STATED CHOICE DESIGN FACTORS AFFECTING THE PROBABILITY OF NONTRADING AND THE PRECISION OF VALUE OF TIME ESTIMATES: A META-ANALYSIS OF INDIVIDUAL PARTICIPANT DATA

Alex Mitrani
Steer Davies Gleave
alex.mitrani@sdgworld.net

Abstract

An approach to stated choice design is presented which takes into account the characteristics of real stated choice data, including nontrading responses and the impact they can have on value of time estimates. A meta-analysis of individual participant data from 51 stated choice surveys was undertaken to develop models of how characteristics of stated choice designs affect the proportions of nontraders and the precision of value of time estimates. An optimization model was developed to find the parameter values for a distribution of boundary values of time that would maximize the expected precision of a future value of time estimate.

Keywords: stated choice; nontrading; value of time

1. INTRODUCTION

Stated choice exercises involving pairs of options offering trade-offs principally between time and cost are often used to estimate the value of travel time, a parameter that is crucial for reliable travel demand forecasting as part of the planning and appraisal of high value transportation projects. The objective of this research was to find ways of improving the precision of such value of time estimates by optimizing the experimental designs regarding the expected probability of nontrading.

Nontrading may result from always choosing a labelled alternative, or lexicographic answering such as always choosing with reference to just one attribute, for instance always choosing the fastest route (S Hess, Rose, & Polak, 2010). In this article the term "nontraders" will be used to refer to either respondents who always choose the cheapest route (cost-sensitive nontraders), or those who always choose the fastest route (time-sensitive nontraders).

Most existing methods of efficient design of choice experiments are based on models which do not explicitly take into account nontrading or the impact it can have on the results (ChoiceMetrics, 2012; Sawtooth Software, 2016). Whilst these tools are extremely valuable and useful, a need was perceived for a complementary approach that would optimize simple stated choice designs considering the expected level of nontrading that they would produce and the impact that this would be expected to have on the precision of the estimated value of time.

Choice models such as the logit model are best suited to compensatory decision making where people trade off the attributes of the alternatives against each other. This apparently does not occur with non-traders, whose choices vary in response to a limited subset of the attributes, making it impossible to evaluate their sensitivity to the others. For instance, in the case of value of time experiments with just time and cost attributes, for the respondents who always chose the slowest option or always chose the fastest option, only an upper or lower bound for the value of time can be established, respectively (Börjesson, Fosgerau, & Algers, 2012; S Hess, Stathopoulos, & Daly, 2012). This bound corresponds to the lowest or highest "boundary value" (Fowkes & Wardman, 1988) or "bid" (Börjesson et al., 2012) that was tested in the design. However, in the case of the cost-sensitive nontraders their values of time can be expected to have a theoretical lower bound of zero.

Fosgerau et al proposed a simple test for the degree to which the distribution of values of time is supported by the data (Mogens Fosgerau, Hjorth, & Lyk-Jensen, 2007), which involves applying the choice model to quantify the range of the probability of choosing the slowest and least expensive alternative – this range should ideally extend from 0 to 1. When there are non-traders, the support interval will be a narrower subsection of the [0, 1] range. When nontraders are mixed together with other groups of responses it is possible to estimate values of time despite limited support for the distribution of the values of time. However, the results of doing so may be biased (S Hess et al., 2010, 2012; Rose, Hess, & Collins, 2013). Hess et al showed that with the logit model, the problem of nontraders is non-trivial and they recommended that approaches should be developed to explicitly recognise the diversity of decision rules that may be evident in the data (S Hess et al., 2010). Further research has shown that the impacts of heterogeneous decision rules on results such as value of time estimates can be very significant, and that therefore it is unwise to ignore the presence of nontrading in choice data (S Hess et al., 2012; Rose et al., 2013).
Several researchers have reached the conclusion that nontrading responses may be expressions of

genuine preferences and that therefore it is unreasonable to exclude them from the analysis (Börjesson et al., 2012; Kaa, 2006; Killi, Nossum, & Veisten, 2007). Chorus (Chorus, 2014) comments that it is unlikely to find true lexicographic decision making in real life and therefore sequences of choices that appear lexicographic are probably the result of steep indifference curves combined with insufficient attribute variation in the stated choice design, which is consistent with the findings of Killi et al (Killi et al., 2007). Along the same lines, recent research in Sweden and Denmark has shown that the identification of the high end of the value of time distribution is very important to be able to correctly estimate the mean of the distribution without having to resort to strong assumptions (Börjesson et al., 2012; Mogens Fosgerau et al., 2007). Fosgerau and Börjesson (Mogens Fosgerau & Börjesson, 2015) have shown how experimental designs that are focused on an expected value of time can bias the results of logit models towards the reference values used for the design. They recommend that practitioners should use designs that can capture a broad range of parameter values and can be used with multiple model specifications, in particular non-parametric methods that can be used to find the most appropriate type of model.

All of this evidence implies that on designing choice experiments to estimate the value of time, it is important to stretch the range of boundary values of time high enough. However, the higher it is stretched the bigger the gaps between the boundary values of time will become, assuming a fixed number of choice cards per respondent. Increasing the size of the gaps between the boundary values of time can be expected to reduce the precision of the value of time estimates, ceteris paribus. Therefore it is expected that there will be an optimal range to cover, as was already pointed out by Ojeda et al (Ojeda-Cabral, Hess, & Batley, 2016).

