

Impact of violations of model assumptions: the role in model estimation of non-trading, changing preferences, lexicography, and inconsistent choices

John M. Rose, Stephane Hess and Andrew T. Collins

*John M. Rose**

The University of Sydney
john.rose@sydney.edu.au

Stephane Hess

University of Leeds
S.Hess@its.leeds.ac.uk

Andrew T. Collins

The University of Sydney
andrew.collins@sydney.edu.au

Version: 1 June 2011

Abstract

Many econometric models exist that may be used to analyse choice data. These however are all based on specific assumptions made by the analyst in relation to how individual respondents make their choices. The main assumption is that respondents behave as if they are using a linear additive fully compensatory utility process. In practice however, it may be difficult to satisfy these assumptions. Accepting that some respondents may exhibit alternative behaviour, the question becomes what impact such 'non-conformist' behaviour may have on estimated models. The present paper presents an in-depth analysis using simulated data to study the impact of such behaviour on model estimates.

Keywords: stated choice experimental design, mixed multinomial logit models, non-trading, lexicographic behaviour, model outcomes, inconsistent behaviour

* Corresponding author

The Faculty of Economics and Business
Room 205, Building C37 (Newtown Campus),
The University of Sydney,
Sydney NSW, 2006,
T: +61 2 9351 0168,
F: +61 (0)2 9351 0088,
E: john.rose@sydney.edu.au

1. INTRODUCTION

Once collected, it is customary to fit a discrete choice model, typically in the form of a logit or probit model, to stated choice (SC) data aggregated over the responses obtained from multiple individuals. Despite lingering concerns with the response quality of SC data (e.g., Verplanken and Aarts 1999 and Fujii and Garling 2003), the utility specifications of these models are often treated in a way that ignores the possibility that different individuals may have applied different processing or decision rules when completing the SC survey. By ignoring this possibility, the resulting model outputs are estimated on the assumption of a single homogeneous information processing strategy (IPS) being used by the entire sampled population. In practical terms, this equates to an assumption that every attribute of the SC experiment is considered as relevant, at least to some degree, by every individual within the sampled population. The fact that some respondents may completely ignore a certain attribute, or process it jointly with some other related attribute, may potentially lead to biased model outputs.

Despite not being accounted for in the majority of SC modelling projects, the concept that different respondents may display heterogeneous IPSs whilst undertaking SC tasks is now well recognised within the literature. A significant proportion of this research stream has demonstrated the presence of heterogeneous IPSs in various SC data sets, particularly with regards to whether subsets of respondents tend to ignore specific attributes or add up composite attributes over the course of SC experiments. For a recent overview of the field, see for example Hensher et al. (2007) and Hess and Hensher (2010) and the references therein. In extreme cases, research has indicated the presence of lexicographic choice behaviour in segments of respondents completing SC tasks (cf. Salensminde 1994, 2002). Further, respondents have been shown in several instances to have exhibited inconsistent behaviour over the course of the experiment due to learning or fatigue effects at the

beginning and end of the survey instrument respectively (see e.g., Rose and Black 2006). Finally, numerous studies have shown a proportion of respondents who refuse to trade between alternatives during the course of a SC experiment, particularly in the presence of a no-choice, status quo or reference alternative, independent of the attributes shown as part of the experiment (see e.g., Hess et al. 2010).

In the present paper, we focus in particular on four specific behavioural traits, namely non-trading, lexicographic choice, inconsistent choice behaviour and dynamic switching between behavioural rules over time. Neither of these traits is in our opinion dealt with adequately in standard practice. To test the impact of such traits on the behavioural outputs of choice models, we conduct a large scale simulated data analysis in which the proportion of the different traits is varied across case studies. In this way, we are able to test the impacts of such behaviour on the estimated model outputs such as the correctness of parameter estimates, the robustness of parameter estimates, the model log-likelihoods and willingness-to-pay (WTP) measures.

The remainder of the paper is organised as follows. In the next section, we provide a brief discussion of the specific rules that are explored within this paper. In Section 3, we outline two case studies and discuss the results derived from Monte Carlo simulations based on these case studies. Finally, we provide concluding remarks in Section 4.

2. DECISION RULES

In this section, we briefly outline four different processing or decision rules which are of interest in this paper. The list we provide is far from complete with numerous other decision rules that respondents may adopt over the course of an experiment being possible.

Nevertheless we believe that the rules we discuss below represent the dominant rules that are likely to exist within most SC datasets.

2.1 Linear additive fully compensatory behaviour

Random utility maximisation (RUM) models such as the family of GEV models (e.g. Multinomial, Nested, and Cross-Nested Logit) and GEV mixture models (e.g. Mixed Multinomial Logit, Mixed Nested Logit, etc.) are based on the notion of linear additive compensatory behaviour, where gains in one attribute are traded against losses in another attribute. The utility of an alternative is computed on the basis of all considered attributes and the associated sensitivities (to be estimated) and the probability of choosing an alternative increases with the utility of that alternative relative to the utility of other alternatives.

