Koster, Paul; Koster, Hans

Conference Paper

Commuters' preferences for fast and reliable travel

53rd Congress of the European Regional Science Association: "Regional Integration: Europe, the Mediterranean and the World Economy", 27-31 August 2013, Palermo, Italy

Provided in Cooperation with:
European Regional Science Association (ERSA)

Suggested Citation: Koster, Paul; Koster, Hans (2013) : Commuters' preferences for fast and reliable travel, 53rd Congress of the European Regional Science Association: "Regional Integration: Europe, the Mediterranean and the World Economy", 27-31 August 2013, Palermo, Italy, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:
https://hdl.handle.net/10419/124087

Commuters' Preferences for Fast and Reliable Travel

BY PAUL R. KOSTER* AND HANS R.A. KOSTER

DEPARTMENT OF SPATIAL ECONOMICS, VU UNIVERSITY AMSTERDAM

TINBERGEN INSTITUTE AMSTERDAM

May 14, 2013

SUMMARY — Traffic congestion contributes to longer travel times and increased travel time variability. We account for the dynamic nature of travellers' choices, by deriving a closed-form solution for the costs of travel time variability. The resulting travel delay cost function is linear in the mean travel delay. Then, we use a semiparametric estimation approach to analyse observed and unobserved heterogeneity in the value of travel time and reliability. Using data from a stated choice experiment, we show that there is substantial heterogeneity in the willingness to pay for fast and reliable travel. About 5-25 percent of the heterogeneity in the value of time and reliability is attributable to observed characteristics of individuals, implying that unobserved heterogeneity is much more important than heterogeneity related to observable characteristics. It is furthermore shown that schedule delay costs are on average 24 percent of the total costs of travel delays.

JEL–code — C13, C14, R40, R41

Keywords — local maximum likelihood estimation, heterogeneity, semiparametric Logit, value of time, value of reliability.

* Corresponding author. Department of Spatial Economics, VU University, De Boelelaan 1105 1081 HV Amsterdam, e-mail: p.r.koster@vu.nl. This study is financially supported by Rijkswaterstaat Centre for Infrastructure and benefited from a NICIS-KEI research grant. Stefanie Peer, Yin-Yen Tseng, Jasper Knockaert and Erik Verhoef are thanked for their contribution to the development of the survey. Part of the work was done while the first author was visiting the Danish Technical University, Department of Transport. We like to thank Mogens Fosgerau, Katrine Hjorth, Thijs Dekker, Jan Rouwendal, Erik Verhoef, Piet Rietveld and Vincent van den Berg for their valuable comments. An early version of this paper was presented at the International Choice Modelling Conference 2011 and participants are thanked for their constructive comments.

I. Introduction

More than 50 percent of the world population is concentrated in cities and this number is steadily increasing. Apart from the positive agglomeration economies, cities also imply negative externalities such as traffic congestion, air pollution, noise and accidents. This paper studies the costs of traffic congestion. Traffic congestion contributes to longer commuting trips and increased variability of travel times.[1] There is a vast literature on the valuations of reductions in travel time and improvements in reliability, but the literature still faces some substantial challenges (see Wardman et al., 1998; Bates et al., 2001; Lam and Small, 2001; Mackie et al., 2001; Small, 2012).

One of the foremost challenges is to model the dynamics of travellers' choice, because travellers cope with travel time variability by departing earlier or later from home (Gaver, 1968; Knight, 1974; Noland and Small, 1995; Fosgerau and Karlström, 2010). More precisely, they trade-off their expected costs of being earlier than their preferred arrival time (schedule delay early) against their expected costs of being later than their preferred arrival time (schedule delay late). Any model ignoring these dynamics will likely overestimate the costs of travel time variability, because anticipating behaviour of travellers is not accounted for. Although the dynamics of travellers' choice are considered to be important, these are often ignored in transport models used for policy evaluation. Noland and Small (1995), Fosgerau and Karlström (2010) and Fosgerau and Engelson (2011) therefore developed reduced form cost functions that can be incorporated in static congestion models. Instead of employing a fully dynamic equilibrium model, the costs of travel time variability are then related to the standard deviation or variance of travel times.

A second challenge is to correctly deal with preference heterogeneity. Preference heterogeneity is of key importance when evaluating the effect of transport economic policies for several reasons. Previous research has showed that the distributional effects of congestion pricing policies strongly depend on the heterogeneity in individuals' preferences. (Arnott et al. 1988; Arnott et al. 1994; Lindsey, 2004; Van den Berg and Verhoef, 2011). Further, heterogeneity is important when evaluating the benefits of private provision of highways, because profits depend on the marginal willingness to pays for reductions in travel time and travel time variability of travelers. (Mills, 1981; Winston and Yan, 2011; Tan and Yang, 2012). Heterogeneity is also important when one studies the effect 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 mean

[1] See for example Carrion and Levinson (2012) and Li et al. (2010) for recent overviews of the literature on travel time unreliability.

willingness to pay values may lead to biased welfare estimates and incorrect policy recommendations.

The contribution of this paper is twofold. First, we derive an easy-to apply reduced-form cost function for travel under unreliable conditions, assuming that travel times are log-normally distributed and the mean and standard deviation of travel delay are linearly related. These assumptions are in line with our data, as well with previous research by Pu (2010) and Peer et al. (2012). Estimated values of schedule delay early and late can therefore directly be translated into the costs of travel time variability. More specifically, our resulting reduced-form cost function turns out to be linear in the mean travel delay. Since mean delays are generally available in transport statistics we consider this as a very useful result. Our expression for the costs of travel time variability accounts for the dynamic nature of travellers' decision, but remains easy to apply in static congestion models (see for example Winston and Yan, 2011).

Second, we propose an econometric framework to analyse heterogeneity conditional on observed individual characteristics, as well as unobserved heterogeneity, using panel data from a stated choice experiment.[2] More specifically, we estimate a semiparametric discrete choice model, where the estimates of value of time and value of schedule delay depend on observed and unobserved individual characteristics. We obtain semiparametric distributions of preferences, by employing local-likelihood estimation methods introduced by Tibshirani and Hastie (1987), Fan et al. (1995) and Fan et al. (1998), and by assuming that individuals who are similar in terms of socio-economic characteristics will have more similar preference parameters. To analyse how the estimated parameters relate to demographic characteristics, we will regress the estimated preference parameters on individual characteristics.[3]

Our econometric approach estimates a semiparametric panel latent class model and 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 unobserved and observed heterogeneity. Third, because we use kernel smoothing

[2] Choice experiments are often used to estimate preference parameters (Brownstone and Train, 1998). There may be several reasons to prefer stated preference over revealed preference, for example because of collinearity between the variables of interest. In transport economics, the cost and time component of a trip are highly correlated. We refer to Hensher (2010) for a recent discussion on the validity of stated preference willingness to pay estimates.

[3] 'Local' techniques are often applied in the hedonic price literature, where parameters are estimated *conditional* on the geographic location (also known as geographically weighted regression) (McMillen and Redfearn, 2010). A similar approach is used in the literature on hedonic pricing (Bajari and Kahn, 2005; Bajari and Benkard, 2005).

techniques, we avoid the well-known curse of multidimensionality. Nevertheless, we allow for *all* interactions between individual characteristics and preferences (Horowitz and Savin, 2001; Bontemps et al., 2008; McMillen and Redfearn, 2010). One may argue that many studies have interacted preferences with individual characteristics. However, the latter approach typically does not allow for all interactions between individual characteristics and is computationally intensive 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 potential bias due to the correlation of errors of individuals' choices. It results in a unique semiparametric distribution of preferences for each unique combination of individual characteristics in our dataset.

The data to estimate commuters' value of travel time and value of arriving at the preferred arrival time at work is obtained from a stated choice experiment held among participants of a real-world rewarding experiment to combat congestion. We estimate the willingness to pay values (WTP) for reductions in travel time (α), schedule delay early (β) and late (γ). It is shown that there is substantial heterogeneity in α, β and γ. We also link heterogeneity to observable individual characteristics, by regressing the estimated parameters on individual characteristics. We find that individuals with high incomes have a higher value of times and schedule delay costs. Women have 25 percent higher values of schedule delay and presence of young children in the household increases the β with 33 percent, as young children usually impose strong scheduling constraints. Our estimated total delay costs are about 25 percent of the total trip costs (excluding vehicle and fuel costs) and scheduling costs are 24 percent of the total delay costs.

The paper continues as follows. In Section II we specify the utility function and derive an easy-to-use expression for the costs of travel time variability. Section III introduces the econometric set-up. Section IV discusses the design and set-up of the stated choice experiment, followed by the empirical results in Section V. Section VI concludes.

II. Utility specification and the costs of travel time variability

A. Utility specification

We developed a stated choice experiment to collect data about the preferences of morning-commuters participating in a peak-avoidance project. Travellers receive a reward if they travel outside the morning peak. They trade-off earlier or later arrivals and shorter travel times with the monetary reward. We assume that the deterministic part of the utility (v_{inj}) of individual i facing choice n and choosing alternative j, is explained by three types of variables: expected reward r_{inj}, expected travel time tt_{inj} and expected schedule delay

sd_{inj}. Schedule delay is defined as the deviation of an arrival time from the preferred arrival time pat_i. Vickrey (1969) and Small (1982) introduced this scheduling model, where arrivals different from pat_i result in a disutility. Their model was extended to a model with stochastic travel times by Noland and Small (1995), where each departure time from home can have multiple outcomes for the arrival time at work. The expected utility of individual i that makes choice n by choosing alternative j is defined as follows:

$$u_{inj} = \frac{1}{\xi_i} \cdot v\left[\delta_i; r_{inj}, tt_{inj}, sd_{inj}\right] + \epsilon_{inj}. \quad (1)$$

Equation (1) shows that utility is an additive function of the deterministic part and the random component ϵ_{inj}. The δ_i's are individual-specific and are only the same for respondents with the same characteristics and ξ_i is the scale of the utility which is normalised to one for identification purposes. In order to capture travel time variability, each departure time from home (dt) has M possible outcomes of the travel time resulting in M corresponding arrival times. The schedule delay of mass point $m = 1 \ldots M$ is given by the following equations:

$$\tilde{e}_{injm} = \max\left(0; pat_i - dt_{inj} - g_{injm}\right), \quad (2)$$

$$\tilde{l}_{injm} = \max\left(0; dt_{inj} + g_{injm} - pat_i\right), \quad (3)$$

where the ~ indicates observed values, g_{injm} is the travel time, $\tilde{e}_{injm}$ is the schedule delay early, and $\tilde{l}_{injm}$ is schedule delay late. These equations define the scheduling model introduced by Vickrey (1969) and Small (1982). Schedule delay disutility is a piecewise linear function of arrival time. So, besides the disutility of travel time, there is additional disutility of not arriving at pat_i, where the marginal disutility of being early and late are valued differently by travellers. When scheduling preferences are nonlinear, these linear functions may serve as an approximation. In our choice experiment, every departure time has two possible outcomes for the travel time ($M = 2$). Both arrival times have a corresponding probability Π_{inj} and $1 - \Pi_{inj}$ respectively. Following Noland and Small (1995), the model variables are the averaged values over these two mass points and are calculated as:

$$r_{inj} = \Pi_{inj}\tilde{r}_{inj1} + (1 - \Pi_{inj})\tilde{r}_{inj2}, \quad (4)$$

$$tt_{inj} = \Pi_{inj}\tilde{tt}_{inj1} + (1 - \Pi_{inj})\tilde{tt}_{inj2}, \quad (5)$$

$$e_{inj} = \Pi_{inj}\tilde{e}_{inj1} + (1 - \Pi_{inj})\tilde{e}_{inj2}, \quad (6)$$

$$l_{inj} = \Pi_{inj}\tilde{l}_{inj1} + \left(1 - \Pi_{inj}\right)\tilde{l}_{inj2}. \quad (7)$$

The deterministic part of the utility is then given by the sum of the expected reward, expected travel time and expected schedule delay variables, multiplied by the individuals' preference parameters:

$$v_{inj} = \delta_i^r r_{inj} + \delta_i^{tt} tt_{inj} + \delta_i^e e_{inj} + \delta_i^l l_{inj}. \tag{8}$$

We are interested in the ratio's of the parameters δ_i. These are defined as follows:

$$\alpha_i = -\frac{\delta_i^{tt}}{\delta_i^r}, \tag{9}$$

$$\beta_i = -\frac{\delta_i^e}{\delta_i^r}, \tag{10}$$

$$\gamma_i = -\frac{\delta_i^l}{\delta_i^r}. \tag{11}$$

where α_i denotes the value of time, β_i is the value of schedule delay early and γ_i the value of schedule delay late.

B. Congestion and scheduling costs

In this subsection we derive the costs of travel time variability assuming a parametric distribution of travel delays. Fosgerau and Karlström (2010) derive the optimal expected costs assuming a general standardised travel delay distribution. A major advantage of assuming a parametric distribution is that we are able to express the costs of travel time variability in a convenient closed-form function.