Attribute non-attendance may be a cause of apparently lexicographic answering including nontrading (S Hess et al., 2010; Stephane Hess, Stathopoulos, Campbell, O'Neill, & Caussade, 2013; Rose et al., 2013). However, with binary choice stated choice experiments limited to 2-3 attributes, attribute non-attendance due to excessive cognitive burden on respondents was not expected to be a significant issue (Wardman & Ibáñez, 2013). Most nontrading in these cases was expected to be due to steep indifference curves, insufficient scales of the attributes, and policy-response bias in some cases. This sort of nontrading must be minimized by careful design of the stated choice experiments, and this research aims to leverage evidence from a large quantity of value of time studies to help improve future stated choice designs.

2. META-ANALYSIS

Meta-analysis should ideally consider all relevant data. Examples of meta-analysis in transportation include studies of values of time and value of time multipliers (Abrantes & Wardman, 2011; Wardman & Chintakayala, 2012). Meta-analysis with individual participant data uses the original disaggregate data for each study. This type of meta-analysis is widely used in health research but very few other instances of its use in transport research or with stated choice data have been found (Meyerhoff, Mørkbak, & Olsen, 2014; Mitrani, 2013). Meta-analysis of individual participant data offered a way of finding specific ways of improving stated choice designs, by modelling the effects of varying the characteristics of the experimental designs on the rates of nontrading and the precision of the value of time estimates.

2.1. Data

There are three principal types of bias that can affect this type of research: publication bias, selection bias and bias related to unavailable data (Ahmed, Sutton, & Riley, 2012). The scope of the meta analysis reported here was limited to data from studies undertaken by Steer Davies Gleave plus the data for California State Route 91 that was made publicly available as part of the mlogit package for R (Croissant, 2015). Only a few of these studies have been published. Care was taken to seek out and include all relevant datasets that were available within the company, restricting the search to binary route choice exercises for car drivers. Out of 52 datasets considered for inclusion only 1 dataset was excluded, resulting in a sample of 51 datasets that is summarized in Table 1. The excluded dataset appeared to have been randomly generated and had not been used in the related project for this reason.

TABLE 1 Summary of Meta-Analysis Datasets (Year and Country)

Year (Datasets)	Country (Datasets)
2000 (1)	Australia (2)
2001 (2)	Brazil (3)
2002 (2)	Canada (1)
2003 (3)	Chile (2)
2004 (1)	Colombia (10)
2006 (2)	Ghana (1)
2008 (3)	Italy (1)
2009 (3)	Mexico (12)
2010 (6)	New Zealand (1)
2011 (5)	Peru (2)
2012 (4)	Puerto Rico (1)
2013 (7)	Russian Federation (1)
2014 (4)	South Africa (1)
2015 (7)	United Kingdom (3)
2016 (1)	United States (10)

The main potential bias is related to availability: due to resource limitations and data confidentiality issues the scope excludes studies by other consultants and researchers which may use different design methodologies and may work in other parts of the world. Therefore it would be very interesting to extend the scope of this work in future or for other researchers to see what results are obtained using their datasets and methodologies.

Metadata collected for each study was limited to the country and the year. The stated choice data included the respondent id, the choice card id, the alternative id, time, cost (in local currency, nominal values), choice, and in some cases a route-specific constant representing a different type of road. The data were cleaned to remove undecided choices and incomplete responses (people who only responded to a fraction of the choices before abandoning the interview). Incomplete responses were removed to avoid potential confusion between the number of cards in the choice exercise design and the number of cards included in incomplete responses. Before data cleaning there were 311,758 choices from 38,738 respondents. After data cleaning there were 303,639 choices from 38,101 respondents: approximately 3% of the choices and 2% of the respondents were removed. Several variables were added to the database to enable the analysis to account for the differences between years and countries: GDP per capita in purchasing power parity US dollars (Wolfram|Alpha, n.d.-a), consumer price inflation data to adjust from each year to 2015 prices (Wolfram|Alpha, n.d.-b), exchange rate data (International Monetary Fund, n.d.), purchasing power parity conversion factors (World Bank, n.d.), and Gini coefficients (Alméstica, 2004; OECD, 2016; The World Bank, 2016).

Table 2 presents summary statistics for the stated choice designs, in terms of boundary values of time, at the individual participant level. Analyzing the data at this level takes into account that many studies include several stated choice designs. A boundary value of time for a choice between two alternatives is defined as follows:

$$BVTT = -\frac{(c_2 - c_1)}{(t_2 - t_1)} \tag{1}$$

where the subscripts denote alternatives, alternative 2 is faster and more expensive, alternative 1 is slower and less expensive, t is time and c is cost.

Two sets of data were identified. Set A consists of all the datasets that were analyzed without any problems. Set B consists of 3 datasets that produced some questionable results in some parts of the analysis that will be commented on in due course. It was not clear whether this data should be excluded or not, therefore it was flagged so that results could be considered both with it and without it.

