WORKING PAPER

ITLS-WP-11-14

Does the choice model method and/or the data matter?

By

David A Hensher, John M Rose and Zheng Li

July 2011

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-11-14

TITLE: **Does the choice model method and/or the data matter?**

ABSTRACT: The opportunity to have seven data sets associated with a stated choice experiment that are very similar in content and design is rare, and provides an opportunity to look in detail at the empirical evidence within and between each data set in the context of a range of discrete choice estimation methods, from multinomial logit to latent class to scale multinomial logit to mixed logit, and the most general model, generalized mixed multinomial logit that accounts for preference and scale heterogeneity. Given the problems associated with data from different countries and time periods, we estimate separate models for each data set, obtaining values of travel time savings that are then updated post estimation to a common dollar for comparative purposes. We also pooled all data sets for a scaled MNL model, treating each data set as a set of three separate utility expressions, but linked to the other data sets through scale heterogeneity. This is not behaviourally appropriate with MNL, latent class or mixed logit. The main question investigated is whether there exists greater synergy in the willingness to pay evidence within model form across data sets compared to across model forms within data sets. The evidence suggests that there is a relatively greater convergence of evidence across the choice models, with the exception of generalized mixed logit, after controlling for data set differences; and there is strong evidence to suggest that differences between data sets do matter.

KEY WORDS: ***Multiple data sets; stated choice; discrete choice models; scale; MNL; latent class; mixed logit; scale MNL; generalized mixed logit; Australia; New Zealand.***

AUTHORS: David A Hensher, John M Rose and Zheng Li

Acknowledgements: Discussions with Joffre Swait are appreciated.

CONTACT: INSTITUTE of TRANSPORT and LOGISTICS STUDIES (C37)
The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9351 0071
Facsimile: +612 9351 0088
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

DATE: July 2011

1. Introduction

This paper is motivated by the often asked question as to whether there exists greater synergy in the willingness to pay (WTP) evidence within model form across comparable data sets compared to across model forms within data sets. The opportunity to investigate this issue is rare because of the lack of a number of data sets that are similar. We have collected, over a 10 year period, in Australia and New Zealand, seven very similar data sets whose centerpiece is a stated choice (SC) experiment described by three unlabelled alternatives and 16 choice sets in the context of a sample of commuters choosing between a bundle of trip time and cost attributes associated with tolled and non-tolled routes. In all data sets the SC design is a pivot design, in which the reference (or status quo) alternative is a representation of a recent commuting trip, and the two other alternatives are constructed as variations around the attribute levels associated with the experienced trip. All data sets include running cost and toll, as well as free flow; however non-free flow time is defined as either a single attribute 'congested time' or two separate attributes 'slowed down time' and 'stop/start/crawling time'. An overview of each data set is presented in Table 1.

In this paper we estimate five classes of choice models: multinomial logit (MNL), latent class (LC), scaled multinomial logit (SMNL), mixed multinomial logit (MMNL), and the most general model, generalized mixed multinomial logit (GMMNL) that accounts for preference and scale heterogeneity. We also test for constrained and unconstrained distributions associated with the random parameters in MMNL and GMMNL. The main output of interest is the valuation of travel time savings, which initially is estimated without any adjustment in the time period and country, but *ex post* estimation, is converted to a common currency time period $AU2010. Using the conversion for the cost attributes in *ex ante* estimation is not appropriate since it runs the risk of distorting the behavioural trade-offs that are evident at the time of each survey.

The paper is organized as follows. We begin with an overview of the seven data sets, and then present the estimated models. The values of travel time savings are then calculated and contrasted within and between data sets with respect to each of the choice models. A multivariate model is presented to reveal the underlying structural relationships between the evidence that form the basis of answering the question posed on whether there exists greater synergy in the WTP evidence within model form across data sets compared to across model forms within data sets. Conclusions highlight the main contribution.

2. Data sources

Seven SC data sets from five Australian and two New Zealand tollroad studies that were conducted between 1999 and 2008 are used in this paper. The choice experiments involved each sampled respondent[1] answering 16 choice scenario questions. In each choice question, the respondent was required to make a choice among three alternatives, one described by a recent trip and two alternatives defined by attribute levels pivoted off of the recent (or reference) trip profile. Pivoting offers more realism in the stated preference experiment since hypothetical alternatives are defined relative to the reference alternative (status quo), giving better specificity in the context of the choice task (Train and Wilson 2008). The seven surveys were conducted as computer aided personal interviews (CAPI). We briefly describe each data set below, with further details given in Appendix A.

[1] In some data sets, both car commuters and non-commuters were sampled. This paper focuses on car commuters.

Table 1: A summary of seven tollroad studies

	Number of sampled car commuters **	Year of data	Country	Recent toll used in reference trip (proportion of sample)	Attributes					Socioeconomics Variables				
					Free flow minutes (mins)	Slowed down time (mins)	stop/start/ crawling (mins)	Congestion time (mins)	Running cost ($)	Toll Cost ($)	Age	Gender (proportion female)	Annual Personal Income (thousands)	Hours worked per week
Study 1	280	2008	Australia	0.36%	13.29 (11.43)	11.52 (9.82)	13.59 (14.59)	-	2.95 (2.30)	5.80*	39.44 (13.01)	0.43	54.75 (30.76)	37.79 (13.65)
Study 2	147	2000	Australia	33.30%	23.35 (15.52)	-	-	16.78 (15.95)	2.72 (2.15)	3.05 (0.82)	42.48 (9.84)	0.31	82.82 (38.27)	39.94 (15.60)
Study 3	152	1999	New Zealand	0%	13.76 (19.82)	5.99 (7.89)	4.18 (5.21)	-	1.98 (3.93)	-	40.74 (10.22)	Not Collected	26.51 (11.15)	34.29 (15.65)
Study 4	304	2005	Australia	7.89%	12.19 (9.73)	16.66 (9.74)	10.92 (11.09)	-	1.76 (1.33)	2.23 (0.34)	42.30 (11.86)	0.34	57.56 (28.23)	41.04 (13.77)
Study 5	57	2004	Australia	33.30%	38.86 (20.83)	-	-	34.93 (19.95)	7.03 (3.75)	4.28 (2.02)	42.73 (11.70)	0.11	82.55 (34.32)	42.56 (12.81)
Study 6	243	2004	Australia	75.30%	22.53 (12.35)	-	-	31.80 (19.28)	2.44 (1.66)	2.72 (1.36)	41.70 (11.26)	0.36	87.46 (33.41)	41.91 (11.60)
Study 7	115	2007	New Zealand	0%	27.96 (16.31)	-	-	9.89 (9.73)	4.38 (2.43)	-	48.02 (12.26)	0.63	48.10 (24.57)	41.92 (13.83)

Notes: the standard deviation is given in parenthesis. * Only one sampled car commuter paid toll. ** Model estimation has a multiple of 16 times the number of observations.

Study 1

The first study is the most recent, undertaken in 2008, using a D-efficient experimental design structured to increase the statistical performance of models with relatively smaller samples than are required for other less-efficient (statistically) designs such as orthogonal designs (see e.g., Rose *et al.* 2008). In total, 280 car commuters (with less than 120 minutes' trip length) were sampled.

The three alternatives shown in each choice set were described in terms of free flow time, slowed down time, stop/start/crawling time, running cost, toll cost, and travel time variability. Compared to the other data sets, this data set (Study 1) is unique in terms of how travel time variability[2] is portrayed, where each alternative has three travel scenarios - 'arriving *x* minutes earlier than expected', 'arriving *y* minutes later than expected', and 'arriving at the time expected'. Each time is associated with a corresponding probability[3] of occurrence to indicate that travel time is not fixed but varies from time to time. For all attributes except the toll cost, minutes arriving early and late, and the probabilities of arriving on-time, early or late, the values for the SC alternatives are variations around the values for the current trip.

In contrast to Study 1 (see Figure 1), where three arrival scenarios along with their probabilities of occurrence for a trip were presented in the choice experiments, the other six studies defined the trip time variability attribute as *plus* or *minus* a level of trip time associated with a trip (see Figure 2 for an example). Despite the innovation in the trip variability attribute in data set 1, we exclude trip time variability in all model estimation given previous evidence using these data sets that, with the exception of data set 1, the variability attribute was poorly specified and often not statistically significant.

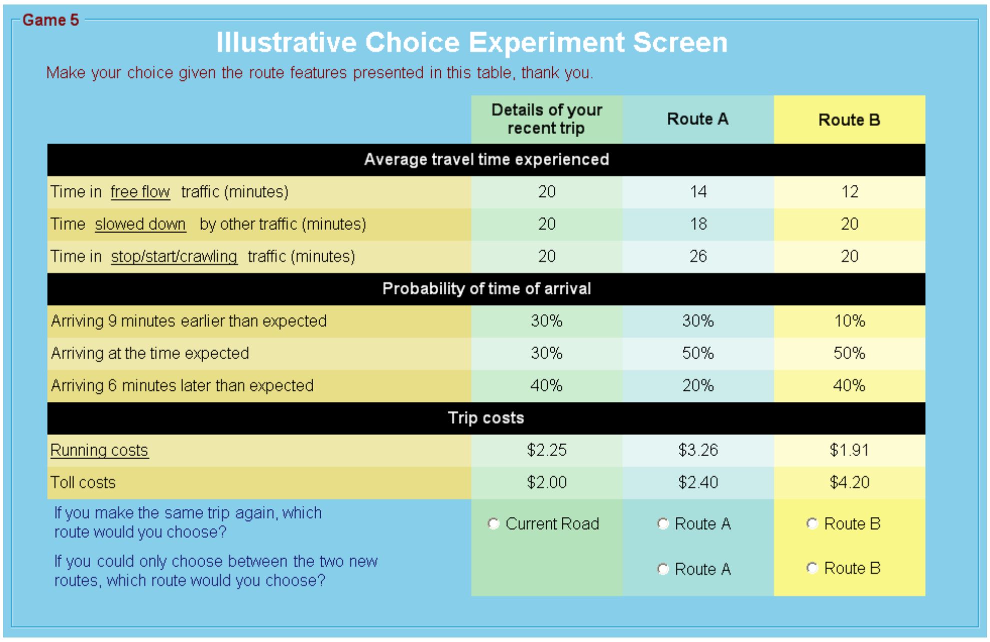
Studies 2-7

Study 2 (Australia, 2000), Study 3 (New Zealand, 1999), Study 4 (Australia, 2005), Study 5 (Australia, 2004), Study 6 (Australia, 2004) and Study 7 (New Zealand, 2007) have used a survey similar to that shown in Figure 2. With the exception of Studies 2 and 3 where an orthogonal design was used, a D-efficient design was used. For all studies, the trip cost is disaggregated into the running cost and the toll cost. Studies 3 and 4 have three time components, i.e., free flow time, slowed down time, and stop/start/crawling time; while the last two components are combined into congestion time in Studies 2, 5, 6 and 7.

