Journal of Transport Geography 62 (2017) 236–246

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Mitigating the impact of errors in travel time reporting on mode choice modelling

Silvia F. Varotto[a,*,1], Aurélie Glerum[b], Amanda Stathopoulos[b,2], Michel Bierlaire[b], Giovanni Longo[a]

[a] Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, Via Valerio 6/4, 34127 Trieste, Italy
[b] Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Station 18, 1015 Lausanne, Switzerland

ARTICLE INFO

Keywords:
Reported travel time
Measurement error
Latent variables
Hybrid choice
Mode choice
Discrete choice

ABSTRACT

Travel time is a major component in understanding travel demand. However, the quantification of demand and forecasting hinges on understanding how travel time is perceived and reported. Travel time reporting is typically subject to errors and this paper focuses on the mitigation of their impact on choice models. The aim is to explain the origin of these errors by including elements of travel behaviour (e.g., activities during the trip), which have been shown to significantly affect mode choices and commuting satisfaction. Based on responses from a revealed preferences survey, we estimate a mode choice model that treats travel time as a latent variable and incorporates different sources of data along with information on travel activities. Employing these multiple – sometimes incongruent – sources of information in the choice model appears to be beneficial. Results from comparing a logit model assuming error-free inputs and the integrated hybrid model revealed significant impacts on the generated policy scenarios. The model results also contributed to identifying the main travel activity features that affect travel time reporting, providing indications that can assist in understanding and mitigating the impact of imprecise measurements.

1. Introduction

Travel behaviour models typically rely on data afflicted by errors, both in measurement (e.g., software or researcher imputation error) and reporting (e.g., over/under-estimation by traveller). The impact of these errors on choice model outputs has been extensively investigated since the 1970s (McFadden, 2000). Several studies (Bhatta, 2011; Brownstone and Small, 2005; Daly and Ortúzar, 1990; Ghosh, 2001; Ettema et al., 2012; McFadden and Talvitie, 1977; Ortúzar and Ivelic, 1987; Reid and Small, 1976; Small et al., 2005; Walker et al., 2010) have shown that key forecasting indicators such as value of time (VOT) are quite sensitive to the accuracy of travel attributes and to individual-specific explanatory variables. Parameter estimation might be significantly biased when temporally aggregated travel times (Reid and Small, 1976) and spatially aggregated level of service measurements (Ortúzar and Ivelic, 1987; Daly and Ortúzar, 1990) are used instead of individual measurements. Moreover, measurements calculated by researchers (or software) and those reported by users typically differ and result in significantly different model outputs (Brownstone and Small, 2005; McFadden and Talvitie, 1977; Small et al., 2005). Most of these studies have shown in empirical applications that errors in travel behaviour measurements can downward bias VOT up to 50%. Since this indicator is often used for the cost-benefit appraisal of transport projects, errors in travel behaviour measurements can result in significantly lower estimations of willingness to pay of individuals to reduce their trip duration.

Despite the relevant impact of these errors, few attempts to explain their origin have been made. The presence of multiple measurements of travel variables, and a lack of consensus on which to rely on or how to reconcile these different origins has given rise to important debates in the travel behaviour field. One approach would be to seek to identify the most revealing measurement input and disregard other (inconsistent) ones when constructing models. The question then arises of which measurement is most likely to effectively drive the choices of respondents. Research has addressed the question of modelling with reported versus calculated data for various aspects: quality of service

* Corresponding author.
E-mail addresses: s.f.varotto@tudelft.nl (S.F. Varotto), aurelie.glerum@epfl.ch (A. Glerum), a-stathopoulos@northwestern.edu (A. Stathopoulos), michel.bierlaire@epfl.ch (M. Bierlaire), giovanni.longo@dia.units.it (G. Longo).
[1] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, P.O. Box 5048, 2600 GA Delft, The Netherlands.
[2] McCormick School of Engineering, Northwestern University, Technological Institute, 2145 Sheridan Rd., Evanston, IL 60208, USA.

http://dx.doi.org/10.1016/j.jtrangeo.2017.05.016
Received 26 July 2016; Received in revised form 31 May 2017; Accepted 31 May 2017

(Bordagaray et al., 2014; Eboli and Mazzulla, 2011) and level of crowding in public transport (Li and Hensher, 2013), travel distances (Witlox, 2007), life-style and life situation (Scheiner and Holz-Rau, 2007), walking behaviour (Lin and Moudon, 2010), safety of driving patterns of older adults (Molnar et al., 2013), and travel time inputs (Carrion, 2013; Curl et al., 2015; Li, 2003; Peer et al., 2014).

In this research, we focus on results related to travel time measurements while the debate extends to a broader set of travel variables such as cost, income, etc. To date, most studies rely on software calculated travel time, typically considering these measurements as "objective", and unaffected by errors. In line with this thinking, self-reported travel times are seen as uninformative, or biased. However, reported measurements can be argued to provide better indications of the perceived values that are ultimately driving the choice process (Daly and Ortúzar, 1990; McFadden and Talvitie, 1977). Indeed, it is well established that travellers often overestimate or underestimate the actual travel time of their journey and this perceived value could influence the travel decisions. A few studies have attempted to control for these differences in travel time measurements by using advanced choice models (Diaz et al., 2015; Ribeiro et al., 2014), but the formulations proposed are not suitable in case a variable is available only for users who made a certain choice (e.g., reported travel time for the chosen alternative) and gained limited insight into the potential sources of error affecting reported travel time.