Let $F[T]$ be the cumulative density function and $F'[T]$ the probability density function of travel delays T. Define the mean travel time of individual i by μ_i and define h_i as the headstart, which is defined by $pat_i - dt_i$. Individuals' willingness to pay values are defined by equations (9), (10) and (11). Let $C[h_i] = f_i + d_i[h_i]$, where $C[h_i]$ are the expected total costs of a trip (only related to travel time and scheduling, while ignoring fuel and vehicle costs). Furthermore, f_i denotes the free-flow travel costs, where $f_i = \alpha_i g_i$ and g_i denotes the individual-specific free-flow travel time which depends on individuals' work and residential locations. Following Noland and Small (1995), the expected delay costs $d[h_i]$ are given by the sum of congestion time costs and the costs of expected schedule delay early and late:

$$d_i[h_i] = t_i + s[h_i], \tag{12}$$

where $t_i = \alpha_i \mu_i$ are the time costs of mean travel time delay. The expected scheduling costs $s[h_i]$ are given by:

$$s_i[h_i] = \beta_i \int_0^{h_i} (h_i - T)F'[T]\mathrm{d}T + \gamma_i \int_{h_i}^{\infty} (T - h_i)F'[T]\mathrm{d}T. \tag{13}$$

Let's formulate the first assumption regarding travellers' behaviour:

ASSUMPTION A. *We assume that people optimise their headstart h_i using the travel time distribution $F[T]$, where h_i^* denotes the optimal headstart.*

This assumption may seem restrictive, because travellers are assumed to have a perfect perception of the travel time distribution. However, it is shown that if travellers tend to overestimate both the probabilities of short and long travel time, this will result in an optimal headstart close to h_i^* (Koster and Verhoef, 2012). For analytical convenience we therefore assume that travellers maximize their expected utility. In line with Fosgerau and Kärlstrom (2010), it is shown in Appendix A that delay costs $d_i[h_i]$ are given by:

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i) \int_{\frac{\gamma_i}{\beta_i+\gamma_i}}^{1} F^{-1}[x]\mathrm{d}x, \tag{14}$$

where F^{-1} is the inverse cumulative density function. The last integral captures the size of the tail of the travel delay distribution to the right of the $100\,\gamma_i/(\beta_i + \gamma_i)$ percentile (Small, 2013). The disadvantage of equation (14) is that one needs simulation to approximate the integral. However, if we make an additional assumption we may avoid the need for simulation and arrive at a closed-form expression for $d[h_i^*]$.

ASSUMPTION B. *We assume that travel delays follow a log-normal distribution with shape parameter τ_i and scale parameter θ_i.*

This seems to be a very plausible candidate distribution for travel times as will be shown in Section IV.B and also by Pu (2010). We then may formulate the following proposition:

PROPOSITION 1: *Under Assumptions A and B there is a closed-form solution for the expected delay costs that is a function of α_i, β_i, γ_i, θ_i and μ_i and is given by:*

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\mu_i \Phi\left[\theta_i + \Phi^{-1}\left[\frac{\beta_i}{\beta_i + \gamma_i}\right]\right]. \tag{15}$$

Proof. See for the proof, Appendix A. ■

$\Phi[\,\cdot\,]$ is the cumulative standard normal distribution and $\Phi^{-1}[\,\cdot\,]$ is the inverse of this distribution. Furthermore, $\mu_i = \mathrm{e}^{\tau_i + \theta_i^2/2}$ is the mean travel delay in hours. If $\theta_i \to 0$, there is no variation in travel times and (15) reduces to $\alpha_i\mu_i$. So, given observations on μ_i and estimates for θ_i, α_i, β_i and γ_i, we can approximate the total delay costs per trip and the scheduling costs s_i.

The standard deviation of travel delays is often unknown, but measurements of μ_i are available (see Section IV.B). In line with Peer et al. (2012), we show that there is a strong empirical relation between the standard deviation of travel delays and the mean travel delay. So, it seems reasonable to make the following assumption.

ASSUMPTION C. *We assume that the mean and the standard deviation of delay are linearly related, implying that the coefficient of variation $\sigma_i/\mu_i = \eta$ is a constant.*

This leads to the second proposition:

PROPOSITION 2. *Under Assumption A, B and C, θ_i is independent of the mean travel delay. The delay costs are linear in the mean travel delay with $d[h_i^*] = (\alpha_i + \zeta[\beta_i, \gamma_i, \eta])\mu_i$, and $\zeta[\beta_i, \gamma_i, \eta] = -\beta_i + (\beta_i + \gamma_i)\Phi\left[\sqrt{\ln[1+\eta^2]} + \Phi^{-1}[\beta_i/(\beta_i + \gamma_i)]\right]$.*

Proof. In Appendix B we show that if the mean and the standard deviation are linearly related, θ_i is given by:

$$\theta_i = \sqrt{\ln[1+\eta^2]}. \tag{16}$$

This means that θ_i is independent of mean delay and therefore (15) is linear in the mean delay. The parameter η is the marginal increase in the standard deviation of travel times if the mean delay increases with one hour and will be estimated in Section IV.B. By substituting (16) in (15) we obtain a convenient closed-form expression for $d[h_i^*]$ that depends linearly on the mean travel delay. ■

This result is easily applicable in static transport models. These models usually have a cost function which is linear in the mean delay ($\alpha_i\mu_i$). We arrive at a cost function of $(\alpha_i + \zeta[\beta_i, \gamma_i, \eta])\mu_i$ which implies that a simple increase of the value of travel time of $\zeta[\beta_i, \gamma_i, \eta]$ will be sufficient to incorporate the costs of travel time variability in these models.

In our empirical application, we are interested in the relative importance of delay costs in the total costs of a trip. The share of the time costs due to delay is given by:

$$\psi_i[h_i^*] = \frac{d[h_i^*]}{C[h_i^*]}. \tag{17}$$

Furthermore we estimate the share of the costs of travel time variability in the total delay costs of a trip as:

$$\omega_i[h_i^*] = \frac{d[h_i^*] - \alpha_i\mu_i}{d[h_i^*]} = \frac{\zeta[\beta_i, \gamma_i, \eta]}{\alpha_i}. \tag{18}$$

The above equations have the advantage that they are independent of the marginal utility of income δ_i^r. Equation (18) is also independent of the mean travel delay μ_i.

III. Econometric setup

A. Local estimation

Given observations on free flow travel time g_i, mean travel delays μ_i and an estimate of θ_i, we still need to estimate the preference parameters. We consider the following econometric model:

$$E(y|\boldsymbol{x}) = G[H[\boldsymbol{z};\boldsymbol{x};\boldsymbol{\delta}];\boldsymbol{z}], \tag{19}$$

where y is a dichotomous dependent variable, $\boldsymbol{x}$ is a matrix of explanatory variables, $\boldsymbol{z}$ is a matrix of observed individual characteristics, and $\boldsymbol{\delta}$ is a matrix of parameters that is unknown. Many widely used parametric models have this form, including linear regression, Probit, Logit and Tobit models (Horowitz and Härdle, 1996). Often, it is assumed that $G[\cdot]$ is known and that $H[\cdot]$ is $\boldsymbol{\delta x}$.

In this paper we estimate $G[\cdot]$ and $H[\cdot]$ using flexible estimation techniques. The most popular application for estimating $G[\cdot]$ is the Mixed Logit model that allows for unobserved heterogeneity in preferences, often given an assumption on the multivariate distribution of preferences $\boldsymbol{\delta}$ (see, among others, Revelt and Train, 1998; Brownstone and Train, 1998; McFadden and Train, 2000; Small et al., 2005; Harding and Hausman, 2007). Other methods have been used to estimate $H[\cdot]$ nonparametrically, such as local polynomial methods (Fosgerau, 2007), a Box-Cox type Logit model (De Lapparent et al., 2002) or smoothing cubic splines (Fukuda and Yai, 2010), to name just a few. We will estimate $G[\cdot]$ semiparametrically using Panel Latent Class estimation but let the shape of $G[\cdot]$ depend semiparametrically on the characteristics $\boldsymbol{z}$. Furthermore, we estimate $H[\cdot]$ as a semiparametric function of $\boldsymbol{z}$. We will combine Panel Latent Class estimation with so-called Local Logit estimation.

Tibshirani and Hastie (1987) and Fan et al. (1998) introduced Local Likelihood estimation. The term 'local' implies that each local point (e.g. individual, observation) is treated as a reference point. Conditional on the local point, a vector of weights is determined, reflecting the (multidimensional) distance between the local point and the other points in the dataset. We let the kernel weight depend on individual characteristics $\boldsymbol{z}$, implying 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 semiparametric distribution of preferences is the same. The bandwidths λ determine the degree of smoothing. Since we will only include categorical variables we have $0 \leq \lambda \leq 1$. For strong smoothing (a high bandwidth), the weights are uniform and the model reduces to the (standard) Panel Latent Class model. For weak smoothing (a low bandwidth), the model becomes a saturated model

resulting in a separate estimation of a Panel Latent Class models with $\mathcal{C}$ classes for each unique combination of individual characteristics.

This is an appealing feature of the estimation setup is that both the the Local Logit model ($\mathcal{C} \to 1, \lambda < 1$), the Binary Logit ($\mathcal{C} \to 1, \lambda \to 1$) and the Panel Latent Class model ($\lambda \to 1, \mathcal{C} > 1$) are special cases of our model. The (restrictive) Binary Logit model results in equal preferences for all individuals, the Local Logit Model estimates unique preferences for each unique combination of individual characteristics, but ignores unobserved heterogeneity. The Panel Latent Class model estimates equal preference distributions for all individuals, and hence ignores heterogeneity related to observable characteristics. The Local Panel Latent Class model estimates a unique preference distribution for each unique combination of individual characteristics and therefore allows for both observed and unobserved heterogeneity.

More formally, we have a balanced panel with I individuals making N choices each. The probability that individual i chooses alternative j^* for choice n is defined as the Logit probability $P_{in}^* = e^{v_{inj^*}} / \sum_{j=1}^{J} e^{v_{inj}}$. The probability for a sequence of choices of individual i is then given by the weighted sum of the sequences of choices of each class and is given by:

$$P_i^*(\boldsymbol{\delta}_i; p_i; \boldsymbol{x}_i) = \sum_{c=1}^{\mathcal{C}} p_{ic} \prod_{n=1}^{N} P_{in}^*(x_{in}; \delta_{ic}), \tag{20}$$

where δ_{ic} indicates the $R \times 1$ vector of preference parameters of individual i for latent class $c = 1 \dots \mathcal{C}$, where R is the number of preference parameters to be estimated, p_{ic} is the class probability of class c and x_{in} is a $1 \times R$ vector of explanatory variables. Furthermore, $\boldsymbol{\delta}_i$ is the $R \times \mathcal{C}$ matrix with preference parameters and p_i is the $\mathcal{C} \times 1$ vector of corresponding class probabilities. These class probabilities determine the shape of the distribution of preferences. Finally, $\boldsymbol{x}_i$ is the $N \times R$ matrix with explanatory variables. We do not allow for unobserved heterogeneity in all parameters $\boldsymbol{\delta}_i$, because that would lead to unstable and unrealistic estimates. More specifically, we do not allow for unobserved heterogeneity in the reward parameter, implying that $\delta_{ic}^r = \delta_i^r$. Because the parameters of interest (α_i, β_i and γ_i) are ratios, we do not see this as a major problem.

To include observed preference heterogeneity, we estimate how individual characteristics $\boldsymbol{z}$ affect the preference parameters $\boldsymbol{\delta}_i$ and the class probabilities p_i. We condition on Q individual characteristics and $\boldsymbol{z}$ is a $I \times Q$ matrix with characteristics. The preference parameters $\boldsymbol{\delta}_i$ and the class probabilities p_i depend in a nonlinear way on $\boldsymbol{z}$. This means that all interactions of the different variables in $\boldsymbol{z}$ are modelled implicitly. The

vector of preference parameters of reference individual $\hat{\imath} = 1, \dots, I$ can then be estimated by maximising the local likelihood:

$$\{\widehat{\boldsymbol{\delta}}_{\hat{\imath}}, \hat{p}_{\hat{\imath}}\} = \underset{\boldsymbol{\delta}_i, p_i}{\operatorname{argmax}} \sum_{i=1}^{I} w_{\hat{\imath}}[\boldsymbol{z}; \lambda] \cdot \log P_i^*(\boldsymbol{\delta}_i; p_i; \boldsymbol{x}_i). \tag{21}$$

This shows that the local log-likelihood is calculated by taking the log of the probability of the chosen sequence, multiplied by a $I \times 1$ vector of kernel weights $w_{\hat{\imath}}[\boldsymbol{z}; \lambda]$ which will be defined later. The log of the probability of the chosen sequence depends on the independent variables and the preference parameters, and the local likelihood needs to be maximised in order to get a local estimate for the preference parameters $\boldsymbol{\delta}_i$ and the class probabilities p_i. The kernel weights depend on the socio-economic 'distance' of an individual compared to the other individuals, which is governed by the $Q \times 1$ vector of bandwidths λ. Estimation of a Panel Local Latent Class model for these variables implies the estimation of a Panel Latent Class model for each unique combination of individual characteristics.