For the current paper, we explore the impact of a number of different types of linear additive compensatory behaviour upon the outputs obtained from estimated logit models. Firstly, we test the impact of linear additive compensatory behaviour where respondents maintain the same preferences over the entire course of the experiment. Secondly, we test the impact of respondents dynamically switching preferences (i.e., marginal utilities) over the course of the experiment. In both cases, the unobserved effects over the population are assumed to be randomly drawn from an extreme value distribution and hence both behaviour types remain within the logit model family of behaviour.

2.2 Non-trading

The second decision rule considered herein is non-trading behaviour. We define non-trading behaviour as behaviour whereby a respondent continually selects a particular alternative over the course of the SC experiment, independent of the attribute levels shown. We examine a particular form of non-trading behaviour where respondents refuse to choose any alternative other than that of the status quo alternative, or more precisely, a reference alternative (e.g., a

recent trip that was made; see Rose et al. 2008, Train and Wilson 2008 or Hess and Rose 2009 for a review of such experiments).

Hess et al. (2010) provide three possible explanations for such behaviour in SC studies. The first explanation they offer reflects the possible existence of some respondents displaying extreme preferences, which may for example take the form of high mode allegiance and inertia. A second explanation offered by Hess and Hensher (2010) is the possible use of a heuristic (i.e., non-utility maximising) decision process by the respondent, arising from misunderstanding, boredom or fatigue during the SC exercise. The third explanation for non-trading behaviour is that it may reflect a form of political or strategic behaviour on behalf of some respondents which can express itself especially in the case of controversial topics such as the building of new toll roads (see e.g., Kuriyama 2005). Hess et al. (2010) discuss the importance of distinguishing between these potential causes of such behaviour due to the differing implications that each cause has on model estimation. In the present context we are solely interested in the impact of such behaviour on model estimates.

2.3 Lexicographic choice behaviour

For the purposes of this paper, we define lexicographic decision behaviour as existing when decision makers evaluate the alternatives of a SC experiment only on the basis of a subset of the total number of attributes shown (see e.g., Salensminde 1994, 2001; Rosenberger 2003; Blume et al. 2006). For the purposes of this paper, we examine two types of lexicographic decision rules. The first type involves respondents choosing the alternative that is most attractive in terms of a single attribute (e.g., choose the alternative with the lowest price). The second type of lexicographic decision rule we examine involves respondents choosing the alternative that is best or most attractive based on a composite attribute made up by summing

several related attributes (e.g., respondents choose the alternative that offers the best overall price based on toll and running costs).

As with non-trading behaviour, Hess et al. (2008) suggest several reasons for respondents adopting lexicographic decision rules during a SC survey. Firstly, lexicographic decision processes may reflect the true preferences of some respondents (i.e., they may be extremely time sensitive and select the fastest mode independent of cost). Secondly, lexicographic decision processes may be an artefact of the SC experiment itself, where the attribute level ranges of the attributes may not be sufficiently wide enough to ensure trading away from one or more attributes within the experiment. This may reflect individual-specific thresholds that may not have been met given the attribute level ranges used in the experiment (e.g., Swait 2001). Finally, lexicographic decision processes might be used in complex or overly long surveys as a way for respondents to simplify the SC tasks they are asked to complete.

2.4 Inconsistent choice behaviour

The last decision rule examined within this paper is that of inconsistent choice behaviour (see e.g., Salensminde 2001). Inconsistent choice behaviour occurs where decision makers are observed to violate one or more of the axioms of rational choice behaviour. For the present paper, we examine a specific form of inconsistent choice behaviour where respondents are observed in one choice task to select an alternative with a time-cost ratio benefit relative to all other alternatives of value X but later reject an alternative with time-cost ratio benefit relative to all other alternatives of a value greater than X in a subsequent choice task. The presence of such choices in a SC experiment signal i. possible changes to respondents' decision rules over the course the experiment, ii. the possible presence of highly complex utility functions, and iii. the possibility that respondents are being inattentive to the survey. Although all three

causes are of concern, the later possibility represent the greatest difficulty for analysts as it firstly raises concerns as to the internal validity of any SC data that has such observations present, and secondly and more importantly, goes against the traditional view held by random utility theory (upon which most discrete choice studies are built) which states that any error in the model arises from the analyst's inability to observe all influences upon a decision makers choice and not from mistakes made by respondents.

3. CASE STUDIES

In this section, we describe the results from two different case studies. In each of the case studies, we simulate different decision rule strategies, which we systematically vary via experimental designs. We then draw different proportions of respondents displaying these different choice styles and compare and contrast the estimation results obtained from over 25,000 estimated models against the ideal assumption that all respondents act in a linear additive compensatory fashion.

To test the influence of the presence and proportion of the different respondent choice types, we calculate the mean square error (MSE) across runs and the expected mean square error (EMSE) of the parameter estimates and willingness to pay (WTP) outcomes. Unlike the MSE statistic which captures statistical evidence of any potential biases in the individual parameters (or WTP), the EMSE provides a single summary statistic of the overall bias and variances across all parameter estimates (or WTP outcomes) which may be used to easily compare different response strategies. The MSE and EMSE statistics are given in Equations (4) and (5) respectively for the parameter estimates (similar calculations are used for WTP).

$$MSE = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_k^{(r)} - \beta_k)^2 \quad (4)$$

$$EMSE = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}^{(r)} - \beta)' (\hat{\beta}^{(r)} - \beta), \quad (5)$$

where $\hat{\beta}_k^{(r)}$ is the parameter estimate for attribute k obtained at sample iteration r , and β_k is the known prior parameter for attribute k used in constructing the Monte Carlo simulation for respondents assumed to have acted in a linear additive compensatory fashion.