TABLE 2 Summary Statistics for the Stated Choice Designs (PPP $ in 2015 prices)

Variable	Set A	Set B	N	Mean	Std. Dev.	Min	Max
Boundary values of time (number of stated choice cards)	1	0	36540	8.06	1.92	3.00	12.00
Average boundary value of time ($ / hour)	1	0	36540	0.24	0.13	0.01	1.15
Median boundary value of time ($ / hour)	1	0	36540	0.21	0.11	0.01	0.85
Maximum boundary value of time ($ / hour)	1	0	36540	0.52	0.32	0.02	3.53
Minimum boundary value of time ($ / hour)	1	0	36540	0.06	0.05	0.00	0.39
Standard deviation of boundary values of time ($ / hour)	1	0	36540	0.16	0.11	0.00	1.44
Coefficient of variation (dimensionless)	1	0	36540	0.66	0.19	0.07	2.01
Non-parametric skew (dimensionless)	1	0	36540	0.18	0.15	-0.58	0.71
Boundary values of time (number of stated choice cards)	0	1	1561	5.89	1.21	3.00	9.00
Average boundary value of time ($ / hour)	0	1	1561	0.26	0.08	0.12	0.38
Median boundary value of time ($ / hour)	0	1	1561	0.21	0.04	0.11	0.26
Maximum boundary value of time ($ / hour)	0	1	1561	0.56	0.26	0.24	0.95
Minimum boundary value of time ($ / hour)	0	1	1561	0.07	0.02	0.04	0.13
Standard deviation of boundary values of time ($ / hour)	0	1	1561	0.19	0.10	0.06	0.35
Coefficient of variation (dimensionless)	0	1	1561	0.69	0.18	0.38	0.92
Non-parametric skew (dimensionless)	0	1	1561	0.19	0.20	-0.27	0.42
Boundary values of time (number of stated choice cards)	1	1	38101	7.97	1.94	3.00	12.00
Average boundary value of time ($ / hour)	1	1	38101	0.24	0.12	0.01	1.15
Median boundary value of time ($ / hour)	1	1	38101	0.21	0.11	0.01	0.85
Maximum boundary value of time ($ / hour)	1	1	38101	0.53	0.32	0.02	3.53
Minimum boundary value of time ($ / hour)	1	1	38101	0.07	0.05	0.00	0.39
Standard deviation of boundary values of time ($ / hour)	1	1	38101	0.16	0.11	0.00	1.44
Coefficient of variation (dimensionless)	1	1	38101	0.66	0.19	0.07	2.01
Non-parametric skew (dimensionless)	1	1	38101	0.18	0.15	-0.58	0.71

Table 3 presents summary statistics for the degree of non-trading, at the study level. Overall cost-sensitive non-trading is almost twice as common as time-sensitive non-trading, and approximately two thirds of respondents trade between routes.

TABLE 3 Summary Statistics for Nontrading

Variable	Set A	Set B	N	Mean	Std. Dev.	Min	Max
Cost-sensitive non-traders (%)	1	0	48	0.177	0.098	0.039	0.458
Time-sensitive non-traders (%)	1	0	48	0.129	0.105	0.002	0.410
Traders (%)	1	0	48	0.694	0.131	0.300	0.913
Cost-sensitive non-traders (%)	0	1	3	0.593	0.121	0.501	0.730
Time-sensitive non-traders (%)	0	1	3	0.175	0.118	0.091	0.311
Traders (%)	0	1	3	0.231	0.125	0.141	0.374
Cost-sensitive non-traders (%)	1	1	51	0.200	0.137	0.039	0.730
Time-sensitive non-traders (%)	1	1	51	0.129	0.106	0.000	0.410
Traders (%)	1	1	51	0.671	0.169	0.141	0.913

2.2. Modelling

The value of travel time (VTT) is defined as follows.

$$VTT = \frac{\beta_t}{\beta_c} = \beta_0 \tag{2}$$

where β_t is the marginal utility of time, β_c is the marginal utility of cost, and β_0 is an alternative abbreviation for the value of time.

The normalized margin of error of the value of time estimate was selected as the variable to be minimized and is defined as follows:

$$\mathrm{NME}_{VTT} = \frac{1.96 * SE_{VTT}}{VTT} \tag{3}$$

where SE_{VTT} is the standard error of the value of time estimate and VTT is the mean value of time.

Two types of models were used to model the values of time and their margins of error: the random utility logit model (RU model hereafter) and the random valuation mixed logit model with multiplicative error structure (RV model hereafter). The RV model has been used on most of the recent European national value of time studies as it has been found to have several desirable characteristics for modelling simple time-cost trade-off data (Börjesson & Eliasson, 2014; Farideh, Flügel, Samstad, & Killi, 2010a, 2010b; Mogens Fosgerau et al., 2007; Ojeda Cabral, Batley, & Hess, 2016). Ojeda et al offer a detailed comparison of various RU and RV model formulations (Ojeda Cabral et al., 2016), which has been used as a basis for the following summary.

For both the RU and RV approaches we will consider a slower, less expensive option 1 and a faster, more expensive option 2. Alternative 2 will be chosen when $U_2 > U_1$, otherwise alternative 1 will be chosen. The RU and RV approaches differ mainly in how the random error is modelled: as random variation in utility in the former approach and as random variation in the values of time in the latter approach.

The utility equations for the RU models are formulated as follows:

$$U_i = \beta_i * \left(\frac{\beta_t}{\beta_c} * t_i + c_i\right) + \varepsilon_i \tag{4}$$

where the subscript i indicates the alternative (1 or 2), U is utility, t is time, c is cost, and ε is the random error term which is independently, identically distributed extreme value for the ordinary logit model (Train, 2009).

The utility equations for the RV models are formulated as follows:

$$U_1 = \mu * BVTT * \varepsilon_1 \tag{5}$$
$$U_2 = \mu * VTT * \varepsilon_1$$

where μ is a scale factor and ε is the random error term.

A logarithmic transformation is applied to make the models easier to estimate, resulting in the following:

$$U'_1 = \ln(U_1) = \mu' * [\ln(BVTT) + \ln(\varepsilon_1)] \tag{6}$$
$$U'_2 = \ln(U_2) = \mu' * [\ln(VTT) + \ln(\varepsilon_2)]$$

where μ' is a scale factor.