B. Kernel functions

The individual-specific weights are based on the multi-dimensional 'distance' between individuals in their characteristics. This distance between individuals in terms of their characteristics is calculated using a kernel function. When the difference in individual characteristics between individual $\hat{\imath}$ and another individual becomes smaller, the other individual has a higher kernel weight in the local regression of $\hat{\imath}$ (and vice versa).

We include Q variables in the kernel function. Variable $q = 1 \dots Q$ has a corresponding kernel function $K_q(\cdot)$ and bandwidth λ_q (we discuss issues with respect to bandwidth selection in the next subsection). We employ a mixed kernel function consisting of ordered categorical variables, as well as dichotomous variables. A general specification of the kernel weights is given by:

$$w_{\hat{\imath}}[\boldsymbol{z}; \lambda] = \prod_{q=1}^{Q} K_q(\lambda_q; z_{\hat{\imath}q} - z_q). \tag{22}$$

In this equation $\boldsymbol{z}$ is the $I \times$ Q matrix with observed characteristics and $z_{\hat{\imath}q} - z_q$ is the distance of individual $\hat{\imath}$ to all other individuals for the $I \times 1$ vector z_q. The kernel weight $w_{\hat{\imath}}[\,\cdot\,]$ is a distance metric that decreases in the 'distance' between individual $\hat{\imath}$ and the other individuals. For a fixed number of individuals I and number of choices N, and given a vector of bandwidths λ, adding individual characteristics to the kernel leads to a lower kernel weight $w_{\hat{\imath}}[\,\cdot\,]$.

In 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 *and* that has the possibility to become a constant. We use ordered categorical and dichotomous individual characteristics. Following Racine and Li (2004) and Hall et al. (2007), the kernel function is:

$$K_q\left(\lambda_q; z_{\hat{\imath}q} - z_q\right) = \begin{cases} 1, & \text{if } z_{\hat{\imath}q} = z_q \\ \lambda_q^{|z_{\hat{\imath}q} - z_q|}, & \text{if } z_{\hat{\imath}q} \neq z_q \end{cases}. \tag{23}$$

In equation (23), the bandwidth λ_q has to be between zero and one.[4] One can verify that if λ_q equals one, $K_q(\cdot)$ is a constant, implying that the variable q has no effect on the estimated preference parameters.

C. Selection of the bandwidths and the number of latent classes

The first assumption we make is that $\lambda_1 = \lambda_2 = \cdots = \lambda_Q = \lambda$, so the bandwidth is equal for all variables included in the weight matrix. A univariate bandwidth simplifies the analysis and results in substantial savings in computation time.[5]

For low values of λ and high values of $\mathcal{C}$, estimates for α, β, and γ will deteriorate quickly and become unrealistic, due to the curse of multidimensionality. When λ increases, more observations in the 'neighbourhood' are taken into account in the local estimation of $\hat{\imath}$. Larger bandwidths may create a larger bias when the underlying function is nonlinear (Fan and Gijbels, 1996). A lower bandwidth leads to a better model fit and therefore to a higher value of the likelihood function, but increases the variance of the estimator.

There are several statistical criteria to determine the optimal bandwidth λ^* and latent classes $\mathcal{C}^*$. These criteria are usually based on the trade-off between fit and the number of parameters. We tried cross-validation methods, but they tend to under smooth the semiparametric functions to be estimated, leading to unrealistic values for the average willingness to pays $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$. In Appendix D, we show that when $\lambda < 0.4$ these mean values become unrealistically large. Similarly, if $\mathcal{C} > 6$, the results become instable. To jointly determine λ^* and $\mathcal{C}^*$, we therefore define a number of *economic* criteria that have to be met for $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$. These are listed in Table 1.

---

[4] It is also possible to include continuous individual characteristics by using a Gaussian or exponential weighting function. The weights are then a product of the different characteristics with different kernel functions (Racine and Li, 2004).

[5] This is because the computation time increases exponentially in Q. We refer to Yang and Tschernig (2002) for a discussion on multivariate bandwidths. The analysis of multivariate bandwidth optimisation and possible shortcuts to reduce computation times are beyond the scope of this paper.

TABLE 1: ECONOMIC CRITERIA FOR BANDWIDTH AND LATENT CLASSES SELECTION

Criterion	Implication
Value of time should be higher than value schedule delay early	$\bar{\alpha}(\lambda^*, \mathcal{C}^*) > \bar{\beta}(\lambda^*, \mathcal{C}^*)$
Mean willingness to pay values should be positive	$\bar{\alpha}(\lambda^*, \mathcal{C}^*) > 0, \bar{\beta}(\lambda^*, \mathcal{C}^*) > 0, \bar{\gamma}(\lambda^*, \mathcal{C}^*) > 0$
Mean willingness to pay values should be not far away from the results from Binary Logit (BL) model. We set $\iota = 0.25$.	$\bar{\alpha}(\lambda^*, \mathcal{C}^*) < (1+\iota)\alpha_{BL}, \bar{\beta}(\lambda^*, \mathcal{C}^*) > (1+\iota)\beta_{BL}$, $\bar{\gamma}(\lambda^*, \mathcal{C}^*) > (1+\iota)\gamma_{BL}$

First, it is economically infeasible to have willingness to pay values that are lower than zero, so we impose that at least the average willingness to pay values should be larger than zero. Second, a general finding in the literature is that the mean value of time is larger than the mean value of schedule delay early. Otherwise, the average traveller would prefer longer trips over arriving too early. This will also be an outcome of the Binary Logit model. We therefore impose that $\bar{\alpha} > \bar{\beta}$. Third, we impose that $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$, should be not too far from the original estimates of the Binary Logit estimates. We assume that the difference between the WTP values for the Binary Logit model and the mean WTP values obtained by the semiparametric models will not be larger than 25 percent. Given these economic criteria, we will choose λ^* and $\mathcal{C}^*$ in such a way that the log-likelihood function is maximised.

We are fully aware that these economic criteria are to some extent arbitrary. We therefore have investigated the results for a wide range of combinations of λ^* and $\mathcal{C}^*$ in Appendix D. In general, our conclusions are hardly influenced by relaxing or changing these criteria.

D. Unobserved versus observed heterogeneity

Given the local latent class models, we have different WTP parameters for each individual-class combination. To distinguish between unobserved and observed heterogeneity, we regress the estimated parameters on an individual constant. For example, for the value of time:

$$\hat{\alpha}_{ic} = \phi_i + \upsilon_{ic}. \tag{24}$$

where ϕ_i is a parameter to be estimated and υ_{ic} denotes the error term. We define $\mathcal{H}(\alpha) = \sum_{ic}(\phi_i - \bar{\alpha}_{ic})^2 / \sum_{ic}(\hat{\alpha}_{ic} - \bar{\alpha}_{ic})^2$, where $\mathcal{H}$ denotes the share of variation that is attributable to observed heterogeneity. In the (extreme) case where $\mathcal{C} = 1$ and $\lambda < 1$, all variation is attributable to observable characteristics ($\mathcal{H} = 1$). In the (extreme) case where $\lambda = 1$ and $\mathcal{C} > 1$, all variation is related to unobserved heterogeneity ($\mathcal{H} = 0$).

To learn about how preference heterogeneity is related to individual characteristics, we regress estimated parameters on individual characteristics. For example, we may expect that the value of time is generally higher for people that have higher incomes. Hence:

(25) $$\hat{\alpha}_{ic} = \kappa z_i + \upsilon_{ic}.$$

where z_i are individual characteristics such as income, age and gender, and κ is a vector of parameters to be estimated. To facilitate interpretation, we assume that the individual characteristics are linearly related to the estimated parameters.

IV. Experimental setup and data

A. Set-up and experiment

We developed a stated choice experiment to collect data about the preferences of morning-commuters participating in a peak-avoidance project. In order to reduce congestion these commuters are rewarded if they do not travel between cameras A and B during the morning peak (6:30-9:30). These cameras are placed on a congested highway.

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. The preferred arrival time of the traveller is given as a reminder. It is based on previous questions in the questionnaire and is 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 has been 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 where made. First, we made explicit what can be important for travellers: departure time, probabilities, travel times, arrival times and rewards. Second, we used a 'vertical' setup for the presentation of unreliability to emphasise that a departure time results in two possible travel times, arrival times and rewards. Third, travel times of the separate parts of the trip are given. This is to show the respondents why they (do not) receive the reward and how the travel time is built

Your preferred arrival time if there is no delay is: 8:40.

	Alternative 1		**Alternative 2**	
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 QUESTION

up from the different components. Fourth, we use bold values for the variables that are (potentially) important and need to be compared by respondents.

The attribute values for travel times are pivoted around the average travel time of the respondent to enhance realism (see Hensher, 2010). Arrivals at Camera A are spread over the whole peak hour to have sufficient variation in arrival times. Several other constraints are put on the design attributes to enhance realism. These are described in detail in Knockaert et al. (2011). The efficiency of the experimental design has been pre-tested using extensive simulation such that the design is able to recover a broad range of parameters.

We excluded respondents who stated that they answered randomly and for whom no observed characteristics were available. For each individual we have information on gross monthly income, level of education, gender, age, household composition (single, children, etc.) and their residential and working location. The summary statistics of these individual characteristics are presented in Table C1 in Appendix C. 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 underrepresented (17 percent). To calculate the delays for each individual, we use door-to-door GPS measurements of Peer et al. (2011).

B. Testing assumptions

It is shown in Section III.B that the delay costs, and in particular scheduling costs, are a function of mean delay. More specifically, it will be assumed that the travel delays are log-normally distributed. Previous empirical evidence of a lognormal distribution of highway travel times is given by Pu (2010). To calculate the delays for each individual, we use GPS measurements of Peer et al. (2011). These measurements have the advantage that they concern door-to-door trips. We have 6,231 observations on trips for 397 individuals. In Figure 2 it is shown that the travel delays seem to be approximately log-normally distributed for our population.

It may be that over the whole population travel delays are not log-normally distributed, while at the individual level they are, for example because people are not located randomly over space. As we are interested in the distribution of the mean delay *per individual*, we also perform a skewness-kurtosis test for each individual for which we have at least eight observations on commuting trips. It appears that for about 70 percent of the individuals, we cannot reject log-normality of travel delays. However, because of the relatively low number of observations per individual the latter result is somewhat suggestive.

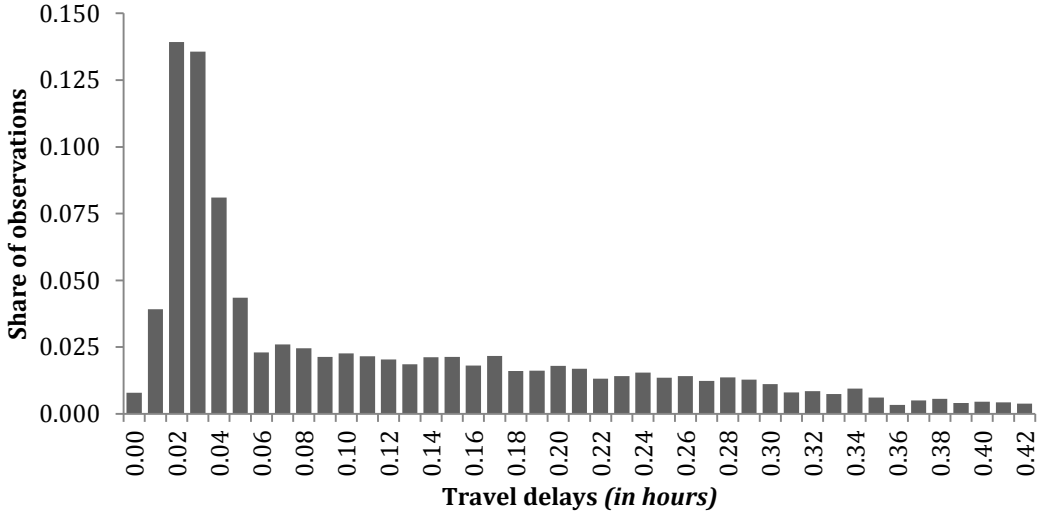

FIGURE 2 — DISTRIBUTION OF TRAVEL DELAYS

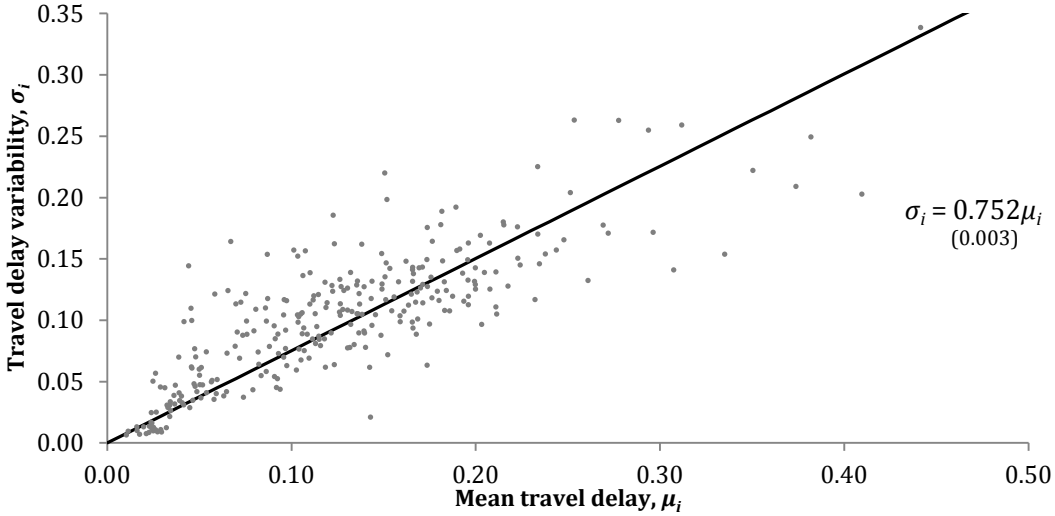

FIGURE 3 — MEAN TRAVEL DELAYS AND TRAVEL DELAY VARIABILITY

We also assume that the standard deviation in travel delays σ_i is linearly related to the mean travel delay μ_i, so $\sigma_i = \eta\mu_i$ resulting in delay costs that are linear in the mean travel delay. We estimate the mean travel delay and the standard deviation of travel delays for each individual for which we have at least eight observations. Figure 3 shows that there is a clear positive and linear relationship between the mean travel delay and the travel delay variability, suggesting that our assumption is reasonable. We find that $\hat{\eta} = 0.752$.

