The individual local estimates are substituted in Equation (7) to obtain the estimated likelihood $\hat{LL}_{\bar{n}}$ for each individual. Summing over all individuals gives the estimated global log-likelihood of the model:

$$\hat{GLL}(C) = \sum_{\bar{n}=1}^{N} \log(\hat{LL}_{\bar{n}}). \tag{9}$$

3.2. Kernel functions

The kernel weights in Equation (8) reflect the multidimensional 'distance' between individuals based on their characteristics. This distance is calculated using a multidimensional kernel function. When differences in the characteristics between individual $\bar{n}$ and another individual n become smaller, individual n has a higher kernel weight in the local likelihood estimation of $\bar{n}$ (and vice versa). The vector of bandwidths $\lambda = \lambda_1, ..., \lambda_Q$ determines the degree of smoothing.

We include Q variables in the kernel function. Individual characteristic $q = 1...Q$ has a corresponding kernel function $K_q(.)$, and bandwidth or smoothing parameter λ_q. A specification of the kernel weights is then given by:

$$w_{\bar{n}}[\mathbf{z}; \lambda] = \prod_{q=1}^{Q} K_q(z_{\bar{n}q} - z_q; \lambda_q), \tag{10}$$

where we use z_q to denote the $N \times 1$ vector with values for characteristic q for each individual, and $z_{\bar{n}q} - z_q$ as the difference in characteristics between individual $\bar{n}$ and all the other individuals. As is common practice, we use the restriction $\lambda_q = \bar{\lambda}$, implying that the bandwidth is the same for all variables included in the kernel function. For a fixed number of individuals N and number of choices T, and given bandwidth $\bar{\lambda}$, adding individual characteristics to the kernel (increasing Q) leads to lower kernel weights in Equation (10). When individuals share the same characteristic q, $z_{\bar{n}q} - z_q = 0$, resulting in a kernel weight $K_q(.)$ that is equal to 1.

For our analysis we only use categorical individual characteristics. Racine and Li (2004) show that for these variables one needs a kernel function that has the possibility to be an indicator function as well as a constant. Following Racine and Li (2004) and Hall et al. (2007), the kernel function is assumed to be of the form:

$$K_q(z_{\bar{n}q} - z_q; \bar{\lambda}) = \begin{cases} 1, & \text{if } z_{\bar{n}q} = z_q \\ \bar{\lambda}^{|z_{\bar{n}q} - z_q|} & \text{if } z_{\bar{n}q} \neq z_q. \end{cases} \tag{11}$$

The kernel weights are equal to one when $\bar{\lambda} = 1$. Since we will only include categorical variables we have $0 \leq \bar{\lambda} \leq 1$. For strong smoothing, the kernel weights are uniform ($\bar{\lambda} = 1$), and the model collapses to the standard panel latent class model. For weak smoothing ($\bar{\lambda} = 0$), the model is a saturated model resulting in a separate estimation of a panel latent class model for each unique combination of individual characteristics.

The appealing feature of the estimation set-up is that the standard logit model ($C = 1, \bar{\lambda} = 1$), the local logit model ($C = 1, 0 \leq \bar{\lambda} < 1$), and the panel latent class model ($C > 1, \bar{\lambda} = 1$) are special cases of our model. The standard logit model results in identical preference parameters for all individuals, whereas the local logit model estimates unique preferences for each unique combination of individual characteristics. Both these models may lead to biased estimates because unobserved heterogeneity is ignored. The local panel latent class model estimates unique preference distributions for each combination of individual characteristics and therefore allows for a flexible modelling of both observed and unobserved heterogeneity. The next section discusses the bandwidth selection and the selection of the number of classes.

3.3. Selection of bandwidths and the number of latent classes

There are two remaining issues that deserve attention: the selection of the number of classes and the selection of the bandwidth parameter. The first issue is to determine the number of classes. Here we use the normalised entropy criterion (NEC), which was introduced by Celeux and Soromenho (1996) and further developed by Biernacki et al. (1999). The advantage of this criterion is that it only relies on the ex-post processing of the estimated parameters and does not need the calculation of the number of parameters as for example the Bayesian Information Criterion does.[3] The NEC criterion is derived by observing that the

---

[3] Dekker et al. (2014) show how to derive the number of parameters for the local multinomial logit model, but determining the number of parameters in a local mixture model is challenging and is left as a topic for future research.

