CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

A Systematic Test for Identifying Inconsistency, Learning and Fatigue Effects in Stated Choice Surveys

Ali Rezaei
Zachary Patterson

June 2015

CIRRELT-2015-27

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

UNIVERSITÉ LAVAL McGill UNIVERSITÉ Concordia UNIVERSITY ÉTS UQÀM Université du Québec à Montréal HEC MONTRÉAL POLYTECHNIQUE MONTRÉAL Université de Montréal

A Systematic Test for Identifying Inconsistency, Learning and Fatigue Effects in Stated Choice Surveys

Ali Rezaei[1,*], Zachary Patterson[1,2]

[1] Department of Geography, Planning and Environment, Concordia University, 1455 de Maisonneuve W., H 1255-26 (Hall Building), Montreal, Canada H3G 1M8

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

Abstract. Surveys focusing on choice behaviour, and in particular, Discrete Choice Experiments (DCEs) are widely used in studies across many disciplines, such as marketing research, transportation, economics, and environmental studies. Investigation of inconsistency in responses to choice tasks in DCEs has received a fair bit of attention. The phenomenon of inconsistency of responses across choice tasks in DCEs is well recognized. It has been attributed to many causes including, respondent inattention, the complexity of choice tasks, as well as to learning and fatigue effects. The latter two are known commonly as ordering effects. In the behavioural modelling literature concentrating on understanding ordering effects, there is mixed evidence as to whether (and which) of these ordering effects are observed. The most commonly used econometric model to evaluate ordering effects is the Heteroscedastic Multinominal Logit (HMNL). In this research we evaluate ordering effects using a newly proposed systematic consistency test (SCT) that detects inconsistent choices in complex surveys. The SCT is based on the assumption that each respondent has a given preference structure and that her/his choices should be consistent with this structure across their choices. As such, choices that are not consistent with an individual's observed preference structure are identified as *inconsistent with his/her own choices*. In the paper, the SCT procedure is applied to examining learning and ordering effects in data from three DCEs. The results of the procedure are then compared with the results of the H-MNL model used to analyse ordering effects on the same data. Our analysis shows that the results of the SCT approach are similar to the H-MNL approach, but that the SCT approach provides researchers with information about which choice tasks are the source of inconsistency in data set.

Keywords. Discrete Choice Experiments (DCE), inconsistent behaviour, learning, fatigue, ordering effect, scale, Systematic Consistency Test (SCT).

Acknowledgements. The authors would like to acknowledge support from the Canada Research Chairs Program, the Canadian Social Sciences and Humanities Research Council (SSHRC) as well as the Canadian Foundation for Innovation (CFI).

---

* Corresponding author: a.rezaaei@gmail.com

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2015

1. INTRODUCTION

Discrete Choice Experiments (DCEs) are a common tool used to understand the factors affecting people's choices, and the degree to which different factors affect these choices. Central to the Discrete Choice Experiment is a series of hypothetical choice tasks presented to respondents who are asked to choose their preferred alternative. Respondent choices are then analysed using discrete choice statistics that allow evaluation of statistical significance of different factors as well as the magnitude of the influence. Critical to the suitability of the use of discrete choice statistics in this context is that the preferences of individuals are stable across choice tasks. Many have been concerned, however, by whether decision strategies (and thereby observed preference structures) change as a respondent proceeds through multiple choice tasks (Swait, et al., 2001). The phenomenon of preference structures appearing to change over the course of choice tasks is referred to as ordering effects and can be caused by learning effects, i.e. 1) institutional learning, which says most respondents of a DCE have never faced an SP survey before and as they proceed through tasks they become more familiar with the nature of choice questions(Bateman, et al., 2008); 2) value learning, which relates to discovering and learning one's own preferences as one progresses through choice tasks, such as attitude to risk (Braga, et al., 2005; Bateman, et al., 2008); or fatigue (boredom) effects that work in opposite direction of learning effects. That is, as a respondent moves through tasks the variance of utility functions increases and choices become less deterministic (Bradley, et al., 1994; Hess, et al., 2012). Both of these effects may cause inconsistencies in the structure of preferences and consequently between one's choices (Caussade, et al., 2005). In general, these effects are not mutually exclusive and some researchers have investigated them jointly. Therefore, some researchers have suggested a U-shaped progression of inconsistency in choice tasks. They argue that as a respondent proceeds through the tasks learning effects decrease, while fatigue effects increase (Caussade, et al., 2005; Chung, et al., 2010). Also, some studies have shown respondent consistency depends on the complexity of choice tasks, such as the number of alternatives, number of attributes, their levels, ranges and design efficiency (Swait, et al., 2001; Caussade, et al., 2005; Hensher, 2006; Louviere, et al., 2008). Moreover, different surveying techniques have been found to result in different learning and fatigue effects patterns i.e. internet-based surveys, might result in weaker learning effects and much stronger fatigue effects compared to mail survey (Savage, et al., 2008).