As well as calculating the MSE and EMSE for the parameter estimates and willingness to pay outcomes, we also compute and report results based on the average t -ratios for each of the Monte Carlo models.

3.1 Case Study 1

The first case study consists of a simple SC experiment involving the choice between two alternatives, each described by two attributes, cost and time. An efficient experimental design consisting of 18 choice situations blocked into two versions was generated. As such, each simulated respondent was assumed to have performed nine choice tasks. The utility specification for the case example is given as:

$$V_j = \beta_1 cost_j + \beta_2 time_j, \quad j=1,2, \quad (6)$$

where $cost_j \in \{1,3,5\}$ and $time_j \in \{20,30,40\}$. The two parameters in the case study are specified as random parameters drawn from normal distributions, i.e., $\beta_1 \sim N(\mu_1, \sigma_1)$ and $\beta_2 \sim N(\mu_2, \sigma_2)$, the assumed parameter values of which are given in Table 1. The use of the

Normal distribution is justified despite issues of sign violation given the exploratory nature of this paper.

Table 1: Case study 1 simulation assumptions

Behaviour (master data sets)	Cost (β_1)		Time (β_2)		Notes
	μ_1	σ_1	μ_2	σ_2	
i. Compensatory	-0.3	0.1	-0.05	0.02	Applied to all 9 choice tasks
ii. Lexicographic (cost)	-	-	-	-	Always choose lowest cost alternative
iii. Lexicographic (time)	-	-	-	-	Always choose lowest time alternative
iv. Inconsistent choice behaviour	-0.3	0.1	-0.05	0.02	Parameters apply to 6 choice tasks; the remaining 3 choice tasks display inconsistent behaviour
v. Compensatory with changing sensitivities	-0.3	0.1	-0.05	0.02	Applied to choice task 1-3
	-0.5	0.2	-0.04	0.02	Applied to choice task 4-9
vi. Compensatory with changing sensitivities	-0.3	0.1	-0.05	0.02	Applied to choice task 1-3
	-0.5	0.2	-0.04	0.02	Applied to choice task 4-6
	-0.75	0.3	-0.075	0.03	Applied to choice task 7-9

Six different data sets, each consisting of 1,000 simulated respondents, were generated. Simulated respondents in each data set were assumed to have displayed a particular strategy in making their observed choices. The simulated strategies included i. linear additive compensatory behaviour, ii. lexicographic behaviour with regards to cost iii. lexicographic behaviour with regards to time, iv. inconsistent choice behaviour over the course of the experiment, and v. and vi. two different linear additive compensatory behaviour choice strategies where a respondent's marginal utilities change over the course experiment. Descriptions of these behaviours and the assumptions made in simulating the data are summarised in Table 1. To construct respondents with inconsistent choice behaviour, first the time-cost benefit ratio for linear additive compensatory respondents was calculated. Next the choices in three of the choice tasks were switched so as to be inconsistent in terms of the time-cost benefit ratios calculated for earlier choice tasks.

Different subsets of the simulated data sets were combined according to an orthogonal experimental design to form eight master data sets. The orthogonal design is shown in Table 2, where the value 1 denotes the presence of a behavioural trait in a master data set and 0 the

exclusion from that master data set. According to the orthogonal design, between two and six different respondent processing types were represented in any one master data set.

Table 2: Orthogonal design used to assign Case Study 1 master data sets

Master data set	i. Compensatory	ii. Lexicographic (cost)	iii. Lexicographic (time)	iv. Inconsistent choice behaviour	v. Compensatory with changing sensitivities	vi. Compensatory with changing sensitivities	Number of prop. variations
1	0	0	0	1	1	1	10
2	1	0	0	0	0	1	7
3	0	1	0	0	1	0	7
4	1	1	0	1	0	0	10
5	0	0	1	1	0	0	7
6	1	0	1	0	1	0	10
7	0	1	1	0	0	1	10
8	1	1	1	1	1	1	13

From each master data set, a series of one hundred bootstrap runs were performed drawing 180 respondents each. To test not only the influence of the presence of respondents applying different processing rules on model results, the proportions of the different types of respondents were also systematically varied. For master data sets with two respondent types, seven different proportions of respondents of each processing type were used with variations of between 0.2 to 0.8. Thus, for example, a single Monte Carlo simulation for Master data set 2 might involve 100 bootstrap runs, each drawing 36 respondents (a 0.2 proportion) from data set i. (compensatory) and 144 respondents (a 0.8 proportion) from data set vi. (linear additive compensatory behaviour where the respondent's marginal utilities change over the course experiment). A second simulation might involve 100 bootstrap runs drawing 108 respondents (a 0.6 proportion) from data set i. and 72 respondents (a 0.4 proportion) from data set vi. For master data sets with three information processing styles, ten different variations were used with proportions varying systematically between 0.1 and 0.75. Finally, for the master data set involving each type (master set 8), 13 different versions were used, with proportions varying between 0.1 and 0.5. The number of proportion variations per master data set is shown in the

last column in the Table. As such, a total of 74 bootstrap runs were performed representing a total of 7,400 MMNL model estimations, each for a sample of 1,620 observations for 180 respondents.