The value of time (VTT) is parameterized to introduce random heterogeneity and the possibility of incorporating a dummy variable for road type (e.g. expressway vs. normal road), as follows:

$$VTT = e^{\beta_0 + \delta_2 * d_2 + u} \tag{7}$$

where the base level of the value of time is β_0, road standard dummy variable d_2 takes value 0 for the base level and 1 for a higher standard, δ_2 is a parameter to be estimated to isolate the effect of a higher standard of road on the value of time, u is a random variable to represent the heterogeneity of the values of time which is normally distributed with mean 0 and standard deviation σ.

The expectation of the value of time at its base level, without considering the effect of any dummy variables, is given by:

$$E(VTT) = e^{\beta_0} * e^{\sigma^2/2} \tag{8}$$

where σ is the standard deviation of the random heterogeneity u and all other terms have been defined previously.

It can be expected that the RV models will produce different results to the RU models because they model random variation as being related to the values of time, whereas the RU models associate the random error with utility, in such a way that it cancels out of the value of time calculation. Furthermore, the RV models can be expected to produce a better fit to the data in terms of the log likelihood of the models because of both the heterogeneous element of the value of time and the multiplicative error structure (M. Fosgerau & Bierlaire, 2009). The objective of including both types of model is to compare the model that represents the state of the art for simple time and cost choice exercises (Stephane Hess, Daly, Dekker, Cabral, & Batley, 2016) with the model that has been most widely used (RU), not to compare like with like.

All models were estimated in Stata (StataCorp, 2013). The RV models were implemented using the *mixlogit* command (Hole, 2007) and 1000 Halton draws, having successfully reproduced an initial version of the RV model implemented in BIOGEME (Bierlaire, 2003). The simple support check proposed by Fosgerau et al (Mogens Fosgerau et al., 2007) was implemented by applying the choice models to the estimation data and calculating the range of probabilities for choosing the slower and less expensive alternative. The following table presents selected results for the individual studies: unique identifiers (ID), sample sizes (N), % traders (NT), % time-sensitive non-traders (TSNT), % cost-sensitive non-traders (CSNT), 95% confidence interval percentage errors on the values of time (NME_{VTT}), support intervals as percentages of the [0, 1] interval and the set to which each study was assigned. Where appropriate, results are presented both for the random utility (RU) and the random valuation (RV) models.

TABLE 4 Selected details of individual studies

ID	N	NT	TSNT	CSNT	NME_{VTT} RU	NME_{VTT} RV	Support RU	Support RV	Set
37	714	9%	4%	5%	6%	11%	56%	50%	A
25	510	10%	6%	4%	15%	33%	37%	48%	A
27	250	12%	5%	6%	75%	172%	49%	48%	A
22	741	15%	10%	4%	8%	13%	67%	76%	A
6	1242	16%	2%	13%	4%	7%	75%	81%	A
26	849	16%	9%	7%	13%	139%	16%	16%	A
38	156	16%	2%	14%	16%	19%	97%	76%	A
28	403	19%	5%	14%	9%	26%	80%	48%	A
35	478	19%	6%	13%	8%	15%	68%	71%	A
50	1003	20%	8%	12%	6%	11%	72%	80%	A
7	727	20%	5%	16%	4%	7%	58%	65%	A
10	547	21%	9%	12%	4%	20%	98%	90%	A
11	697	22%	4%	18%	4%	14%	91%	88%	A
36	3295	22%	12%	10%	3%	8%	74%	91%	A
33	607	22%	15%	7%	4%	16%	86%	95%	A
2	1187	23%	7%	16%	8%	20%	82%	46%	A
1	1989	23%	3%	20%	8%	7%	85%	73%	A
34	183	23%	12%	11%	6%	16%	70%	68%	A
30	652	24%	10%	15%	3%	7%	40%	40%	A
5	1096	27%	8%	19%	14%	59%	50%	39%	A
40	452	28%	17%	10%	4%	15%	60%	58%	A
39	596	28%	13%	15%	7%	33%	65%	50%	A
32	537	29%	9%	20%	7%	17%	40%	44%	A
43	483	30%	0%	30%	3%	5%	81%	72%	A
15	210	30%	13%	17%	8%	25%	100%	95%	A
24	421	31%	8%	23%	25%	82%	95%	41%	A
18	2801	32%	12%	20%	2%	4%	82%	77%	A
45	630	33%	15%	18%	4%	12%	92%	81%	A
47	132	33%	3%	30%	101%	253%	89%	17%	A
19	1933	34%	15%	19%	17%	15%	29%	58%	A
14	266	34%	24%	10%	40%	87%	19%	42%	A
52	1066	34%	4%	30%	23%	11%	42%	50%	A
42	609	34%	18%	16%	4%	15%	80%	64%	A
41	1028	35%	30%	5%	3%	14%	65%	59%	A
13	356	36%	11%	25%	15%	40%	94%	57%	A

