Quantifying errors in travel time and cost by latent variables.

Juan Manuel Lorenzo Varela - CTS
Maria Börjesson – CTS
Andrew Daly – CTS / ITS, Leeds

Abstract

Travel time and travel cost are key variables for explaining travel behaviour and deriving the value of time. However, a general problem in transport modelling is that these variables are subject to measurement errors in transport network models. In this paper we show how to assess the magnitude of the measurement errors in travel time and travel cost by latent variables, in a large-scale travel demand model. The case study for Stockholm commuters shows that assuming multiplicative measurement errors for travel time and cost result in a better fit than additive ones; however, when measurement errors are modelled, the estimated time and cost parameters are robust to the modelling assumptions. Moreover, our results suggest that measurement errors in our dataset are larger for the travel cost than for the travel time, and that measurement errors are larger in self-reported travel time than software-calculated travel time for car-driver and car-passenger, and of similar magnitude for public transport. Among self-reported travel times, car-passenger has the largest errors, followed by car-driver and public transport, and for the software-calculated times, public transport exhibits larger errors than car. These errors, if not corrected, lead to biases in measures derived from the models, such as elasticity and values of travel time.

Keywords: Hybrid choice models; Latent variables; Error quantification; Measurement error models; RP Value of Time, Self-reported indicators

1. INTRODUCTION

Although considerable research has been devoted to measurement errors in the econometric literature, far less attention has been paid to measurement errors in discrete choice modelling and transportation. Recent studies in the transportation field have shown that when measurement errors exist in discrete choice models, the explanatory variable becomes correlated with the error term, and endogeneity problems arise, analogous to those of their linear counterparts (Díaz et al., 2015; Vij and Walker, 2016). To reduce bias arising from measurement errors, statistical models that can be used to accommodate errors in explanatory variables have become increasingly popular; and among them, the Hybrid Choice Model (HCM) is the modern workhorse in discrete choice analysis.

Parameter bias due to measurement errors in input variables has been highlighted as a substantial problem in the appraisal of policy. For instance, there are reasons to expect that travel cost variables have substantial errors, which attenuate the cost parameters in transport models and lead to under-estimation of the response to pricing measures in appraisal. Moreover, errors in the time and the cost variables are one major reason for collecting Stated Preference (SP) data for value of time estimation, leading to other problems, such as: reference dependence and gain-loss asymmetry (De Borger and Fosgerau, 2008; Börjesson and Eliasson, 2014; Börjesson and Fosgerau, 2015; Hess et al., 2017).

The aim of this paper is to explore the capabilities of the HCM framework to quantify the magnitude of the errors in the key explanatory variables in large-scale travel demand modelling. Quantification of the measurement errors in input data will help identifying the least reliable variables, aiding modellers to concentrate efforts where they are most needed. Hence, we expect our findings to be of interest not only to discrete choice modellers, but also to transport planning practitioners.

The first application of the HCM to account for measurement errors can be found in Walker et al. (2010), where a latent variable approach is introduced to deal with error-prone travel times. Using the same methodology, Guevara (2015) and Varotto et al. (2017) investigate how the time parameter changes when accounting for measurement errors. Furthermore, Walker et al. (2010) and Vij and Walker (2016) use Monte Carlo experiments to show that the estimated parameters converge to their true value as the model accounts for measurement errors in the input variables. However, these studies do not treat cost variables as latent, or model more than one latent variable at a time; and, to date, no study on large-scale transport models has quantified measurement errors in both time and cost variables; important biases therefore could not be detected. Nor do these studies use self-reported travel times and costs, which are a feature of some travel surveys and which give useful additional information about biases.

In this study, we use a HCM to account for measurement errors in the time and cost variables in a large-scale mode choice model estimated on National Travel Survey (NTS) data. First, we explore the sensitivity of parameter estimates to different modelling assumptions. We show that assumptions regarding the distributions of the latent variables and the measurement error impact the estimated error in the latent attributes. However, the parameter estimates of the utility function are robust, given that errors are modelled. Second, we show how goodness-of-fit measurements can be used to rank different measurement error model formulations. In our case study, we find that the multiplicative measurement error model provides a better fit to the observed data than the classical additive measurement error model used in previous studies. Third, we show how the measurement errors in the time and the cost variables can be compared using a multiplicative measurement error formulation, and we find that the cost variables have larger measurement errors. Fourth, we present the policy implications of these findings, including impact on the model elasticities and Values of Travel Time (VoT).

The rest of the paper is structured as follows: Section 2 presents the modelling framework and the modelling assumptions to be tested. Section 3 provides an overview of the data. Section 4 presents the application of the framework in a case study. Section 5 gives some model properties and Section 6 concludes.

2. METHODOLOGY

2.1. Hybrid Choice Model

We take the equation framework of Walker et al. (2010) as our starting point. These authors treat the true value of the explanatory variable suffering from measured errors as a latent variable (X), known only up to a distribution $f_X(X; \theta)$, where θ is a set of estimated parameters, and the measured value of the variable (X) is used as an indicator (I) . They define a mode choice model, with the choice probability of alternative i conditional on the set of parameters β:

$$P(i|\beta, X). \tag{1}$$

Since X is unknown, it is necessary to integrate the conditional choice probability over the distribution of X:

$$P(i|\beta, \theta) = \int P(i|\beta, X) f_X(X; \theta) dX \quad . \tag{2}$$

The measurement equation assumes that the distribution of the indicator (I) conditional on X and a set of estimated parameters λ is

$$I \sim f_M(I|X; \lambda). \tag{3}$$

Putting the pieces together, the likelihood function of the choice model is

$$L(i, I|\beta, \theta, \lambda) = \int P(i|\beta, X)\, f_X(X; \theta) f_M(I|X; \lambda)\, dX \qquad (4)$$

The unknown parameters (β, θ, λ) can be estimated using maximum likelihood estimation, using observed choices, observed characteristics of the alternatives and individuals, and the indicator variables.

The model schematics are shown by Figure 1, where observed variables, indicators and choices are represented by rectangular boxes, whilst unobserved variables such as utilities and latent variables are represented by ellipses. In addition, structural equations are represented by continuous lines and measurement equations by dashed lines.
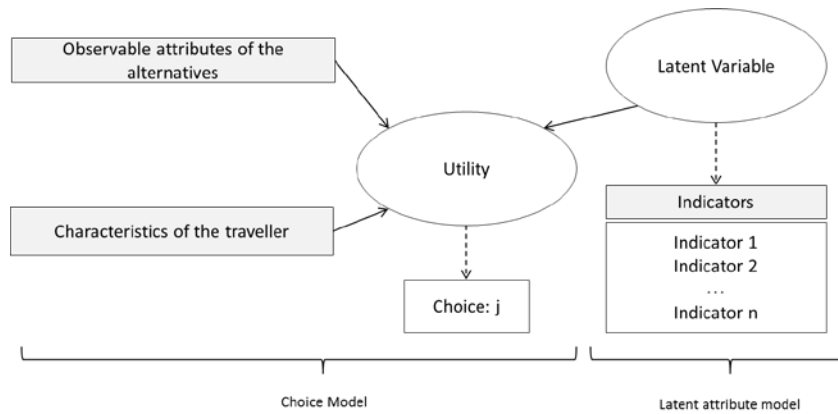

Figure 1. Integrated Choice and Latent Variable framework[1]

In this paper we go beyond the work of Walker et al. (2010) by further developing the specification of the model, including both the latent variable distribution (using different distributions) and the measurement relationship between the measured travel time and the latent variable (better capturing the sources of error). We show how parameter estimates of the measurement equations can be used not only to assess the goodness-of-fit of the postulated theories, but also to provide insights into the magnitude and nature of the errors in the explanatory variables.

2.2. HCM modelling assumptions

Several assumptions are required to implement the mode choice model defined in Section 2.1 Assumptions include the functional form of the choice model defined by (1), the prior distribution for each latent variable, f_X, and the measurement equation (3). These assumptions and their potential effects on the results are discussed in subsections 2.2.1 to 2.2.4.