V. Estimation results

A. Baseline results

In this section we discuss the estimation results. We present the marginal effects, so the average value of time $\bar{\alpha}$, the value of schedule delay early $\bar{\beta}$ and the value of schedule delay late $\bar{\gamma}$ and also the parameters that relate to the costs per trip ($\bar{f}$, $\bar{t}$, $\bar{s}$, $\bar{\psi}$ and $\bar{\omega}$). In what follows, we exclude the lower and upper one percent of the estimated values, to avoid the possibility that our results are driven by outliers.[6] Table 2 presents the results.

TABLE 2 — THE ESTIMATED AVERAGE WILLINGNESS TO PAY VALUES AND DELAY COSTS

		(1)	(2)	(3)	(4)
		Logit	Latent Class	Local Logit	Local Latent Class
Panel 1: Marginal effects					
Value of time	$\bar{\alpha}$	35.050	28.373	42.365	31.484
		(–)	(0.856)	(0.444)	(0.582)
Value schedule delay early	$\bar{\beta}$	23.217	22.152	26.543	23.650
		(–)	(0.682)	(0.341)	(0.620)
Value schedule delay late	$\bar{\gamma}$	17.162	18.534	19.339	17.037
		(–)	(1.178)	(0.351)	(0.855)
Panel 2: Costs per trip					
Free-flow travel costs	$\bar{f}$	20.037	15.848	24.196	18.129
		(0.420)	(0.933)	(0.637)	(0.753)
Delay time costs	$\bar{t}$	4.533	3.566	5.660	4.076
		(0.599)	(1.057)	(0.832)	(0.881)
Scheduling costs	$\bar{s}$	1.199	0.98	1.384	1.041
		(0.599)	(1.34)	(0.74)	(1.124)
Share total delay costs ÷ total trip costs	$\bar{\psi}$	0.229	0.256	0.228	0.251
		(0.511)	(0.577)	(0.511)	(0.575)
Share scheduling costs ÷ total delay costs	$\bar{\omega}$	0.209	0.242	0.206	0.238
		(–)	(1.066)	(0.142)	(0.879)
Latent classes	$\mathcal{C}^*$	1	6	1	3
Bandwidth	λ^*	1.0	1.0	0.4	0.4
LL($\mathcal{C}^*, \lambda^*$)		-2,719	-2,422	-2,638	-2,321

Note: We exclude the lower and upper one percent for all reported values. This does not affect the mean estimates. The coefficient of variation (e.g. $\sigma_{\bar{\alpha}}/\bar{\alpha}$) is between parentheses. The share of delay costs in the total trip costs is calculated as $\psi_{ic} = (t_{ic} + s_{ic})/(f_{ic} + t_{ic} + s_{ic})$. The share of expected scheduling costs with respect to total delay costs is calculated as $\omega_{ic} = s_{ic}/(t_{ic} + s_{ic})$.

[6] The results are very similar if we include the lower and upper one percent, although the coefficient of variation is somewhat higher.

We first estimate a standard Binary Logit specification, without heterogeneity (so the coefficient of variation is equal to zero for $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ and $\bar{\omega}$). It is shown that the value of time is € 35 per hour. The value of schedule delay late and early are respectively € 23 and € 17 per hour. These WTP-values are higher than found in the literature (see for example: Brownstone and Small, 2005; Li et al., 2010). There may be 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 average commuters. Second, it is very likely that travellers are less sensitive to rewarding incentives than to the payment of a congestion toll. This difference in valuation of gains and losses is a common finding in prospect theory studies. The pattern $\bar{\beta} > \bar{\gamma}$ is the same for all models and is remarkable, since usually $\bar{\gamma} > \bar{\beta}$ is found in the literature (Lam and Small, 2001; Brownstone and Small, 2005; Li et al., 2010). However, 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, so this result is not too surprising. Another observation is that the average value of time is higher than the value of schedule delay late. An explanation may be that this is due to a selection effect of participants who have lower values of schedule delay. However, our estimated value of schedule delay early is about 65-75 percent of the estimated value of time, which is slightly higher than earlier previous in the literature (Li et al. 2011). If there is a sample selection bias, one would also expect a low value of schedule delay early.

It is shown that the average delay time costs are about € 4.50 per trip and the scheduling costs are about € 1.20 per trip. The overall share of delay costs (time and scheduling) in the total trip costs (including free-flow travel time) is about 23 percent. Scheduling costs are not so important: only 21 percent of the delay costs is attributable to scheduling costs, in line with Fosgerau and Karlström (2010). For the Binary Logit model, the variation in f_i, t_i, s_i and ψ_i is entirely determined by the differences in mean delay.

In Specification (2) we estimate a Latent Class model with six latent classes. In Figure D1, it is shown that for a higher number of latent classes, the results become unstable. Also, the Bayesian Information Criterion is minimised for six latent classes. Not allowing for unobserved heterogeneity seems to lead to an upward bias for the value of time. For example, $\bar{\alpha}$ is about € 7 lower in the Latent Class specification. The average values of schedule delay early and late are very similar. Allowing for unobserved heterogeneity also implies that the share of scheduling costs in the total delay costs ω, is about 5 percentage points higher.

Specification (3) only allows for heterogeneity related to observable individual characteristics. In Figure D2, Appendix D, we show that WTP values become negative and very large when $\lambda < 0.4$.[7] We therefore set the bandwidth $\lambda = 0.4$. It is shown that the average WTP values are similar to Specification (2), except for the value of time: $\bar{\alpha}$ is about € 15 higher. The Likelihood score suggests that Specification (3) is inferior to Specification (2), as the likelihood is higher for the latter Specification. In other words, unobserved heterogeneity seems to be more important than observed heterogeneity.

A preferred specification would allow for both observed and unobserved heterogeneity. In Specification (4) we therefore estimate a Local Latent Class model, which allows for both types of heterogeneity. The bandwidth and number of latent classes are determined by maximising the likelihood, given the economic bandwidth selection criteria (see Table 1).[8] We did some robustness checks with respect to the economic criteria (for example, we adjust ι and impose that $\bar{\alpha} > \bar{\gamma}$), but this hardly impacts the chosen bandwidth and number of latent classes. It is shown that the average willingness to pay values are very similar to Specification (1). The coefficients of variation are in between Specification (2) and (3). Nevertheless, the model fit is better than in Specification (2), suggesting that Specification (2) may have picked up some random variation.

B. Investigating heterogeneity

Although the average estimates are interesting, we are particularly interested in the heterogeneity in the estimated parameters. Table E1 in Appendix E shows the correlations between the estimated parameters for Specifications (3) and (4). For Specification (3) there is a strong positive correlation between $\hat{\alpha}_{ic}$, $\hat{\beta}_{ic}$ and $\hat{\gamma}_{ic}$. This suggests that observable characteristics have a similar impact on the willingness to pay values. For example, individuals with high incomes are likely to have a higher value of time, as well as a higher value of schedule delay early. If we, however, allow for unobserved heterogeneity the correlation becomes essentially zero or even negative, suggesting that unobserved heterogeneity dominates observed heterogeneity. It is also shown that the share of scheduling costs in the total delay costs ($\hat{\omega}_{ic}$) is negatively correlated with the value of schedule delay early and late in Specification (3), but positively correlated in Specification

[7] Also if we exclude outliers for lower bandwidths the results become unrealistic and unstable.

[8] In Appendix D, Figures D3-D6, we analyse the likelihood and value of time for different combinations of λ and $\mathcal{C}$. It is shown that the results become unrealistic when λ is low and $\mathcal{C}$ is large.

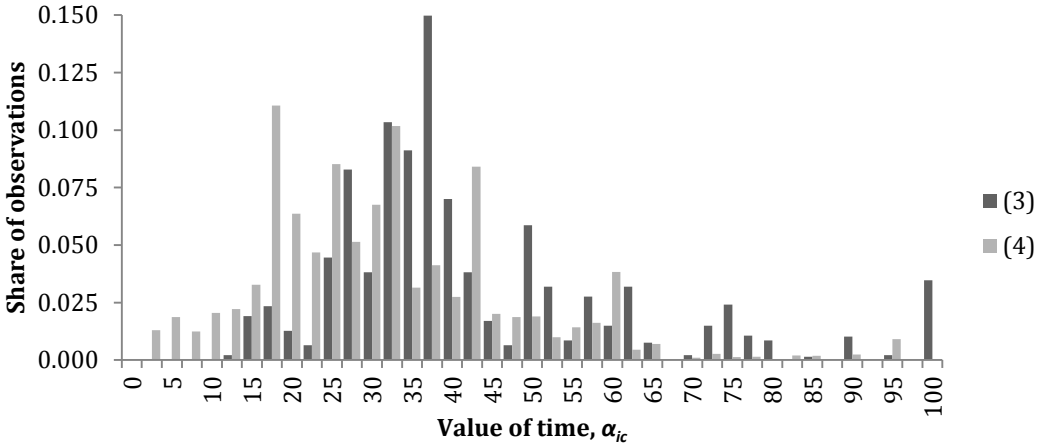

FIGURE 4 — ESTIMATED DISTRIBUTION FOR THE VALUE OF TIME

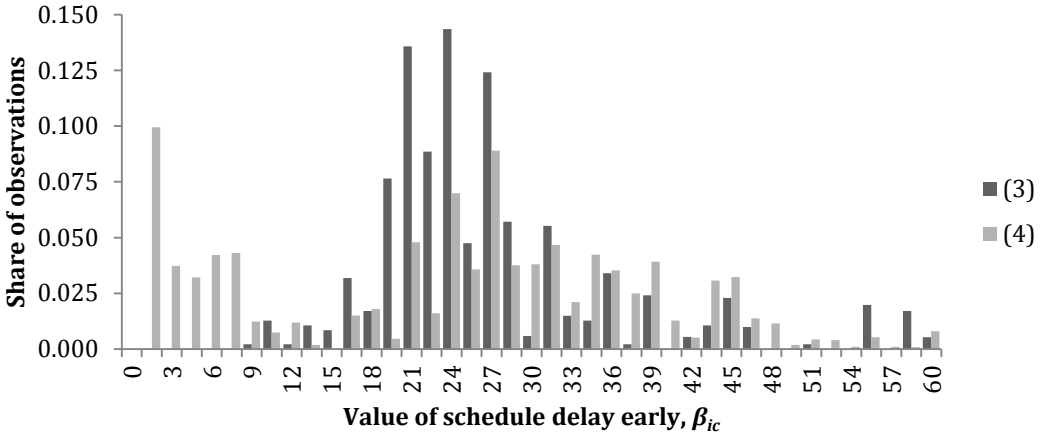

FIGURE 5 — ESTIMATED DISTRIBUTION FOR THE VALUE OF SCHEDULE DELAY EARLY

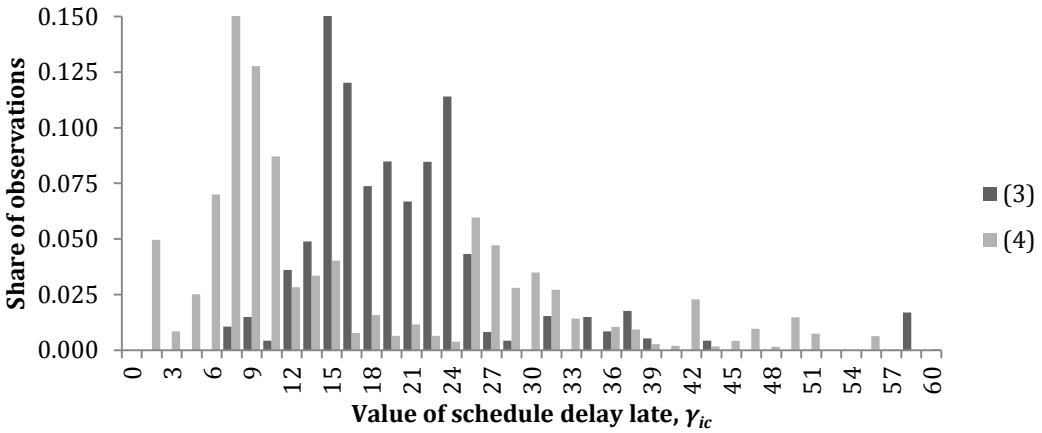

FIGURE 6 — ESTIMATED DISTRIBUTION FOR THE VALUE OF SCHEDULE DELAY LATE

(4). In Specification (3) a higher $\hat{\beta}_i$ or $\hat{\gamma}_i$ implies a relatively higher $\hat{\alpha}_i$ (positive correlation), leading to a lower $\hat{\omega}_i$. In Specification (4) there is hardly correlation between $\hat{\alpha}_{ic}$, $\hat{\beta}_{ic}$ and $\hat{\gamma}_{ic}$, so a higher $\hat{\beta}_{ic}$ or $\hat{\gamma}_{ic}$ implies a higher $\hat{\omega}_{ic}$ (see equation (18)).