Results for the case study can be found at <http://sydney.edu.au/business/itls/staff/johnr/papers> in Tables S1 and S2. Table S1 reports the results of a Seemingly Unrelated Regressions Equations (SURE) model where the MSEs of the parameter estimates obtained from the 74 sets of Monte Carlo runs are used as the dependent variables in the series of equations estimated against the proportions of the types of choice behaviour as the independent variables. Also shown in the table are the results of a linear regression model of the EMSE calculated over the 74 sets of Monte Carlo runs against the different choice behaviour styles as the independent variables.

From the table, it is clear that lexicographic choice behaviour has resulted in statistically significant biases for all four parameter estimates suggesting that the unmodelled presence of such choice decision rules within a sample may bias other parameters within the model and not just the parameter for the attribute to which the behaviour is specifically applied to. The positive parameter estimates suggest larger biases in the parameter estimates as the proportion of respondents displaying that type of behaviour increases within the sample. The increased bias in the standard deviation parameters also suggests that the standard deviation parameters in MMNL models may be affected quite significantly by the presence of lexicographic behaviour. This is possibly due to the model explaining lexicographic behaviour through extreme sensitivities. No other statistically significant biases were found for any of the decision rule types. The regression results for the EMSE are consistent with the results of the SURE model for the MSEs of the individual parameters. The results for this model suggest that lexicographic choice behaviour has statistically significant biases in terms of the overall model results.

Table S2 presents the results for a SURE model where the dependent variables used in the system of equations are the average t -ratios for the four parameter estimates calculated over the Monte Carlo bootstrap runs. The average t -ratios across runs were used as the dependent variable. Given that the mean time and cost parameters are negative, a statistically significant negative parameter estimate in the SURE model represents an improved t -ratio for that parameter. Likewise, statistically significant positive parameter estimates for the two standard deviation parameters in the SURE model represent improved t -ratios for these parameters. For the mean cost parameter, only lexicographic behaviour has a statistically significant influence on the t -ratios relative to the other behaviours. In this instance, as the proportion of respondents in the dataset displaying lexicographic behaviour towards costs increases, the t -ratio for the mean cost parameter goes down in absolute value. The magnitude of influence of lexicographic behaviour on the t -ratio of time however is much larger, but is statistically insignificant. Both types of lexicographic behaviour also impact upon the standard deviation parameter for cost, however in this case, leading to improved t -ratios. This is a slightly misleading result, and needs to be put into context. By increasing the share of lexicographic respondents, and by allowing the Mixed Multinomial Logit (MMNL) model to accommodate these respondents through extreme sensitivities, we increase the scope for retrieving random heterogeneity, hence also leading directly to a higher standard deviation parameter with constant standard errors resulting in a larger t -ratio. Increasing the proportions of respondents with linear additive compensatory behaviour but changing sensitivities on the other hand result in worsening t -ratios for this parameter.

Looking at the mean travel time parameters, for two of the behaviours, increasing the number of respondents results in worsening t -ratios: inconsistent choice, and the first sample of respondents with linear additive compensatory behaviour but changing sensitivities., For the standard deviation parameter of the travel time sensitivity, both samples of linear additive

compensatory behaviour but changing sensitivities increase the t -ratios, which is likely to be for the same reasons as discussed above. Neither form of lexicographic behaviour impact upon the t -ratios for the mean travel time parameter, but both increase the t -ratios for the standard deviation travel time parameter, again for the same reasons as discussed above.

In order to understand the relative impact that each decision type has in terms of biasing the model outputs, we calculate the elasticities based on the two SURE models and present these in Table 3. Values in bold represent estimates corresponding to statistically significant effects. The elasticities for the parameter estimates suggest that an increase of one percent of respondents within the sample acting in a lexicographic manner will result in a greater than one percent increase in biasing the parameter estimates obtained. Similar results are observed for respondents whose sensitivity to the various attributes changes during the course of the experiment, although for the cost standard deviation parameter, the impact is relatively elastic. Of the three decision types explored in the first case study, those displaying inconsistent choice behaviour have the smallest impact on biasing the parameter values, whilst the impact overall of additional respondents acting in a lexicographic manner has the largest overall impact.

Table 3: Parameter and T-ratio elasticities for Case Study 1

(a) <i>Parameters</i>				
	Cost (μ_1)	Cost (σ_1)	Time (μ_2)	Time (σ_2)
ii. Lexicographic (cost)	9.686	2.307	6.075	5.421
iii. Lexicographic (time)	10.521	3.094	5.983	5.047
iv. Inconsistent choice behaviour	-0.402	1.018	-0.901	-0.306
v. Compensatory with changing sensitivities	2.015	0.534	-1.491	-1.682
vi. Compensatory with changing sensitivities	6.333	16.351	9.943	4.879
(b) <i>T-ratios</i>				
	Cost (μ_1)	Cost (σ_1)	Time (μ_2)	Time (σ_2)
ii. Lexicographic (cost)	-0.059	0.527	-0.322	0.573
iii. Lexicographic (time)	-0.031	0.372	-0.312	0.606
iv. Inconsistent choice behaviour	0.012	0.054	-0.223	-3.608
v. Compensatory with changing sensitivities	0.015	-0.390	-0.485	-2.138
vi. Compensatory with changing sensitivities	-0.015	-0.985	-0.091	-2.560