---

[1] Here we show the possibility of multiple indicators for each latent variable; previous work has used just a single indicator for each of these.

2.2.1 – Functional form of the choice model

We take the choice model in (1) to be the Multinomial Logit (MNL) including time and cost variables for all modes. A full description of the model utility functions is provided in Appendix 1.

2.2.2 – Latent variable distributions

Latent variables are known only up to the distributions f_X. In practice, most latent variables are assumed to have normal priors, see for instance Walker et al. (2010), Díaz et al. (2015), Vij and Walker (2016) or Varotto et al. (2017), even in cases where the represented variable cannot take negative values. The use of normal distributions is supported by Bartholomew et al. (2011) in their review of the general linear latent variable model; where the authors argued that the latent variable prior distribution can essentially be arbitrary, as the choice of f_X seemed to have very little effect on the parameter estimates. However, Walker et al. (2010) points out the importance of sensitivity analyses of distributional assumption of structural and measurement equations for future research. In this paper we are interested also in the measurement errors in the latent variables themselves, and in their distribution, not only in how the relevant assumptions impact the parameter estimates.

In this paper we model time and cost as latent variables. We explore how the assumptions of the distribution of the priors (normal and lognormal) impact the results. These variables cannot take negative values; for this reason, multiplicative error models can only be implemented when assuming positive, e.g. lognormally distributed, priors for the latent variables. Hence, we test a model with normal priors implemented with additive errors against a model with lognormal priors implemented with multiplicative errors.

- *Normal distribution as prior*

The true value of the latent variable $X \sim N(\mu, \sigma^2)$, for individual n, is

$$X_n = \mu + \sigma * \delta_n, \tag{5}$$

where δ_n is a draw from a standard normal distribution, $\delta_n \sim N(0,1)$, and μ, σ are parameters to be estimated.

- *Lognormal distribution as prior*

As an alternative prior, we select the lognormal distribution such that the true value of the latent variable $X \sim LN(\mu, \sigma^2)$ is given by

$$X_n = \exp(\mu + \sigma * \delta_n), \tag{6}$$

where again $\delta_n \sim N(0,1)$, and μ and σ are the parameters to be estimated. We expect that this distribution is more consistent with our data for times and costs

because it is not symmetric and has support on the interval $(0, +\infty)$ which is consistent with the possible values of time and cost variables.

2.2.3 – Latent variable measurement model

In this section we define the measurement relationships that will connect the latent and observed values of a variable.

- *Classical measurement error model*

The most common error model assumes that the magnitude of the error is independent of the value of the variable. The indicator I_n for individual *n* is

$$I_n = \alpha \ + \lambda \ * X_n \ + \varepsilon, \qquad \text{with} \quad \varepsilon_n \sim N(0, \sigma_\varepsilon^2), \tag{7}$$

where α is the offset and λ the scale parameter, both of which are fixed for normalisation purposes, X_n is the latent attribute, ε_n is a normally distributed error component with expected value zero, and σ_ε is a parameter to be estimated.

- *Multiplicative measurement error model*

An alternative to the classical error model is to assume that errors are proportional to the latent value. The indicator I_n for individual *n* is in this case

$$I_n = \alpha \ * X_n{}^\lambda * e^{\varepsilon_n}, \qquad with\ \varepsilon_n \sim N(\mu_\varepsilon, \sigma_\varepsilon{}^2) \tag{8}$$

where, $\alpha \ > 0$ and λ are parameters fixed for normalisation purposes, and the random error, e^{ε_n}, has expected value equal to one, $E[e^{\varepsilon_n}] = 1$. This condition is similar to the additive errors having mean zero. In our case, the expected value of a lognormally distributed error component will equal one if its parameters fulfil the following condition

$$\mu_\varepsilon \ = -\frac{1}{2}\sigma_\varepsilon{}^2 \ , \tag{9}$$

The multiplicative error formulation seems particularly suitable for modelling travel times, as psychological research has found evidence suggesting that perceived time follows a power function of the clock time, see Roeckelein (2000).

The multiplicative error model can be transformed into linear form by taking the natural logarithm,

$$\log(I_n) = \log(\alpha) + \lambda * \log(X_n) + \varepsilon_n, \tag{10}$$
$$with \quad \varepsilon_n \sim N\left(-\frac{1}{2}\sigma_{error\_ln}^2, \sigma_{error\_ln}^2 \ \right).$$

Now, we can only take the log of I_n and X_n if these variables only take positive values. For this reason, we adopt a latent variable prior distribution that has

support only within the positive semi-infinite interval $(0, +\infty)$. This is a natural constraint for travel times and costs variables.

2.2.4 - Latent variables with multiple indicators and errors

The HCM framework does not limit the number of indicators for a latent variable; in fact, it is desirable to have more than one indicator if available because that adds information. Hence, the models in this study have been designed to use the two available indicators for travel time: self-reported and software-simulated (i.e. modelled by transport networks). To be able to use the self-reported travel time we have included total travel time in the utility function for public transport, and not different travel time components.

The self-reported travel time is assumed to have an error distribution that is independent from the other indicators. However, some of the software-simulated variables will have common measurement error distributions. For instance, the simulated travel times for car as driver and car as passenger will be identical and have identical errors for a given trip. Similarly, car cost indicators are assumed to be proportional to the travel distance, which again will be identical for both driver and passenger alternatives. Table 1 shows the available indicators for each latent variable and whether the distribution of the measurement error is independent or shared with other indicators.

Table 1. Overview of indicators and measurement error distributions

Variable	Mode	Indicators		Measurement error	
Travel time	Public transport	-	Self-reported values (available only for the chosen mode)	-	Independent
		-	Calculated values from the assignment model	-	Independent
	Car (Driver and Passenger)	-	Reported values (available only for the chosen mode)	-	Independent
		-	Calculated values from the assignment model	-	Shared distribution for car as driver and car as passenger
Travel cost[2]	Car (Driver and Passenger)	-	Calculated cost based on distance from assignment model	-	Shared distribution for car as driver and car as passenger

2.3. Research questions and models

After presenting the components of the HCM model and how the different modelling assumptions are implemented, this section shows how those elements are combined to explore the capabilities of the HCM framework to quantify the magnitude of the errors in the key explanatory variables in large-scale travel demand modelling. We do this by addressing three Research Questions (RQ).

---

[2] Public transport cost is not represented as a latent variable because of the very simple fare structure used in Stockholm.

RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and prior distributions?

To explore whether parameter estimates are dependent on the modelling assumptions we set up two models with different measurement error modelling assumptions. M_0, implements the classical additive measurement error model, whilst M_1 assumes the alternative multiplicative measurement error model. The two models also apply different distributional assumption of the latent variables prior distribution, for the reasons explained in Section 2.2.2.

- M_0 – Time variables for all motorised modes are modelled as latent variables with normal prior distributions. Measurement equations are additive.

$$U_i = \sum_{k \in S_i} \beta_{ik} * x_{ik} + \sum_{k \in S_i} \gamma_{ik} * LV_{ik} + ASC_i + \varepsilon_i, \tag{11}$$

$$LV_{ik} = \mu_{ik} + \sigma_{ik*}\phi \qquad with \ \phi \sim N(0, 1^2) \tag{12}$$

$$I_{ik} = LV_{ik} + \eta_{ik} \tag{13}$$

$$\eta_i = \sigma_{\varepsilon i} * \phi' \qquad with \ \phi' \sim N(0, 1^2), \tag{14}$$

where S_i is the set of explanatory variables for alternative i, β_{ik} and γ_{ik} are parameters for variable k and alternative i, x_{ik} are observable explanatory variables, LV_{ik} are latent variables, I_{ik} are the observed indicators, ASC_i is the alternative-specific constant, and ε_i is the random variation of the unobserved variables.