Figures 4-6 present the estimated nonparametric distributions for the values of time, schedule delay early and late for Specifications (3) and (4).[9] It is shown that these distributions seem not to follow any conventional distribution and do not have similar shapes for the different WTP-values. For example, the distribution of $\hat{\alpha}_{ic}$ is more or less unimodal, while the distributions for $\hat{\beta}_{ic}$ and $\hat{\gamma}_{ic}$ are bimodal. Figure 1 shows that there are some extreme values for the value of time (> € 100) in Specification (3). One may argue that these values may explain the high average value of time in Table 1. However, if we exclude value of times higher than € 90, the average value of time is almost identical ($\bar{\alpha} = 41.64$). For Specification (4), about 90 percent of the observations have a value of time that is lower than € 50 per hour, which seems reasonable. Similarly, about 90 percent of the observations have a value of schedule delay early and late that is lower than € 40.

Figure 6 presents the relationship between the Local Logit and Local Latent Class specification. As expected, there is a positive correlation ($\rho = 0.33$) between the value of time obtained in Specification (3) and Specification (4), but due to the importance of unobserved heterogeneity, this relationship is certainly not one-to-one.

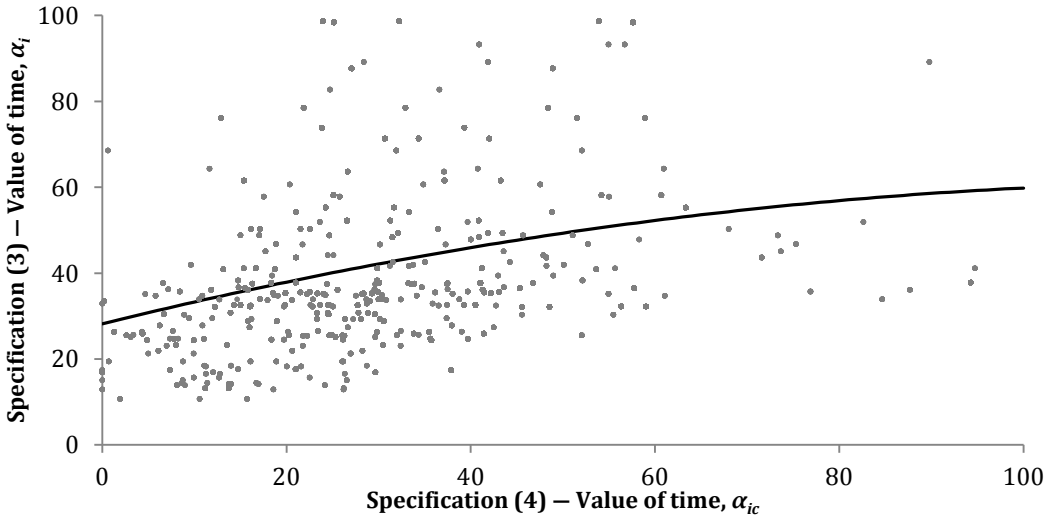

FIGURE 6 – VALUE OF TIME FOR SPECIFICATIONS (3) AND (4)

[9] Distributions for other estimated parameters are presented in Appendix E.

C. Explaining heterogeneity

Based on these results, it seems that unobserved heterogeneity is the most important source of variation in the estimated parameters. To test this, we determine the portion of heterogeneity that is attributable to observable characteristics by a regression of the estimated parameters on individual fixed effects. Further, we regress the estimated parameters on individual characteristics to explain heterogeneity. Tables 3 and 4 present the second stage results based on the Local Latent Class specification.[10]

TABLE 3 – ANALYSING HETEROGENEITY, WEIGHTED REGRESSIONS
(Dependent variables: $\hat{\alpha}_{ic}$, $\hat{\beta}_{ic}$ and $\hat{\gamma}_{ic}$)

	(5)	(6)	(7)
	$\hat{\alpha}_{ic}$	$\hat{\beta}_{ic}$	$\hat{\gamma}_{ic}$
Panel 1: Observed vs. unobserved heterogeneity			
Observed heterogeneity, $\mathcal{H}$	0.233	0.094	0.065
Number of individuals	487	487	487
Panel 2: Analysing observed heterogeneity			
Income €2500-€3500	1.143 (1.114)	3.340 (0.939)***	3.796 (0.650)***
Income €3500-€5000	10.204 (1.172)***	5.517 (0.963)***	4.307 (0.698)***
Income >€5000	8.930 (1.330)***	4.742 (1.036)***	6.565 (0.760)***
Education – vocational	5.701 (0.995)***	1.730 (0.787)**	-0.862 (0.715)
Education – bachelor degree	8.320 (0.850)***	2.552 (0.642)***	-0.470 (0.692)
Female	1.868 (0.716)***	1.281 (0.424)***	2.330 (0.366)***
Single	-2.770 (0.687)***	-3.559 (0.453)***	-1.749 (0.464)***
No children	-2.649 (0.656)***	1.329 (0.341)***	-2.458 (0.385)***
Young children	8.025 (0.911)***	5.246 (0.466)***	1.389 (0.353)***
Children at primary school	-3.390 (0.771)***	-1.947 (0.419)***	-3.825 (0.346)***
Age 25-50	1.335 (0.724)*	-0.399 (0.582)	-2.506 (0.576)***
Age>50	-3.631 (0.865)***	0.324 (0.617)	-0.169 (0.621)
Constant	17.527 (1.466)***	16.529 (1.042)***	16.071 (0.941)***
$\mathcal{R}^2/\mathcal{H}$	0.635	0.587	0.466
Number of individuals	487	487	487

Note: The reference category is a married (or living together) male person younger than 25 that has children that are all older than 12 years and a monthly income higher than €2500 and an educational degree lower than vocational. $\mathcal{R}^2/\mathcal{H}$ is the share of observed heterogeneity that is explained by a linear relationship of individual characteristics z. The standard errors are clustered at the individual level and are between parentheses.

*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

[10] We also have estimated seemingly unrelated regressions (SUR), because errors v_{ic} are likely to be correlated across regressions. It is well-known that SUR only improves efficiency and not the consistency of the estimated coefficient. As it is not possible to cluster standard errors in SUR, we prefer the results from standard Ordinary Least Squares.

About 23 percent of heterogeneity is related to observed characteristics for the value of time. This is even lower for the value of schedule delay early and late (respectively 9.4 and 6.5 percent). This strongly suggests that unobserved heterogeneity is more important than observed heterogeneity. One may argue that we do not have information on all relevant individual characteristics, so that this is an underestimate. However, we have data on at least the most important observable demographic characteristics of individuals (income, education, etc.). Including less important variables is therefore unlikely to lead to very different conclusions. In contrast, we think that this is more likely to have an overestimate of the importance of observed heterogeneity, as we are not able to allow for unobserved heterogeneity in the reward parameter δ_i^r. We also perform a regression of the estimated willingness to pay values on individual characteristics, enabling us to investigate how individual characteristics relate to the WTP-values.[11] For highly educated people the value of time is more than € 8 higher compared to low educated people. The effect of education on the value of time is much stronger than on the value schedule delay early and the income effect is strong: switching from an income lower than € 2,500, to the highest income class increases the value of time with € 10, which is about 36 percent higher. Also, the increase of value schedule delay early and late is substantial and about € 5 (24 percent). Education has a positive effect on WTP-values even if we control for income effects. For higher educated people the value of time is more than € 8 higher compared to less educated people. This implies that high-income and highly-educated households have a willingness to pay that is about 10 percent higher than the average. The effect of education on the value of time is much stronger than on the value schedule delay early and late, confirming that highly educated people tend to have more flexible job starting times (Arnott et al., 1990; Golden, 2001). Females have higher WTP-values than males (about € 1.50). This is a common finding in the literature, which could reflect that females are more often responsible for the children in the household, and therefore have more scheduling constraints (Kwan, 1999; Lam and Small, 2001; Brownstone and Small, 2005; Schwanen, 2008).[12] Travellers that are single have lower WTP-values, especially for the value of schedule delay early and late. Previous research already showed that time budgets decrease because of the presence of children in the household (Becker, 1985; Browning, 1992). Indeed, having young children

[11] We also estimated similar models where we take the logarithm of the WTP-values, but the results are qualitatively similar.

[12] If we repeat the analysis for Specification (3), it may be shown that almost all effects are much stronger. Hence, ignoring unobserved heterogeneity leads to an overestimate of the impact of individual characteristics on the willingness to pay values.

increases the value of time, value of schedule delay early and late with respectively € 8.03, € 5.25 and € 1.39. The effect is stronger for the value of schedule delay early, reflecting that commuters with young children face more scheduling cost in the early morning. The results for people with children at primary school is surprising, as we would expect a positive coefficient, given the fact that scheduling constraints are more stringent because of the fixed starting times of schools (Schwanen and Ettema, 2009). It may be due to the fact that people start working full-time again when their children go to secondary school, implying *stronger* scheduling constraints.

We find that the average willingness to pay values are higher compared to the previous literature. It has been argued that this is partly due to the overrepresentation of highly educated and high income households. Using the estimates of Table 3 we may weigh the average value of time, schedule delay early and schedule delay late based on demographic characteristics of the Dutch population (See Table C1). In Table E2, Appendix E the results are presented. It is shown that the value of time is about 35 percent lower (€ 22.68). Because the effects of income and education in particular are less strong for the value of schedule delay early and late, the population-weighted average values are similar to the sample estimates.

Table 4 analyses parameters related to the costs of the trip, given our assumptions in Section II. It is first shown that the share of heterogeneity explained by observable features is generally higher. As $\hat{t}_{ic}$, $\hat{s}_{ic}$, $\widehat{\psi}_{ic}$ and $\widehat{\omega}_{ic}$ are functions of the mean travel delay μ_i (which is only different across individuals and not across latent classes), this is not too surprising. For $\hat{t}_{ic}$ and $\hat{s}_{ic}$ we again find a statistically significant income and education effect. For example, switching from an income lower than € 2,500 to the highest income class increases the delay costs per trip with € 1.323 and increases the scheduling costs per trip with € 0.149.

For the regression of $\widehat{\psi}_{ic}$ on individual characteristics, we find that the impact of individual characteristics is limited and is mainly explained by differences in the mean travel delay. Nevertheless, females tend to experience a higher share of total delay costs. There is also some evidence that high incomes have relatively lower delay costs, likely because income has a relatively stronger effect on $\hat{\alpha}_{ic}$ than on $\hat{\beta}_{ic}$ and $\hat{\gamma}_{ic}$, so that the free-flow travel costs become relatively more important. The regression of $\widehat{\omega}_{ic}$ on individual characteristics shows that singles have relatively lower scheduling costs. A plausible explanation is that there are no other people in the household that impose scheduling constraints on them. Also highly educated people experience relatively lower scheduling costs, again, because of flexible work-start hours. Older people experience relatively higher

TABLE 4 – ANALYSING HETEROGENEITY, WEIGHTED REGRESSIONS BASED ON SPECIFICATION (4)
(Dependent variables: $\hat{t}_{ic}$, $\hat{s}_{ic}$, $\hat{\psi}_{ic}$ and $\hat{\omega}_{ic}$)

	(5)	(6)	(7)	(8)
	$\hat{t}_{ic}$	$\hat{s}_{ic}$	$\hat{\psi}_{ic}$	$\hat{\omega}_{ic}$
Panel 1: Observed vs. unobserved heterogeneity				
Observed heterogeneity, $\mathcal{H}$	0.622	0.307	0.757	0.049
Number of individuals	397	397	397	397
Panel 2: Analysing observed heterogeneity				
Income €2500-€3500	0.080 (0.467)	0.167 (0.108)	-0.019 (0.030)	0.047 (0.012)***
Income €3500-€5000	1.217 (0.583)**	0.270 (0.123)**	-0.054 (0.032)*	0.005 (0.012)
Income >€5000	1.323 (0.595)**	0.311 (0.130)**	-0.056 (0.032)*	0.013 (0.013)
Education – vocational	0.988 (0.481)**	-0.006 (0.127)	0.020 (0.036)	-0.042 (0.011)***
Education – bachelor degree	1.852 (0.402)***	0.222 (0.107)**	0.036 (0.031)	-0.044 (0.010)***
Female	0.887 (0.351)**	0.150 (0.081)*	0.052 (0.016)***	0.013 (0.004)***
Single	-0.338 (0.519)	-0.025 (0.110)	-0.039 (0.023)*	-0.046 (0.007)***
No children	-0.511 (0.342)	0.061 (0.086)	-0.015 (0.017)	0.010 (0.004)**
Young children	1.554 (0.455)***	0.371 (0.105)***	0.001 (0.018)	-0.027 (0.004)***
Children at primary school	-0.722 (0.381)*	-0.150 (0.092)	-0.003 (0.018)	0.033 (0.005)***
Age 25-50	-0.585 (0.430)	-0.324 (0.110)***	-0.027 (0.021)	0.017 (0.007)**
Age>50	-1.392 (0.470)***	-0.319 (0.127)**	-0.011 (0.025)	0.036 (0.008)***
Constant	2.224 (0.601)***	0.840 (0.162)***	0.275 (0.039)***	0.236 (0.015)***
$\mathcal{R}^2/\mathcal{H}$	0.198	0.140	0.058	0.516
Number of observations	397	397	397	397

Note: See Table 3.

scheduling costs, possibly because older people have more job responsibilities and therefore have tighter schedules.