ID	N	NT	TSNT	CSNT	NME_{VTT} RU	NME_{VTT} RV	Support RU	Support RV	Set
46	416	36%	19%	17%	6%	28%	44%	43%	A
49	1001	37%	4%	33%	8%	9%	56%	61%	A
4	661	37%	13%	24%	21%	37%	42%	50%	A
51	740	39%	2%	37%	13%	13%	56%	59%	A
48	694	39%	26%	13%	4%	17%	65%	66%	A
44	414	44%	32%	13%	5%	38%	72%	52%	A
20	843	45%	30%	14%	9%	41%	41%	47%	A
21	210	46%	41%	5%	5%	25%	70%	60%	A
29	81	47%	14%	33%	15%	66%	60%	57%	A
16	247	52%	6%	46%	16%	44%	36%	31%	A
12	1206	59%	39%	20%	17%	237%	29%	26%	A
17	657	60%	23%	37%	6%	15%	32%	32%	A
3	730	63%	12%	50%	2900%	20321%	19%	4%	B
31	709	70%	39%	31%	28%	230%	10%	14%	A
9	681	82%	9%	73%	287%	445%	16%	11%	B
8	177	86%	31%	55%	496%	4253%	19%	8%	B

As expected, the RU models tended to produce tighter confidence intervals on the mean value of time than the RV models. Also as expected, the RV models produced very significant improvements in goodness of fit as measured by the log likelihood, with an average improvement of 964, a maximum of 5604 and a minimum of 58. The proportion of the [0, 1] interval on which the value of time distribution is supported varies considerably, although it tends to diminish as the proportion of non-traders increases. There are a few cases where the confidence interval on the value of time increased explosively, particularly for the RV models. The most severe instances were identified as Set B, and correspond to instances where the lognormal distribution has had to accommodate significant proportions of both cost-sensitive and time-sensitive non-traders, and has done so by increasing the variance to obtain significant densities beyond both extremes of the range of boundary values. These are all cases with 50% or more cost-sensitive nontraders. These anomalous results obtained with the log-normal distribution in the presence of significant proportions of non-traders are consistent with the findings of Dumont et al (Dumont, Hess, & Daly, 2013) based on their analysis of the Danish value of time study data (Mogens Fosgerau et al., 2007), and also consistent with the comments of Fosgerau et al (Mogens Fosgerau et al., 2007) who described how in some cases the calculated mean values were higher than the mean boundary values that were tested.

To get the modelling dataset, the RU and RV models were applied for 4 variations of each dataset: the whole dataset, just the traders, traders plus time-sensitive nontraders, and traders plus cost-sensitive non-traders. The NME_{VTT} was transformed by taking the natural logarithm before modelling it using ordinary least squares regression. This ensured that the regression model would not predict any negative values for the percentage error which is positive by definition. The regression model coefficients are presented in Table 5.

TABLE 5 OLS Regression models for LN($\mathbf{NME_{VTT}}$)

Variable	RU Sets A + B Coef. (z)	RU Set A Coef. (z)	RV Sets A + B Coef. (z)	RV Set A Coef. (z)
LN(sample size)	-0.54 (-5.76)	-0.42 (-5.83)	-0.55 (-7.45)	-0.48 (-6.79)
% of time-sensitive non-traders	2.68 (2.35)	1.04 (1.68)	4.57 (7.57)	3.83 (6.90)
% of cost-sensitive non-traders	3.49 (6.49)	2.41 (5.18)	3.67 (4.91)	2.1 (4.68)
Mean spacing of BVTT (2015 PPP$)	6.03 (3.94)	5.35 (3.28)	5.99 (3.61)	5.3 (3.38)
Non-parametric skew	-2.26 (-2.96)	-1.11 (-1.76)	-1.84 (-2.00)	-
Constant	0.31 (0.51)	-0.5 (-1.02)	0.78 (1.53)	0.2 (0.43)
R^2	0.42	0.31	0.45	0.42
Root Mean Square Error	0.93	0.72	0.99	0.72
Number of cases	204	192	204	192

NOTE: Cost-sensitive non-traders are type 1, traders are type 2, and time-sensitive non-traders are type 3.

The regression model results show that as sample size increases, the percentage error decreases, but not linearly. The proportions of both time-sensitive and cost-sensitive non-traders increase the percentage error. The cost-sensitive non-traders have greater impact in the RU models, whereas the time-sensitive nontraders have greater impact in the RV models. Increasing the average interval between boundary values of time also increases the percentage error on the value of time estimates. Non-parametric skew is the difference between the mean and the median, divided by the standard deviation. Increasing the non-parametric skew was associated with lower percentage error on the value of time estimates. To operationalize this model it was necessary to model the probability of both types of non-trading behavior in terms of design variables under the control of the practitioner. Due to the log-transformed dependent variable it was necessary to use the root-mean-square error to apply the model (Gould, 2014), as follows:

$$\widehat{\mathrm{NME}_{VTT}} = e^{\Sigma \beta x} * e^{\sigma^2/2} \qquad (9)$$

where $\widehat{\mathrm{NME}_{VTT}}$ is the predicted normalized marginal error on the value of time, $\sum \beta x$ is the vector of regression coefficients multiplied by their corresponding variables, and σ is the root-mean-square value of the regression model.

Trading behavior (cost-sensitive non-trading, time-sensitive non-trading, or trading) was modelled using a multinomial logit model, with trading as the base category. The design variables that were found to be the most significant were the natural logarithms of the minimum and maximum boundary values of time. Categories of GDP per capita were used to differentiate between different contexts. The categories were defined as terciles of the GDP per capita distribution at the respondent level. The model parameters are shown in Table 6.