- M_1 – Alternatively, time variables for all motorised modes are modelled as latent variables with lognormal prior distributions. Measurement equations are multiplicative.

$$U_i = \sum_{k \in S_i} \beta_{ik} * x_{ik} + \sum_{k \in S_i} \gamma_{ik} * LV_{ik} + ASC_i + \varepsilon_i, \tag{15}$$

$$LV_{ik} = exp(\mu_{ik} + \sigma_{ik} * \phi) \qquad with \ \phi \sim N(0, 1^2) \tag{16}$$

$$I_{ik} = LV_{ik} * \eta_{ik} \tag{17}$$

$$\eta_i = exp\left(-\frac{1}{2}\sigma_{\varepsilon i}^2 + \sigma_{\varepsilon i} * \phi'\right) \qquad with \ \phi' \sim N(0, 1^2) \tag{18}$$

Since there are different possible assumptions regarding the measurement error models and prior distributions, we need to select the measurement error model producing the most accurate parameter estimates, this triggers RQ2.

RQ2: What measurement error model should we use?

Measurement error models describe the relation between the observed and the true values; hence, it seems sensible to compare the goodness of fit for each of the competing measurement error models and select the one that better reproduces the observed data. We use the parameter estimates from the measurement equations of models M_0 and M_1, and we present and discuss common goodness-of-fit measurements, including: Analysis of simulated residuals; Bayes factor; and fit indexes (BIC and AIC). A description of these goodness-of-fit measurements is provided in Section 2.4.

Results from RQ2 will show that the multiplicative error formulation provides a better fit to the observed data. So, the next question regards how to quantify the errors in input variables.

RQ3: How large are the errors in our input variables?

To answer RQ3, we use M_2, a HCM where travel time for car driver, car passenger and public transport, as well as cost variables for the car alternatives are modelled as latent variables with a multiplicative error model. As mentioned earlier, the multiplicative measurement error formulation assumes that measurement errors are heteroskedastic with standard errors proportional to the latent variable. These properties make it easy to compare the magnitude of the measurement errors for different variables, even if the latent variables have different ranges and/or units; e.g. time and cost variables.

Results from RQ3 will shed some light on the common expectation that cost indicators are more error-prone than time indicators, as the modeller typically lacks cost data based on individual characteristics; such as the type of fuel, car, driving behaviour or fare for public transport.

2.4. Testing

Despite the increasing popularity of the HCM framework, discussion of the accuracy of measurement of latent variables in the HCM framework is largely absent in transportation research, see Motoaki and Daziano (2015). The traditional goodness-of-fit measurements of discrete choice models (e.g. likelihood ratio test and ρ^2) cannot be used to assess the model fit of hybrid choice models, and there is no consensus about how to test the HCM goodness-of-fit.

Section 2.4.1 describes how we evaluate the measurement model fit and Section 2.4.2 describes the overall goodness-of-fit measures that we apply to our HCM models.

2.4.1. Measurement model fit

We suggest and apply three tests of the fit of the measurement model. First, to visually inspect the normality assumption (since our two error models (7) and (10) both apply normally distributed error terms) we plot the residuals using a Quantile-Quantile plot (QQplot). A QQplot is a graphical technique for determining whether two data sets come from populations with a common distribution. In this plot, the quantiles of the first data set are plotted against the quantiles of the second data set, and if the two data sets come from a population with the same distribution, the points should follow a straight line.

Second, we complement the QQplots with a density plot, based on simulated residuals from the two measurement models (7) and (10).

Third, we use the Shapiro-Wilk normality test to provide a quantitative measurement of the normality of residuals. It tests the null hypothesis that the population of the sample is normally distributed; Hence, if the p-value is less than the chosen significance level, the null hypothesis is rejected (Royston, 1982).

2.4.2 Model selection

Studies in structural equation modelling (SEM) have established a number of goodness-of-fit measurements. For instance: the chi-square statistic, the root-mean-square error of approximation (RMSEA), or fit indices such as the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). Below we describe the three tests that we apply in this paper.

- *Bayes Factor (BF)*

Bayes factor (Kass and Raftery, 1995) is a well-known statistic in Bayesian hypothesis testing and model comparison. The Bayes factor (BF_{10}) is the ratio of the likelihood probability of two competing hypotheses, and measures how well M_1 predicts the data relative to M_0. In this study, M_1 is a hybrid choice model with multiplicative error assumptions, and M_0 is a hybrid choice model following an additive error model,

$$BF_{10} = \frac{\int p(\theta_1|M_1)\,p(i|\theta_1,M_1)d\theta_1}{\int p(\theta_0|M_0)\,p(i|\theta_0,M_0)d\theta_0} = \frac{p(M_1|i)}{p(M_0|i)}\frac{p(M_0)}{p(M_1)} = \frac{p(i|M_1)}{p(i|M_0)} \qquad (21)$$

where $p(M_1|i)$ and $p(M_0|i)$ are the model posterior probabilities given data i, $p(M_0)$ and $p(M_1)$ are the model a priori probabilities, θ_0 and θ_1 are vectors of the model parameters, $p(i|M_1)$ is the probability of observing i under the assumption that errors in variables follow a multiplicative measurement error model, and $p(i|M_0)$ is the probability of choosing i under the assumptions of the additive measurement error model.

If instead of the Bayes factor integral, the likelihood corresponding to the maximum likelihood estimate of the parameter for each statistical model is used, then the test becomes a classical likelihood-ratio test. Unlike a likelihood-ratio test, this Bayesian model comparison does not depend on any single set of parameters, as it integrates over all parameters (with respect to the respective priors). Kass and Raftery (1995) showed that it is useful to consider twice the natural logarithm of the Bayes factor and interpret the resulting statistic as per the following table:

Table 2. Indicative Bayes Factor thresholds

BF_{10}	$2\log BF_{10}$	Evidence against M_0
<1	<0	Negative (support M_0)
1-3	0-2	Not worth more than a bare mention
3-20	2-6	Positive (support M_1)
20-150	6-10	Strong
>150	>10	Decisive

These values are given as suggestions and not as strict rules.

- *Bayesian information criterion (BIC)*

In this study we calculate the BIC as a stand-alone measurement of the model performance, defined as

$$BIC = \log(n)\,k - 2\log(\hat{L}) \tag{22}$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, n is the sample size, and k is the number of parameters to be estimated. The BIC index makes it easy to rank multiple models, as the lower BIC value, the better the model fit.

- *Akaike information criterion (AIC)*

The AIC is another comparative measure of the relative quality of statistical models for a given dataset. The AIC measurement is given by,

$$AIC = 2k - 2\log(\hat{L}) \tag{23}$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, and k is the number of parameters to be estimated. AIC is interpreted in the same way as BIC, the model with the lowest AIC is preferred.

Some authors have strongly advised against the use of these goodness-of-fit measurements when evaluating Hybrid Choice Models. For instance, Barrett (2007) and Ropovik (2015), state that fit indexes add nothing to the analysis, and Hayduk et al (2007) argue that fit index thresholds can be misleading and subject to misuse. Moreover, Motoaki and Daziano (2015), showed through a Monte Carlo experiment that the behaviour of SEM fit assessment tools did not work as expected for the HCM. Furthermore, these goodness-of-fit measurements are based on the model's final likelihood; hence, when applied to HCM with different number of latent variables and or different measurement equation formulations, the goodness-of-fit measurements might provide counterintuitive results.

Despite of these warnings, it is common to find fit indices reported in current literature despite these criticisms. In this study, the Bayes factor, BIC and AIC goodness-of-fit indices are reported to inform further discussions.

3. DATA

We estimate a mode choice model for commuting[3] trips in Stockholm using PythonBiogeme. The model is estimated on 3777 trips, for which the attributes

---

[3] Commuting trips in this study include work and education purposes.

of the chosen and non-chosen modes were computed using the assignment model in TransCad. The final sample includes 41% trips by PT, 49% by car (driver and passenger) and 11% by walk and cycle. This distribution is broadly representative of commuting, where around 80% of trips to/from the inner city are made by PT in the peak hour but most trips to other areas are made by car.