The sample size for car commuters ranges from 57 (Study 5) to 280 (Study 1). Given that the sampled New Zealand's car commuters had no tolling experience before they were interviewed, toll costs are only available for Australian studies.

[2] Given this difference, the paper focuses only on the value of travel time savings.

[3] The probabilities are designed and hence exogenously induced to respondents, similar to other travel time reliability studies.


Game 5

Illustrative Choice Experiment Screen

Make your choice given the route features presented in this table, thank you.

	Details of your recent trip	Route A	Route B
Average travel time experienced			
Time in free flow traffic (minutes)	20	14	12
Time slowed down by other traffic (minutes)	20	18	20
Time in stop/start/crawling traffic (minutes)	20	26	20
Probability of time of arrival			
Arriving 9 minutes earlier than expected	30%	30%	10%
Arriving at the time expected	30%	50%	50%
Arriving 6 minutes later than expected	40%	20%	40%
Trip costs			
Running costs	$2.25	$3.26	$1.91
Toll costs	$2.00	$2.40	$4.20
If you make the same trip again, which route would you choose?	Current Road	Route A	Route B
If you could only choose between the two new routes, which route would you choose?		Route A	Route B

Figure 1: Type 2 design used in Study 1


Practice Game

Make your choice given the route features presented in this table, thank you.

	Details of your recent trip	Route A	Route B
Time in free flow traffic (minutes)	15	21	12
Time slowed down by other traffic (minutes)	10	10	8
Time in stop/start/crawling traffic (minutes)	2	2	3
Trip time variability (minutes)	+/- 8	+/- 9	+/- 8
Taxi fare	$30.70	$27.63	$18.42
Toll costs	$4.00	$0.00	$0.70
If you make the same trip again, which route would you choose?	Current Road	Route A	Route B
If you could only choose between the two new routes, which route would you choose?		Route A	Route B

Figure 2: An example of Type 1 design used in other studies

3. Model estimation and evidence

Table 2 provides a summary of the overall goodness of fit of each of the models that are set out in detail in Appendix B Table B1 (seven parts). Table 3 summarises the scaled MNL model in which all data sets are combined in a model with 21 (7 by 3) utility expressions, three per data set.

At the outset it is important to indicate that we chose the same set of attributes and functional form for inclusion in all models and data sets for the following reasons[4]. Firstly, we opted for a

[4] A referee made the comment that "...you are likely to obtain similar results across similar studies if a consistent model specification is used. However, this consistent specification is by definition a compromise solution, and if the optimal specification was used for each dataset, then larger differences may well arise." While we do not disagree with the general position, which would have to be proven, we have opted for the consistent model for the reasons explained in the text. The development of 'optimal' models, 49 in total, would be a worthy separate study.

common functional form so that we could compare the models without confounding the interpretation. Secondly, it is common in practice for analysts to estimate simpler models, and then to use the same variables in a more complex model where preference and/or scale heterogeneity effects are accounted for or latent classes. Thirdly, we excluded an attribute where it was only available in a particular form in one data set (e.g., trip time variability in data set 1), but not until we checked whether its exclusion would have an impact on the WTP estimates of interest[5]. Finally, we took the decision to exclude socioeconomic effects after we found that the value of travel time savings (VTTS) estimates within each data set for the various models did not change as a result of the independent addition of income and gender[6]. Importantly what we found when including income is that it did not significantly change the mean VTTS estimate, and hence the relative evidence on VTTS does not change.

Before presenting the main findings, we provide an overview of the more advanced methods that are now available to account for scale and taste heterogeneity[7].The SMNL and GMMNL models build on the specifications of the MMNL model developed in Train (2003) and Hensher and Greene (2003) amongst others, and the GMNL model first operationalised in Fiebig *et al.* (2009). The MMNL model version of interest is:

$$\text{Prob}(choice_{it} = j \,|\, \mathbf{x}_{it,j}, \mathbf{z}_i, \mathbf{v}_i) = \frac{\exp(V_{it,j})}{\sum_{j=1}^{J_{it}} \exp(V_{it,j})} \quad \textbf{(1)}$$

where

$V_{it,j} = \boldsymbol{\beta}_i'\mathbf{x}_{it,j}$

$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{v}_i$

$\mathbf{x}_{it,j}$ = the K attributes of choice j in choice situation t faced by individual i,

$\mathbf{v}_i$ = a vector of K random variables with zero means and known (usually unit) variances and zero covariances.

---

5 For data set 1, we estimated the model with and without travel time variability. Given that the scheduling model (see e.g., Small et al. 1999) is the state of practice model for valuing travel time variability and VTTS, a series of scheduling models with travel time variability variables (i.e., expected schedule delay early/late) were estimated. For example, the mean VTTS under the scheduling model within a MMNL framework with constrained triangular distributions is $21.02 per person hour, which is similar to $21.39 per person hour reported in Table 4 from the MMNL model without variability. Using the Wald test, we calculated the confidence levels for the two mean VTTS, where the VTTS from the scheduling model with travel time variability is within the range of $17.17-24.74 per person hour, and the VTTS from the model without variability is within the range of $17.37-25.30 per person hour. Hence we can reject that these two mean VTTS are statistically significantly different at the 95 percent confidence interval. The VTTS under the scheduling model within a latent class model (two classes) is $11.93 per person hour in contrast to $12.70 per person hour reported in Table 4. Under the 95 percent confidence interval, the mean VTTS from the model with travel time variability is within the range $7.09-14.59 per person hour, and the mean VTTS from the model without income within the range $9.72-15.49 per person hour. Again, we can reject that the hypothesis that the two means are statistically significantly different at the 95 percent confidence interval. This enabled us to use the same set of attributes (recognising the two attributes were combined in some data sets), in the comparisons.

6 To illustrate, we estimated the models which have the income variable in the utility function for the reference alternative, and found that income has a marginal effect on the estimated VTTS and model fit.

Using Dataset 1 as an illustrative example, the mean VTTS under the MMNL model (constrained triangular) with income is $20.83 per person hour, which is similar to $21.39 per person hour reported in Table 4 from the MMNL model without the income variable; and this MMNL model with income has a Bayes Information Criterion (BIC) value of 1.4934 (vs. 1.4843 from the corresponding model without income). Using the Wald test, we also calculated the confidence levels (95 percent) for two mean VTTS, where the VTTS from the model with income is within the range of $17.09-25.11 per person hour, and the VTTS from the model without income is within the range of $17.37-25.30 per person hour. Hence we can reject that these two mean VTTS are statistically significantly different at the 95 percent confidence interval. The latent class model (two classes) with income produces the VTTS of $13.41 per person hour, which is slightly higher than $12.70 per person hour reported in Table 4, where the BIC value for this latent class model with income is 1.2451 (vs. 1.2319 from the corresponding model without income). For the 95 percent confidence interval, the mean VTTS from the model with income is within the range of $9.32-20.84 per person hour, and the mean VTTS from the model without income is within the range of $9.72-15.49 per person hour. Again, we can reject the hypothesis that the two means are statistically significantly different at the 95 percent confidence interval.

7 For an overview of latent class and MMNL models, see Greene and Hensher (2003).

The unobserved heterogeneity in the preference parameters of individual i is embodied in $\mathbf{\Gamma v}_i$. Structural parameters are the constant vector, $\boldsymbol{\beta}$, the $K\times M$ matrix of the nonzero elements of the lower triangular Cholesky matrix, $\mathbf{\Gamma}$. A number of interesting special cases are straightforward modifications of the model. Specific non-random parameters are specified by rows of zeros in $\mathbf{\Gamma}$.

Scale heterogeneity across choices can be built into the model by random alternative-specific constants. The preceding is modified as equation (2) (see Feibig *et al.* 2009, Keane 2006, Greene and Hensher 2010).

$$\boldsymbol{\beta}_i = \sigma_i\boldsymbol{\beta} + \sigma_i \mathbf{\Gamma v}_i \quad \textbf{(2)}$$

The additional terms not yet defined are:

σ_i = $\exp[\bar{\sigma} + \tau w_i]$, the individual specific standard deviation of the idiosyncratic error term

w_i = the unobserved heterogeneity, standard normally distributed

$\bar{\sigma}$ = a mean parameter in the variance

τ = the coefficient on the unobserved scale heterogeneity

The model in all forms is estimated by maximum simulated likelihood. Fiebig *et al.* (2009) and Greene and Hensher (2010) discuss details of normalisations required for identification. In particular: (i) to identify $\bar{\sigma}$ which is not identified separately from τ, we normalize σ_i so that $E[\sigma_i^2] = 1$, by setting $\bar{\sigma} = -\tau^2/2$ instead of zero; (ii) to ensure $\tau \geq 0$, the model is fit in terms of λ, where $\tau = \exp(\lambda)$ and λ is unrestricted.