TABLE 6 Logit models for type of trading behavior

Variable	MNL, Sets A + B Coef. (z)	MNL, Set A Coef. (z)
LN Minimum BVTT x type 1	0.23 (1.83)	0.16 (1.25)
LN Maximum BVTT x type 1	-0.23 (-1.19)	-0.28 (-1.57)
Constant type 1	-0.76 (-1.86)	-1.13 (-3.31)
LN Maximum BVTT x type 3	-0.92 (-3.73)	-0.89 (-3.47)

Variable	MNL, Sets A + B Coef. (z)	MNL, Set A Coef. (z)
LN Minimum BVTT x type 3	0.48 (3.49)	0.47 (3.39)
Gini Index x type 3	0.04 (2.67)	0.04 (2.57)
Constant type 3	-2.67 (-4.26)	-2.94 (-4.30)
Pseudo R^2	0.25	0.27
Number of cases	38101	36540
Log-likelihood	-31347	-29204

NOTE: Cost-sensitive non-traders are type 1, traders are type 2, and time-sensitive non-traders are type 3.

Some of the parameters were not significantly different from zero but were nevertheless maintained in the model because their sign and size were plausible. The key effects that this model captures are that increasing the maximum boundary value of time reduces time-sensitive non-trading, while increasing the minimum boundary value of time increases cost-sensitive non-trading. These two variables are both under the control of the practitioner and can be varied to produce improved stated choice designs. The Gini index was used to reflect differences between countries. It is plausible that higher income inequality may have been associated with less support for the high-end of the value of time range in many stated choice designs, hence a positive relation between the Gini index and the proportion of time-sensitive nontraders. The GDP per capita was tested in earlier versions of these models, and produced similar goodness of fit in terms of log likelihood, but the sign of the parameter was counterintuitive, suggesting less time-sensitive non-trading with higher levels of GDP per capita.
Considering the value of time percentage error and trading behavior models together, the percentage error can be reduced by reducing the proportions of nontraders, which can be done by reducing the minimum boundary value of time and / or increasing the maximum boundary value of time. However, for a fixed number of choices per respondent this implies increasing the average interval between boundary values, which will tend to increase the percentage error on the value of time. An optimization model was developed to narrow the search for optimal boundary value of time distributions.

3. Optimization

The two models presented in the previous section were applied in a spreadsheet and a standard solver add-in was used to find a set of boundary values of time that would minimize the percentage error on the value of time estimate, given a lognormal distribution for the boundary values of time and a set of restrictions. The lognormal distribution produces only positive boundary values, is positively skewed, and can readily push the high-end of the distribution. Furthermore, several recent European national value of time studies have found that the lognormal distribution best represents the value of time distributions found in practice (Börjesson & Eliasson, 2014).

The calculations use the inverse of the selected probability function to define boundary values of time that cover most of the distribution. The optimization process systematically modifies the parameters of the distribution to minimize the expected percentage error on the value of time. Restrictions were specified for the maximum of the minimum boundary value of time, the minimum of the maximum boundary value of time (guaranteeing a minimum acceptable range) and a minimum acceptable interval between boundary values. The Generalized Reduced Gradient nonlinear algorithm (Fylstra, Lasdon, Watson, & Warren, 1998) was used, with multiple starting points to improve the chances of finding a global optimum. Using this specification the solver succeeded in finding solutions for sets of boundary values that are expected to be global optima and which minimize the expected percentage error of the value of time. Figure 1 shows four cumulative distribution functions for boundary values of time generated by this process, differing by the type of choice model (RU or RV) and by the inclusion or exclusion of outliers (Set B).

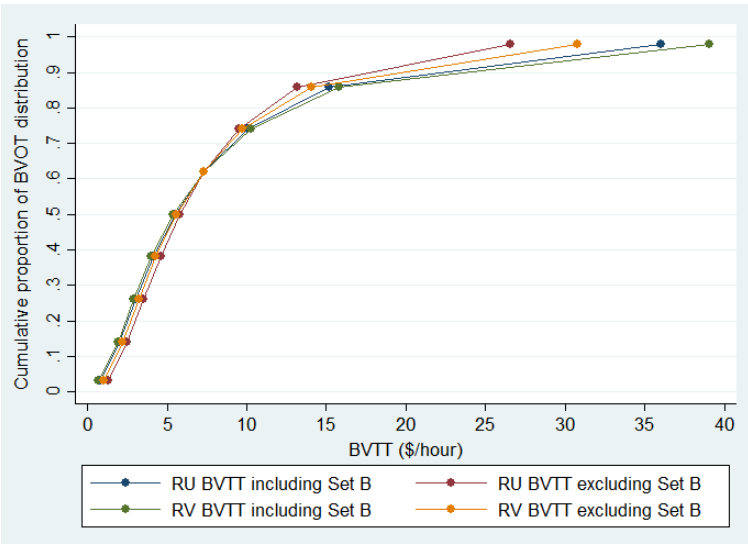

FIGURE 1 Cumulative distribution functions for boundary values of time generated by the non-linear optimization process, showing differences between the RU and RV models, and the effect of including the cases identified as outliers

The inclusion of Set B pushes out the high-end of the optimized distribution of boundary values of time, as does the use of the RV models instead of the RU models. The distribution of boundary values that stretches the highest is that which includes Set B and uses the RV models. Even so, it does not stretch as high as the designs used in the recent Swedish national value of time study which included boundary values of time as high as PPP $62/hour in 2016 prices, and managed to reduce the proportion of high-end non-traders to 1.1% (Börjesson & Eliasson, 2014; Börjesson et al., 2012). The results produced by the RV models for the datasets in Set B would be highly questionable if we were concerned about the mean values of the values of time, and we would have had to censor the value of time distributions to obtain sensible mean values if this were the case (Börjesson et al., 2012; Mogens Fosgerau et al., 2007). However, the focus of this article is the percentage error on values of time, not the values of time themselves. Furthermore, removing Set B or preferring the RU results would seem unadvisable, both because of the relevance of the data points in Set B and also considering the evidence in favor of trying to provide as much as possible support for the high end of the value of time distribution (Börjesson et al., 2012).