We estimate the models using the 2005/06 Swedish National Travel Survey. Respondents were randomly selected from the Swedish population aged between 6 and 84 years. Twenty-seven thousand interviews were conducted, with a response rate of 68 per cent. The respondents were asked to report all the trips they made on a randomly selected survey day. For each reported trip they were asked to indicate their main travel mode, start and end times, trip purpose, and origin and destination addresses. Also, the survey contains socio-economic variables such as gender, household type, number of cars in the household, etc. which we have included in our models as discussed below and shown in Section 2.1. Further details of the survey design are given by Trafikanalys (2007).

The data had to be processed in order to extract the variables necessary to define the utility functions. First, the Origin-Destination (OD) matrices were constructed, based on zone sizes between 0.1 and 1 km^2, with smaller zones located in the inner city and larger zones in the outskirts. Second, the distances between each reported origin and destination were calculated. Third, the travel times for each alternative mode were computed, and fourth the travel costs were calculated for the chosen and the unchosen alternatives.

3.1. Stockholm Public Transport network overview

The Stockholm PT network considered in this study covers the modes bus, metro, train and tram. The bus network provides a dense coverage of the entire Stockholm county area. In the assignment model bus services are separated into two mutually exclusive groups: the first group includes the inner-city buses, with an average speed of 15 km/h; and the second includes the commuting buses, with an average speed of 25 km/h. Metro and commuting train services, both have a radial configuration with centre in Stockholm central station (T-Centralen). Metro services have an average speed of 35 km/h, whilst the average speed for commuting train services is 65 km/h. Finally, the average speed of other rail services, mainly trams, is 25 km/h. Assumed speeds are in accordance with statistics provided by the Stockholm Public Transport manager, SL (SLL 2009).

3.2. Mode availability

There are five alternative modes in the data: car driver, car passenger, public transport, bicycle, and walk. The car-driver mode is assumed to be available if the trip-maker has a driving licence, whilst the car-passenger mode is always considered to be available.

Availability for car driver is designed to allow observations where leased and/or hired cars are used. To capture the additional attractiveness of the car

driver alternative for households that own one or more cars, we add a variable equal to the minimum of the number of cars in the household and the number of driving licences. To reduce the attractiveness of the car driver alternative we add a dummy variable that is one if the household members have no access to a private car, whilst car competition is modelled by a dummy variable that is one if there are more licences than cars in the household. The car passenger alternative is available for all, because many observed car passengers do not have access to a car in the household. However, to increase the attractiveness of car as passenger for households with access to a private car, we include a dummy variable indicating household car ownership.

The bicycle mode is assumed to be available if the one-way home-to-work distance is less than 20km, while the walk mode is assumed to be available if the one-way home to work distance is less than 10km. The public transport alternative is available if the journey takes less than 220 minutes of which a maximum of 20 minutes can be allocated for access, and another 20 minutes for egress. In this study, access to/from public transport is considered only by foot, because of the low percentage (0.2%) of observations accessing PT by bicycle in our data.

3.3. Cost variables

The Stockholm Public Transport manager, SL, applies a unit fare structure to all public transport modes and travel distances within the area of study.

To calculate the cost of a trip by public transport, the following logic was followed. First, a rational traveller who decides to commute regularly by public transportation will use a monthly discount ticket. Second, it is assumed that the traveller will commute daily, and therefore will make at least 40 trips per month (4 weeks of 5 working days and 2 trips per day). Third, the decision to buy a monthly ticket is based on commuting costs; therefore, any leisure trip will have zero marginal cost. Fourth, depending on whether the traveller is a student or not, the student or the full-price ticket will be used.

Car driving costs are assumed to be proportional to the distance travelled, multiplied by a factor that represents the marginal cost of travel by car. For this study, 1.8 SEK/km is used as recommended by the Swedish Transport Administration. This value is used nationwide in transport demand models, including the national demand model used by the Government to produce CBA, and has produced reliable traffic volume forecasts (Eliasson et al. 2013; West et al. 2016). Half of this value accounts for fuel costs, whilst the rest covers depreciation etc. (Trafikverket 2015). While other, lower, values are used in other countries, we have retained the standard Swedish practice as it has worked well in similar contexts and there may therefore be perception differences between Sweden and other countries. All reasonable assumptions on this point are likely to be within the margins of estimation error.

Furthermore, our models allow the possibility to have different costs as car driver and as car passenger. For this, different cost sharing assumptions have been tested using the formulation suggested by Fox et al. (2009).

$$V(Cost)_{CD} = \beta_{Cost} CarCost_{OD} \left(1 - \frac{S(O_{CD} - 1)}{O_{CD}}\right) \tag{24}$$

$$V(Cost)_{CP} = \beta_{Cost} CarCost_{OD} \left(\frac{S}{O_{CP}}\right), \tag{25}$$

where S is the cost sharing factor between driver and passengers, O_{CD} is the mean occupancy for car drivers and O_{CP} is the mean occupancy for car passengers. After testing, we apply $S = 1$ (i.e. equal sharing), with values extracted from the data of $O_{CD} = 1.15$ and $O_{CP} = 2.54$.

3.4. Sources of measurement error

Transport modelling typically uses either self-reported travel times and travel costs, or level of service times and cost calculated by a modelled transport network. In the latter approach, simplifications and heuristics such as the ones described in the previous section are used, whilst in the former misperceptions and self-justifications are likely to alter the true value of the reported variables (Peer et al., 2014)

The impact of measurement errors in the input variables is best understood by comparing the distributions of self-reported and calculated travel times

Figure 2 shows 'violin plots' for self-reported and calculated travel times for public transport and car. A violin plot is similar to a box plot, but more informative as it also shows the probability density of the data (it is always symmetric). From these plots is clear that reported indicators experience a rounding effect, as we can observe multimodal distributions, with peaks in multiples of 15 minutes. In addition, we have plotted the mean, first and third quartiles of the distributions in the inner part of the violin plots.
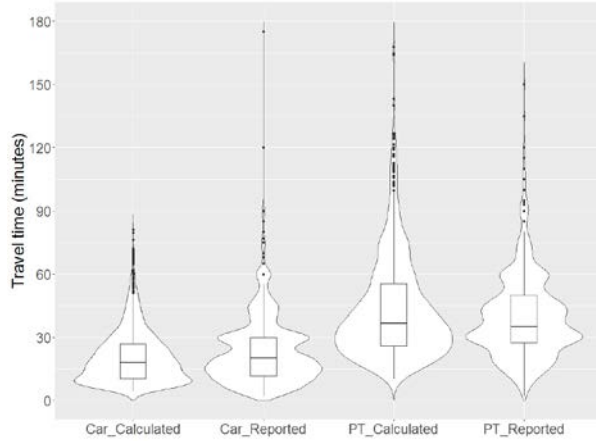

Figure 2. Violin plot for of reported and calculated travel times for car and public transport.

The figure demonstrates discrepancies between reported and calculated travel times. We can for instance observe how rounding reported travel times create multimodal distributions, with peaks located on multiples of 15 minutes. The rounding effect has been studied by Rietveld (2002), where the author shows that rounding is a rule rather than an exception, and rounding has larger

impacts than just affecting the variance of travel times. Given the large scale at which rounding takes place, it may affect averages computed on the basis of national surveys when probabilities of rounding upward and downward do not cancel.

4. RESULTS

Parameter estimates, log-likelihood and goodness-of-fit results for the models are reported in Table 4. We can observe that time and cost parameters are all negative and so the model is micro-economically consistent. The rest of this section presents detailed answers to the research questions described in Section 2.

RQ1: Do parameter estimates depend on assumptions regarding the measurement error model and prior distributions?

To answer this question, we focus on whether different modelling assumptions modify the marginal utilities of time and cost multiplied by the scale, in other words, the γ parameters in equations (11) and (15). Looking at the parameter estimates from models M_0 (normal priors and additive error model) and M_1 (lognormal priors and multiplicative error model), we can observe that two of the parameter values differ by between 1 and 2 standard deviations. We conclude that parameter estimates of the choice model are reasonably robust to these modelling assumptions. However, the parameter estimates differ more substantially from those in the MNL model.