Comparing the elasticities for the t -ratios, with the exception of the time standard deviation parameter with respect to inconsistent choice behaviour and respondents changing sensitivities, all elasticities are relatively inelastic suggesting a one percent increase in respondents within a sample employing any of the decision rules will result in a less than one percent bias in the t -ratios. In terms of the elasticity for the time standard deviation parameter with respect to sampled respondents displaying inconsistent choice behaviour, it is worth noting that the parameter estimate for this segment is statistically insignificant and hence this elasticity should be treated with extreme care. The parameter estimates for the segments of respondents who change sensitivities are statistically significant suggesting that this type of behaviour might have the largest potential impact on biasing the t -ratios of discrete choice models.

3.2 Case Study 2

A similar process as described for the first case study was used to construct the simulated data for the second case study. Unlike the first case study however, the attributes of the SC experiment used in the analysis were captured from real data collected from 844 respondents. The original data were collected in Sydney in 2004 and consisted of commuter, non-commuter and heavy vehicle segments (see Hess et al. 2008 for a recent application using the non-commuter segment of the data). The first alternative of the SC experiment was a reference alternative, using respondent provided attribute levels for a recent trip. The remaining two alternatives represented competing hypothetical routes (SC1 and SC2). As such, the reference alternative remained invariant across the 16 choice situations with only the levels of the hypothetical SC alternatives varying, although the levels of the reference alternative did differ between respondents. The alternatives were described in terms of free flow travel time (FFT), slowed down travel time (SDT), travel time variability (VAR),

running costs (RC) and toll costs (TC). The base utility specification for this case study is given as Equation (4):

$$\begin{aligned}
 U(ref) &= \beta_{ref} + \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC \\
 U(SC1) &= \beta_{SC1} + \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC \\
 U(SC2) &= \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC
 \end{aligned} \tag{7}$$

where $\beta_{FFT} \sim N(\mu_1, \sigma_1)$, $\beta_{SDT} \sim N(\mu_2, \sigma_2)$, $\beta_{VAR} \sim N(\mu_3, \sigma_3)$, whilst β_{RC} and β_{TC} are assumed to be fixed parameters.

Data for ten different types of choice behaviour were simulated as part of the case study. Simulated choices were based on assumed linear additive compensatory choice behaviour, non-trading where respondents always select the reference alternative, lexicographic behaviour with respect to each of the design attributes, lexicographic behaviour with respect to combined time (TIME = FFT + SDT) and combined cost (COST = RC + TC), and linear additive compensatory choice behaviour where composite travel times (Equation 8) and costs (Equation 9) were assumed to be used in the choice process.

$$\begin{aligned}
 U(ref) &= \beta_{ref} + \beta_{Time} Time + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC \\
 U(SC1) &= \beta_{SC1} + \beta_{Time} Time + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC \\
 U(SC2) &= \beta_{Time} Time + \beta_{VAR} VAR + \beta_{RC} RC + \beta_{TC} TC,
 \end{aligned} \tag{8}$$

$$\begin{aligned}
 U(ref) &= \beta_{ref} + \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{Cost} COST \\
 U(SC1) &= \beta_{SC1} + \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{Cost} COST \\
 U(SC2) &= \beta_{FFT} FFT + \beta_{SDT} SDT + \beta_{VAR} VAR + \beta_{Cost} COST,
 \end{aligned} \tag{9}$$

where $\beta_{Time} \sim N(\mu_4, \sigma_4)$, and β_{Cost} is assumed to be a fixed parameter.

The parameter values used in the simulation of the choice decision rule types are shown in Table 3. The parameter values for the three linear additive compensatory behavior types were obtained from models estimated on the real data sets.

As with case study 1, different subsets of the simulated data sets were assigned to form 16 master data sets according to an orthogonal experimental design (shown in Table 5). Values of 1 in the table denote the presence of simulated data in a master data set and a 0 the exclusion from that master data set. Based on the orthogonal design, each master data set has between three and 10 different choice rule types.

Again, similar to the first case study, a series of 100 Monte Carlo bootstrap runs were performed for each set, each consisting of 180 respondents. In each Monte Carlo bootstrap run, different proportions of respondents were drawn. For master data sets with three choice behaviour types, ten different variations of the proportions of each response type were used, with values between 0.1 and 0.75. For master data sets with four types of response types, the proportion of response types were varied between 0.1 and 0.5 using 13 different variations of proportions whilst master data sets with five behaviour types varied the response type proportions between 0.1 and 0.6 using 11 different variations of proportions. For master data sets with six choice decision rule types, the proportions of each type of respondent in the data set were varied between 0.1 and 0.5 using 16 different variations of proportions within the sample. A single Monte Carlo run with 100 bootstrap runs was performed on the final Master data set assuming equal representation of each decision rule type (i.e., 18 respondents from each decision rule were drawn during the Monte Carlo simulation). The last column of Table 4 shows the number of variations applied to each master data set. In total, 178 bootstrap runs were performed representing a total of 17,800 MMNL model estimations.