An extension of interest herein is to allow τ to be a function of a series of dummy variables that identify the presence of scale heteroscedasticity between different data sets (the model in Table 3). This is a simple but important extension as follows: $\tau = \tau + \eta d_s$ where η is a data-set specific scale parameter and d_s =1 for data set s and zero otherwise, with s=1, 2…, S-1. Combining all terms, the simulated log likelihood function for the sample of data is shown in equation (3) (See Greene and Hensher 2010).

$$\log L = \sum\nolimits_{i=1}^{N} \log\left\{\frac{1}{R}\sum\nolimits_{r=1}^{R} \prod\nolimits_{t=1}^{T_i} \prod\nolimits_{j=1}^{J_{it}} P(j,\mathbf{X}_{it},\boldsymbol{\beta}_{ir})^{d_{it,j}}\right\} \quad \textbf{(3)}$$

where

$\boldsymbol{\beta}_{ir}$ = $\sigma_{irs}\boldsymbol{\beta} + \sigma_{irs}\mathbf{\Gamma v}_{ir}$,

σ_{irs} = $\exp[-(\tau + \eta d_s)^2/2 + (\tau + \eta d_s)w_{ir}]$,

$\mathbf{v}_{ir}$ and w_{ir} = the R simulated draws on v_i and w_i,

d_{itj} = 1 if individual i makes choice j in choice situation t and 0 otherwise,

and

$$P(j,\mathbf{X}_{it},\boldsymbol{\beta}_{ir}) = \frac{\exp(\mathbf{x}'_{it,j}\boldsymbol{\beta}_{ir})}{\Sigma_{j=1}^{J_{it}}\exp(\mathbf{x}'_{it,j}\boldsymbol{\beta}_{ir})} \quad \textbf{(4)}$$

In models with random parameters we have allowed for preference heterogeneity in all travel time attributes, but imposed preference homogeneity on the cost parameter. Although this may have some impact on other model results, it is the belief that in simulating VTTS, treating the cost parameter as random will have a far worse impact on the evidence as shown by Daly *et al.* (2010) if both the numerator and denominator are drawn from random parameter distributions. They state "…we show that some popular distributions used for the cost coefficient in random coefficient models, including normal, truncated normal, uniform and triangular, imply infinite moments for the distribution of WTP, even if truncated or bounded at zero. [And]…that relying on simulation approaches to obtain moments of WTP from the estimated distribution of the cost and attribute coefficients can mask the problem by giving finite moments when the true ones are infinite." The authors also show that using finite mixture such as latent class models can assure that the distribution of WTP has finite moments.[8]

The best fit model form on log-likelihood and Bayes information criterion (BIC) varies between the data sets, with latent class outperforming other models for four of the seven data sets[9]. As expected, MNL is not a good performer, with significant improvement in fit as we move to SMNL, MMNL) and GMMNL. Figure 3 presents the difference between the log-likelihood at convergence of each model relative to the best fit model for each data set. In addition to the MNL models, we see that the SMNL model and the mixed logit model with constrained distributions on the parameters have the greatest deviation from the best fitting model, although the exception is data set 3 where the SMNL model is the best fit. There is clear evidence that in all cases, a MMNL model with unconstrained random parameter distribution is a significant improvement in fit over MMNL with a constrained distribution, despite behavioural concerns about sign changes in parameter estimates across the full distribution.

The statistical gains in moving from SMNL to GMNL are extremely variable, with noticeable improvement in fit in some data sets (i.e., DS1, DS2, DS6), but little gain otherwise, but with SMNL noticeably better on DS3. The evidence to support gains by including both scale and preference heterogeneity in GMMNL is not overwhelming in general, even recognising that the efforts to find the preferred GMMNL model requires careful selection of starting values and evaluation of the number of draws. We have found that using starting values from the equivalent MMNL model (in contrast to MNL estimates on a subset of comparable parameters) is far superior in estimation time and in picking a 'winning' set of parameter estimates.

[8] We thank a referee for raising this point, although the referee suggested we estimate all relevant models with random parameters for cost.

[9] In all data sets a two-class latent class model had the best BIC.

Table 2: Summary of overall goodness of fit of all models

	DS1 Australia 2008	DS2 Australia 2000	DS3 NZ 1999	DS4 Australia 2005	DS5 Australia 2004	DS6 Sydney 2004	DS7 NZ 2007
	Log-likelihood						
MNL	-3435.78	-1867.75	-1714.52	-2670.61	-855.87	-3037.75	-1639.61
SMNL stand alone	-2826.91	-1512.91	***-1378.93***	-2371.24	-808.32	-2903.67	-1466.58
Latent Class	***-2713.27***	-1393.81	-1380.96	-2335.29	***-739.06***	***-2759.91***	***-1237.06***
MMNL (t,1)	-3303.25	-1478.43	-1700.81	-2568.82	-831.86	-2886.65	-1570.45
MMNL (t)	-2961.94	-1364.55	-1637.34	-2412.04	-802.73	-2816.82	-1533.71
GMNL (t,1 or n,1)	-2783.69	-1396.92	-1627.66	***-2286.04***	-789.82	-2799.32	-1454.39
GMMNL (t)	-2955.97	**-1350.83**	-1630.61	-2334.16	-792.12	-2794.24	-1535.37
	Bayes Information Criterion (BIC)						
MNL	1.5432	1.6014	1.4262	1.1068	1.9068	1.5711	1.7985
SMNL stand alone	1.2733	1.3029	**1.1532**	0.9855	1.8099	1.5043	1.6145
Latent Class	***1.2319***	1.2149	1.1709	0.9794	***1.6882***	***1.4388***	***1.3814***
MMNL (t,1)	1.4843	1.2704	1.4147	1.0651	1.8542	1.4934	1.7331
MMNL (t)	1.3372	1.1801	1.3721	1.0058	1.8052	1.4617	1.6916
GMMNL (t,1 or n,1)	1.2559	1.2077	1.3611	***0.9522***	1.7769	1.4527	1.6054
GMMNL (t)	1.3384	***1.1751***	1.3732	0.9772	1.7969	1.4544	1.7016

Note: best model associated with each data set is bolded and italicised (t,1) is a constrained triangular distribution, (t) is an unconstrained distribution, and (n,1) is a constrained normal distribution, 500 Halton draws

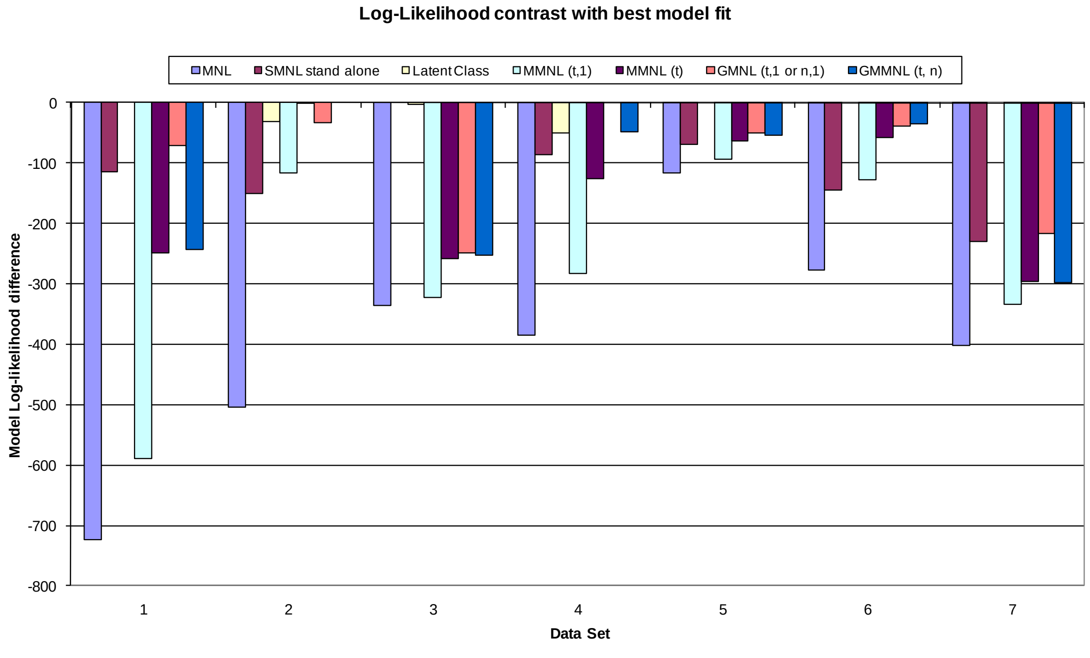


Figure 3: Log-likelihood differences relative to best fit model

In Table 3, we present the SMNL results for a single model using all data sets, each with their own utility expressions, but recognising heterogeneity in scale overall and heteroscedastic differences (in the GMMNL scale factor) between data sets. We have included reference-specific constants for each data set as well as data-set specific parameter estimates for slowed down time, stop start time and congested time; however we have constrained the parameter estimate for free flow time to be generic across all data sets. This is a requirement explained in Louviere *et al*. (2000) as data enrichment in which pooling the seven choice data sources necessitates an equality restriction on at least one common parameter, while controlling for the scale factors.

