Sustainability of Travel Demand Forecasting Models

Wafaa Saleh and Séona Farrell
School of the Built Environment, Napier University, UK

Abstract

There is increasing research work on travel demand forecasting using different sources of data including stated preference and revealed preference data. There are a number of problems that can be associated with data collection, sampling methods and types of data used in such models. For example, the problem of repeated measurements in stated preference data has generally been recognised in the literature but ignored in practice (Bates & Terzis, 1997). This problem has been investigated and a number of correction factors have been suggested. Other error sources include the scale factor and taste variations. In this paper five randomly selected samples are drawn from a larger data set and used to investigate the reliability of the forecasting models using the segmented models relative to the full data set model. All segments have the same sample size and the same number of observations per respondent. A reliability indicator has been derived to investigate the variations in the coefficient estimates. This indicator shows a range of variations in the estimates. Further research is clearly needed in this area.

Keywords

Stated preference; error sources; forecasting; reliability and correction factors.

INTRODUCTION

Discrete choice analysis is extensively used in the transport field to represent the choices made between a finite set of alternatives, for example, a set of alternative departure times. More specifically, it is often used to investigate and forecast individual travel behaviour. Ben-Akiva & Lerman (1985) provide a detailed overview of discrete choice analysis.

There is increasing research work on travel demand forecasting using different sources of data including stated preference and revealed preference data. There are a number of problems that can be associated with data collection, sampling methods and the types of data used in models (see Bradley & Daly, 1993; Ortúzar & Willumsen, 2001).

DESCRIPTION OF THE DATA

A questionnaire survey was used as a part of a study to investigate the potential impacts of congestion charging on the travel behaviour of Edinburgh commuters. The sample was drawn from employees working in the city centre of Edinburgh. In the questionnaire respondents were presented with three sets of congestion charging scenarios related to departure time choice. In total 211 questionnaires were collected. Only respondents who drove to work (not necessarily just regular drivers) completed the departure time stated choice section of the questionnaire.

Thus, 94 of the 211 respondents were eligible to complete the departure time choice scenarios (Farrell 2004).

Each respondent was presented with seven departure time scenarios; thus, a total of 658 observations were possible. After removing respondents who did not provide all required information, e.g. some socio-economic data, the dataset was reduced to 632 observations.

Disaggregate choice based models were developed and estimated to investigate the effect of variable congestion charging levels on departure time choice. Three departure time alternatives were defined for the departure time model. Two of the alternatives represented a change in departure time from the respondents' usual departure time for work (earlier than usual and later than usual departure time), while the third alternative represented the respondents' current departure time choice (but with changes in travel time and cost). The changes in departure time ranged from 30 minutes earlier to 30 minutes later than usual. Furthermore, the charging levels varied from £2 to £5.50.

RELIABILITY OF DEPARTURE TIME MODEL ESTIMATES

In this paper five randomly selected samples were drawn from the full data set and were used to investigate the reliability of the model. Each data segment is made up of 126 observations from the full data set. Each respondent provided seven observations. Coefficient estimates of departure time choice models for each of the segments were calibrated. Table 1 shows the coefficient estimates, the ρ^2, and the likelihood values for each of the segments as well as for the models estimated with the full data.