The aim of this paper is to explain discrepancies between reported and calculated travel times and to explicitly capture the impact of these different travel time measurements on mode choices. The research hypotheses are detailed in Section 2.4. The issue of poorly reported travel times is explored in a real mode choice case study for a university campus in Trieste (Italy). The paper is structured as follows. Section 2 provides a literature review of approaches to model transportation choices with multiple measurements of travel time and formulates the research questions. Section 3 presents the transport mode choice case study. Section 4 presents the methodology and specification of the logit and the hybrid choice models. Section 5 discusses the estimation results obtained by using the extended software package BIOGEME (Bierlaire and Fetiarison, 2009). Section 6 presents the validation and policy analysis. Section 7 gives conclusions and suggestions for future research, discussing the limitations concerning the dataset used and possible extensions.

2. Literature review

The literature review focuses on studies proposing comparative choice model structures to analyse travel time measurements and to quantify the impact of these measurements on policy outputs. Section 2.1 introduces the concept of time perception in psychology and transportation. Section 2.2 presents current approaches to deal with reported and calculated travel time measurements in choice models. Section 2.3 summarizes literature to support a model framework which has the potential to account for different types of errors and missing values in travel time measurements. Section 2.4 discusses the research gaps and hypotheses that are addressed in this study. While we acknowledge that the debate surrounding travel time measurements is relevant to other factors influencing travel decisions (Bordagaray et al., 2014; Eboli and Mazzulla, 2011; Li and Hensher, 2013; Lin and Moudon, 2010; Scheiner and Holz-Rau, 2007), it is beyond the scope of the paper to make parallel assessments for other types of variables. An important reason for this is the lack of consistent model approaches and methods to compare inputs for most of these research fields.

2.1. Subjective perspectives on time

The subjective nature of temporal judgment has been established in psychological research. Evidence from the literature has suggested perceived time as a power function of the clock time (Roeckelein, 2000). Block (1985) proposed a cognitive model in which the duration experienced was influenced by several elements, such as activities during time periods and subject's characteristics. In addition, Hornik (1992) found that good mood led to retrieving biased memories of time congruent with the mood. Following these studies, the interest of transportation researchers in travel time has increased. Bates et al. (2001) argued that it was likely that travellers were maximizing utility according to their own divergent views of the travel time distribution notwithstanding actual measurements. Consequently, travellers differed in their optimal choices depending on the degree of distortion of their subjective distribution with regards to the actual measurement distribution. Rietveld (2002) noted that in travel surveys most respondents applied rounding of departure and arrival times to multiples of 5, 15 and 30 min. A possible explanation for this effect is that scheduled activities force people to plan their trips in advance which provide them with anchor points for their memory afterwards. These findings should be integrated into transportation models.

2.2. Reported and calculated travel times in transportation models

The effect of multiple travel time measurements on choice models has originally been investigated in studies combining revealed preference (RP) and stated preference (SP) data. For instance, Small et al. (2005) and Brownstone and Small (2005) noted that VOT estimates using SP data (based on reported travel times) corresponded to less than half of VOT estimates based on RP data (relying on calculated travel times in real traffic) when choosing a congestion-free lane. They concluded that travellers overstate the travel time they experience in congestion in a SP experiment, due to either emotional aversion to traffic delays or over-estimation of the actual impact of tolled lanes. Similar findings were shown by Ghosh (2001) in a different congestion pricing project, in which the median VOT from SP responses was half to one third of RP values, depending on the model form.

Interestingly, recent studies relying on advanced travel tracking technology have reached opposite conclusions on the role and value of reported travel time data in choice modelling. Carrion (2013) analysed the role of reported and GPS-derived travel time in route choice, concluding that the goodness of fit statistics favoured the models with reported measurements, compared to those with calculated measurements. Peer et al. (2014) compared reported travel time by respondents to travel time measured by GPS and camera data in real traffic, noting that reported measurements were overestimated. However, this distortion (expressed as a ratio between reported and calculated travel time) did not seem to influence departure time choices in SP and RP settings. Therefore, they concluded that the reported travel time was affected by errors and did not represent the actual travel time perceived by respondents.

Early models included some subjective information to improve RP models. Ghosh (2001, sec. 5.2) included an 'excess time savings' term defined as perceived minus actual time savings in a RP mode choice model. People with more positive time saving biases were more likely to select the toll option, but the variable did not alter the VOT estimate. Recent studies have attempted to control for these differences in travel time measurements by using advanced choice models. Despite the significant differences in the travel time data available, Ribeiro et al. (2014) found similar model performances using GPS and self-reported travel times in mixed logit models which accounted for taste heterogeneity between individuals (panel effect) and random travel time parameters. They concluded that the choice of adequate model specification when using reported data allows results to align with those based on more precise GPS data. Therefore, the development of advanced models using reported travel times is particularly promising. Diaz et al. (2015) conducted an econometric analysis to identify the most suitable model structures that could deal with discrepancies between calculated and experienced measurements using synthetic and real data. They included a measurement equation directly in the utility