global log-likelihood (Equation (9)) can be separated in a classification log-likelihood term $\hat{CL}(C)$, and a (Shannon) entropy term $\hat{E}(C)$ that measures the overlap of the individual posterior latent class probabilities $\hat{v}_{nc}$ (which are introduced below, see Greene and Hensher (2003)). Hathaway (1986) shows that the log-likelihood can be written as:

$$G\hat{L}L(C) = \hat{CL}(C) + \hat{E}(C). \tag{12}$$

The classification log-likelihood and the entropy term are given by:

$$\hat{CL}(C) = \sum_{c=1}^{C} \sum_{n=1}^{N} \hat{v}_{nc} \log\left(\hat{p}_{nc}\hat{P}_{n|c}\right), \tag{13}$$

where $\hat{P}_{n|c}$ is Equation 6 evaluated at the estimated parameters, and

$$\hat{E}(C) = -\sum_{c=1}^{C} \sum_{n=1}^{N} \hat{v}_{nc}\log\hat{v}_{nc}. \tag{14}$$

The individual posterior class membership probabilities are given by:

$$\hat{v}_{nc} = \frac{\hat{p}_{nc}\hat{P}_{n|c}}{\sum_{d=1}^{C} \hat{p}_{nd}\hat{P}_{n|d}}. \tag{15}$$

The hats indicate that the functions are evaluated at the locally estimated parameters. A proof that (12) is indeed given by the sum of (13) and (14) is given in Appendix A. Celeux and Soromenho (1996) and Biernacki et al. (1999) then propose the NEC-criterion for the selection of the number of classes. This criterion is given by:

$$NEC(C) = \frac{\hat{E}(C)}{G\hat{L}L(C) - G\hat{L}L(1)}. \tag{16}$$

where $\hat{LL}(1)$ is the estimated log-likelihood for the local logit model which has one class. The NEC-criterion needs to be minimized in order to determine the optimal number of classes. When C increases, the overlap between the posterior class membership probabilities $\hat{v}_{nc}$ will increase and will (all else equal) increase the (Shannon) entropy term resulting in an increase in the NEC. The denominator of (16) is always positive because models with a higher number of classes will always have a better fit. Increasing the number of classes will therefore increase the denominator of (16) resulting in a decrease in the NEC, implying that the NEC criterion balances model fit with the overlap of the class probabilities. An additional check is needed to test the latent class model against the logit model ($C = 1$). Biernacki et al. (1999) show that when $NEC(C^*) < 1$ the latent class model is preferred, whereas otherwise the logit model is preferred.

For the selection of the bandwidth parameter we have experimented with several strategies, such as the use of a leave-one-individual-out cross-validation criterion, or a rule-of-thumb bandwidth using Silverman's rule. These strategies either resulted in severe over

or undersmoothing. Therefore we decided to use eye-balling to determine the bandwidth, since statistical strategies did not result in plausible outcomes. This means that when too extreme outliers were present (values > €100), the bandwidth was considered to be too low. The development of a statistical bandwidth selection criterion for local mixture models is beyond the scope of this paper and is left for future research.

4. Estimation results

4.1. Baseline results

This section discusses the estimation results. We first report the average VOT, the average VSDE and the average VSDL for the sample. Then, we investigate the distributions of the estimated parameters and examine the correlation of the estimated parameters with individual characteristics. Table 2 reports the results for the average estimates. We estimate the standard errors using a bootstrap procedure. We re-sample individuals rather than observations, because observations are not independent, while the assumption that individuals are randomly drawn from the population seems reasonable. We calculate for each of the 500 bootstrap runs randomly-drawn bootstrap weights for each individual.[4] We then estimate the standard error by taking the standard deviation of the sampled distribution for the parameter of interest (see Guan, 2003).

Table 2: The estimated average values of time and schedule delays

	(1) Binary Logit	(2) Latent Class	(3) Local Logit	(4) Local Latent Class
VOT	35.050	30.978	38.186	34.668
	(4.158)	(7.561)	(10.622)	(13.26)
VSDE	23.218	22.837	24.516	23.768
	(2.211)	(13.506)	(5.217)	(14.228)
VSDL	17.162	15.922	17.974	16.549
	(1.621)	(6.36)	(3.976)	(6.998)
Latent classes	1	2	1	2
Bandwidth $\bar{\lambda}$	1	1	0.6	0.6
LL(C,$\bar{\lambda}$)	-2,719	-2,590	-2,689	-2,524

Note: The bootstrapped standard errors (500 replications) are between parentheses.

---

[4]Horowitz (2001) argues that only when the asymptotic bias is essentially zero, bootstrapped standard errors are correct, which is the case when the bandwidth approaches zero. He therefore suggests to use a lower smoothing parameter to estimate the standard errors than when estimating the parameters of interest, as to minimise differences between true and coverage probabilities of bootstrapped confidence intervals. Furthermore, for low values of the bandwidth, confidence intervals would approach infinity due to outliers. We therefore interpret our confidence bands as indicative, but are aware of the fact that they might be too small.