VI. Conclusions

In this paper we first derive an easy to apply closed-form solution for the costs of travel time variability assuming that travellers choose their departure time optimally and travel times follow a log-normal distribution, leading to an easy-to-implement cost function that is only a function of the mean delay. We also allow for preference heterogeneity using semiparametric estimation techniques, which is of key importance when evaluating the benefits of infrastructure investments.

We use data from a stated choice experiment held among participants of a real-world rewarding experiment to combat traffic congestion. A semiparametric estimation approach is used to analyse observed and unobserved heterogeneity in the value of travel time, schedule delay early and late. We also estimate the delay costs and the scheduling costs. It is shown that there is substantial heterogeneity in the willingness to pay for travel time. For

example, high-income and highly-educated households have a willingness to pay that is about 10 percent higher. 5-25 percent of the heterogeneity in the value of time is attributable to observed characteristics of individuals, implying that unobserved heterogeneity is much more important than heterogeneity related to observable characteristics. It is furthermore shown that delay costs are important. About 25 percent of the trip costs (ignoring fuel costs and vehicle costs) is attributable to delay costs. A large share of these delay costs is attributable to additional travel time, as scheduling costs are in our data only 24 percent of the total costs of travel delays.

References

Arnott, R., de Palma,A., Lindsey, R., (1988). Schedule delay and departure time decisions with heterogeneous commuters. *Transportation Research Record* 1197: 56-67.

Arnott, R., de Palma,A., Lindsey, R., (1994). The Welfare Effects of Congestion Tolls with Heterogeneous Commuters. *Journal of Transport Economics and Policy* 28 (2): 983-992.

Bates, J., Polak, J., Jones, P., Cook, A. (2001). The Valuation of Reliability. *Transportation Research E 37(2-3): 191-229.*

Bajari, P., Benkard, C.M. (2005). Demand Estimation with Heterogeneous consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy 113(6): 1239-1276.*

Bajari, P., Kahn, A.E. (2005). Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy 113 (6): 1239-1276.*

Becker, G.S. (1985). Human Capital, Effort, and the Sexual Division of Labour. *Journal of Labour Economics 3 (1), 533-558.*

Bontemps, C., Simioni, M., Surry, Y. (2008). Semiparametric Hedonic Price Models: Assessing the Effects of Agricultural Nonpoint Source Pollution. *Journal of Applied Econometrics 23 (6): 825-842.*

Browning, M. (1992). Children and Household Economic Behaviour. *Journal of Economic Literature 30(3): 1434-1474.*

Brownstone, D., Train, K. (1998). Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics 89 (1-2): 109-129.*

Brownstone, D., Small K.A. (2005). Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A 39 (4): 279-293.*

Carrion, C., Levinson, D. (2012). Value of Travel Time Reliability: Review of Current Evidence. *Transportation Research Part A 46 (4): 720-741*.

De Lapparent, M., De Palma, A., Fontan, C. (2002). Nonlinearities in the Valuation of Time Estimates. *Proceedings of the PTRC Annual Meeting 2002.*

Fan, J., Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. *In: Monographs and Statistics and Applied Probability 66. Chapman and Hall, England*.

Fan, J., Heckman, N.E., Wand, M.P. (1995). Local Polynomial Regression for Generalized Linear Models and Quasi-Likelihood Functions. *Journal of the American Statistical Association 90 (429): 141-150.*

Fan, J., Farmen, M., Gijbels, I. (1998). Local Maximum Likelihood Estimation and Inference. *Journal of the Royal Statistical Society B 60 (3): 591-608.*

Fosgerau, M. (2007). Using Nonparametrics to Specify a Model to Measure the Value of Time. *Transportation Research A 41 (9): 842-856.*

Fosgerau, M., Karlström, A. (2010). The Value of Reliability. *Transportation Research Part B 44(1): 38-49.*

Fosgerau, M., Engelson, L. (2011). The Value of Travel Time Variance. *Transportation Research Part B 45(1): 1-8.*

Fukuda, D., Yai, T. (2010). Semiparametric Specification of the Utility Function in a Travel Mode Choice Model. *Transportation 37 (2): 221-238.*

Gaver, D.P. (1968). Headstart Strategies to Combat Congestion. *Transportation Science* 2 (2), 172-181.

Golden, L. (2001). Flexible Work Schedules. Which Workers Get Them? *American Behavioural Scientist 44(7): 1157-1178.*

Greene, W.H. (2003). Econometric Analysis (5th edition). *New York: McMillan*.

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

Harding, M.C., Hausman, J. (2007). Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Nonlinear Least Squares. *International Economic Review 48 (4): 1311-1328.*

Hastie, T., Tibshirani, R. (1987). Local Likelihood Estimation. *Journal of the American Statistical Association 82 (398), 559-567*.

Hensher, D.A. (2010). Hypothetical Bias, Choice Experiments and Willingness to Pay. *Transportation Research B 44 (6): 735-752.*

Horowitz, J.L., Härdle, W. (1996). Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates. *Journal of the American Statistical Association 91 (436): 1632-1640*.

Horowitz, J.L. Savin, N.E. (2001). Binary Response Models: Logits, Probits and Semiparametrics. *Journal of Economic Perspectives 15 (4): 43-56.*

Huang, H.-J (2000). Fares and Tolls in a Competitive System with Transit and Highway: The Case with Two Groups of Commuters. *Transportation Research E 36(4): 267-284.*

Knight, T.E. (1974). An Approach To The Evaluation of Changes in Travel Time Unreliability: A 'Safety Margin' Hypothesis. *Transportation* 3, 393-408.

Knockaert, J., Peer, S., Tseng, Y.-Y, Verhoef, E.T., Koster, P.R. (2011). Spitsmijden IID Scientific Report. *Forthcoming at www.spitsmijden.nl.*

Koster, P.R., Tseng, Y.-Y. (2010). Stated Choice Experimental Designs for Scheduling Models. *In: Hess, S., Daly, A. (eds.) Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference. Bingley, UK: Emerald.*

Koster, P.R., Verhoef, E.T. (2012). A Rank-dependent Scheduling Model. *Journal of Transport Economics and Policy 46 (1):123-138.*

Kwan, M.-P. (1999). Gender, the Home–Work Link and Space–Time Patterns of Nonemployment Activities. *Economic Geography 76 (4): 370–394.*

Lam, T.C., Small, K.A. (2001). The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E 37 (2-3): 231-251.*

Li, Z., Hensher, D.A., Rose, J. (2010). Willingness to Pay for Travel Time Reliability in Passenger Transport: A Review and Some New Empirical Evidence. *Transportation Research E 46 (3): 384-403.*

Lindsey, R. (2004). The Welfare-Distributional Impacts of Congestion Pricing of a Road Network. *In: Proceedings of the 39th Annual Conference of the Canadian Transportation Research Forum: 2004 Transportation Revolutions, Calgary, Canada 9-12 May 2004, 149-163.*

Mackie, P.J. Jara-Diaz, S. Fowkes, A.S (2001). The Value of Travel Time Savings in Evaluation. *Transportation Research E 37(2-3): 91-106*

McFadden, D.L., Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics 15 (5): 447-470.*

McMillen, D.P., Redfearn, C.L. (2010). Estimation and Hypothesis Testing for Nonparametric Hedonic Price Functions. *Journal of Regional Science 50 (3): 712-733.*

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

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

Peer, S., Koopmans, C.C., Verhoef, E.T. (2012). Prediction of travel time variability for cost-benefit analysis. *Transportation Research Part A*, *46 (1): 79-90*.

Peer, S., Knockaert, J., Koster, P.R., Tseng, Y.-Y, Verhoef, E.T. (2011). Door-to-door Travel Times in RP Departure Time Choice Models: An Approximation Method Based On GPS Data. *Tinbergen Discussion Paper 11-180/3*.

Pu, W. (2010). Analytic Relationships Between Travel Time Reliability Measures. *Transportation Research Record 2254: 122-130*.

Racine, J. S.,Li, Q. (2004). Nonparametric Estimation of Regression Functions with Both Categorical and Continuous data. *Journal of Econometrics 119 (1): 99-130*.

Revelt, D., Train, K. (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics 80 (4): 647-657.*

Schwanen, T. (2008). Struggling with time: Investigating Coupling constraints*. Transport Reviews 28 (3): 337-356.*

Schwanen, T., Ettema, D. (2009). Coping with Unreliability when Collecting Children: Examining Parents' Behaviour with Cumulative Prospect Theory*. Transportation Research A 43 (5), 511-525.*

Sherman, R.P. (1993). The Limiting Distribution of the Maximum Rank Correlation Estimator. *Econometrica 61 (1): 123-138.*

Small, K.A. (1982). The Scheduling of Consumer Activities: Work Trips. *American Economic Review 72 (3): 469-479.*

Small, K.A. (2012). Valuation of Travel Time. *Economics of Transportation 1 (1-2): 2-14.*

Small, K.A, Winston, C., Yan, J. (2005). Uncovering the Distribution of Motorists' Preferences of Travel Time and Travel Time Reliability. *Econometrica 73 (4): 1367-1382.*

Tan, Z.J., Yang, H. (2012). The Impact of User Heterogeneity on Road Franchising. *Transportation Research Part E* 48 (5): 958-975.

Van den Berg, V., Verhoef, E.T. (2011). Winning or Losing from Dynamic Bottleneck Congestion Pricing? The Distributional Effects of Road Pricing with Heterogeneous Values of Time and Schedule Delays. *Journal of Public Economics 95 (7-8): 983-992.*

Van den Berg, V., Verhoef, E.T. (2013). Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. *Forthcoming in Transportmetrica A.*

Verhoef, E.T., Small, K.A. (2004). Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users. *Journal of Transport Economics and Policy* (38), 127-156.

Vickrey, W.S. (1969). Congestion Theory and Transport Investment. *American Economic Review 59 (2): 251-261.*

Wardman, M. (1998). The Value of Travel Time. *Journal of Transport Economics and Policy 32 (3): 285-316.*

Winston, C., Yan, J. (2011). Can Privatization of U.S. Highways Improve Motorists' Welfare? *Journal of Public Economics* 95 (7-8): 993-1005.

Appendix A. Deriving the expected delay and scheduling costs

The first part of this Appendix shows the derivation of equation (14). The second part focuses on the derivation of equation (15).

Derivation of equation (14). Following Noland and Small (1995), the expected delay costs $d[h_i]$ are given by the sum of the costs of travel time and the costs of expected schedule delay early and late:

$$d[h_i] = t_i + s_i[h_i], \tag{A1}$$

where $t_i = \alpha_i \mu_i$ are the costs of delay travel time. The expected scheduling costs $d[h_i]$ are given by:

$$d[h_i] = \alpha_i \mu_i + \beta_i \int_0^{h_i} (h_i - T) F'[T] dT + \gamma_i \int_{h_i}^{\infty} (T - h_i) F'[T] \mathrm{d}T.$$

This leads to:

$$d[h_i] = \alpha_i \mu_i + \beta_i \left(h_i F[h_i] - \int_0^{h_i} T F'[T] dT \right) + \gamma_i \left(-h_i (1 - F[h_i]) + \mu_i - \int_0^{h_i} T F'[T] dT \right).$$

Using integration by parts this reduces to:

$$d[h_i] = (\alpha_i + \gamma_i)\mu_i - \gamma_i h_i + (\beta_i + \gamma_i)\int_0^{h_i} F[T]\mathrm{d}T.$$

Assumption A implies that travellers optimise their headstart h_i. The first order condition is given by:

$$\frac{\partial d[h_i]}{\partial h_i} = -\gamma_i + (\beta_i + \gamma_i)F[h_i]. \tag{A2}$$

Solving (A2) for h_i gives the optimal headstart:

$$h_i^* = F^{-1}\left[\frac{\gamma_i}{\beta_i + \gamma_i}\right], \tag{A3}$$

where F^{-1} is the inverse of the cumulative density function. An increase in β_i will lead to a lower optimal headstart whereas an increase in γ_i will increase the optimal headstart. Substituting the optimal headstart in the expected cost function yields:

$$\begin{aligned} d[h_i^*] &= (\alpha_i + \gamma_i)\mu_i - \gamma_i h_i^* + (\beta_i + \gamma_i)\int_0^{h_i^*} F[T]\mathrm{d}T, \\ &= (\alpha_i + \gamma_i)\mu_i - \gamma_i h_i^* + (\beta_i + \gamma_i)\left(h_i^* \cdot F[h_i^*] - \int_0^{h_i^*} T \cdot F'[T]\mathrm{d}T\right). \end{aligned}$$

Using $F[h_i^*] = \gamma_i/(\beta_i + \gamma_i)$ this reduces to:

$$\begin{aligned} d[h_i^*] &= (\alpha_i + \gamma_i)\mu_i - (\beta_i + \gamma_i)\int_0^{h_i^*} T \cdot F'[T]\mathrm{d}T, \\ &= (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\int_{h_i^*}^{\infty} F^{-1}[F[T]] \cdot F'[T]\mathrm{d}T. \end{aligned}$$

Using integration by substitution we find:

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\int_{\frac{\gamma_i}{\beta_i+\gamma_i}}^{1} F^{-1}[k]\mathrm{d}k. \tag{A4}$$

which is equal to equation (14). This expression was earlier derived by Fosgerau and Kärlstrom (2010), but then for a standardised distribution of travel times.