4. RESULTS AND DISCUSSION

The findings support the hypothesis that there is a trade-off at the design stage between stretching the distribution of BVTT and trying to maximize the precision of the VTT estimates. A spreadsheet-based optimization tool has been developed to optimize a lognormally-distributed set of BVTT taking into account the number of choice cards, the sample size, the expected proportions of non-traders and the effect of these on the precision of the value of time estimate.

This research is perhaps the first instance of an approach to stated choice design that explicitly accounts for the expected proportions of nontraders in the data and their expected effect on the precision of value of time estimates. A key advantage of the approach proposed here is that it will help to minimize the proportion of nontraders by design, before the start of the data collection. This does not preclude additional measures such as using specific questions or analytical techniques to try to determine which nontraders are genuinely expressing preferences.

This research also points to several promising opportunities for further research. The meta-analysis models will become more robust as further datasets are added, especially cases with broader distributions of boundary values of time. The scope of the meta-analysis sample could be broadened by developing means of collaboration between different organizations while respecting confidentiality requirements. Alternatively, the findings of this research could be checked using other meta-analysis datasets.

It is hoped that the evidence provided here will prove interesting and useful to many practitioners and will help to improve the quality of stated choice designs and the reliability of value of time estimates.

ACKNOWLEDGMENTS

The author is very grateful to Dr. Manuel Ojeda Cabral for his help with the implementation of the RV models. The author also wishes to thank the anonymous referees whose comments have led to substantial improvements in this research.

REFERENCES

Abrantes, P. A. L., & Wardman, M. R. (2011). Meta-analysis of UK values of travel time: An update. **Transportation Research Part A: Policy and Practice**, *45*(1), 1–17. https://doi.org/10.1016/j.tra.2010.08.003

Ahmed, I., Sutton, A. J., & Riley, R. D. (2012). Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. **BMJ**, *344*(January), 1–10. https://doi.org/10.1136/bmj.d7762

Alméstica, E. V. S. (2004). What happened to the distribution of income in Puerto Rico during the last three decades of the XX Century? A statistical point of view. Retrieved October 26, 2016, from http://economia.uprrp.edu/ensayo 129.pdf

Bierlaire, M. (2003). BIOGEME: A Free Package for The Estimation of Discrete Choice Models. **Swiss Transport Research Conference**, 1–27. https://doi.org/10.1017/CBO9781107415324.004

Börjesson, M., & Eliasson, J. (2014). Experiences from the Swedish Value of Time study. **Transportation Research Part A: Policy and Practice**, *59*(January 2012), 144–158. https://doi.org/10.1016/j.tra.2013.10.022

Börjesson, M., Fosgerau, M., & Algers, S. (2012). Catching the tail: Empirical identification of the distribution of the value of travel time. **Transportation Research Part A: Policy and Practice**, *46*(2), 378–391. https://doi.org/10.1016/j.tra.2011.10.006

ChoiceMetrics. (2012). **Ngene User Manual & Reference Guide**. Sydney, Australia:

ChoiceMetrics. Retrieved from http://www.choice-metrics.com/index.html
Chorus, C. (2014). Capturing alternative decision rules in travel choice models: a critical discussion. In **Handbook of choice modelling** (pp. 295–297). Retrieved from https://books.google.cl/books?hl=en&lr=&id=XghCBAAAQBAJ&pg=PA294#v=onepage&q&f=false
Croissant, Y. (2015). mlogit: multinomial logit model. Retrieved September 29, 2015, from https://cran.r-project.org/web/packages/mlogit/index.html
Dumont, J., Hess, S., & Daly, A. (2013). Are Nontraders the Achilles Heel of the Mixed Logit Model? In **Transportation Research Board 92nd Annual Meeting** (p. 13–3894.). Washington, D.C. Retrieved from http://trid.trb.org/view.aspx?id=1242260
Farideh, R., Flügel, S., Samstad, H., & Killi, M. (2010a). *Den Norske verdsettingsstudien - Tid: Rapporten uten vedlegg*. Retrieved from http://www.toi.no/article29726-29.html
Farideh, R., Flügel, S., Samstad, H., & Killi, M. (2010b). *Den Norske verdsettingsstudien - Tid: Vedleggene*. *Den Norske verdsettingsstudien - Tid*. Retrieved from http://www.toi.no/article29726-29.html
Fosgerau, M., & Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. **Transportation Research Part B: Methodological**, *43*(5), 494–505. https://doi.org/10.1016/j.trb.2008.10.004
Fosgerau, M., & Börjesson, M. (2015). Manipulating a stated choice experiment. **Journal of Choice Modelling**, *16*, 43–49. https://doi.org/10.1016/j.jocm.2015.09.005
Fosgerau, M., Hjorth, K., & Lyk-Jensen, S. V. (2007). Results for Experiment 1. In *The Danish Value of Time Study*.
Fowkes, T., & Wardman, M. (1988). The design of stated preference travel choice experiments: with special reference to interpersonal taste variations. **Journal of Transport Economics and Policy**. Retrieved from http://www.jstor.org/stable/20052833
Fylstra, D., Lasdon, L., Watson, J., & Warren, A. (1998). Design and use of the Microsoft Excel Solver: Interfaces. **INTERFACES**, *28*(5), 29–55. https://doi.org/10.1287/inte.28.5.29
Gould, W. A. (2014). Use poisson rather than regress; tell a friend. Retrieved from blog.stata.com/2011/08/22/use-poisson-rather-than-regress-tell-a-friend/
Hess, S., Daly, A., Dekker, T., Cabral, M. O., & Batley, R. (2016). A framework for capturing heterogeneity , heteroscedasticity , non - linearity , reference dependence and design artefacts in value of time research. **Transportation Research Part B Methodological**, (November).
Hess, S., Rose, J., & Polak, J. (2010). Non-trading, lexicographic and inconsistent behaviour in stated choice data. **Transportation Research Part D: Transport and Environment**, *15*(7), 405–417. Retrieved from http://www.sciencedirect.com/science/article/pii/S1361920910000684
Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., & Caussade, S. (2013). It's not that I don't care, I just don't care very much: Confounding between attribute non-attendance and taste heterogeneity. **Transportation**, *40*(3), 583–607. https://doi.org/10.1007/s11116-012-9438-1
Hess, S., Stathopoulos, A., & Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. **Transportation**. Retrieved from http://link.springer.com/article/10.1007/s11116-011-9365-6
Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. **The Stata Journal**, *7*(3), 388–401. Retrieved from http://www.sheffield.ac.uk/polopoly_fs/1.105585!/file/mixlogit.pdf
International Monetary Fund. (n.d.). Official exchange rate (LCU per US$, period average). Retrieved July 30, 2016, from http://data.worldbank.org/indicator/PA.NUS.FCRF
Kaa, E. Van de. (2006). Assessment of the value of travel time from stated choice surveys: the