Parameter	M0		M1	
	Value	Std err	Value	Std err
IVT Public Transport	-0.101	0.00711	-0.110	0.0083
IVT Car Driver	-0.0595	0.0115	-0.0729	0.0136
IVT Car Passenger	-0.0625	0.0133	-0.0894	0.0178

Table 3. Estimated parameters and standard errors.

Table 4. Estimation results

Model:	Benchmark		M0			M1			M2		
Type of model:	MNL		HCM			HCM			HCM		
			normal priors and additive error model			lognormal priors and multiplicative error model					
Number of latent variables:	-		3			3			5		
Number of draws:	-		5000			5000			20000		
Number of parameters:	16		27			27			32		
Sample size:	3777		3777			3777			3777		
Final log likelihood:	-2386.65		-70600.30			-25869.71			-42815.96		
BIC	4848		141581			52006			85854		
AIC	4827		141413			51838			85686		
Name	Value	t-test	Value	t-test		Value	t-test		Value	t-test	
ASC bicycle	-1.30	-4.00	-1.33	-4.01		-1.33	-4.01		-1.32	-3.96	
ASC car driver	1.76	6.23	2.23	7.12		2.37	7.63		4.03	8.23	
ASC car passenger	-1.93	-6.93	-2.16	-7.00		-2.04	-6.51		-1.00	-2.82	
ASC public transport	1.13	3.21	3.99	8.26		4.30	8.19		5.23	7.69	
Car competition	-0.51	-2.46	-0.60	-2.57		-0.67	-2.81		-0.73	-2.77	
Cost	-0.79	-6.78	-1.28	-8.99		-1.50	-9.36		-1.83	-8.33	
Living on apartment	1.10	5.25	1.18	5.36		1.19	5.52		1.17	5.30	
No car (car driver)	-2.91	-15.04	-3.49	-15.25		-3.54	-14.68		-3.79	-13.86	
No car (car passenger)	-1.15	-5.79	-1.62	-7.14		-1.64	-6.85		-1.66	-6.59	
Time (bicycle)	-0.09	-4.08	-0.11	-4.64		-0.13	-5.22		-0.13	-5.00	
Time (car driver)	-0.08	-10.23	-0.06	-5.08		-0.07	-5.34		-0.09	-6.20	
Time (car passenger)	-0.09	-9.14	-0.06	-4.71		-0.09	-5.01		-0.11	-5.87	
Time (public transport)	-0.06	-16.28	-0.10	-14.23		-0.11	-13.32		-0.11	-11.70	
Time (walking)	-0.06	-7.63	-0.07	-8.16		-0.08	-8.92		-0.07	-8.37	
Winter	-1.63	-4.34	-1.69	-4.46		-1.69	-4.48		-1.69	-4.48	
Woman (car driver)	-1.20	-12.48	-1.44	-12.53		-1.41	-12.06		-1.53	-11.49	
Structural equations											
$\mu_{Car-driver\ cost}$	-	-	-	-		-	-		2.81	114.82	
$\log(\sigma_{Car-driver\ cost})$	-	-	-	-		-	-		0.11	3.34	
$\mu_{Car-passenger\ cost}$	-	-	-	-		-	-		1.87	84.21	
$\log(\sigma_{Car-pssenger\ cost})$	-	-	-	-		-	-		0.17	6.37	
$\mu_{Car-passenger\ cost}$	-	-	19.80	101.85		2.83	287.23		2.84	285.95	
$\log(\sigma_{Car-passenger\ time})$	-	-	2.72	148.98		0.78	-15.40		-0.27	-16.79	
$\mu_{Car-driver\ time}$	-	-	21.90	93.06		2.95	249.53		2.95	257.46	
$\log(\sigma_{Car-driver\ time})$	-	-	2.76	157.37		0.75	-17.51		-0.29	-17.59	
$\mu_{PT\ time}$	-	-	47.70	107.65		3.74	373.76		3.74	373.80	
$\log(\sigma_{PT\ time})$	-	-	3.39	160.45		0.69	-19.81		-0.37	-19.63	
Measurement equations											
$\log(\sigma_{reported\ PT\ time})$	-	-	2.16	33.71		-1.08	-38.60		-1.08	-38.37	
$\log(\sigma_{reported\ Car-driver\ time})$	-	-	2.31	101.44		-0.98	-37.68-		-0.99	-39.39	
$\log(\sigma_{reported\ Car-passenger\ time})$	-	-	2.33	45.54		-0.86	-17.28		-0.87	-17.13	
$\log(\sigma_{calculated\ Car\ time})$	-	-	1.71	49.69		-1.49	-26.72		-1.41	-33.80	
$\log(\sigma_{calculated\ PT\ time})$	-	-	3.01	158.16		-1.10	-37.19		-1.09	-36.84	
$\log(\sigma_{calculated\ Car\ cost})$	-	-	-	-		-	-		-0.70	-11.07	

RQ2: What measurement error model should we use?

Answering RQ2 to determine whether the additive error model is outperformed by a multiplicative formulation requires again models M_0 (normal priors and additive error model) and M_1 (lognormal priors and multiplicative error model). Figures 3 and 4 show the QQplots for the simulated residuals of each time variable under the two different formulations. Based on a visual inspection, lines plotted on the left have a stronger curvature than the ones on the right, indicating that the multiplicative formulation fits the observed data better than the additive.
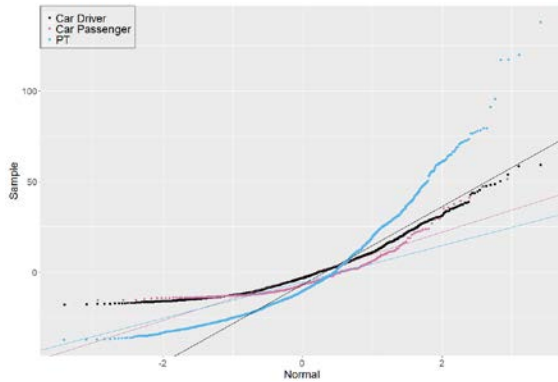
Figure 3.QQplot of time variable residuals following an additive error formulation
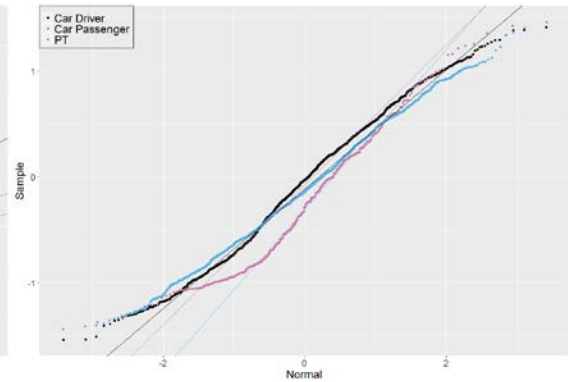
Figure 4. QQplot of time variable residuals following a multiplicative error formulation

In addition, figures 5 and 6 show that the density functions of residuals under the additive formulation are more skewed than the densities of residuals under the multiplicative formulation, confirming the impression given by the QQPlots.
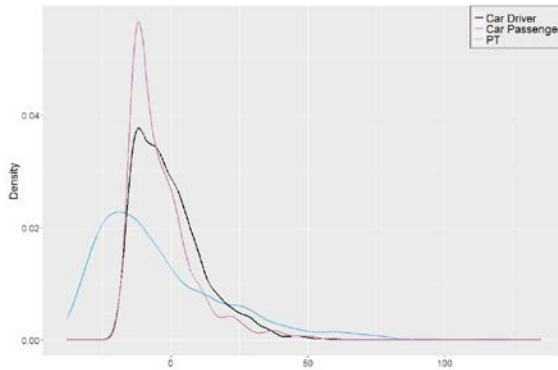

Figure 5. Density plot of time variable residuals following an additive error formulation
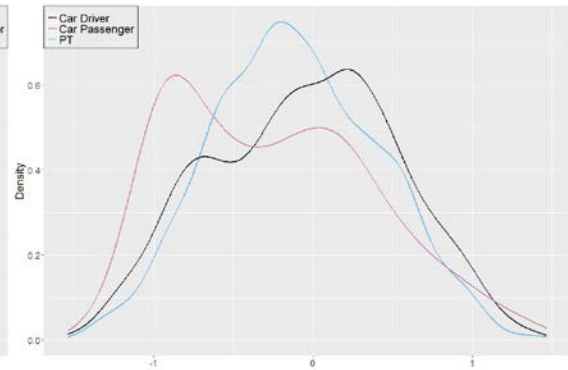
Figure 6. Density plot of time variable residuals following a multiplicative error formulation

Finally, the normality of the residuals is evaluated with the Shapiro-Wilk test p-value.