Proof for Proposition 1. Now we have to proof that equation (A4) simplifies to equation (15) if we are willing to assume that travel delays follow a log-normal distribution with shape parameter τ_i and scale parameter θ_i (Assumption B). Let's define $\mathrm{erf}[b] = \frac{2}{\sqrt{\pi}}\int_0^b \mathrm{e}^{-\nu^2}\mathrm{d}\nu$ as the error function and $\mathrm{erfc}[b] = \frac{2}{\sqrt{\pi}}\int_b^\infty \mathrm{e}^{-\nu^2}\mathrm{d}\nu$ as the complementary error function. T follows a log-normal distribution with $F[T] = \frac{1}{2}\mathrm{erfc}\left[\frac{\tau - \ln[T]}{\theta_i\sqrt{2}}\right]$ and $\mu_i = \mathrm{e}^{\tau_i + \theta_i^2/2}$. The optimal expected costs are then given by:

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\int_{\frac{\gamma_i}{\beta_i+\gamma_i}}^{1} \mathrm{e}^{\tau_i - \sqrt{2}\theta_i \mathrm{erfc}^{-1}[2k]}\mathrm{d}k,$$

where $\mathrm{erfc}^{-1}[b]$ is the inverse of the complementary error function. This reduces to:

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\mu_i \Phi\left[\theta_i + \sqrt{2}\cdot \mathrm{erfc}^{-1}\left[\frac{2\gamma_i}{\beta_i + \gamma_i}\right]\right],$$

where $\Phi[\,\cdot\,]$ is the cumulative standard normal distribution. Next we use $\sqrt{2}\cdot \mathrm{erfc}^{(-1)}[1 - b] = \sqrt{2}\mathrm{erf}^{-1}[b]$ to obtain:

$$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\mu_i \Phi\left[\theta_i + \sqrt{2}\cdot \mathrm{erf}^{-1}\left[\frac{\beta_i - \gamma_i}{\beta_i + \gamma_i}\right]\right],$$

Finally we use $\Phi^{-1}[b] = \sqrt{2}\mathrm{erf}^{-1}[2b - 1]$, to obtain:

(A5) $$d[h_i^*] = (\alpha_i - \beta_i)\mu_i + (\beta_i + \gamma_i)\mu_i \Phi\left[\theta_i + \Phi^{(-1)}\left[\frac{\beta_i}{\beta_i + \gamma_i}\right]\right],$$

which is equal to equation (15). ■

Given estimates on α_i, β_i, γ_i and θ_i and observations on $\mu_i\mu_i$ we can calculate $d[h_i^*]$.

Appendix B. Deriving θ_i using the mean delay μ_i

For a given mean travel time the standard deviation can be approximated by:

(B1) $$\sigma_i = \eta\mu_i$$

The inverse coefficient of variation is given by:

(B2) $$\frac{\sigma_i}{\mu_i} = \eta$$

If travel times follow a log-normal distribution this should be equal to:

(B3) $$\frac{\sigma_i}{\mu_i} = \sqrt{e^{\theta_i^2} - 1}$$

Solving (B2) and (B3) for θ_i gives:

(B4) $$\theta_i = \sqrt{\ln[1 + \eta^2]}$$

This means that θ_i can be calibrated if η is known. The coefficient η depends on the information the travellers have. We use $\hat{\eta} = 0.752$ (See Section IV.B). A sensitivity analysis for different values of η is given in Appendix F.

Appendix C. Descriptive statistics

TABLE C1 — DESCRIPTIVES OF THE INDIVIDUAL CHARACTERISTICS

	Sample, Mean	Population, Mean
Income <€2500	0.066	0.380
Income €2500-€3500	0.340	0.366
Income €3500-€5000	0.401	0.121
Income >€5000	0.193	0.134
Education – Primary or Secondary	0.043	0.353
Education – Vocational	0.117	0.324
Education – Bachelor Degree or higher	0.840	0.323
Female	0.237	0.505
Single	0.165	0.364
No Children	0.430	0.657
Young Children(<5 years)	0.212	0.110
Children at Primary School	0.263	0.079
Children at Secondary School (>12 years)	0.095	0.154
Age<25	0.014	0.111
Age 25-50	0.761	0.441
Age>50	0.224	0.449
Number of Individuals	487	13,039,488

Note: The population data is from 2010 and for people older than 18 years obtained from Statistics Netherlands. Mean delay for the population is obtained from KIM (2009).

Appendix D. Bandwidth selection

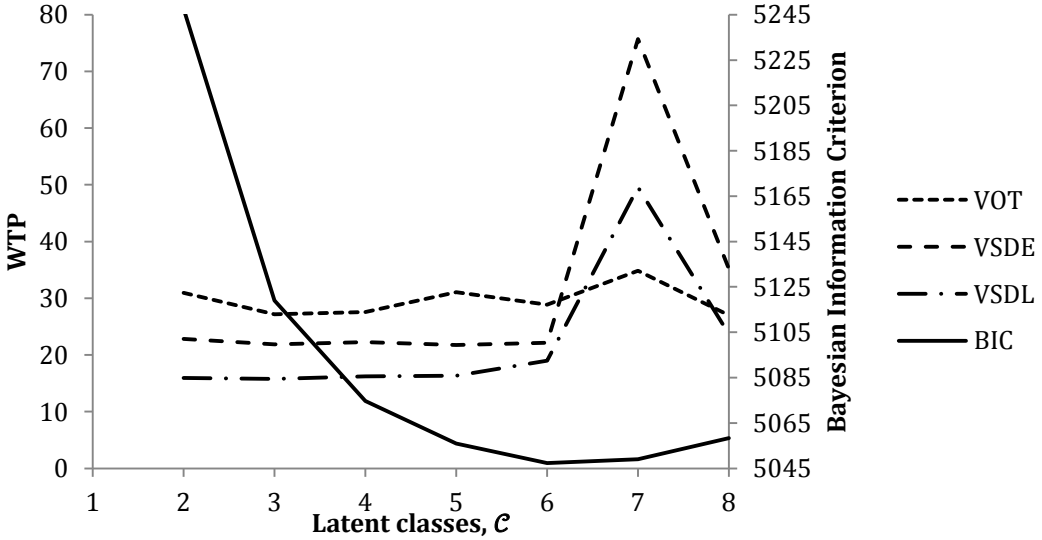

FIGURE D1 — THE CHOICE OF $\mathcal{C}$ FOR SPECIFICATION (2)

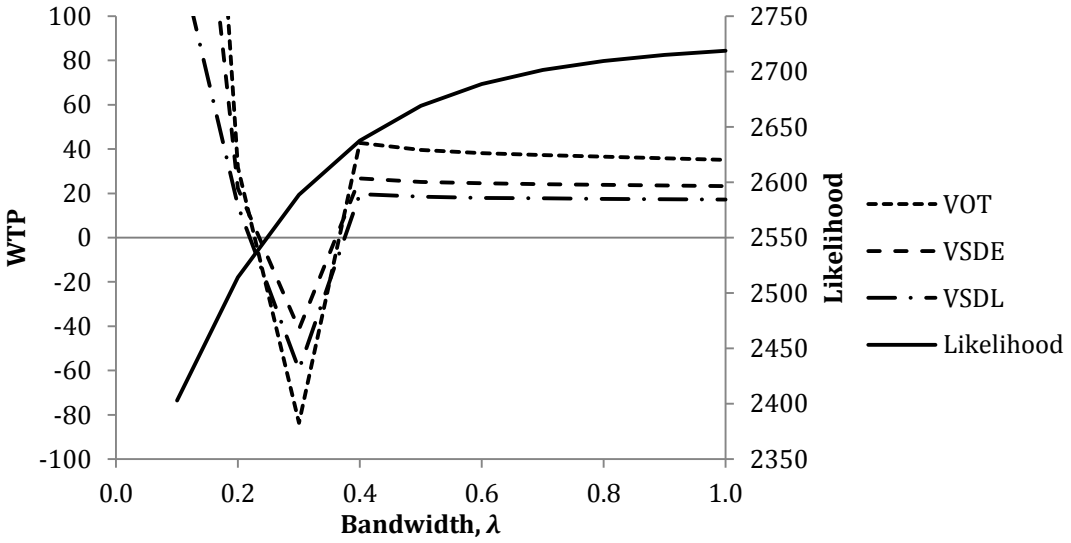

FIGURE D2 — THE CHOICE OF λ FOR SPECIFICATION (3)

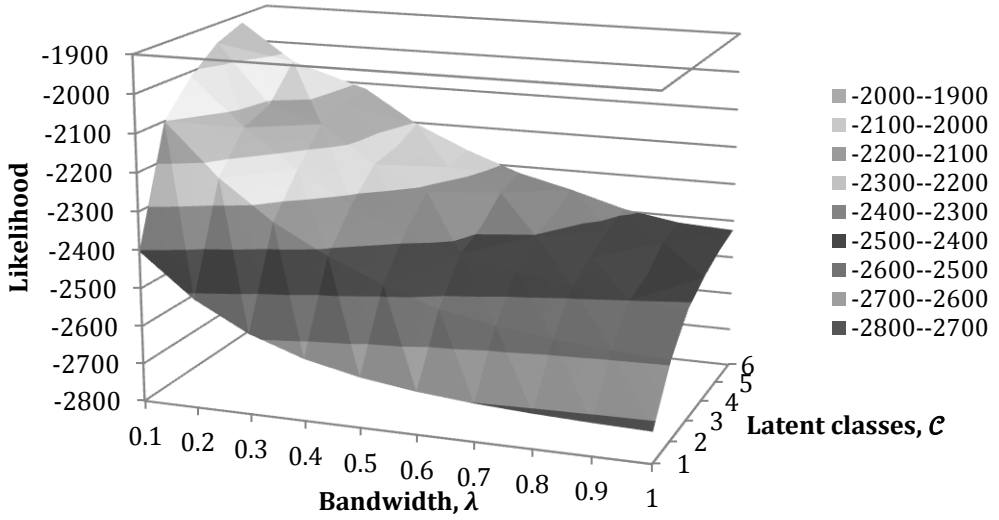

FIGURE D3 — LIKELIHOOD GIVEN λ AND C FOR SPECIFICATION (4)

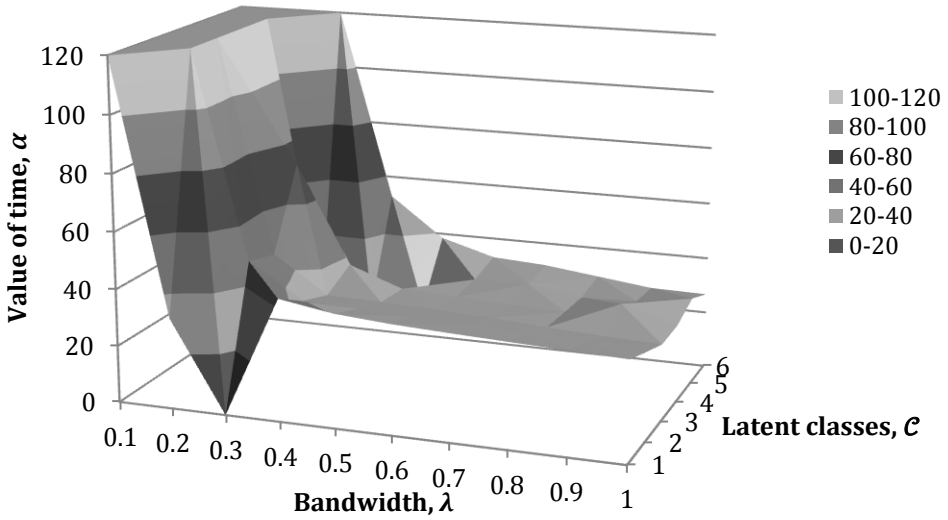

FIGURE D4 — THE VALUE OF TIME GIVEN λ AND $\mathcal{C}$ FOR SPECIFICATION (4)