Column (1) of Table 2 reports the result for the standard binary logit specification without heterogeneity. The VOT is estimated at €35 per hour and the VSDE and VSDL are estimated at €23 and €17 per hour respectively. These values are higher than found in the literature (see for example: Brownstone and Small (2005), Abrantes and Wardman (2011) and Li et al. (2010)). There may be at least two reasons for this. First, on average we have a high share of high-income travellers in our sample and since these have a lower marginal utility of income they are less sensitive to rewards than the average Dutch commuter. Second, it is very likely that due to loss aversion travellers are less sensitive to rewarding incentives than to the payment of a congestion toll. These higher values do not necessarily imply that the estimates are subject to hypothetical bias. Recent estimates by Peer et al. (2014) using data from the same experiment show that the estimated values are close to estimates based on revealed preference data.

The pattern VSDE>VSDL is the same for all models and is remarkable, since usually VSDL>VSDE is found in the literature (Lam and Small, 2001; Brownstone and Small, 2005; Li et al., 2010). This may be caused by the fact that we have a relative high share of individuals with an early preferred arrival time in our dataset, because we only analyse the preferences of individuals participating in the rewarding experiment. Another finding is that the estimated average VOT is higher than the estimated average VSDL. An explanation may be that this is due to a selection effect of participants who have lower values of schedule delays. Our estimated average VSDE is about 65-75 percent of the estimated VOT, which is slightly higher than in the literature (Li et al., 2010).

Column (2) of Table 2 shows the result for the latent class model with two latent classes. When we allow for unobserved heterogeneity, the estimated average VOT is about €31, which is slightly lower compared to the previous specification. The average VSDE and VSDL are very similar to the binary logit model. Although the average estimates are very similar, the standard error of the estimates is much higher, which suggests that there is substantial heterogeneity in the estimated parameters. The model seems also be able to better explain choices, as the log- likelihood increases substantially compared to the binary logit model.

In column (3) of Table 2, we only allow for heterogeneity related to observable individual characteristics. Based on eye-balling, we set the bandwidth $\bar{\lambda} = 0.6$ (for lower values, the standard errors become unrealistically large due to outliers). The average values are similar to the previous specifications. There is a likelihood improvement of about 30 points compared to the binary logit model, but the model has a lower fit than the latent class model, as the likelihood is higher for the latter specification. In other words, unobserved heterogeneity seems to be more important than observed heterogeneity. Another interesting observation is that, compared to the latent class model, only the standard error related to the VOT is higher, which might suggest that observed heterogeneity may explain heterogeneity in the VOT, but not so much in the VSDE and VSDL.

A preferred specification allows for the estimation of observed and unobserved heterogeneity. Column (4) of Table 2 then reports the estimation results of the local latent class model. The optimal number of latent classes is 2 and $NEC(2) < 1$ implying that the local latent class model is preferred over the local logit model. Table 2 shows that the estimated

average values of the local latent class model are very similar to the binary logit model in column (1). The standard errors are higher, possibly due to the fact that the model allows for different sources of heterogeneity, or because this more general model has much more degrees of freedom.

4.2. Investigating heterogeneity

Although the average estimates are interesting, we are particularly interested in preference heterogeneity. In what follows, we focus on the results of the preferred specification, which is the Local Latent Class model. Figure 2 reports the estimated semi-parametric cumulative distributions for the values of time, schedule delay early and late (solid line). Probability density functions using histograms are provided in Appendix C. We construct the confidence bands (dashed lines) by simulating the data for each bootstrap replication based on the class probabilities and order for each bootstrap replication the values from small to large and then calculate the standard deviation for each percentile.[5] We note that our confidence bands may be too small because of asymptotic bias (Horowitz, 2001), so they are best interpreted as *local* confidence bands (see Duranton and Overman, 2005). This implies that we can make local statements, i.e. at a certain point in the distribution. However, local confidence bands cannot be used to make statements about the global distribution, because global confidence bands tend to be wider. The upper panel of Figure 2 shows that about 80 percent of the observations have a VOT that is below €40 per hour, which seems reasonable. We do not observe individuals that do not attach any value to time savings. Looking at the upper 20 percent of the distribution reveals that there are quite some individuals that have a relatively high VOT.

The middle panel of Figure 2 shows that the VSDE follows a somewhat irregular distribution. About 10-15 percent of the observations have a VSDE which is close to zero. This is not too surprising as a relatively large share of the participants is highly-educated, which often imply that their jobs are more flexible than blue-collar jobs. Hence, for them it is no problem to be early at work. About 75 percent of the observations has a VSDE that is above €20. The cumulative VSDE distribution also reveals that there are no travellers in our sample that have values between €12.5 and €20 per hour. The estimated probability density function is bi-modal, which is confirmed by the histograms in Appendix C.