Model		M0	M1
Time variable	PT	< 2.2e-16	2.09 e-04
	Car Driver	< 2.2e-16	4.07e-09
	Car Passenger	< 2.2e-16	2.84e-07

Table 5. Shapiro-Wilk test p-value

17

Results from these tests show that p-values are low in all cases; hence, the normality assumption can be rejected for both models at all reasonable levels of confidence. Nevertheless, results from the Shapiro-Wilk test provide a numerical tool to rank the relative likelihood of the residuals produced by the different models. By interpreting the results in this way, we observe that all residuals from the multiplicative formulation M1 are much more likely to follow a normal distribution than any of its additive counterparts. These results support the conclusions drawn from the density and QQplots.

Finally, overall model performance between M_0 and M_1 is compared by means of the Bayes Factor, BIC and AIC values. The test value $2 \log BF_{10}$ is 8.96e+4, which, in accordance with the suggested values by Kass and Raftery (1995), provides "decisive support" for the multiplicative formulation. Moreover, the BIC and AIC criteria, reported in table 4, also favour the use of a multiplicative error model.[4]

RQ 3: How large are the errors in our input variables?

To answer RQ3, we model the time and cost variables under study as latent constructs following a multiplicative error model and compare the shape of the estimated distributions of residuals. As detailed in Section 2 the expected value of the residual distribution is fixed to one for all variables but the lognormal distribution is still flexible enough to model the variance in the data.
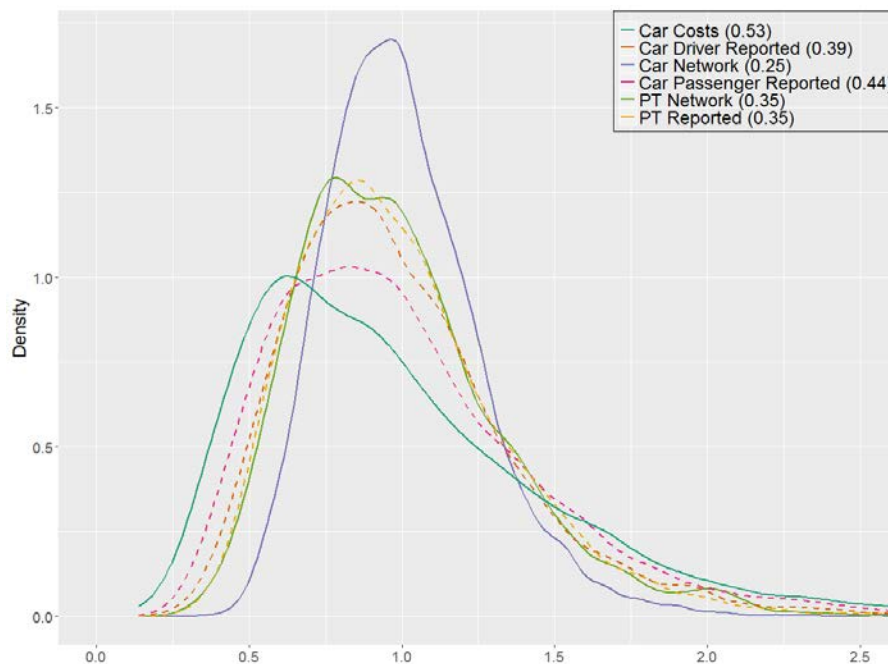

Figure 7. Density plot of simulated residuals from estimated distributions. Numbers between brackets show the standard deviation of the distribution.

---

[4] The likelihood value from the MNL model is not comparable with the HCM values, as discussed in 2.4.2 above.

Figure 7 shows the estimated distribution of the latent variable residuals. The legend gives the standard deviations of the distributions, which, given the mean normalised at 1, represent the coefficients of variation. Results show that among the software calculated variables, car cost residuals have the largest standard deviation of the three (0.53), followed by PT time (0.35) and Car time residuals (0.25).

5. MODEL PROPERTIES AND POLICY IMPLICATIONS

To confirm that the final model gives a reasonable representation of travel by commuters in Stockholm, we calculated the implied elasticities and values of time and compared those with established values.

5.1. Elasticities

Aggregate direct price elasticities for the modes were calculated based on a model simulation with a 10% increase in the variable. Elasticities are calculated as

$$E_{x \to y} = \frac{LOG(D_y) - LOG(D_x)}{LOG(P_y) - LOG(P_x)}, \tag{26}$$

where x refers to the initial state, y to the new state, D is demand, P is price level and E is elasticity.

To further inform this evaluation, results from model M_2 are compared against a MNL model with identical specification of the utility functions, used as a benchmark. Table 6 presents demand elasticities based on parameters from the benchmark and M_2 models. The elasticities indicate that car passenger is the most price sensitive mode, followed by public transport and car driver. This ranking is consistent between the two models. Since the sample population is reasonably representative of commuters in Stockholm, we can take these elasticities as roughly representative of Stockholm commuters.

Table 6. Aggregate direct price arc-elasticities

	Benchmark model (MNL)	M2 (HCM)
Public transport cost +10%	-0,27	-0,98
Car cost +10%		
Car – driver	-0,19	-0,47
Car-passenger	-0,43	-2,1
Car - combined	-0,23	-0,71

Holmgren (2007) presents in its meta-analysis of public transport demand, price elasticity values ranging from −0.009 to −1.32, with a mean value of −0.38. Similar values are reported by Kremers et al (2002) and DfT (2016), where the mean value for short run price elasticities is -0.4. Car price elasticities presented above are around twice the fuel price elasticity, as fuel costs make up around half the considered marginal cost of driving. De Jong and Gunn (2001) report fuel price elasticities for short term commuting trips between -0.16 and -0.52. DfT (2016) suggests that the average fuel price elasticity should lie within the

range -0.25 to -0.35. The Danish national passenger model reports a price elasticity for trips to -0.16 and for mileage -0.42 (Rich and Hansen, 2016). The Swedish national transport model produces the trip price elasticity -0.11 and mileage price elasticity of -0.54 for car.

Looking at the differences between the HCM and the MNL models, the HCM gives larger elasticities for all modes. This finding is consistent with the results in Varotto et al. (2017), where they observe a 65% increase of the time elasticity value when using a HCM formulation. While this leads to values in our work for the public transport and car passenger cost elasticities that are outside the ranges indicated in the literature, these can reasonably be associated with the ways in which costs have been calculated and, in the case of car passengers, with the small sample size. The finding that the omission of measurement error leads to substantial dilution of the coefficient values, both for time and for cost, remains soundly based.

While the elasticity is a somewhat abstract statistic, it should be noted that the responsiveness of forecasts to any changes in cost are proportional to these elasticity values, so any bias wold also apply to more realistic forecast scenarios.