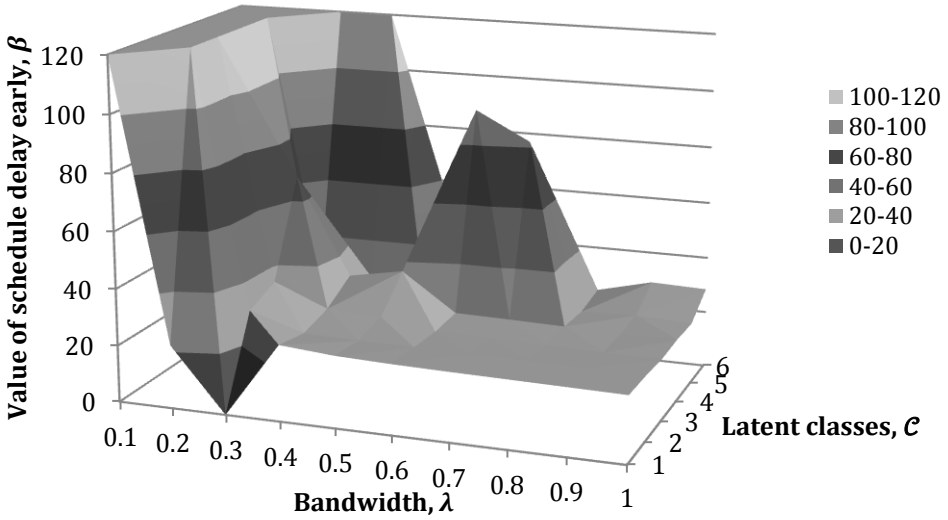

FIGURE D5 — THE VALUE OF SCHEDULE DELAY EARLY GIVEN λ AND $\mathcal{C}$ FOR SPECIFICATION (4)

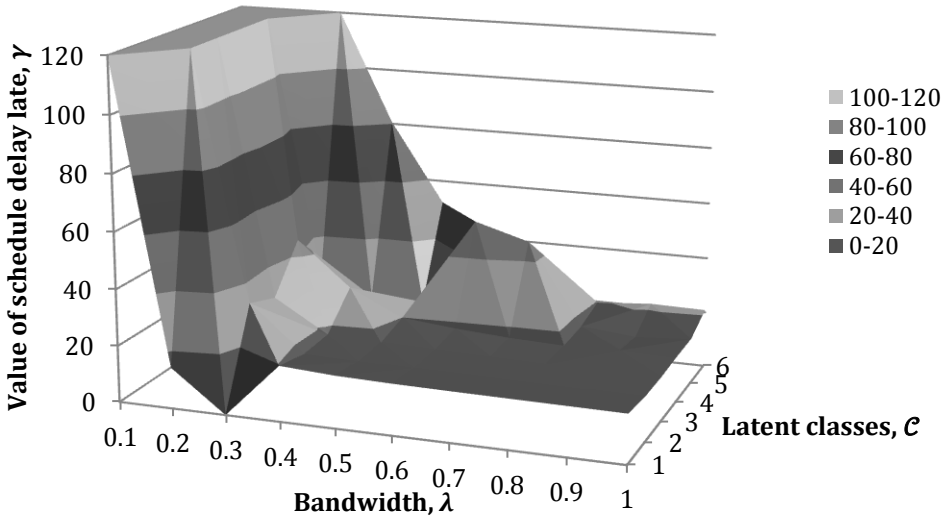

FIGURE D6 — THE VALUE OF SCHEDULE DELAY LATE GIVEN λ AND $\mathcal{C}$ FOR SPECIFICATION (4)

Appendix E. Other results

TABLE E1 — CORRELATION BETWEEN ESTIMATED PARAMETERS

	$\hat{\alpha}_{ic}$		$\hat{\beta}_{ic}$		$\hat{\gamma}_{ic}$		$\hat{f}_{ic}$		$\hat{t}_{ic}$		$\hat{s}_{ic}$		$\hat{\psi}_{ic}$		$\hat{\omega}_{ic}$	
	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)
$\hat{\alpha}_{ic}$	1.000	1.000														
$\hat{\beta}_{ic}$	0.964	0.065	1.000	1.000												
$\hat{\gamma}_{ic}$	0.864	-0.463	0.941	0.313	1.000	1.000										
$\hat{f}_{ic}$	0.667	0.750	0.613	0.033	0.517	-0.349	1.000	1.000								
$\hat{t}_{ic}$	0.631	0.582	0.610	0.017	0.527	-0.287	0.524	0.501	1.000	1.000						
$\hat{s}_{ic}$	0.536	-0.308	0.560	0.397	0.536	0.675	0.445	-0.197	0.969	0.132	1.000	1.000				
$\hat{\psi}_{ic}$	0.080	-0.255	0.126	0.127	0.121	0.334	-0.276	-0.429	0.560	0.293	0.632	0.623	1.000	1.000		
$\hat{\omega}_{ic}$	-0.580	-0.672	-0.370	0.358	-0.138	0.840	-0.468	-0.504	-0.386	-0.431	-0.199	0.597	0.041	0.382	1.000	1.000

TABLE E2 — THE POPULATION-WEIGHTED WILLINGNESS TO PAY AND DELAY COSTS

		(1)	(2)
		Sample	Population
Value of time	$\bar{\alpha}$	35.050	22.678
Value schedule delay early	$\bar{\beta}$	23.217	21.054
Value schedule delay late	$\bar{\gamma}$	17.162	16.023

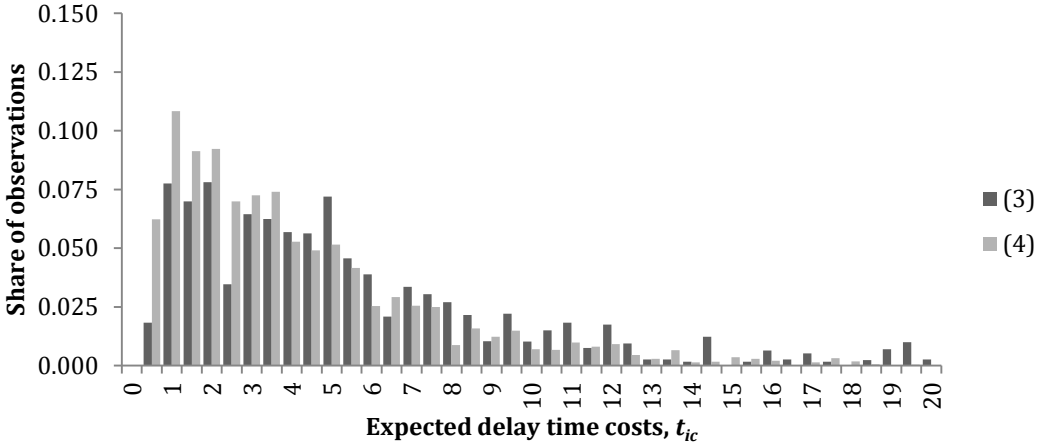

FIGURE E1 — ESTIMATED DISTRIBUTION FOR DELAY TIME COSTS PER TRIP

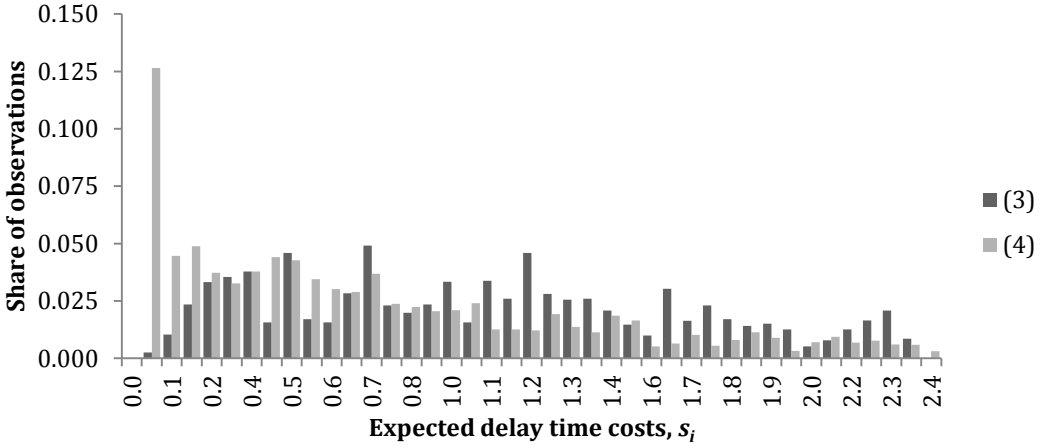

FIGURE E2 — ESTIMATED DISTRIBUTION FOR THE SCHEDULING COSTS PER TRIP

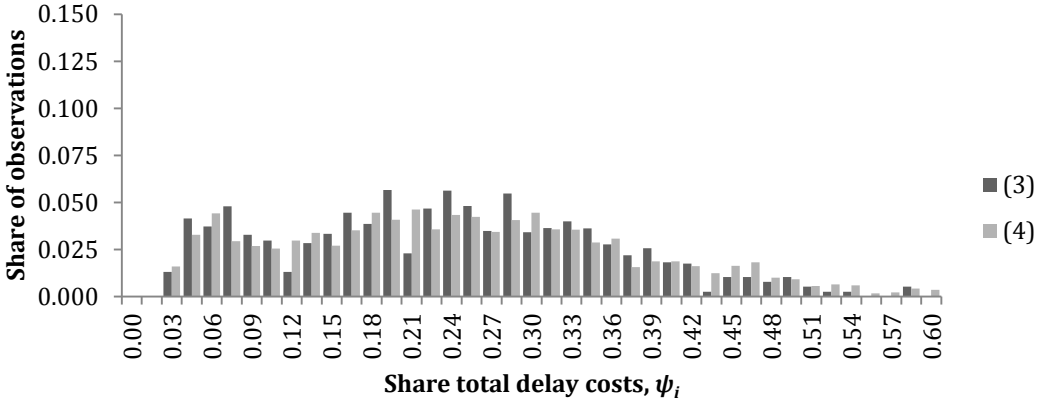

FIGURE E3 — ESTIMATED DISTRIBUTION FOR SHARE OF DELAY COSTS IN TOTAL TRAVEL COSTS

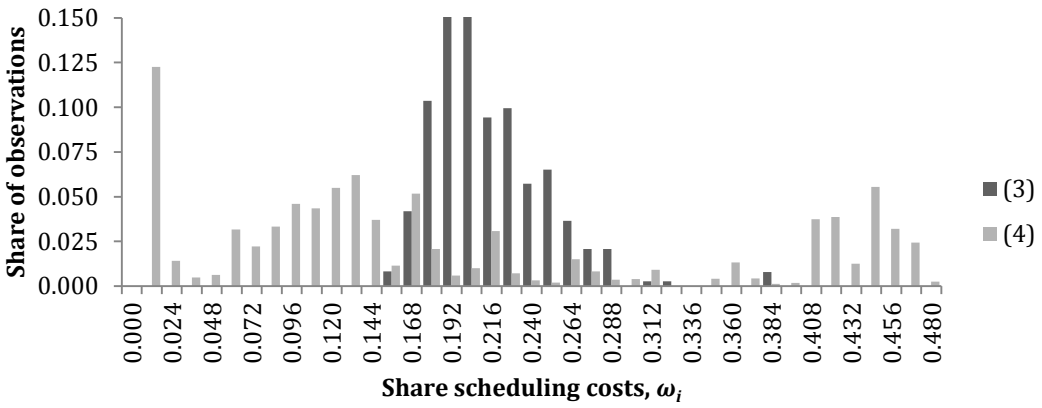

FIGURE E4 — ESTIMATED DISTRIBUTION FOR THE SHARE OF SCHEDULING COSTS IN TOTAL DELAY COSTS

Appendix F. Sensitivity analysis of the calibration of $\hat{\eta}$

TABLE F1 — SCHEDULING COSTS FOR DIFFERENT VALUES OF $\hat{\zeta}$ AND $\hat{\eta}$

		(F1)	(F2)	(F3)
		Local Latent Class	Local Latent Class	Local Latent Class
		$\hat{\eta} \cdot 1.00$	$\hat{\eta} \cdot 0.75$	$\hat{\eta} \cdot 1.25$
Panel 1: Costs per trip				
Free-flow travel costs	$\bar{f}$	18.129	18.129	18.129
		(0.753)	(0.753)	(0.753)
Delay time costs	$\bar{t}$	4.076	4.076	4.076
		(0.881)	(0.881)	(0.881)
Scheduling costs	$\bar{s}$	1.041	0.844	1.197
		(1.124)	(1.113)	(0.135)
Share total delay costs ÷ total trip costs	$\bar{\psi}$	0.251	0.242	0.257
		(0.575)	(0.570)	(0.579)
Share scheduling costs ÷ total delay costs	$\bar{\omega}$	0.238	0.210	0.257
		(0.879)	(0.920)	(0.853)
Marginal increase in the σ due to an increase in μ	$\hat{\eta}$	0.752	0.564	0.940

Note: See Table 2. Specification (F1) is identical to Specification (4).