The lower panel of Figure 2 shows the distribution of the VSDL. In contrast to the distribution of VSDE, nearly all observations are positive and significantly different from zero. About 75 percent of the observations is in between €15 and €25, which seems plausible. The VSDL histogram in Appendix C shows that the estimated VSDL distribution also has a bi-modal shape.

In all cases, there is a strong positive correlation between mean values (at the individual

---

[5]More specifically, to estimate the cumulative distribution for the VOT, VSDE and VSDL we generate for each individual 100 values and assign the different values according to the class probabilities. We order the values from small to large to obtain the cumulative distribution. We then estimate the cumulative distributions for each replications and calculate the standard deviation for each centile to obtain the 95 percent confidence interval

Figure 2: Cumulative sample distributions of estimated VOT,VSDE and VSDL
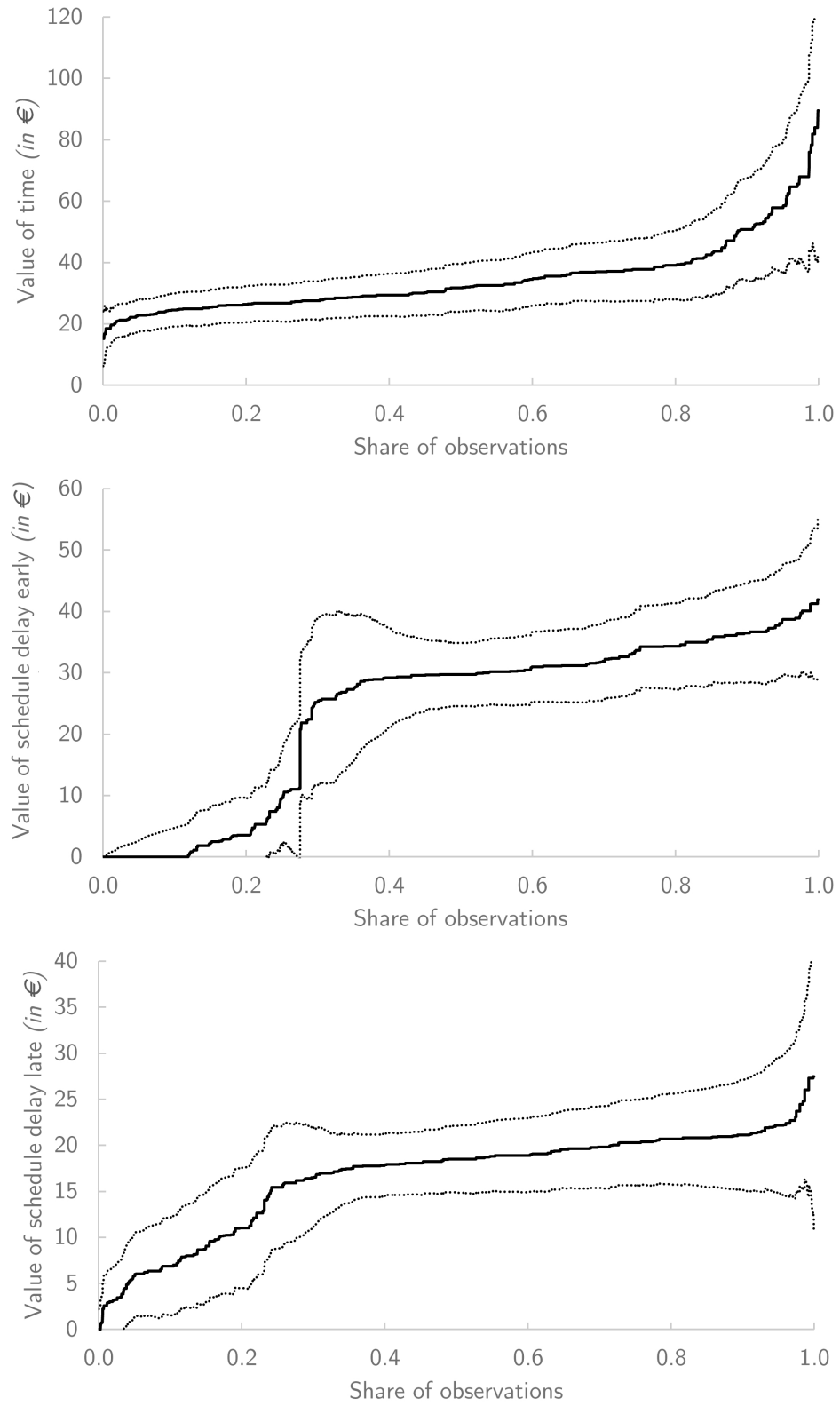
level) for the local latent class model and the outcomes from the local logit model. The correlation for the VOT, VSDE and VSDL is respectively 0.920, 0.948 and 0.956.