5.2. Values of time

The Value of Time (VoT) is the sum of the marginal utility of time (the sum of the opportunity value of time and the direct utility of travel time), divided by the marginal utility of money (DeSerpa, 1971; Jara-Díaz, 2003). The VoT can be calculated from the model parameters as

$$VoT_i = \frac{\partial V_i / \partial x_{Time}}{\partial V_i / \partial x_{Cost}}. \tag{27}$$

In the particular case where the utility function is linear in time and logarithmic in cost, the VoT is given by,

$$VoT_i = \frac{\beta_{IVT_i}}{\beta_{Cost}} * Cost \tag{28}$$

Table 7 shows the VoT estimates for motorised transport modes and compares values from M_2 with the benchmark model (MNL). In addition, two different VoT from M_2 are presented in Table 7 for the car alternatives. One is calculated based on the mean value from the estimated latent cost distributions, and the other uses the average value of the cost variables with measurement errors.

From these values we can observe two things. First, VoT estimates in Table 7 suggest that differences between the benchmark and M_2 models are not only a scale issue, as the ratio of parameters is not affected by scale changes.

Second, when we compare the VoT provided by the benchmark MNL model against the ones from the HCM – M_2, we observe smaller VoT, between 15 to 45%, when time and cost are treated as latent variables simultaneously. These

results suggest that time and cost parameters are poorly estimated when using RP data, as a consequence of measurement errors. So, by modelling time and cost as latent variables with a multiplicative error formulation, as done in M_2, we observe a 130% increase of the cost parameter; a 4% increase of the car-driver time parameter, a 23% increase of the car-passenger time parameter, and the time parameter for public transport doubles, when compared against the estimates from the benchmark MNL model.

We can interpret these results as another indication of which variables have the largest measurement errors, as the larger the error in the variable, the larger the bias towards zero in the associated parameter. This phenomenon is known as regression dilution; see for instance Díaz et al., 2015. Hence, when the model accounts for errors in variables, we should expect a larger percentage increase in the parameter associated with the more contaminated variables.

For comparison, the VoT for commuting trips, estimated on SP data are 98 SEK/h for car and 53 - 72 SEK/h for public transport (Börjesson and Eliasson (2014). For car, the SP VoT is lower than that resulting from our benchmark model, but for public transport, it is higher than that resulting from our the M2 model. Now, comparison with SP does not lend any conclusive evidence as to whether the baseline or the M2 estimates is closest to the true VoT. Stated preference data is not subject to measurement errors, but it has other problems such as reference dependency and gain-loss asymmetry. Moreover, the VoT estimated on SP data are not directly comparable to the travel survey data because VoT for, say car drivers, is estimated on data collected from car drivers only (and the values of time for PT are estimated for PT users only). However, in a mode choice model as in this study, the estimated VoTs are representative for all travellers (not only the car drivers or PT users).

Table 7. Values of Time [SEK/h]

Mode	Benchmark model	M2 model	
		Latent average cost	Average cost with measurement errors
Car Driver	155	87	67
Car Passenger	55	44	25
Public transport	68	-*	59

* Pt cost not treated as latent.

Previous studies reporting VoT from HCMs accounting for measurement errors (Walker, 2010; Varotto, 2017) only treat the time as latent variables. They found higher time parameters, which translates into a higher VoT. However, our results indicate that the VoT estimates based on RP data are already too high due to large measurement errors in the cost variables. Hence, even higher VoT by means of more advanced modelling defies intuition and might lead to less accurate forecasts, particularly when pricing measures are being analysed.

6. CONCLUSIONS

Errors in time and cost variables are a well-known problem. In this paper, we show how parameter estimates differ substantially between models that

account for measurement errors in explanatory variables and those that do not, as well as how modelling assumptions for HCM impact parameter estimates.

All tests performed, formal and informal, favour the use of a multiplicative error models for variables with support only on a semi-infinite interval and our dataset. However, the estimated time and cost parameters are reasonably robust to modelling assumptions. This is interesting because most of the HCM applications in recent studies did not consider a measurement equation other than additive, and it is possible that better results could have been achieved by the use of a multiplicative error model. As shown in this paper, we can increase the performance of our models, and our understanding of the error source in the input variables by testing different modelling assumptions of the latent variable prior distributions and the measurement error formulation.

Adopting a multiplicative error formulation not only gives a better fit; it also provides a consistent framework that can be applied to any variable which has support only on a semi-infinite interval, in other words, variables that can only take one sign, such as: time, cost, income, etc. Moreover, as measurement errors are modelled as a factor rather than an absolute value, the estimated measurement error distributions are directly comparable among variables and indicators. Applying a multiplicative error formulation to our case study, results from M_2 suggest that cost indicators have proportionately larger measurement errors than any time indicators. Furthermore, estimated measurement error distributions of the different indicators indicate that;

1. Errors in self-reported time indicators are larger than errors in software calculated indicators for car-driver and car-passenger, and of similar magnitude for public transport.
2. Among self-reported time indicators, car-passenger has the largest errors, followed by car-driver and public transport.
3. Among calculated time indicators, public transport has higher errors than car. This result suggests that the modelling assumptions of the public transport network are worse than the ones for the car network. This finding is not surprising as the number of modelling assumptions for the public transport network is larger than for the car network, increasing the chances of introducing measurement errors.

In this framework, the shape of the error distribution and its mean value are imposed by the modeller. In this case, the lognormal probability density function is particularly useful to represent right-skewed data. However, in a lognormal distribution the mode and median will always be smaller than the mean. Because these characteristics of the distribution are unchangeable, meaningful claims about the frequency of an indicator to over or under-report cannot be made from the shape of the estimated error distributions.

A drawback of using advanced models such as the HCM formulation presented in this study is the increased difficulty of dealing with confounding effects. For instance, the interaction between taste variations and measurement errors in the input variables: the more flexibility we build into our models, the higher the

risk for these undesired interactions. These interactions are dangerous for two main reasons; first, we do not know their magnitude and second, data to help us evaluate them it is very scarce. Despite the challenges, the HCM framework offers a promising tool to explore the magnitude of the interactions between measurement errors in variables and taste variations. (Vij and Walker, 2016)

Table 4 shows that BIC and AIC indices yield counterintuitive values when used to compare models with different number of latent variables. Nevertheless, BIC and AIC values when comparing models with equal number of latent variables, M_0 and M_1, are in agreement with the other tests carried out (QQplots, density plots and Shapiro-Wilk test) and favour the use of a multiplicative error formulation.

Simulated price elasticities for the MNL model yield values reasonably consistent with the international literature. However, the HCM accounting for measurement errors in time and cost variables generates significantly higher elasticities. This finding is consistent with the results in Varotto et al. (2017).

Finally, our results indicate that the VoT estimates based on RP data are too high due to large measurement errors in the cost variables. Hence, even higher VoT by means of more advanced modelling defies intuition and might lead to less accurate forecasts, particularly when evaluating pricing measures.

7. REFERENCES

Bartholomew, D.J. Knott, M., Moustaki, I. 2011 John Wiley and Sons, Ltd. Latent Variable Models and Factor Analysis: A Unified Approach, 3rd Edition. ISBN: 978-0-470-97192-5

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

Borger, B. de and Fosgerau, M. 2008. The trade-off between money and travel time: A test of the theory of reference-dependent preferences, J. of Urban Economics, Vol. 64, pp. 101-115.

Börjesson, M. and Eliasson, J. 2014. Experiences from the Swedish Value of Time study. Transportation Research A, 59, 144–158.

Börjesson, M. and Fosgerau, M. 2015. "Response time patterns in a stated choice experiment," Journal of Choice Modelling, vol. 14, pp. 48-58, 2015.

Börjesson, M., and Kristoffersson, I. 2017. The Swedish congestion charges: ten years on: - and effects of increasing charging levels. CTS working paper 2017:2. http://swopec.hhs.se/ctswps/abs/ctswps2017_002.htm

De Jong, G., Gunn, H., 2001. Recent evidence on car cost and time elasticities of travel demand in Europe. Journal of Transport Economics and Policy 35 (2), 137–160

De Serpa, A. (1971). A Theory of the Economics of Time. The Economic Journal, 81(324), 828-846. doi:10.2307/2230320".

Diaz, F., Cantillo, V., Arellana, J., Ortúzar, J. de D., 2015. Accounting for stochastic variables in discrete choice models. Transp. Res. B Methodol. 78, 222–237.