The results for the case study are available from <http://sydney.edu.au/business/itls/staff/johnr/papers> in Tables S3 and S5. Unlike the first case study, all decision rule types occur simultaneously in only one of the 178 model runs. Further, despite data sets IX and X being generated under the utility specification assumptions given in Equations (8) and (9) respectively, the actual Monte Carlo simulations were estimated using the utility specification given in Equation (7). As such, simulations with samples drawn from data sets IX and X represent a misspecification of the true utility functions of these respondents, where the aim is to gauge the effect of this misspecification. The first set of results, given in Table S3, report the outcomes from a SURE model regressing the parameter MSEs obtained from the 178 Monte Carlo runs against the various proportions of the choice behaviour types. Also shown in the table are the results of a linear regression model of the EMSE calculated over the 178 Monte Carlo runs. Bold values in the table represent the effect of a particular trait on the associated parameter (e.g., the effect of lexicographic behaviour towards time on the time parameter). As a first observation, it should be noted that some of the parameters are statistically significant and negative, hence showing that an increase in the given proportion leads to a reduction in bias. This is to be expected as an increase in a given proportion means a decrease in another proportion. If the trait for which the proportion is decreased has a bigger impact than the trait for which the proportion is increased, then we would expect to see a reduction in the bias. As such, the observed impacts on model results may at times appear to be counter-intuitive when in fact they are not. In the remainder of this discussion, we will thus focus on factors that lead to increases in bias.

As is to be expected, the presence of non-traders (respondents who always select the reference alternative) produces a statistically significant bias for the alternative specific constant (ASC) of the reference alternative. The parameter may be interpreted as follows. If

10 percent of the sample are non-traders (as described), the reference alternative ASC will be biased by 0.5875 (i.e., 5.875×0.1) relative to what it would have been if the entire sample was acting in a linear additive compensatory fashion. The presence of non-trading respondents also has a statistically significant, although much smaller in magnitude, bias upon the ASC for the second constant as well as biasing the travel time variability mean and standard deviation parameters.

Lexicographic decision rules as applied to specific attributes statistically impact upon the parameter estimates for that attribute, both in terms of the mean, and where relevant, standard deviation parameter estimates. As with the first case study, the results here suggest an increase in the bias in the parameter estimates associated with a particular lexicographic choice rule as the proportion of respondents displaying that type of behaviour increases within the sample. Once more, the fact that the presence of lexicographic rules in a sample may influence the standard deviation also raises questions about what precisely standard deviation parameters obtained from MMNL models are actually retrieving. Also of note is the fact that some of the lexicographic decision rules appear to increase the bias in model ASCs.

Lexicographic behaviour where a respondent also combines the two travel time attributes also appears to have a statistically significant influence upon parameter estimates, although the influence appears to be largely confined to the affected attributes (i.e., they do not appear to bias the cost parameters) and the model ASCs. Additionally, respondents displaying this type of behaviour in a sample appear to increase bias in the parameter estimates. Interestingly, an increase in the sample of respondents displaying lexicographic behaviour towards the combined travel cost attribute results in a statistically significant reduction in the bias for the time parameters and the constant for the reference alternative but has no discernable impact upon the cost parameters. Increasing the number of respondents in the sample who act in a linear additive compensatory manor but who combine the travel cost

attributes also has significant impacts upon the biases of the parameters in the data. In particular, the presence of such respondents appears to decrease bias for all other parameter estimates. Behaviour where respondents combine the two cost attributes but apply linear additive compensatory behaviour also appears to increase the bias in the two cost coefficients and the first ASC but not the time parameters. One possible reason for this is that the magnitudes of the two separate time parameters are fairly similar and not very different from the combined parameter; this could help facilitate estimation of a joint parameter.

Table 8 presents the outcomes from a SURE model regressing the bias in the WTP estimates calculated at the means of the travel time parameter distributions obtained from the 178 Monte Carlo runs (i.e., we do not look at the full parameter distributions, simply the ratios of the mean travel time parameters over the fixed cost parameters). This once again shows that lexicographic decision rules where respondents attend to only a subset of attributes tend to increase the bias in one or more WTP values within the sample. Of particular note is the influence of non-trading behaviour where respondents always choose the reference alternative irrespective of the attribute levels of the alternatives under consideration which would be expected to influence mainly the model ASCs, but which also, as shown here, can also bias the WTP outcomes derived from the model. This raises questions as to whether removing the respondents from the data prior to estimation and inflating the ASCs afterwards might be a valid strategy.

Table S4 presents the results for a SURE model regressing the average t -ratios for the various parameter estimates calculated over the Monte Carlo bootstrap runs. The presence of all decision rule types appears to influence the t -ratios relative to those obtained from a sample behaving in a linear additive compensatory manner, sometimes for the better, but on other occasions for the worse. For example, the presence of respondents in the simulated samples who act in a linear additive compensatory fashion but combine the two costs produce

a very large positive and statistically significant (positive) effect on the (negative) t -ratios for the two cost parameters relative to those respondents who act in a linear additive compensatory manner following the estimated utility specification. This suggests that the increasing presence of such respondents in the data result in smaller t -ratios for the two cost attributes relative to samples with greater numbers of respondents acting in a linear additive compensatory manner assumed in the model estimation. The other interesting result from Table S5 is the influence that non-traders (i.e., those who always choose the same alternative, in this case represented by the reference alternative) have upon the t -ratios of the various parameter estimates within the model. Once again, it is clear that the presence of non-traders within a data set may affect not just the model ASCs but also the t -ratios of other parameters, sometimes positively and sometimes negatively. This makes dealing with the presence of non-traders a far more difficult task than simply removing them from the analysis.