4.3. Explaining heterogeneity

Similar to Bajari and Kahn (2005), we investigate how individual demographic characteristics relate to the estimated VOT, VSDE and VSDL. We regress the estimated parameters on individual characteristics.[6] To allow for as much flexibility as possible, we use a non-parametric regression of the form:

$$VOT_{nc} = f(z_n) + \xi_{nc}, \tag{17}$$

where $f(\cdot)$ is a non-parametric function of individual characteristics z_n, and ξ_{nc} is an error term, which is assumed to be uncorrelated to z_n. Because we do not observe VOT_{nc} directly, we use the estimated values $\hat{VOT}_{nc}$ (see Bajari and Kahn, 2005). One may interpret ξ_{nc} as an individual-class-specific taste shock. We do not make any functional form assumptions on ξ_{nc}, in contrast to for example Berry et al. (1995) and Petrin (2002). We estimate $f(\cdot)$ by local linear regression (LLR) techniques and calculate the kernel weights using the kernel function in Equation (10). The kernel weights are then multiplied with the estimated class probabilities. We again assume a bandwidth of 0.6. The standard errors are obtained by bootstrapping the whole estimation procedure: for each of the 500 bootstrap runs, we resample individuals, then estimate the VOT, VSDE and VSDL for each individual/latent-class combination using Equation (9) and then regress the estimated values on individual characteristics using the non-parametric regression of (17). For presentation purposes we only present the mean effect and the related bootstrapped standard error of the estimates (McMillen and Redfearn, 2010).[7] The results are reported in Table 3. We first report simple weighted least squares (WLS) estimates, where it is assumed that $f(\cdot) = \zeta z_n$ and the weights are given by the corresponding class probabilities. Then, we present the results using local linear regression techniques.

In the WLS specifications (columns (1)-(3)), we find substantial income effects: switching from an income lower than €2,500, to the highest income class increases the VOT with more than €10. The value schedule delay early and delay are also positively related to income. Conditional on the income effect, we also find an effect of education. For highly educated people, the VOT is more than €4 higher compared to low educated people. The effect of education on the VSDE is much less pronounced and statistically insignificant. We find that having younger children leads to a higher VOT, but surprisingly not to higher scheduling costs. Travellers that are single seem to have fewer scheduling constraints. The negative coefficients for the VSDE and VSDL for people that have children at primary school

---

[6]Because observed heterogeneity is related to many individual characteristics, it is not insightful to generate two-dimensional plots. For example, the unconditional effect of age on VOT, VSDE and VSDL may also be caused by a generally higher income of older people

[7]We also estimated similar models where we take the logarithm of the values, but the results are qualitatively similar

Table 3: VOT, VSDE and VSDL and individual characteristics

	WLS			LLR		
	(1)	(2)	(3)	(4)	(5)	(6)
	VOT	VSDE	VSDL	VOT	VSDE	VSDL
Income 2500-3500	2,925*	2,217***	1,572***	2,551	3,626*	1,689
	(1,562)	(0,724)	(0,556)	(2,329)	(1,958)	(1,382)
Income 3500-5000	8,02**	3,68***	2,634**	6,72*	5,554**	2,914
	(3,49)	(1,321)	(1,054)	(3,609)	(2,445)	(2,198)
Income >5000	10,2***	4,412**	3,299**	5,117	5,287	2,319
	(3,846)	(1,865)	(1,309)	(3,876)	(3,385)	(2,631)
Education - high	4,388**	1,287	1,206*	5,116**	1,736	1,598
	(1,849)	(0,866)	(0,667)	(2,408)	(1,896)	(1,019)
Female	2,203	1,855*	1,121	1,864	1,906	1,509
	(1,894)	(1,004)	(0,726)	(2,142)	(1,711)	(1,329)
Single	-2,265	-2,608**	-1,799*	-0,725	-3,69	-0,982
	(2,429)	(1,248)	(0,994)	(3,435)	(2,265)	(1,8)
No children	-1,201	0,343	0,451	-2,146	-2,403	-1,176
	(2,263)	(1,177)	(0,907)	(3,276)	(2,583)	(1,574)
Children <5 years	5,422**	1,704	0,517	2,474	-1,027	-1,37
	(2,687)	(1,405)	(0,943)	(3,545)	(2,843)	(1,766)
Children 5-12 years	-3,407	-2,02*	-1,697**	-8,473**	-4,882*	-2,302
	(2,363)	(1,188)	(0,844)	(3,46)	(2,674)	(1,687)
Age <25	-2,713	-2,073**	-1,079*	0,116	-1,621	-0,318
	(1,699)	(0,848)	(0,622)	(2,284)	(1,726)	(1,257)
Age >50	-3,537	-0,078	0,876	-2,295	0,074	1,217
	(2,327)	(1,116)	(0,92)	(2,559)	(1,695)	(1,231)
Constant	25,766***	19,671***	13,302***	23,836***	28,234***	16,836***
	(3,261)	(1,812)	(1,223)	(4,904)	(3,86)	(2,4)