DfT. 2016. Update to TAG Unit M2 to reflect updated values of travel time savings. Unit M2 – Variable Demand Modelling. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/542504/webtag-unit-m2-variable-demand-modelling-forthcoming-change-november-2016.pdf

Eliasson, J., Börjesson, M., van Amelsfort, D., Brundell-Freij, K., Engelson, L., 2013. Accuracy of congestion pricing forecasts. Transportation Research A 52, 34-46.

Fox, J. Daly, A. Patruni, B. 2009. "Improving the Treatment of Cost in Large Scale Models". Association for European Transport paper repository.

Guevara, C.A. 2015. Critical Assessment of Five Methods to Correct for Endogeneity In Discrete-Choice Models. Transportation Research Part A - Policy and Practice 82, 240–354

Hess, S., Daly, A., Dekker, T., Ojeda Cabral, M. and Batley, R. (2017) A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research, Transportation Research B, Vol. 96, pp. 126-149, Feb.

Holmgren, J. 2007. Meta-Analysis of Public Transport Demand. Transportation Research A, Vol. 41, No. 10, pp. 1021–1035.

Jara-Díaz, S.R. 2003. On the goods-activities technical relations in the time allocation theory. Transportation 30: 245. doi:10.1023/A:1023936911351

Kass, R. E. and Raftery, A. E. 1995. Bayes factors. Journal of the American Statistical Association, 90, 773–795

Kremers, H., P. Nijkamp, and P. Rietveld. 2002. A Meta-Analysis of Price Elasticities of Transport Demand in a General Equilibrium Network. Economic Modelling, Vol. 19, No. 3, pp. 463–485.

Peer, S., Knockaert, J., Koster, P., Verhoef, E. 2014. Over-reporting vs. overreacting: Commuters' perceptions of travel times, Transportation Research Part A: Policy and Practice, Volume 69, pp 476-494, ISSN 0965-8564, https://doi.org/10.1016/j.tra.2014.07.005.

Rich, J., Hansen, C.O. The Danish national passenger model – Model specification and results. 2016. European Journal of Transport and Infrastructure Research, 16 (4), pp. 573-599.

Rietveld, P., 2002. Rounding of arrival and departure times in travel surveys: an interpretation in terms of scheduled activities. J. Transp. Stat. 5 (1), 71–81.

Roeckelein, J.E., 2000. The Concept of Time in Psychology: A Resource Book and Annotated Bibliography. Greenwood Press, Westport.

Royston, P. 1982. An extension of Shapiro and Wilk's W test for normality to large samples.Applied Statistics, 31, 115–124.

Schafer, D., and Purdy, K. 1996. Likelihood Analysis for Errors-in-Variables Regression with Replicate Measurements. Biometrika, 83(4), 813-824. Retrieved from http://www.jstor.org/stable/2337285

SLL 2009. Trafikplan 2020 – Externremiss.
http://www1.sl.se/pdf/trafikplan%202020_klickbar.pdf

Trafikanalys. 2007. '2005 – 2006 The National Travel Survey'. RES.

Trafikverket. 2015. Samhällsekonomiska principer och kalkylvärden för transportsektorn: ASEK 5.2. Kapitel 13 Operativa trafikeringskostnader för persontrafik.
http://www.trafikverket.se/contentassets/823481f052a74a388149213638
3eb01b/filer/13_trafikeringskostnad_person_a52.pdf

Varotto,S.F. Glerum, A. Stathopoulos, A. Bierlaire, M. Longo, G. 2017.Mitigating the impact of errors in travel time reporting on mode choice modelling, In Journal of Transport Geography, Volume 62, 2017, Pages 236-246, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2017.05.016.

Vij, A. and Walker, J., 2016. How, when and why integrated choice and latent variable models are latently useful. Transp. Res. B Methodol. 90, 192–217.

Walker, J., Li, J., Srinivasan, S., Bolduc, D., 2010. Travel demand models in the developing world: correcting for measurement errors. Transportation Letters 2, 231–243.

Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved". Biometrika. 34 (1–2): 28–35. doi:10.1093/biomet/34.1-2.28. MR 0019277.

West, J., Börjesson M., Engelson M. 2016. Accuracy of the Gothenburg congestion charges forecast. Transportation Research Part A. vol. 94, pp 266-277

8. APPENDIX 1

This appendix presents the explanatory variables entering the utilities of the models and gives an example of the full model specification.

Table A1.1. Benchmark model full specification.

Parameter	Walk	Bicycle	PT	Car Driver	Car Passenger
ASC Bicycle		1	-		
ASC Car Driver	-	-	-	1	-
ASC Car Passenger	-	-	-	-	1
ASC Public Transport	-	-	1	-	-
InVehicle T*ime – Bicycle*	-	bc_tt_min	-	-	-
Winter	-	winter	-	-	-
Car Competition	-	-	-	ccomp	-
Household no car (*Car Driver*)	-	-	-	nocars	-
InVehicle T*ime – Car Driver*	-	-	-	c_tt_min	-
Woman (Car Driver)	-	-	-	woman	-
Cost	-	-	l_ptcost	l_cdcost	l_cpcost
Household no car (*Car Passenger*)	-	-	-	-	nocars
InVehicle T*ime – Car Passenger*	-	-	-	-	c_tt_min
Total TravelTime – Public Transport	-	-	pt_ttt_min	-	-
Living on apartment block	apartment	-	-	-	-
"In − Vehicle" Time – Walking	wa_tt_min	-	-	-	-

apartment Dummy variable. It takes the value 1 if the user lives in an apartment and 0 otherwise

bc_tt_min Bicycle travel time in minutes

carow Car ownership. The variable takes the minimum value between the number of cars in the household and the number of driver licenses.

ccomp Car competition dummy variable. It takes the value 1 if the number of cars in the household is less than the number of driver licenses and 0 otherwise.

l_cdcost Logarithmic transformation of the car driver costs in SEK. Costs are proportional to the distance travelled, multiplied by a factor that represents the marginal cost of travel by car[5]. Furthermore, different cost sharing assumptions have been tested using the formulation suggested by Fox et al.(2009).

$$V(Cost)_{CD} = \beta_{Cost} CarCost_{OD} \left(1 - \frac{S(O_{CD}-1)}{O_{CD}}\right) \tag{27}$$

$$V(Cost)_{CP} = \beta_{Cost} CarCost_{OD} \left(\frac{S}{O_{CP}}\right) \tag{28}$$

where:

S is the cost sharing factor

---

[5] Marginal cost of travel by car assumed 1.8 SEK/km as recommended by the Swedish Transport Authority (Trafikverket 2015).

O_{CD} is the mean occupancy for car driver observations
O_{CP} is the mean occupancy for car passenger observations

Final parameters used in the case study are S = 1, $O_{CD} = 1.17$ and $O_{CP} = 1.55$.

l_cpcost	Logarithmic transformation of the car passenger costs in SEK. The cost for the passenger alternative is calculated following the same procedure described above.
c_tt_min	Car travel time in minutes
nocars	Dummy variable. It takes the value 1 if the user´s household do not own any car and 0 otherwise
l_ptcost	Logarithmic transformation of the cost of the public transport alternative in SEK. The cost of the monthly ticket has been divided by 40 trips, assuming 4 weeks of 5 working days and 2 trips per day. Depending if the traveller is a student or not, the student or the full ticket has been used. No distinction has been made between the people who reported that they own a discount ticket and the ones who do not; this was done to prevent confirmation biases as PT for the people who normally commute by public transport will be cheaper than for others.
pt_ttt_min	Total travel time in minutes for public transport. Including: Access and egress time, first waiting time, in-Vehicle travel time and transfer time.
wa_tt_min	Walking travel time in minutes
winter	Dummy variable. It takes the value 1 if the trip has been made between weeks 47 and 14 and 0 otherwise
woman	Dummy variable. It takes the value 1 if the user is a woman and 0 otherwise