Table 3: A summary of modelling results for the pooled SMNL model (seven data sets)

	DS1	DS2	DS3	DS4	DS5	DS6	DS7
Parameters:							
Reference constant	1.249 (-13.86)	0.4397 (12.55)	0.4323 (7.62)	0.3128 (7.94)	-0.6297 (-6.55)	0.0901 (3.60)	0.3689 (11.45)
Free flow time	-0.1405 (-26.99)	-0.1405 (-26.99)	-0.1405 (-26.99)	-0.1405 (-26.99)	-0.1405 (-26.99)	-0.1405 (-26.99)	-0.1405 (-26.99)
Slowed down time	-0.1195 (-9.71)	-	-0.1138 (-6.53)	-0.3476 (-20.82)	-	-	-
Stop start time	-0.1383 (-14.37)	-	-0.2575 (-11.62)	-0.2793 (-19.99)	-	-	-
Congested time	-	-0.1381 (-21.74)	-	-	-0.1393 (-11.33)	-0.1462 (-21.08)	-0.1651 (-17.10)
Cost	-0.7744 (-13.16)	-0.7267 (-24.64)	-1.2088 (-10.52)	-1.2103 (-20.86)	-0.2814 (-11.97)	-0.6699 (-22.56)	-0.8988 (-23.61)
Heteroscedasticity in GMMNL scale factor	-0.0005 (-0.01)	0.2568 (4.02)	0.4088 (3.87)	0.5039 (10.63)	-0.5735 (-3.62)	0.1353 (2.42)	-
Variance parameter in scale	1.0130 (74.35)						
Model fit							
Log-likelihood Bayes Information Criterion (BIC)	13237.0 1.2901						
No. of observations	20768						

We have a statistically significant parameter estimate for the overall τ (t-ratio of 74.35). What this suggests is that we have identified the presence of unobserved scale heterogeneity when this is fed into the calculation of the standard deviation σ_{ir}, or the individual-specific standard deviation of the idiosyncratic error term, equal to $\exp(-\tau^2/2 + \tau w_{ir})$, assuming an estimate for $w_{ir,}$ the unobserved heterogeneity is standard normally distributed, the 'mean of the standard deviation' and the 'standard deviation of the standard deviation' are such that the overall influence is significantly different from unity. When we allow for heteroscedasticity in the GMMNL scale factor between the seven data sets, through the inclusion of a dummy variable for D-1 data sets i.e., $\exp(-(\tau + \eta d_s)^2/2 + (\tau + \eta d_s)w_{ir})$, we obtain the standard deviation of scale for data sets 1 to 7 as respectively 0.982 (1.35), 0.934 (1.66), 1.038 (3.62), 1.014 (2.90), 0.990 (0.45), 0.996 (1.81), and 0.989 (1.39). The figures in brackets are the standard deviations around the mean standard deviation of scale. Hence the mean range is 0.934 to 1.014, a maximum difference between pairs of data sets of 16 percent. Given the standard deviations of the scale standard deviations, we are not able to discern statistically significant scale differences of a heteroscedastic nature between the data sets, and all we can say is that scale heterogeneity is present but it is not systematically conditioned by any specific data set. This is an interesting finding and an encouraging one for future applications, since it suggests that the presence of scale heterogeneity in studies that combine RP and SC data (or SC data sets) is not spread across the data sets with systematic effects unique to the RP or SC data set. This finding, in many ways, is supportive of previous findings by Hensher *et al.* (2005) who has often claimed that scale differences are linked to specific alternatives (e.g., common public transport modes across RP and SC data) and not within specific data sets. Intuitively this makes sense, since scale is a property of unobserved variance and not of decision structures per se, and thus one expects unobserved influences that are common to the same alternative defined in an RP and an SC setting (e.g. bus and train vs. car) in contrast to differences in unobserved influences associated with SC vs. RP alternatives (i.e., bus, train, car in SC vs. bus, train, car in RP).

4. Willingness to pay output

The mean estimates of value of travel time savings (VTTS) are summarised in Table 4 for the study year and location, as well as in the converted common currency and period ($AU2010), and displayed graphically in Figure 4 for all estimates in $AU2010. The calculation of these estimates involved a weighted average of the VTTS estimates for each travel time component, where the weights represent the incidence of each component of travel time, for each respondent and alternative.

Table 4: Mean VTTS in study year and location ($AU2010 in parenthesis)

Data Set	MNL	SMNL stand alone (scaled)	SMNL pooled (scaled)	MMNL (t,1)	MMNL (t)	GMNL (t,1 or n,1)	GMNL (t)	Latent Class (2 classes)
1, 2008	13.88 (14.42)	10.48 (10.88)	10.15 (10.54)	21.39 (22.21)	14.31 (14.86)	40.82 (42.309)	**47.70** (49.53)	*12.70* (13.19)
2, 2000	11.06 (17.42)	12.67 (19.96)	10.76 (16.95)	17.56 (27.67)	*6.81* (10.73)	**27.45** (43.24)	8.12 (12.79)	11.42 (17.99)
3, 1999	10.79 (11.20)	*3.55* (3.69)	7.94 (8.24)	12.84 (13.33)	9.61 (9.97)	**12.99** (13.48)	12.44 (12.91)	5.60 (5.81)[10]
4, 2005	*13.00* (16.49)	28.00 (33.0)	14.51 (17.10)	18.83 (22.20)	17.80 (20.99)	46.70 (55.05)	**54.27** (63.97)	14.68 (17.30)
5, 2004	15.46 (17.76)	24.02 (27.58)	29.55 (33.93)	15.67 (17.99)	16.18 (18.58)	**36.94** (42.41)	22.49 (25.83)	*15.27* (17.54)
6, 2004	13.95 (16.02)	*10.52* (12.07)	12.82 (14.72)	16.78 (19.27)	14.98 (17.20)	20.81 (23.89)	**21.45** (24.63)	15.11 (17.35)
7, 2005	12.44 (11.02)	13.61 (12.06)	*9.70* (8.60)	14.66 (12.99)	12.65 (11.21)	**19.13** (16.96)	18.67 (16.55)	11.54 (10.23)

Note: bold estimates are the highest in the each data set, italicised are lowest estimates

There are significant variations in the estimates of mean VTTS, as might be expected. What is of particular interest are higher mean estimates from the GMMNL models in all data sets, while the lowest values are spread across all other model forms. Is there a story to tell here? It is far from clear what the behavioural implications are, although it is noteworthy that the mean estimates are more similar within MNL than within any other model form, and diverge most significantly under GMMNL, followed by SMNL. Empirical evidence seems to suggest that scale heterogeneity appears to exert a greater influence on producing differences in mean estimates of VTTS across studies than does preference heterogeneity (as accounted for in MMNL while ignoring scale heterogeneity). If as it appears, this is the empirical situation, then previous studies that have ignored scale heterogeneity have in effect increased the chance of transferability of VTTS when in fact this is misleading as a consequence of failing to recognise scale heterogeneity in the sampled population. The call to recognise that scale matters (e.g., Louviere and Eagle 2006) is certainly reinforced by the evidence herein.

[10] It may initially appear odd that the 2010 estimate is very similar to the 1999 estimate. This is New Zealand data and in $NZ 2010 we have $7.32, but after conversion to $AU2010, it declines significantly.

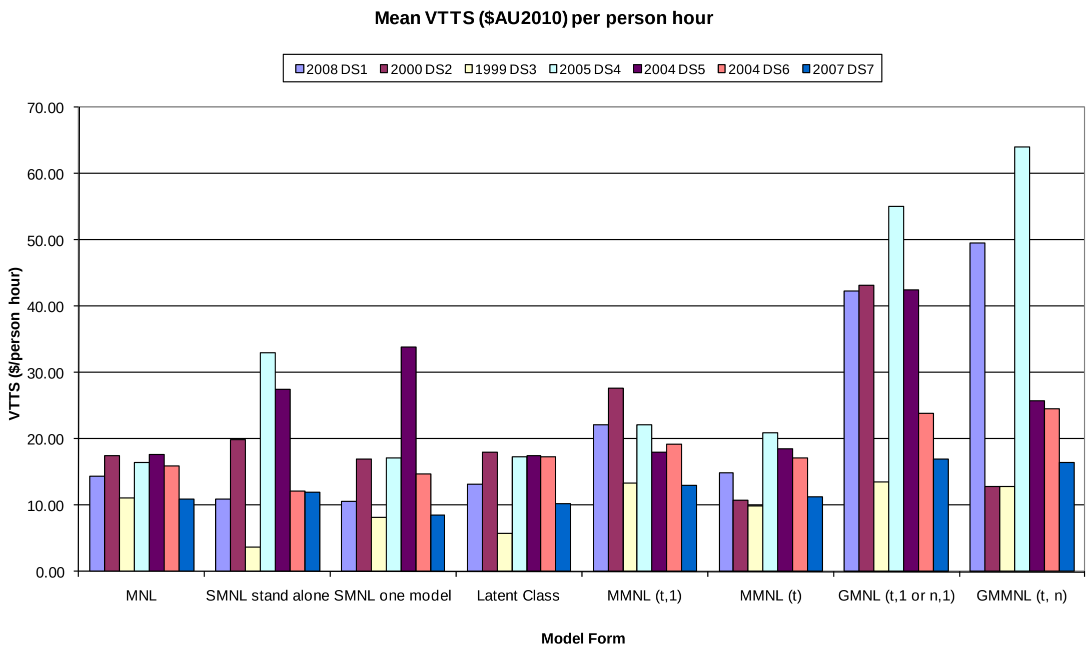


Figure 4: Summary of mean VTTS by model form and data set

We ran a series of linear regression models to establish the role of model and data set in the resulting weighted average estimates of VTTS, and found that the data set is a significant driving force once we controlled for the difference between the VTTS evidence for GMMNL compared to other model forms. The findings are summarised in Table 5. We have included additional variables to account for 'outliers' which improved the initial models, increasing the explanation of the variation in VTTS across the full sample from 44 percent to 76.4 percent. These interaction variables were associated with the GMMNL model and specific data sets 1 and 4, the MMNL (t,1) model and data set 4, and the latent class (LC) model and data set 5 (see Table 5). The resulting model and unstandardised residuals (Figure 5) suggest that once we control for differences between GMMNL and all other model forms, data set differences dominate the sources of systematic variation in mean VTTS, despite latent class and MMNL with constrained random parameter distributions being marginally significant in influence.