Note: The number of observations is 974 (2x487); Weights are given by the class probabilities. Mean estimates for each individual characteristic with bootstrapped standard errors (500 reps) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

is surprising, as we would expect positive coefficients, given the fact that scheduling constraints are likely more stringent because of the fixed starting times of schools (Schwanen and Ettema, 2009). On the other hand, people may start working full-time again when their children go to secondary school (which is the reference category), implying more restrictive scheduling constraints. We also find that younger people seem to have less severe scheduling constraints, which seems to make sense.

The WLS specifications put substantial structure on the data and do not allow for interactions between individual characteristics. In columns (4)-(6) we therefore report the mean effects of the individual characteristics when we use local linear regression techniques that allow for these interactions. Because we put less structure on the data, the standard errors are substantially larger compared to the WLS specifications. We still find a positive income effect: people with an income higher than €3,500 have a VOT that is about €5.50 higher compared to someone who earns less than €2,500. However, this effect is only marginally statistically significant. The effect of income on the VSDE is statistically significant, except for the highest income class for which we have fewer observations. There is a pronounced effect of education, conditional on the income effect. Highly educated people have a VOT that is about €5 higher. To summarise, although the effects of the LLR specifications are somewhat imprecise, the mean effects seem to be broadly in line with the WLS specifications.

5. Conclusion and discussion

We introduced a semi-parametric latent class model that allows for both unobserved heterogeneity in preferences and heterogeneity related to observable individual characteristics. Furthermore, we applied a new entropy criterion for the selection of classes that also may be useful for standard latent class models.

We used data from a stated choice experiment to estimate the VOT, the VSDE and the VSDL. For our dataset, it is shown that, although the estimated average values are hardly affected when ignoring unobserved heterogeneity, there is substantial heterogeneity in the estimated parameters when using more flexible methods. We find that unobserved heterogeneity seems to be most important. Nevertheless, we also find some effects for income, education and household composition on values of time and schedule delays.

The current paper must be viewed as a first step towards the use of flexible local mixture discrete choice modelling and suggests several avenues for further research. First, bandwidth optimisation in local mixture models could be better developed in order to use statistical criteria instead of eye-balling to decide on the degree of smoothing. Closely related to this is the issue of model selection. In order to test the method against other modelling approaches further research may aim at the development of a statistical criterion. Extending the bandwidth selection approach of Dekker et al. (2014) to local mixture models seems to be a promising direction in this respect.

Bibliography

Abrantes, P. A. L. and Wardman, M. R. (2011). Meta-analysis of UK values of travel time: An update. *Transportation Research Part A: Policy and Practice*, 45(1):1–17.

Arnott, R., de Palma, A., and Lindsey, R. (1992). Route choice with heterogeneous drivers and group-specific congestion costs. *Regional Science and Urban Economics*, 22(1):71–102.

Arnott, R., de Palma, A., and Lindsey, R. (1994). The Welfare Effects of Congestion Tolls with Heterogeneous Commuters. *Journal of Transport Economics and Policy*, 28:139–161.

Bajari, P. and Kahn, M. (2005). Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business & Economic Statistics*, 23(1):20–33.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890.

Biernacki, C., Celeux, G., and Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, 20(3):267–272.

Börjesson, M., Fosgerau, M., and Algers, S. (2012). Catching the tail: Empirical identification of the distribution of the value of travel time. *Transportation Research Part A: Policy and Practice*, 46(2):378–391.

Börjesson, M. and Kristoffersson, I. (2014). Assessing the welfare effects of congestion charges in a real world setting. *Transportation Research Part E: Logistics and Transportation Review*, 70:339–355.

Brownstone, D. and Small, K. A. (2005). Valuing time and reliability: Assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39:279–293.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of classification*, 13(2):195–212.

Dekker, T., Brouwer, R., and Koster, P. (2014). Changing with the Tide: Semi-Parametric Estimation of Preference Dynamics. *Land Economics*, 90(4):717–745.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, 72(4):1077–1106.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60:591–608.

Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions. *Journal of the American Statistical Association*, 90:141.