treatment of lexicographic answering. In **Proceedings of the European Transport Conference**. Brussels (Belgium): Association for European Transport (AET). Retrieved from http://repository.tudelft.nl/assets/uuid:f54122a6-8827-4d02-a567-14a82e77b393/288267.pdf
Killi, M., Nossum, Å., & Veisten, K. (2007). Lexicographic answering in travel choice: insufficient scale extensions and steep indifference curves. **European Journal of Transport and Infrastructure Research**, *7*(1), 39–62. Retrieved from http://www.ejtir.tbm.tudelft.nl/issues/2007_01/pdf/2007_01_03.pdf
Meyerhoff, J., Mørkbak, M. R., & Olsen, S. B. (2014). A Meta-study Investigating the Sources of Protest Behaviour in Stated Preference Surveys. **Environmental and Resource Economics**, *58*(1), 35–57. https://doi.org/10.1007/s10640-013-9688-1
Mitrani, A. (2013). Empirical Evidence on Ratio of Means Problem and Estimation of Values of Time with Stated Preference Data. In **TRB 92nd Annual Meeting, January 13-17**. Washington, D.C.: Transportation Research Board. Retrieved from http://amonline.trb.org/2013-1.264263/13-0828-1-1.282203?qr=1
OECD. (2016). Gini (gross income, before taxes). Retrieved October 26, 2016, from http://stats.oecd.org/
Ojeda-Cabral, M., Hess, S., & Batley, R. (2016). Understanding valuation of travel time changes: are preferences different under different stated choice design settings? **Transportation**, (April), 1–21. https://doi.org/10.1007/s11116-016-9716-4
Ojeda Cabral, M., Batley, R., & Hess, S. (2016). The value of travel time: random utility versus random valuation. **Transportmetrica A: Transport Science**, *12*(3), 230–248. https://doi.org/10.1080/23249935.2015.1125398
Rose, J., Hess, S., & Collins, A. (2013). What if my model assumptions are wrong? the impact of non-standard behaviour on choice model estimation. **Journal of Transport Economics and Policy**, *47*(2), 245–263. Retrieved from http://www.ingentaconnect.com/content/lse/jtep/2013/00000047/00000002/art00005
Sawtooth Software. (2016). CBC Questionnaires and Design Strategy. Retrieved August 1, 2016, from http://www.sawtoothsoftware.com/help/lighthouse-studio/manual/index.html?hid_web_cbc_designs_1.html
StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.
The World Bank. (2016). GINI index (World Bank estimate). Retrieved October 14, 2016, from http://data.worldbank.org/indicator/SI.POV.GINI
Train, K. (2009). **Discrete choice methods with simulation**. Retrieved June 6, 2016, from http://eml.berkeley.edu/books/train1201.pdf
Wardman, M., & Chintakayala, P. (2012). *European wide meta-analysis of values of travel time. Final Report*. Retrieved from http://www.significance.nl/papers/2012-European wide meta-analysis of values of travel time.pdf
Wardman, M., & Ibáñez, J. (2013). Testing Contextual and Design Effects on Inter-Urban Motorists' Responses to a Toll Motorway's Travel Time Savings. **Journal of Transport Economics and Policy (JTEP)**. Retrieved from http://www.ingentaconnect.com/content/lse/jtep/2013/00000047/00000001/art00008
Wolfram|Alpha. (n.d.-a). PPP GDP per capita in USD. Retrieved July 30, 2016, from https://www.wolframalpha.com/input/?i=gdp+per+capita+PPP+uk+2003+usd
Wolfram|Alpha. (n.d.-b). USA Consumer Price Inflation. Retrieved July 30, 2016, from https://www.wolframalpha.com/input/?i=consumer+price+inflation+usa+2002+to+2015
World Bank. (n.d.). PPP conversion factor, GDP (LCU per international $). Retrieved July 30, 2016, from http://data.worldbank.org/indicator/PA.NUS.PPP