Table 5: Sources of Systematic Variation in VTTS

Dependent variable: VTTS
All variables are dummy (1,0) specifications

Explanatory variable	Parameter estimate	(*t*-ratio)	mean
Constant	10.6408	(5.32)	
MNL	-1.7616	(-0.92)	0.125
SMNL	0.3706	(0.16)	0.125
Latent Class	-3.1769	(-1.71)	0.125
MMNL (t,1)	3.6807	(1.77)	0.125
MMNL (t)	-1.8736	(-0.98)	0.125
GMNNL (t,1)	17.2517	(4.42)	0.125
Data set 1	5.2819	(2.33)	0.143
Data set 2	8.3921	(3.72)	0.143
Data set 3	-2.6238	(-1.03)	0.143
Data set 4	14.6366	(4.33)	0.143
Data set 5	10.790	(4.56)	0.143
Data set 6	5.6909	(2.38)	0.143
GMMNL(t)*DS1&4	36.1544	(12.63)	0.036
MMNL(t,1)*DS4	-6.7592	(-2.01)	0.018
LC*DS5	15.6784	(7.13)	0.018
Adjusted R^2	0.764		

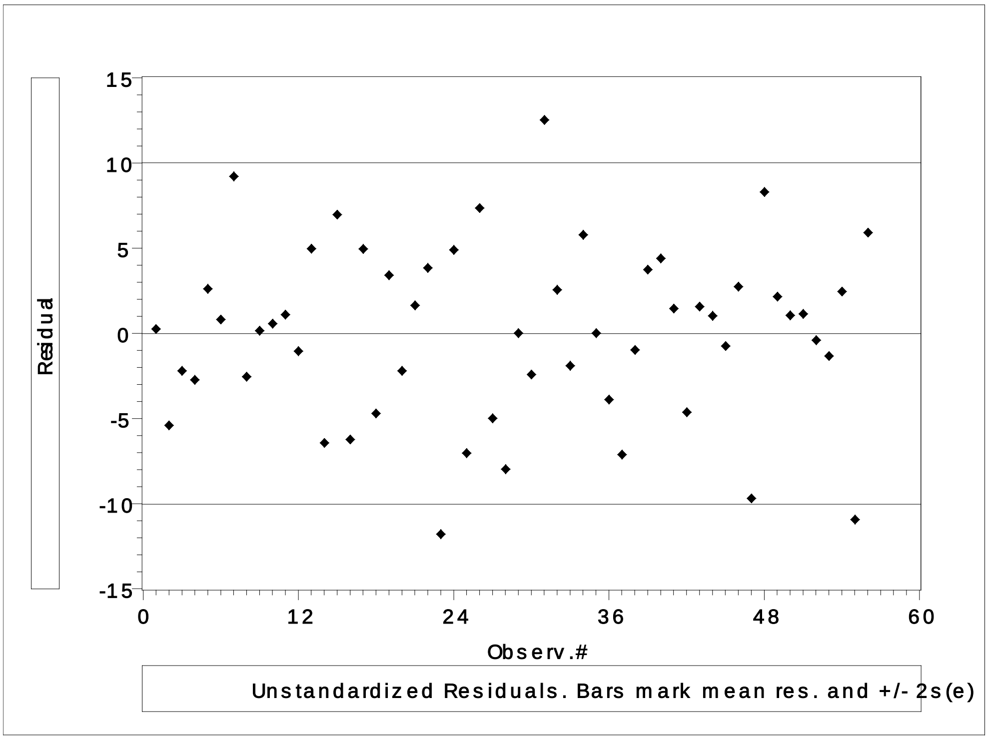


Figure 5: Unstandardised residuals of VTTS model

The key message is that the data set does make a difference, more than the choice of model does, except when using a GMMNL model compared to other choice models. Since GMMNL allows for preference and scale heterogeneity, this finding supports a view that allowing for this broadening of sources of heterogeneity does indeed make a significant difference in mean VTTS. This is a potentially important finding given that we have seven data sets and eight model forms.

Conclusions

We began with the question as to whether there exists greater synergy in the WTP evidence within choice model form across comparable data sets, compared to across model forms within data sets. We believe that the data set drives the differences much more than does the choice model used, with the exception of contrasts between GMMNL and all other choice models. This suggests that the differences in Table 5 across a row appear to be less noticeable than the differences within a column.

As part of the investigation into sources of systematic variation in VTTS, we included the choice model form as a series of explanatory dummy variables, and found that the difference between GMMNL and non-GMMNL models was statistically significant, and given the much higher mean estimates in Table 5, this is not surprising. What is of greater interest is a recognition that when studies using these data sets were undertaken in the past (i.e., in the year that the data was collected) the mean estimates recommended (using, in all situations, MMNL models with constrained triangular distributions) in the local currency at the time are respectively, for data sets 1 to 7, $18.39 (*$21.39*) (2008), $17.31 (*$17.56*) (2000), $10.20 (*$12.84*) (1999), $18.23 (*$18.83*) (2005), $24.88 (*$15.67*) (2004), $18.53 (*$16.78*)(2004) and $14.25 (*$14.66*) (2007)[11]. The equivalent MMNL mean estimates reported in this paper are given in brackets in italics. This range is nowhere inclusive of the GMMNL evidence for data sets 1, 2, 4, and 5. If we are willing to accept the evidence from the GMMNL models, not all of which are much higher (there being up to five exceptions in Table 4), the GMMNL model findings suggest that when we account for scale and preference heterogeneity, we obtain either significantly higher mean estimates or estimates at the higher end of a closer range between choice model forms within a data set. One might speculate that as models become more 'complex' in form, that there is greater variability in the mean estimates of VTTS between data sets. Indeed the evidence supports this. If we order the models by their 'complexity' we would have MNL, then Latent Class, then MMNL, followed by SMNL and then GMNL. We find that the standard deviation of mean VTTS across the seven data sets increases from MNL to GMMNL as per the following sequence (recognizing the small differences in LC and MLuc): 2.80 (MNL), 4.69 (LC), 4.30 (MMNLuc), 5.23 (MMNLc), 10.33 (SMNL), 15.72 (GMMNLuc) and 19.79 (GMMNLc), (noting that uc = unconstrained and c = constrained distributions), closely supporting the view we have adopted.

Thus, as one moves to behaviourally more complex or 'realistic' choice models, the ability to obtain empirical evidence on mean VTTS from one data set to use in the context of another data set, *given the selected choice model*, diminishes. There appear to be two quite startling thresholds, one between MNL and latent class or MMNL, and the other between latent class or MMNL and SMNL or GMMNL. An important question to ask is whether this finding is good or bad? To the extent that the more complex models are more sensitive to the many possible influences such as differences in choice sets, different reference points, different ranges in attribute levels, etc., one might expect these results, and such results should be applauded, at least from a theoretical point of view. Practitioners may not like this finding; however, it is time to start contemplating the nature and extent of differences in key outputs of the myriad of choice models now available to choose from, and to use the evidence from each model and data set to provide some guidance on the implications of choosing one model form and not another form.

There is clearly much research required to establish some unambiguous case for selection of a choice model form that is in a sense the 'best' from the available set. As models increase in complexity in the search for increased behavioural realism, we find that the evidence on willingness to pay becomes increasingly challenging in terms of selection of an appropriate set of values to use in practical applications. There is a sense in the worlds of research and consultancy that MMNL models should always be used on the argument that preference

[11] The difference between the recommended VTTS at the time and estimates herein is due to the inclusion of a few other variables as well as different numbesr of Halton draws where applicable.

heterogeneity matters. Now we have the matter of scale heterogeneity to contend with, which is a less obvious candidate for intuitive support, yet it appears to matter, despite the evidence in the presence of preference heterogeneity (under GMMNL) being controversial.

Allowing for scale heterogeneity in the absence of preference heterogeneity certainly brings the evidence into a range more akin to evidence from choice models such as MNL, latent class and MMNL, but this appears also to be problematic because, with the exception of one data set, it results in lower mean estimates than are obtained from MMNL.

If we were to adopt what might be described as a practitioner's view, the appeal of the latent class model as the best fitting model for four data sets, and close to best fit on the remaining three data sets, suggests that we might draw our VTTS estimates from this model. The mean across all data sets is $14.20 per person hour in $AU2010[12], with a range from $5.81 to $17.99. The Australian data sets deliver $16.67 per person hour and New Zealand evidence is $8.02 per person hour in $AU2010. The New Zealand 2010 figure in the local currency is $10.43. These estimates appear in line with current practice (see Li *et al.* 2010).

While the analysis described in this paper by itself does not definitively answer the question of transferability of evidence, it provides a useful data point, a framework for future empirical studies and some conclusions that suggest the existence of strong differences across regions after accounting for model structure differences. The similarities among the datasets that were used provide a somewhat unique opportunity for the type of analysis conducted here. The need for ongoing research on this issue is fundamental, and it appears that this study, where we have seven comparable data sets, has muddied the water.