Fosgerau, M. (2007). Using nonparametrics to specify a model to measure the value of travel time. *Transportation Research Part A: Policy and Practice*, 41:842–856.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37:681–698.

Guan, W. (2003). From the help desk: bootstrapped standard errors. *The Stata Journal*, 3(1):71–80.

Hall, P., Li, Q., and Racine, J. S. (2007). Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors. *Review of Economics and Statistics*, 89(4):784–789.

Hathaway, R. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics & Probability Letters*, 4(2):53–56.

Hensher, D. A. (2001). The valuation of commuter travel time savings for car drivers: evaluating alternative model specifications. *Transportation*, 28:101–118.

Hjorth, K. and Fosgerau, M. (2012). Using prospect theory to investigate the low marginal value of travel time for small time changes. *Transportation Research Part B: Methodological*, 46(8):917–932.

Horowitz, J. (2001). The bootstrap. In Heckman, J. and Learner, E., editors, *Handbook of econometrics*, volume 5, pages 3159–3228.

Huang, H. (2000). Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters. *Transportation Research Part E: Logistics and Transportation Review*, 36(4):267–284.

Knockaert, J., Ettema, D., Verhoef, E., and Koster, P. (2012). Spitsmijden Gouda-Zoetermeer. Technical report.

Lam, T. C. and Small, K. A. (2001). The value of time and reliability: Measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, 37:231–251.

Li, Z., Hensher, D. A., and Rose, J. M. (2010). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, 46:384–403.

McMillen, D. P. and Redfearn, C. L. (2010). Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions. *Journal of Regional Science*, 50:712–733.

Mills, D. (1981). Ownership arrangements and congestion-prone facilities. *The American Economic Review*, 71(3):493–502.

Noland, R. and Small, K. (1995). Travel-time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes. *Transportation Research Record*, 1493:150–158.

Peer, S., Knockaert, J., Koster, P., and Verhoef, E. (2014). Over-reporting vs. overreacting: Commuters' perceptions of travel times. *Transportation Research Part A: Policy and Practice*, 69:476–494.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy*, 110(4):705–729.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119:99–130.

Schwanen, T. and Ettema, D. (2009). Coping with unreliable transportation when collecting children: Examining parents' behavior with cumulative prospect theory. *Transportation Research Part A: Policy and Practice*, 43:511–525.

Tan, Z. and Yang, H. (2012). The impact of user heterogeneity on road franchising. *Transportation Research Part E: Logistics and Transportation Review*, 48:958–975.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82:559–567.

Van den Berg, V. and Verhoef, E. T. (2011a). Congestion tolling in the bottleneck model with heterogeneous values of time. *Transportation Research Part B: Methodological*, 45:60–78.

Van den Berg, V. and Verhoef, E. T. (2011b). Winning or losing from dynamic bottleneck congestion pricing?. The distributional effects of road pricing with heterogeneity in values of time and schedule delay. *Journal of Public Economics*, 95:983–992.

Van den Berg, V. and Verhoef, E. T. (2013). Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. *Transportmetrica A: Transport Science*, 10(5):377–400.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review*, 59:251–260.

Winston, C. and Yan, J. (2011). Can privatization of U.S. highways improve motorists' welfare? *Journal of Public Economics*, 95:993–1005.

Appendix A. Derivation of the entropy criterion

We start with Equation (12) and then substitute Equations (13) and (14):

$$G\hat{L}L(C) = \sum_{c=1}^{C}\sum_{n=1}^{N} \hat{v}_{nc} \log\left(\hat{p}_{nc}\hat{P}_{n|c}\right) - \sum_{c=1}^{C}\sum_{n=1}^{N} \hat{v}_{nc}\log\hat{v}_{nc} \tag{A.1}$$

This can be rewritten using the following steps:

$$\begin{aligned}
G\hat{L}L(C) &= \sum_{c=1}^{C}\sum_{n=1}^{N}\left(\hat{v}_{nc}\log\left(\hat{p}_{nc}\hat{P}_{n|c}\right) - \hat{v}_{nc}\log\hat{v}_{nc}\right) \\
&= \sum_{c=1}^{C}\sum_{n=1}^{N}\hat{v}_{nc}\log\left(\frac{\hat{p}_{nc}\hat{P}_{n|c}}{\hat{v}_{nc}}\right)
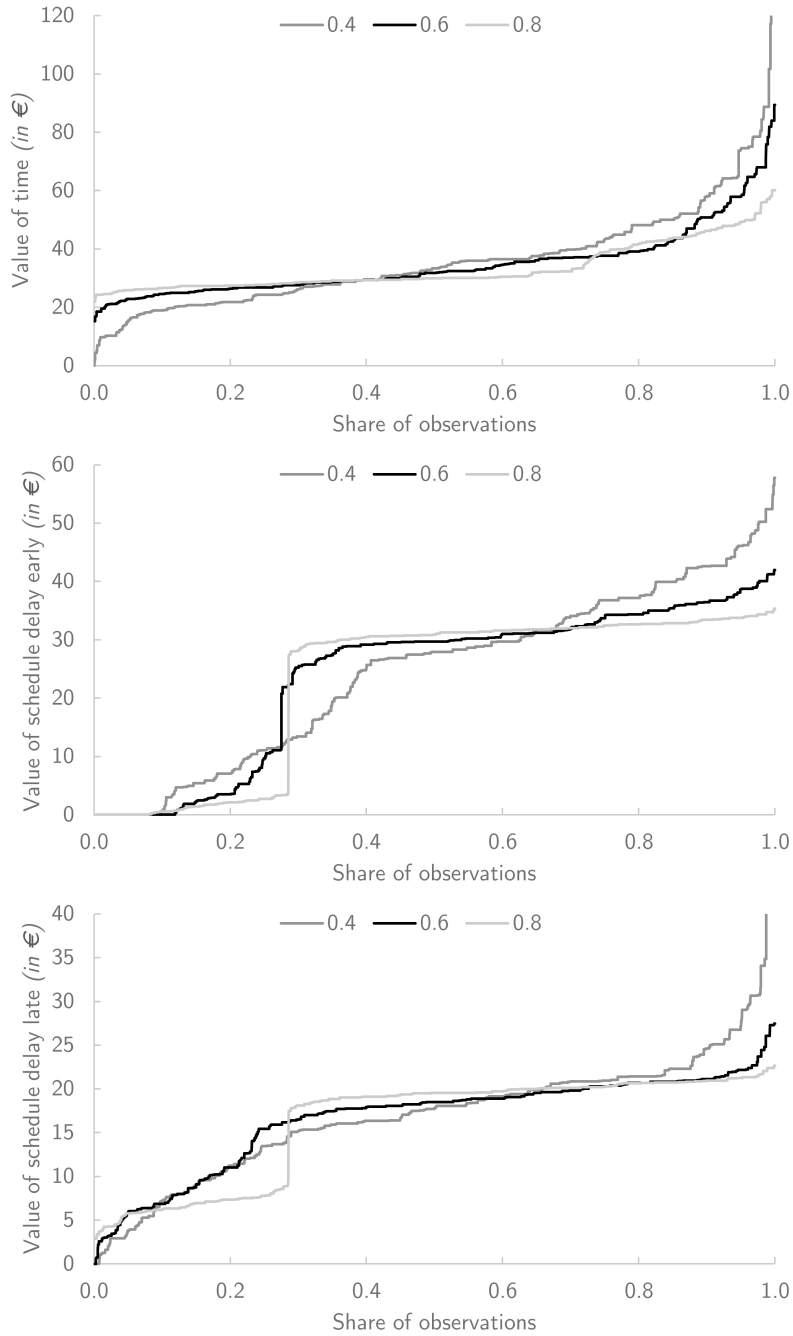\end{aligned} \tag{A.2}$$

Using (15) in the denominator gives:

$$\begin{aligned}
G\hat{L}L(C) &= \sum_{c=1}^{C}\sum_{n=1}^{N}\hat{v}_{nc}\log\left(\sum_{c=1}^{C}\hat{p}_{nc}\hat{P}_{n|c}\right) \\
&= \sum_{n=1}^{N}\sum_{c=1}^{C}\hat{v}_{nc}\log\left(\sum_{c=1}^{C}\hat{p}_{nc}\hat{P}_{n|c}\right) \\
&= \sum_{n=1}^{N}\log\left(\sum_{c=1}^{C}\hat{p}_{nc}\hat{P}_{n|c}\right),
\end{aligned} \tag{A.3}$$

where the last step uses $\sum_{c=1}^{C}\hat{v}_{nc} = 1$. This concludes the proof that the model log-likelihood (Equation 9) can be separated into Equations (13) and (14).

Appendix B. Bandwidth sensitivity analysis

The graphs below show the estimated cumulative distribution functions of the WTP values for bandwidth values $\bar{\lambda} = 0.4$, $\bar{\lambda} = 0.6$ and $\bar{\lambda} = 0.8$. For lower values of $\bar{\lambda}$ the distribution becomes more dispersed and has a longer tail. For higher values of $\bar{\lambda}$ the distribution approaches a distribution with two mass points because there are two latent classes included in the analysis. Note that the results for the other bandwidth generally fall in the local confidence intervals presented in 2.

Appendix C. Estimated probability densities

Value of time *(in €)*

Value of schedule delay early *(in €)*

Value of schedule delay late *(in €)*