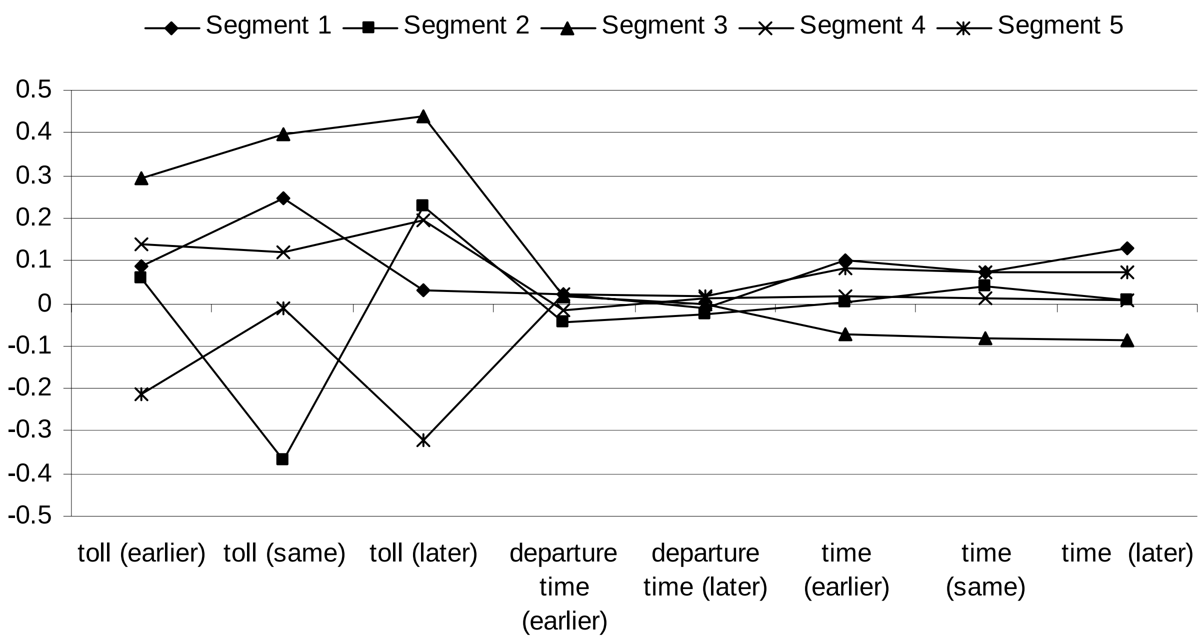
Table 1 Departure time model results

Variables		**Coefficient values (*t-ratios*)**					
		Segment 1	**Segment 2**	**Segment 3**	**Segment 4**	**Segment 5**	**Full model**
Toll	*Earlier*	-.9715 (*-5.3*)	-.9417 (*-5.1*)	-1.177 (*-5.9*)	-1.022 (*-5.7*)	-.6674 (*-4.0*)	-.8832 (*-11.6*)
	Same	-1.090 (*-5.6*)	-.4751 (*-3.2*)	-1.239 (*-5.6*)	-.9620 (*-5.4*)	-.8336 (*-4.7*)	-.8430 (*-11.3*)
	Later	-.9431 (*-5.0*)	-1.142 (*-4.5*)	-1.355 (*-6.8*)	-1.108 (*-5.6*)	-.5904 (*-4.0*)	-.9142 (*-11.7*)
Departure time	*Earlier*	-.04103 (*-1.4*)	.02171 (*0.8*)	-.03883 (*-1.3*)	-.004503 (*-0.2*)	-.04414 (*-1.8*)	-.02187 (*-1.8*)
	Later	-.02564 (*-0.8*)	-.01323 (*-0.4*)	-.03672 (*-1.2*)	-.04745 (*-1.5*)	-.05528 (*-2.2*)	-.03701 (*-2.9*)
Time	*Earlier*	-.2337 (*-2.2*)	-.1353 (*-1.6*)	-.05932 (*-0.7*)	-.1500 (*-1.5*)	-.2156 (*-3.0*)	-.1326 (*-3.7*)
	Same	-.2060 (*-2.0*)	-.1716 (*-2.0*)	-.05105 (*-0.6*)	-.1447 (*-1.5*)	-.2065 (*-2.9*)	-.1324 (*-3.7*)
	Later	-.2586 (*-2.4*)	-.1353 (*-1.6*)	-.04216 (*-0.5*)	-.1364 (*-1.4*)	-.2039 (*-2.9*)	-.1290 (*-3.6*)
$\rho^2(0)$		.3046	.2938	.3731	.3269	.1901	.2437
$\rho^2(c)$		.2879	.2451	.3721	.3032	.1751	.2377
Final likelihood		-96.2588	-97.7614	-86.7849	-93.169	-112.1058	-525.0961
Initial likelihood		-138.4251	-138.4251	-138.4251	-138.4251	-138.4251	-694.3230
N		126	126	126	126	126	632