[12] An unweighted mean for all data sets and all model forms in $AU2010 is $19.93 per person hour. This is coincidentally very similar to the recommended value used from a number of Sydney studies reported by Hensher (2010) of $19.62, but in $AU2004. It should be noted that VTTS for Sydney are typically the highest across all data sets, which are as an unweighted average, $21.67 in $AU2010 in Table 5.

Appendix A: Background to the seven data sets

Study 1

The data is drawn from a study undertaken in Australia in the context of toll vs. free roads, which utilised a stated choice (SC) experiment involving two SC alternatives (i.e., route A and route B), which are pivoted around the knowledge base of travellers (i.e., the current trip). The trip attributes associated with each route are summarised in Table A1.

Table A1: Trip Attributes in Stated Choice Design

Routes A and B
Free flow travel time
Slowed down travel time
Stop/start/crawling travel time
Minutes arriving earlier than expected
Minutes arriving later than expected
Probability of arriving earlier than expected
Probability of arriving at the time expected
Probability of arriving later than expected
Running cost
Toll Cost

Each alternative has three travel scenarios - 'arriving *x* minutes earlier than expected', 'arriving *y* minutes later than expected', and 'arriving at the time expected'. Each is associated with a corresponding probability[13] of occurrence to indicate that travel time is not fixed but varies from time to time. For all attributes except the toll cost, minutes arriving early and late, and the probabilities of arriving on-time, early or late, the values for the SC alternatives are variations around the values for the current trip. Given the lack of exposure to tolls for many travellers in the study catchment area, the toll levels are fixed over a range, varying from no toll to $4.20, with the upper limit determined by the trip length of the sampled trip.

In the choice experiment, the first alternative is described by attribute levels associated with a recent trip; with the levels of each attribute for Routes A and B pivoted around the corresponding level of actual trip alternative with the probabilities of arriving early, on time and late provided. Commuters in a Metropolitan area in Australia were sampled. A telephone call was used to establish eligible participants from households. During the telephone call, a time and location were agreed for a face-to-face Computer Aided Personal Interview (CAPI). In total, 280 commuters (with less than 120 minutes' trip length) were sampled for this study, each responding to 16 choice sets (games), resulting in 4,480 observations for model estimation. The experimental design method of D-efficiency used herein is specifically structured to increase the statistical performance of the models with smaller samples than are required for other less-efficient (statistically) designs such as orthogonal designs (see Rose *et al.* 2008; Rose and Bliemer 2007).

Study 2

The data collected in this study is from a study in Australia in 2000 in the context of route choices of car commuters. All data is entered by the interviewer directly into a decision support system on a laptop. A SC experiment in which the respondent compares the levels of times and costs of a current/recent trip against two alternative opportunities to complete the same trip that are described by other levels of times and costs. The respondent has to choose one of these alternatives. The sample of 147 effective interviews, each responding to 16 choice sets, resulted

[13] The probabilities are designed and hence exogenously induced to respondents, similar to other travel time variability studies.

in 2,352 observations for model estimation. Three trip length segments were investigated: no more than 30 minutes, 31 to 60 minutes, and more than 61 minutes.

The attributes included in the choice experiments are free flow time, congestion time, trip time variability, running cost, toll cost and toll payment options (cash, Electronic/Tag, and Electronic/Licence plate recognition (no tag required)). Attributes except for toll payment of the stated choice alternatives are based on the values for the current trip in terms of travel times and cost (including tolls if paid). In the design of the choice experiment, important considerations are: toll should range from $0 to $16; a longer trip should involve higher toll alternatives; for a current trip without a toll, SC alternatives involving a toll should mostly be faster than the current trip; and it is assumed that the faster the road, the higher the toll; the lower the running costs, the lower the free-flow time; and the lower the slowed down time, the lower the uncertainty.

Study 3

The survey was undertaken in late June and early July 1999, sampling residents of seven cities/regional centres in New Zealand. The main survey was executed as a laptop-based face to face interview in which each respondent was asked to complete the survey in the presence of an interviewer. Each sampled respondent evaluated 16 choice sets, choosing amongst two SC alternatives and the current RP alternative. Given a sample size of 152 car commuters, 2,432 choice observations are yielded.

The design is based on two unlabelled alternatives each defined by six attributes each of four levels (i.e., 4^{12}): free flow travel time, slowed down travel time, stop/start travel time, uncertainty of travel time, running cost and toll charges. Except for toll charges, the levels are *proportions* relative to those associated with a current trip identified prior to the application of the SC experiment. Including the current (i.e., RP) alternative, described by the exact same six attributes as the two SC alternatives, the design starts with six columns of zeros for the last trip attributes followed by six attributes for alternative A and then six attributes for alternative B. . The six attributes for alternative A are orthogonal to the six columns for alternative B, allowing for the estimation of models with complex structures for the random components of the utility expression associated with each of the alternatives (Louviere and Hensher 2001). The levels of the attributes for both SC alternatives were rotated to ensure that neither A nor B would dominate the RP trip, and to ensure that A and B would not dominate each other. The fractional factorial design has 64 rows. Four blocks of 16 were "randomly" allocated to each respondent.

Study 4

This survey, conducted in 2005, sampled 304 car commuters resident in an Australian Metropolitan Area. All data is entered by trained interviewers directly into a CAPI system on laptops, in which a sample of recent/current trips is the setting for establishing individual’s preferences for different combinations of levels of the components of travel time and vehicle trip costs. Each sampled respondent evaluated 16 choice sets (resulting in 4,864 observations over the entire sample), choosing amongst two SC alternatives and the current RP alternative. The two SC alternatives are unlabelled routes. The trip attributes associated with each route are: free flow time, slowed down time, and stop/start/crawling time, travel time variability, toll cost and running cost. For all attributes except the toll cost, the values for the SC alternatives are variations around the values for the current trip. Given the lack of exposure to tolls for many travellers in the study catchment area, the toll levels are fixed over a range, varying from no toll to $8, with the upper limit determined by the trip length of the sampled trip. In addition, the experimental design method of D-efficiency used herein is specifically structured to increase the statistical performance of the models with smaller samples than are required for other less-efficient (statistically) designs such as orthogonal designs.

Study 5

The data collected in this study is from a study undertaken in Australia in 2004, which is entered by trained interviewers directly into a CAPI system on laptops. In the experiment, each car commuter was required to compare the levels of times and costs of a current/recent trip against two SC alternative routes to complete the same trip that is described by other levels of times and costs. Each alternative is described in terms of free flow time, congestion time, trip time variability, running cost and toll cost. The SC alternative values for this attribute are variations around the total trip time. For all other attributes, the values for the SC alternatives are variations around the values for the current trip. A D-efficiency experimental design method was used. During each choice question, the respondent has to choose one of these three alternatives. The process of choosing amongst the three alternatives is repeated 16 times (each time the levels of times and costs associated with the non-current/recent trip alternative are varied). This survey has 57 effective interviews for car commuters, resulting in 921 observations.

Study 6

The data is drawn from a study in Australia in 2004, in the context of car driving commuters making choices from a range of level of service packages defined in terms of travel times and costs, including a toll where applicable. The choice experiment presented respondents with 16 choice sets, each giving a choice between their current (reference) route and two alternative routes with varying trip attributes. The sample of 243 effective interviews, each responding to 16 choice sets, resulted in 3,888 observations for model estimation. To ensure that we captured a large number of travel circumstances, we sampled individuals who had recently undertaken trips of various travel times, in locations where tollroads currently exist. To ensure some variety in trip length, three segments were investigated: no more than 30 minutes, 31 to 60 minutes, and more than 61 minutes (capped at two hours). A telephone call was used to establish eligible participants from households stratified geographically, and a time and location agreed for a face-to-face CAPI. A statistically efficient design that is pivoted around the knowledge base of travellers is used to establish the attribute packages in each choice scenario. The two stated choice alternatives are unlabelled routes. The trip attributes associated with each route are free flow time, congestion time, trip time variability, running cost and toll cost. The attributes of the choice experiment alternatives are based on the values of the current trip. The SC alternative values for this attribute are variations around the total trip time. For all other attributes, the values for the SC alternatives are variations around the values for the current trip. The experimental design has one version of 16 choice sets.

Study 7

The main field survey was undertaken in New Zealand in 2007. The surveys utilised a SC experiment involving one current (RP) trip and two SC alternatives which were shown as unlabelled routes (i.e., route A and route B). The trip attributes associated with each route are free flow time, congestion time, trip time variability, running cost and toll cost. For all attributes except the toll cost, the values for the SC alternatives were variations around the values for the current trip. The SC alternative values for trip time variability, given it applies across repeated trips, were variations around a total trip time. Given the lack of exposure to tolls for many NZ travellers in the study catchment area, the toll levels were fixed over a range, varying from no toll to $4, with the upper limit determined by the trip length of the sampled trip. 115 car commuters were sampled, using a CAPI system. In the experiment, 16 choice sets were presented to each respondent, resulting in 1,864 choice observations for this survey, and the D-efficiency design was used to structure the SC experiment.

Appendix B

Note: all models using simulated likelihood methods used 500 halton draws.