As per the first case study, in order to understand the relative impact that each decision rule type has in terms of biasing the model results, elasticities are generated for each of the SURE model estimates. For reasons of space, these elasticities are reported in Tables S6 to S8 located at <http://sydney.edu.au/business/itls/staff/johnr/papers>. Overall, the majority of the elasticities suggest a relatively inelastic impact upon the parameter bias observed when increasing the proportion of respondents displaying any of the decision rules, however a number of notable exceptions exist. In particular, increasing the proportion of respondents displaying lexicographic behaviour towards a particular attribute appears to have a larger than proportional impact upon biasing the associated parameter estimate, both in terms of the mean and standard deviation parameters, but a less than proportional impact upon the other parameter estimates associated with the other attributes. Combining attributes as a decision rule appears to have a relatively inelastic impact upon biasing the parameter estimates. Interestingly, non-trading behaviour was observed to have a relatively elastic impact upon

biasing the reference alternative ASC and both the mean and standard deviation variance parameters, suggesting that this type of behaviour has the potential to impact significantly upon parameters other than the ASCs.

In terms of the elasticities for WTP biases, lexicographic behaviour appears to have a large relatively elastic impact upon biasing the FFT and SDT time WTP values calculated relative to the fixed running cost parameter. Interestingly, the presence of non-traders within the data set also appears to have a relatively inelastic elasticity for the time SDT WTP values calculated relative to the fixed running cost parameter. Ignoring the ASCs, lexicographic behaviour towards the cost attributes results in an inelastic elasticity towards t -ratio of the mean of the variance parameter, as does non-trading behaviour. Combining the times and costs and acting in either a linear additive or lexicographic manner also have a larger than proportional impact upon biasing the t -ratios of the cost parameters. All other effects, where statistically significant, appear to have a relatively elastic impact upon the t -ratios, save for the ASCs which appear to be particularly impacted upon by the presence of respondents acting in a non-linear compensatory fashion.

4. General Discussion and Conclusions

In this paper, we have examined a number of issues relating to the presence of respondents within SC data sets who utilise various decision making rules. Using simulated data, we have been able to show that the presence of differing information processing strategies within a sample may have statistically significant effects upon a range of model outputs. In particular, we have been able to show that non-linear additive-compensatory behaviour may result in, relative to models estimated purely on data sets consisting of linear additive compensatory behaviour, biased parameter estimates and biased t -ratios.

Our findings further support the need to incorporate heterogeneous IPSs into discrete choice models. Whilst further research is required into how best to elicit either directly from

respondents or indirectly (e.g., Hensher 2008 and Hensher and Puckett 2008) via some form of statistical model (e.g., Swait 2001, Hess and Rose 2007 and Hess and Hensher 2010), the IPS strategies employed by different decision makers represents a clear direction for the literature. Further, testing more complex non-linear utility structures rather than simply assuming linear-additive specifications, such as models that allow for attribute thresholds (e.g., Swait 2001, Cantillo and Ortuzar 2005 and Cantillo et al. 2006) may also assist in detecting the true choice processes used by different respondents over the course of a SC experiment.

Further still, the findings of this paper suggest not only the need for further research into models accounting for different decision rules (as well as a wider use of the models that currently exist), but also the need to construct tests that may be used to determine when a decision maker has adopted a particular decision process. Currently, only in the simplest of experiments can complex decision rules be detected and readily identified and even simple rules can often be difficult to detect in even very simple experiments. Such tests are necessary for any model that takes a deterministic approach to treating heterogeneous IPSs and may also aid models capable of a stochastic treatment of such processes.

Although not conclusive, the results of the case studies suggest that lexicographic behaviour has a significant impact upon the parameter results obtained from discrete choice models as well as the WTP values, whilst respondents who change sensitivities over the course of an SP experiment may also have a larger than proportional impact upon the model outputs. The presence of respondents acting in an inconsistent manner appears to be of less consequence in terms of potential biases, and therefore should be of least concern to practitioners when dealing with this type of data. Interestingly, the presence of non-traders also has the potential to significantly bias the model outputs obtained from SP data, and as such, significant care should be taken when such respondents are present within the data.

In our discussion of heterogeneous choice rules, we identified that several of the decision processes explored within this paper have as a potential cause the SC task itself. As such, our findings stress the need to sufficiently pilot SC experiments prior to going to the field. Our findings also give further impetus to calls for research into better experimental designs that are able to account for, or moderate the influences of heterogeneous decision processes. Understanding the demand characteristics and how these influence the outcomes of SC experiments represents one of the greatest research challenges moving forward into the future.