From the table it can be seen that only the full model and the fifth segmented model have all coefficient estimates statistically significant at the 90% confidence level. In terms of the overall significance of the models (i.e. ρ^2(c) value) the third segmented model has the highest ρ^2(c) value amongst all of the models while the fifth segmented model has the lowest values.

In this paper the full data set model is assumed to be the true model. A comparison of the coefficient estimates of each of the segmented models was made against the true model (see Figure 1).

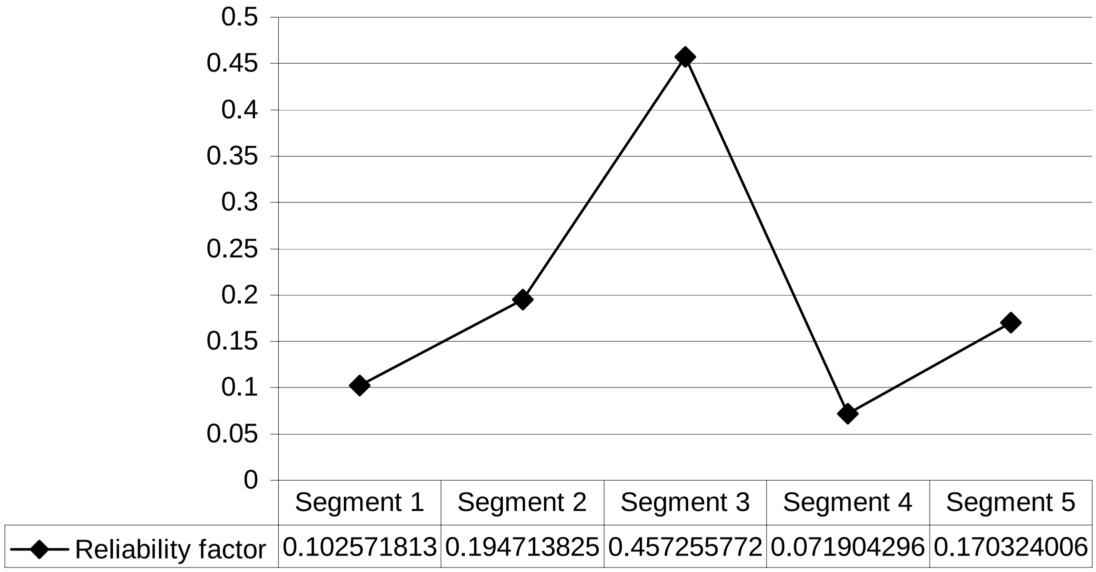
Figure 1 Differences in coefficient estimates between the true model and the five segmented models



A reliability indicator has been derived to investigate the variations in the coefficient estimates. The sum of the squares of the differences between the coefficient estimates of the true model and those of the segmented models were calculated. This indicator shows a range of variations in the estimates (see Figure 2). From the graph it appears that segmented model 4 has the lowest indicator while segmented model 3 has the highest indicator value. That would suggest that model 4 is the most reliable of the segmented models.