Table B1(1): A summary of modelling results for Study 1

	MNL	LC (2 classes) Class 1 (p=0.64)	LC (2 classes) Class 2 (p=0.36)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL(constrained normal)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	0.8892 (17.92)	1.5739 (29.35)	-0.5187 (-15.33)	1.3917 (19.97)	0.8999 (16.90)	1.1237 (17.18)	1.0672 (32.03)	0.8855 (32.13)
Free flow time	-0.0513 (-7.34)	-0.0411 (-2.87)	-0.0714 (-10.23)	-0.0339 (-2.95)	-	-	-	-
Slowed down time	-0.0716 (-9.78)	-0.1125 (-9.30)	-0.0740 (-11.15)	-0.0920 (-7.96)	-	-	-	-
Stop start time	-0.0808 (-14.34)	-0.0736 (-9.06)	-0.0919 (-19.72)	-0.0989 (-18.63)	-	-	-	-
Congested time	-	-	-	-	-	-	-	-
Cost	-0.2938 (-15.10)	-0.8519 (-21.43)	-0.1832 (-15.59)	-0.5050 (-20.30)	-0.3378 (-16.14)	-0.4045 (-17.09)	-0.5359 (-38.87)	-0.3875 (-29.13)
Means for random parameters:								
Free flow time	-	-	-	-	-0.0905 (-10.11)	-0.0780 (-4.63)	-0.2552 (-54.67)	-0.2633 (-40.44)
Slowed down time	-	-	-	-	-0.1165 (-11.32)	-0.1184 (-7.08)	-0.3429 (-44.39)	-0.2540 (-45.50)
Stop start time	-	-	-	-	-0.1562 (-16.13)	-0.1043 (-5.62)	-0.5164 (-109.69)	-0.4064 (-54.80)
Congested time	-	-	-	-	-	-	-	-
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.0905 (10.11)	0.490 (11.23)	0.2552 (54.67)	0.0021 (0.10)
Slowed down time	-	-	-	-	0.1165 (11.32)	0.4905 (10.60)	0.3429 (44.39)	0.1731 (3.86)
Stop start time	-	-	-	-	0.1562 (16.13)	0.7811 (16.55)	0.5164 (109.69)	0.0126 (0.78)
Congested time	-	-	-	-	-	-	-	-
Variance parameter in scale	-	-	-	1.1817 (34.71)	-	-	1.1099 (44.38)	2.0622 (33.08)
Weighting parameter Gamma	-	-	-	-	-	-	0.4965 (22.25)	0.10 (0.36)
Model fit								
Log-likelihood	-3435.78	-2713.27		-2826.91	-3303.25	-2961.94	-2783.69	-2955.97
Bayes Information Criterion (BIC)	1.5432	1.2319		1.2733	1.4843	1.3373	1.2559	1.3384
No. of observations	4480							

Table B1 (2): A summary of modelling results for Study 2

	MNL	LC (2 classes) Class 1 (p=0.76)	LC (2 classes) Class 2 (p=0.24)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	0.5205 (8.77)	0.6249 (7.29)	-0.7432 (-5.13)	0.4962 (9.68)	0.6877 (16.05)	-0.2017 (2.34)	0.5284 (10.49)	0.2152 (4.47)
Free flow time	-0.0962 (-13.02)	-0.1445 (-8.49)	-0.0505 (-3.98)	-0.1589 (-17.47)	-	-	-	-
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-0.1031 (-12.91)	-0.1614 (-9.17)	-0.0927 (-5.61)	-0.2052 (-17.43)	-	-	-	-
Cost	-0.5377 (-21.38)	1.1515 (-17.88)	-0.1837 (-4.56)	-1.0840 (-20.78)	-1.1529 (-40.09)	-0.9798 (-17.95)	-2.0980 (-55.67)	-1.0171 (-28.62)
Means for random parameters:								
Free flow time	-	-	-	-	-0.3248 (-22.64)	-0.0830 (-2.07)	-0.9595 (-187.47)	-0.1138 (-3.66)
Slowed down time	-	-	-	-	-	-	-	
Stop start time	-	-	-	-	-	-	-	
Congested time	-	-	-	-	-0.3536 (-23.55)	-0.1664 (-2.92)	-0.9632 (-147.95)	-0.1749 (-6.35)
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.3248 (22.64)	1.1378 (7.88)	0.9595 (187.47)	1.0 (12.51)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	0.3536 (23.55)	0.6181 (5.93)	0.9632 (147.95)	0.6480 (10.13)
Variance parameter in scale	-	-	-	1.2833	-	-	1.2932 (123.39)	1.0234 (7.54)
Weighting parameter Gamma	-	-	-	-	-	-	0.5009 (9.72)	0.4966 (4.63)
Model fit								
Log-likelihood	-1867.75	-1391.81		-1512.91	-1478.43	-1364.55	-1396.92	-1350.83
Bayes Information Criterion (BIC)	1.6014	1.2149		1.3029	1.2704	1.1801	1.2077	1.1751
No. of observations	2352							

Table B1(3): A summary of modelling results for Study 3

	MNL	LC (2 classes) Class 1 (p=0.73)	LC (2 classes) Class 2 (p=0.27)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	1.2058 (21.87)	1.0286 (13.59)	0.1650 (1.68)	0.7352 (14.65)	1.2149 (21.60)	1.1318 (19.12)	1.0410 (17.44)	0.9715 (30.01)
Free flow time	-0.0840 (-5.29)	-0.0867 (-3.52)	-0.0803 (-3.67)	-0.1533 (-9.27)	-	-	-	-
Slowed down time	-0.0702 (-5.79)	-0.0871 (-3.79)	-0.0624 (-3.89)	-0.1546 (-9.37)	-	-	-	-
Stop start time	-0.1631 (-9.55)	-0.2726 (-7.68)	-0.1042 (-4.84)	-0.4060 (-19.11)	-	-	-	-
Congested time	-	-	-	-	-	-	-	-
Cost	-0.5246 (-12.75)	-3.2051 (-54.17)	-0.3243 (-8.05)	-2.9119 (-67.71)	-0.5393 (-12.85)	-0.7870 (-22.28)	-0.9553 (-27.38)	-0.9256 (-41.05)
Means for random parameters:								
Free flow time	-	-	-	-	-0.0985 (-5.74)	-0.1038 (-5.21)	-0.1481 (-8.95)	-0.1357 (-7.27)
Slowed down time	-	-	-	-	-0.0951 (-5.50)	-0.1215 (-6.20)	-0.2077 (-8.42)	-0.2049 (-7.90)
Stop start time	-	-	-	-	-0.1985 (7.79)	-0.210 (-4.36)	-0.3917 (-8.94)	-0.3612 (-6.98)
Congested time	-	-	-	-	-	-	-	
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.0985 (5.74)	0.2487 (4.92)	0.1481 (8.95)	0.0218 (0.04)
Slowed down time	-	-	-	-	0.0951 (5.50)	0.2048 (3.55)	0.2077 (8.42)	0.0311 (0.03)
Stop start time	-	-	-	-	0.1985 (7.79)	1.1155 (6.30)	0.3917 (8.94)	0.0184 (0.02)
Congested time	-	-	-	-	-	-	-	-
Variance parameter in scale	-	-	-	0.8720 (43.59)	-	-	1.0770 (6.83)	1.1940 (7.69)
Weighting parameter Gamma	-	-	-	-	-	-	0.5001 (1.05)	0.10 (0.01)
Model fit								
Log-likelihood	-1714.52	-1380.96		-1378.93	-1700.81	-1637.34	-1627.66	1630.61
Bayes Information Criterion (BIC)	1.4262	1.1709		1.1532	1.4147	1.3721	1.3611	1.3732
No. of observations	2432							

Table B1(4): A summary of modelling results for Study 4

	MNL	LC (2 classes) Class 1 (p=0.74)	LC (2 classes) Class 2 (p=0.26)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	0.3417 (6.64)	0.6735 (8.84)	-0.8777 (-7.24)	0.2321 (3.25)	0.1426 (2.54)	0.3937 (6.11)	0.4378 (9.62)	0.4247 (11.21)
Free flow time	-0.0762 (-7.95)	-0.1276 (-5.85)	-0.0939 (-5.52)	-0.2167 (-12.66)	-	-	-	-
Slowed down time	-0.1163 (-13.61)	-0.2019 (-11.77)	-0.1282 (-9.03)	-0.2428 (-20.29)	-	-	-	-
Stop start time	-0.1695 (-20.58)	-0.1677 (-12.18)	-0.1813 (-12.66)	-0.2732 (-19.44)	-	-	-	-
Congested time	-	-	-	-	-	-	-	-
Cost	-0.5134 (-25.81)	-1.0461 (-16.96)	-0.2721 (-16.96)	-1.1298 (-30.44)	-0.5548 (-25.77)	-0.5948 (-24.88)	-0.9467 (-55.73)	-1.0949 (-67.20)
Means for random parameters:								
Free flow time	-	-	-	-	-0.1552 (-10.80)	-0.1320 (6.14)	-0.7762 (-161.76)	-0.7684 (-177.47)
Slowed down time	-	-	-	-	-0.1427 (-12.59)	-0.1685 (-10.81)	-0.7352 (-139.39)	-0.9270 (-176.20)
Stop start time	-	-	-	-	-0.2331 (-18.15)	-0.2491 (-12.29)	-0.6911 (-101.68)	-1.3262 (-246.14)
Congested time	-	-	-	-	-	-	-	-
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.1552 (10.80)	0.7068 (12.63)	0.7762 (161.76)	0.0369 (0.55)
Slowed down time	-	-	-	-	0.1427 (12.59)	0.3114 (7.84)	0.7352 (139.39)	0.0165 (0.20)
Stop start time	-	-	-	-	0.2331 (18.15)	0.5982 (11.84)	0.6911 (101.68)	0.0109 (0.14)
Congested time					-		-	
Variance parameter in scale	-	-	-	0.9834 (28.64)	-	-	1.3079 (79.96)	1.4992 (119.30)
Weighting parameter Gamma	-	-	-	-	-	-	0.4974 (11.27)	0.10 (0.02)
Model fit								
Log-likelihood	-2670.61	-2335.29		-2371.24	-2568.82	-2421.04	-2286.04	-2334.16
Bayes Information Criterion (BIC)	1.1068	0.9794		0.9855	1.0651	1.0058	0.9522	0.9772
No. of observations	4864							