Nevertheless, despite showing that non-linear-additive-compensatory behaviour can influence the model outputs obtained from discrete choice models, we caution against treating the results presented as indicative of the precise biases that such behaviour may produce. Indeed, the results presented reflect the inputs into the Monte Carlo simulations we performed, including the parameter priors assumed, the experimental designs used and even the number of choice tasks each simulated respondent was assumed to have reviewed. Rather we wish that our results reflect that the presence of different IPSs within SC data may have unintended impacts upon such model results and highlight that care needs to be taken in interpreting the results obtained from discrete choice models, particularly models that do not account for the possibility of heterogeneous decision processes.

Finally, our findings also suggest a way forward in terms of simulation of discrete choice data used in testing discrete choice models. Whilst Garrow et al. (2010) provide an excellent overview of the simulation process, even allowing for differences in scale, such simulation processes still assume that all respondents act in a linear additive compensatory like fashion. Clearly this is not the case in reality.

References

Blume, L., A., Brandenburger and E., Dekel (2006) 'An overview of lexicographic choice under uncertainty', *Annals of Operations Research*, 19(1), 229-246.

Cantillo, V. and J.deD., Ortuzar (2005) 'A semi-compensatory discrete choice model with explicit attribute thresholds of perception', *Transportation Research Part B*, 39(7), 641-657.

Cantillo, V., B., Heydecker and J.deD., Ortuzar (2006) 'A discrete choice model incorporating thresholds for perception in attribute values', *Transportation Research Part B*, 40(9), 807-825.

Fujii, S. and T., Garling (2003) 'Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis', *Transportation Research Part A*, 37(4), 389-402.

Garrow, L. A., T. D., Bodea, and M., Lee (2010) 'Generation of synthetic datasets for discrete choice analysis', *Transportation*, 37(2), 183-202.

Hensher, D. A. (2008) 'Joint estimation of process and outcome in choice experiments and implications for willingness to pay', *Journal of Transport Economics and Policy*, 42 (2), 297-322.

Hensher, D. A. and S. M., Puckett (2008) 'Behavioural responses of freight transporters and shippers to variable road user charging schemes', in Verhoef, E., B., Van Wee, L., Steg, and M., Bliemer (eds) *Pricing in Road Transport: A Multi-Disciplinary Perspective*, Edward Elgar, 29-63.

Hensher, D. A., J. M., Rose and T., Bertoia (2007) 'The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies', *Transportation Research Part E*, 43(2), 73-89.

Hess, S. and D. A., Hensher (2010) 'Using conditioning on observed choices to retrieve individual-specific attribute processing strategies', *Transportation Research Part B*, 44(6), 781-790.

Hess, S. and J. M., Rose (2009) 'Should reference alternatives in pivot design SC surveys be treated differently?', *Environmental and Resource Economics*, 42(3), 297-317.

Hess, S. and J. M., Rose (2007) 'A latent class approach to modelling heterogeneous information processing strategies in SP studies', paper presented at the *Oslo Workshop on Valuation Methods in Transport Planning*, Oslo.

Hess, S., J. M., Rose and J. W., Polak (2010) Non-trading, lexicographic and inconsistent behaviour in stated choice data. *Transportation Research Part D*, 15(7), 405-417.

Hess, S., J. M., Rose, and D. A., Hensher (2008) 'Asymmetrical preference formation in willingness to pay estimates in discrete choice models', *Transportation Research Part E*, 44(5), 847-863.

Kuriyama, K. (2005) 'Strategic effects on stated preferences for public goods: A theoretical and experimental analysis of the contingent valuation survey', *The Waseda Journal of Political Science and Economics*, 359, 83-92.

Rose, J. M. and I., Black (2006) 'Means Matter, but Variance Matter Too: Decomposing Response Latency Influences on Variance Heterogeneity in Stated Preference Experiments', *Marketing Letters*, 17(4), 295-310.

Rose, J. M., M. C. J, Bliemer, D. A., Hensher, and A. T., Collins (2008) 'Designing Efficient Stated Choice Experiments Involving Respondent Based Reference Alternatives', *Transportation Research Part B*, 42(4), 395-406.

Rosenberger, R. S., G. L., Peterson, A., Clarke and T. C., Brown (2003) 'Measuring dispositions for lexicographic preferences of environmental goods: integrating economics', psychology and ethics. *Ecological Economics*, 44, 63-76.

Salensminde, K. (2002) 'The impact of choice inconsistencies in stated choice studies', *Environmental and Resource Economics*, 23, 403-420.

Salensminde, K. (2001) 'Inconsistent choices in stated choice data: use of the logit scaling approach to handle resulting variance increases', *Transportation*, 28(3), 269-296.

Salensminde, K. (1994) 'The impact of choice inconsistencies in stated choice studies', *Environmental and Resource Economics*, 23(4), 1994, 403-420.

Swait, J. (2001) 'A non-compensatory choice model incorporating attribute cut-offs', *Transportation Research Part B*, 35(10), 903-928.

Train, K. and W. W., Wilson (2008) 'Estimation on stated-preference experiments constructed from revealed-preference choices', *Transportation Research Part B*, 42(3), 191-203.

Verplanken, B. and H., Aarts (1999) 'Habit, attitude and planned behaviour: Is habit an empty construct or an interesting case of goal-directed automatic?', *European Review of Social Psychology*, 10, 101-134.