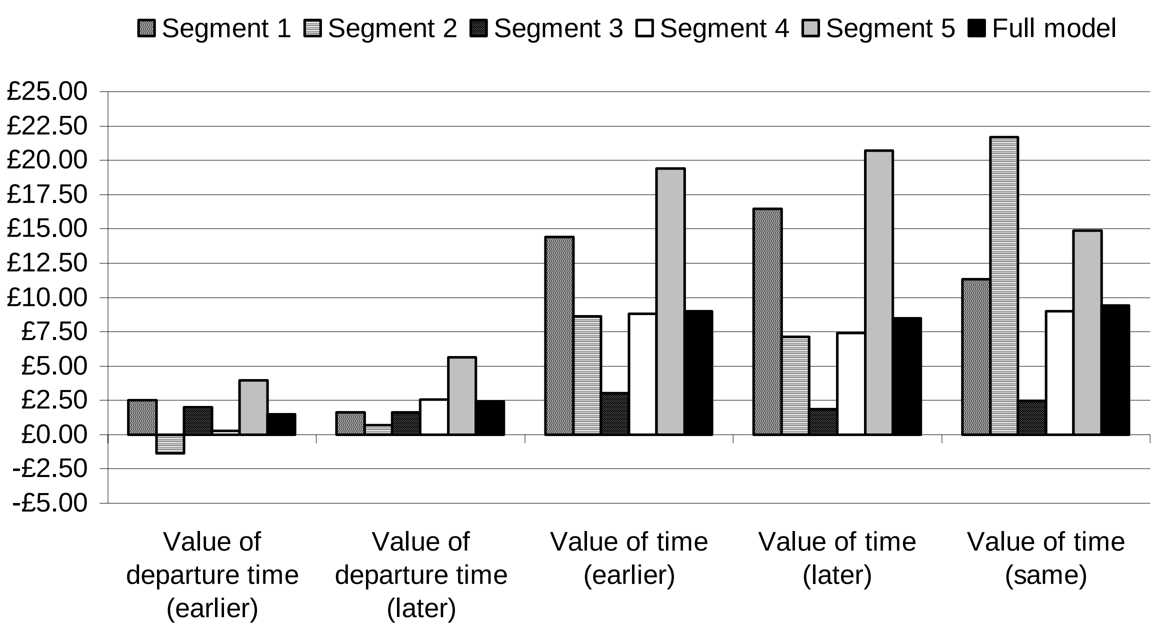
In the absence of the reliability indicator the most obvious model to be selected for carrying out predictions would be segmented model 1, based on the significance of the coefficient estimates as well as the reasonableness of the ρ^2(c) value. Or otherwise it would have been segmented model 3 with the highest ρ^2(c) value. This will have obvious implications on the predictions of these models.

Figure 2 Reliability indicators for the five segmented models



Values of time and values of departure time were estimated for each of the models (see Figure 3). From this data, it seems that the closest VOT and values of departure times to the true model (the model of the total data set) are for segmented model 4, which is similar to the findings based on the reliability indicators. Further research is needed in this area.

Figure 3 Values of departure time and VOT for all models



CONCLUSION

Discrete choice analysis is extensively used in the transport field to represent the choices made between a finite set of alternatives, for example, a set of alternative departure times. More specifically, it is often used to investigate and forecast individual travel behaviour.

There are a number of problems that can be associated with data collection, sampling methods and types of data used in models. These problems have been investigated and a number of correction factors have been suggested for a number of them. In this paper five randomly selected samples were drawn from a larger data set and were used to investigate the reliability of the forecasting models using the segmented models relative to the full data set model. All segments had the same sample size and the same number of observations per respondent. A reliability indicator was derived to investigate the variations in the coefficient estimates. The indicators show a range of variations in the estimates. Further research is therefore recommended in this area.

REFERENCES

Bates, J.J, and Terzis, G. (1997). Stated preference and the ecological fallacy, *Proceedings of the 25th European Transport Forum*, Brunel University, UK.

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Mass.

Bradley, M., and Daly, A. (1993). New analysis issues in stated preference research. *PTRC 21st Summer Annual Meeting*, proceedings of seminar D, University of Manchester Institute of Science and Technology, UK.

Farrell, S. (2004). *Congestion charging levels: Technically optimal or politically feasible?* To be submitted, Unpublished PhD Theses, Napier University.

Ortúzar, J. de D., and Willumsen, L.G. (2001). *Modelling Transport* (3rd edition), Wiley, UK.