Table B1(5): A summary of modelling results for Study 5

	MNL	LC (2 classes) Class 1 (p=0.74)	LC (2 classes) Class 2 (p=0.26)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters :								
Reference constant	-0.5894 (-6.82)	-2.0357 (-11.70)	1.0550 (5.69)	-1.9357 (-5.07)	-0.5615 (-6.34)	-0.6529 (-7.05)	-0.6603 (-7.10)	-0.6595 (-15.09)
Free flow time	-0.0346 (-8.02)	-0.0333 (-6.78)	0.0148 (0.51)	-0.0710 (-4.64)	-	-	-	-
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-0.0356 (7.49)	-0.0495 (-7.72)	-0.0235 (-2.19)	-0.0781 (-4.33)	-	-	-	-
Cost	-0.1360 (-10.59)	-0.1203 (-8.16)	-0.2578 (-6.46)	-0.2439 (-4.69)	-0.1503 (-8.24)	-0.170 (-11.38)	-0.1667 (-11.14)	-0.1855 (-22.25)
Means for random parameters:								
Free flow time	-	-	-	-	-0.0466 (-8.24)	-0.0361 (-3.78)	-0.0956 (-5.31)	-0.0593 (-2.94)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	-0.0546 (-7.18)	-0.0550 (-5.23)	-0.1114 (-7.78)	-0.0823 (-3.03)
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.0466 (8.24)	0.1333 (6.74)	0.0956 (5.31)	0.1168 (2.47)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	0.0546 (7.18)	0.1368 (5.71)	0.1114 (-7.78)	0.0791 (0.86)
Variance parameter in scale	-	-	-	1.3131 (9.28)	-	-	1.6670 (7.63)	1.3549 (3.87)
Weighting parameter Gamma	-	-	-	-	-	-	0.5120 (1.31)	0.10 (0.23)
Model fit								
Log-likelihood	-855.87	-739.06		-808.32	-831.86	-802.73	-789.82	-792.12
Bayes Information Criterion (BIC)	1.9068	1.6880		1.8099	1.8542	1.8052	1.7769	1.7969
No. of observations	912							

Table B1(6): A summary of modelling results for Study 6

	MNL	LC (2 classes) Class 1 (p=0.57)	LC (2 classes) Class 2 (p=0.43)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	0.1184 (2.89)	-0.6953 (-5.74)	0.9125 (8.30)	0.1385 (2.48)	0.1582 (3.66)	0.1077 (2.42)	0.1172 (5.42)	0.1770 (7.79)
Free flow time	-0.0687 (-17.77)	-0.1067 (-16.55)	-0.0215 (-2.68)	-0.1064 (-12.61)	-	-	-	-
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-0.0912 (-28.70)	-0.1125 (-21.56)	-0.0568 (-7.82)	-0.1306 (-14.18)	-	-	-	-
Cost	-0.3519 (-30.76)	-0.2858 (-12.04)	-0.5532 (-9.58)	-0.5103 (-14.82)	-0.3975 (-30.61)	-0.4269 (-30.46)	-0.4237 (-62.65)	-0.4226 (-60.74)
Means for random parameters:								
Free flow time	-	-	-	-	-0.1016 (-16.62)	-0.0940 (-10.60)	-0.1433 (-14.41)	-0.1438 (-10.52)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	-11.72 (-21.22)	-0.1146 (-16.26)	-0.1501 (-14.59)	-0.1558 (-10.76)
Standard deviations for random parameters:								
Free flow time	-	-	-	-	-0.1016 (-16.62)	0.2407 (12.24)	0.1433 (14.41)	0.2257 (8.66)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	-11.72 (-21.22)	0.1936 (12.0)	0.1501 (14.59)	0.0984 (2.31)
Variance parameter in scale	-	-	-	0.7422 (12.44)	-	-	0.9958 (11.63)	1.1827 (12.42)
Weighting parameter Gamma	-	-	-	-	-	-	0.4998 (3.66)	0.0999 (0.73)
Model fit								
Log-likelihood	-3037.75	-2756.91		-2903.67	-2886.65	-2816.82	-2799.32	-2794.24
Bayes Information Criterion (BIC)	1.5711	1.4388		1.5043	1.4617	1.4617	1.4527	1.4544
No. of observations	3888							

Table B1(7): A summary of modelling results for Study 7

	MNL	LC (2 classes) Class 1 (p=0.64)	LC (2 classes) Class 2 (p=0.36)	SMNL	MMNL (constrained triangular)	MMNL (unconstrained triangular)	GMMNL (constrained triangular)	GMMNL (unconstrained triangular)
Non-random parameters:								
Reference constant	0.1031 (1.68)	0.9015 (9.90)	-2.1916 (-12.75)	0.4624 (13.98)	0.0771 (1.19)	0.0596 (0.85)	0.0122 (0.39)	0.1217 (3.42)
Free flow time	-0.0994 (-14.63)	-0.1673 (-13.15)	-0.0768 (-7.79)	-0.1307 (-21.89)	-	-	-	-
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-0.1261 (-9.94)	-0.1423 (-6.27)	-0.1125 (-6.35)	-0.2265 (-17.14)	-	-	-	-
Cost	-0.5132 (-17.56)	-0.9575 (-15.88)	-0.3663 (-9.01)	-0.6333 (-17.33)	-0.5970 (-18.31)	-0.6282 (-18.19)	-0.8036 (-33.67)	-0.5975 (-26.29)
Means for random parameters:								
Free flow time	-	-	-	-	-0.1369 (-13.07)	-0.1327 (-9.18)	-0.4569 (-104.70)	-0.1754 (-12.85)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	-0.1686 (-13.07)	-0.1391 (-3.97)	-0.4088 (-34.88)	-0.2169 (-13.83)
Standard deviations for random parameters:								
Free flow time	-	-	-	-	0.1369 (13.07)	0.2468 (8.47)	0.4569 (104.70)	0.0224 (0.13)
Slowed down time	-	-	-	-	-	-	-	-
Stop start time	-	-	-	-	-	-	-	-
Congested time	-	-	-	-	0.1686 (13.07)	0.7062 (6.70)	0.4088 (34.88)	0.0083 (0.03)
Variance parameter in scale	-	-	-	0.9224 (16.85)	-	-	1.5946 (104.70)	1.0228 (15.23)
Weighting parameter Gamma	-	-	-	-	-	-	0.0378 (0.55)	0.10 (0.01)
Model fit								
Log-likelihood	-1639.61	-1237.06		-1466.58	-1579.45	-1533.71	-1454.39	-1535.37
Bayes Information Criterion (BIC)	1.7985	1.3814		1.6145	1.7331	1.6916	1.6054	1.7016
No. of observations	1840							

References

Daly, A., Hess, S and Train, K. (2010) Assuring finite moments for willingness to pay in random coefficient models, available at http://elsa.berkeley.edu/~train/papers.html

Fiebig, D., Keane, M., Louviere, J., and Wasi, N. (2009) The generalized multinomial logit: accounting for scale and coefficient heterogeneity, *Marketing Science*, published online before print July 23, DOI:10.1287/mksc.1090.0508

Greene, W.H. and Hensher, D.A. (2003) A Latest Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B*, 37, 681-698.

Greene, W.H. and Hensher, D.A. (2010) Does scale heterogeneity across individuals matter? A comparative assessment of logit models, *Transportation*, 37 (3), 413-428.

Hensher, D.A. (2010) Attribute Processing, Heuristics and Preference Construction in Choice Analysis, in Hess, S. and Daly, A. (eds.) *State-of Art and State-of Practice in Choice Modelling*, Emerald Press, UK., 35-70.

Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: state of practice, *Transportation*, 30 (2), 133-176.

Hensher, D.A., Rose, J. and Greene, W. H. (2005) *Applied Choice Analysis: A Primer*, Cambridge University Press, Cambridge.

Li, Z., Hensher, D.A. and Rose, J. M. (2010) Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence, *Transportation Research Part E*, 46(3), 384-403.

Louviere, J. and T. Eagle (2006) Confound it! that pesky little scale constant messes up our convenient assumptions, *Proceedings, 2006 Sawtooth Software Conference, 211-228*, Sawtooth Software, Sequem, Washington, USA.

Louviere, J.J. and Hensher, D.A. (2001) Combining Data Sources (Workshop Report, 9th International Association of Travel Behaviour Research Conference, Gold Coast, Queensland, July.) in Hensher, D.A. and King, J. (eds) *The Leading Edge of Travel Behaviour Research*, Pergamon Press, Oxford, 125-144.

Louviere, J.J., Hensher, D.A. and Swait, J. (2000) Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation, Cambridge University Press, Cambridge.

Rose, J.M. and M.C.J. Bliemer (2007) Stated preference experimental design strategies' in Hensher, D.A. and K. Button, (eds.) *Transport Modelling*, Second Edition, Handbooks in Transport, Vol. 1, Elsevier Science, Oxford, Chapter 8.

Rose, J.M., Bliemer, M.C., Hensher and Collins, A. T. (2008) Designing efficient stated choice experiments in the presence of reference alternatives, *Transportation Research Part B*, 42 (4), 395-406.

Small, K.A., Noland, R.B., Chu, X. and Lewis, D. (1999) Valuation of travel-time savings and predictability in congested conditions for highway user-cost estimation, NCHRP Report 431, Transportation Research Board, National Research Council.

Train, K.E. (2003) Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge.

Train, K. and Wilson, W.W. (2008) Estimation on stated-preference experiments constructed from revealed-preference choices, *Transportation Research Part B*, 42, 191–203.