In general, studies focusing on evaluating ordering effects and inconsistency across choice tasks have employed statistical and econometric approaches to control for inconsistency, in particular by allowing for unobserved preference or scale heterogeneity (Louviere, et al., 2008; Hess, et al., 2012; Czajkowski, et al., 2014).

While a great deal has been learned from these statistical approaches to understanding inconsistency, we believe that inconsistency can also be examined from the perspective of preference-based consumer theory, as suggested by Lancsar & Louviere (2006). In particular, inconsistency can be seen as a violation of rational behaviour and especially a violation of the axiom of transitivity (Sælensminde, 2002) where the axiom of transitivity states that for alternatives A, B and C in a choice set, if A>B and B>C, then A>C (McIntosh, et al., 2002; Rulleau, et al., 2012). An attempt to quantify inconsistency through the identification of intransitivity across choice tasks has been recently proposed by Rezaei & Patterson (2015). The present paper extends this previous work by using their measure of inconsistency to evaluate learning and fatigue effects in DCEs.

As such, it is closely related to Louviere et al.'s (2008) pioneering article in which they assume that high choice consistency in DCEs can be associated with low error variance, low choice variability, low choice uncertainty, or low variance heterogeneity. In order to be able to compare the results of the SCT approach with how other researchers have examined the question of inconsistency and ordering effects (e.g. (Hess, et al., 2012; Czajkowski, et al., 2014; Chen, et al., 2015)) we use H-MNL model structures to account for inconsistency in responses. Also, this paper adds to previous research by applying a systematic approach to investigate inconsistencies between one's choices caused by ordering effects. This will provide analysts with information about where problematic choices are mostly observed in a task sequence.

The remainder of the paper is structured as follows. Section two starts with a simple (two alternatives, two attributes) example to explain what inconsistent mutual relations (choices) mean. In addition, an approach to transform a stated choice data to mutual relations so that the SCT approach can be used to detect inconsistencies is presented. In section three, the dominance approach to find condition profiles and dominance cones, and how they are used in the test procedure, is briefly explained. Section four introduces details of three case studies used to evaluate the use of SCT in the examination of learning and fatigue effects. Section five reports the results of estimating the H-MNL model and running the CST, together with discussion about our findings. Finally, section six provides concluding remarks and offers suggestions relevant for the future investigation of this topic.

2. MODELLING INCONSISTENT BEHAVIOUR

This section draws heavily on the companion paper where the SCT approach was first proposed (Rezaei, et al., 2015). The central idea of the approach is to try to detect inconsistencies across an individual's choices, and in particular whether a violation of the axiom of transitivity (Lancsar, et al., 2006) is observed. Like previous research, the SCT relies fundamentally on the differences between attribute levels of alternatives (Sælensminde, 2002; Hess, et al., 2010; Rose, et al., 2013). As an example, one can think of a binary choice with two alternatives, and two attributes each of which includes three levels, L_1, L_2 and L_3 (where $L_1 < L_2 < L_3$). For the sake of simplicity utility functions are assumed to be monotonically increasing in each of the attributes. This assumption is used to explain the approach, but the approach itself does not require this to be the case as will be explained in section 3.2. The first five columns of Table 1 present an example of attribute levels and decisions made by a respondent in four choice situations.

There are five possible attribute level differences for each attribute: two levels better, one level better, etc. Each choice can be divided into two classes: the alternative is selected (i.e. it was considered better (*b*) than its alternative); or it is rejected (i.e. the alternative was considered worse (*w*)). The differences in attribute levels and decision classes are presented in the last three columns of Table 1. For each choice, the first line presents the attribute levels of alternative one *minus* the attribute levels of alternative 2 (*Alt.1-Alt.2*). The second line presents (*Alt.2-Alt.1*). Evidently, the components of these two lines are symmetric.

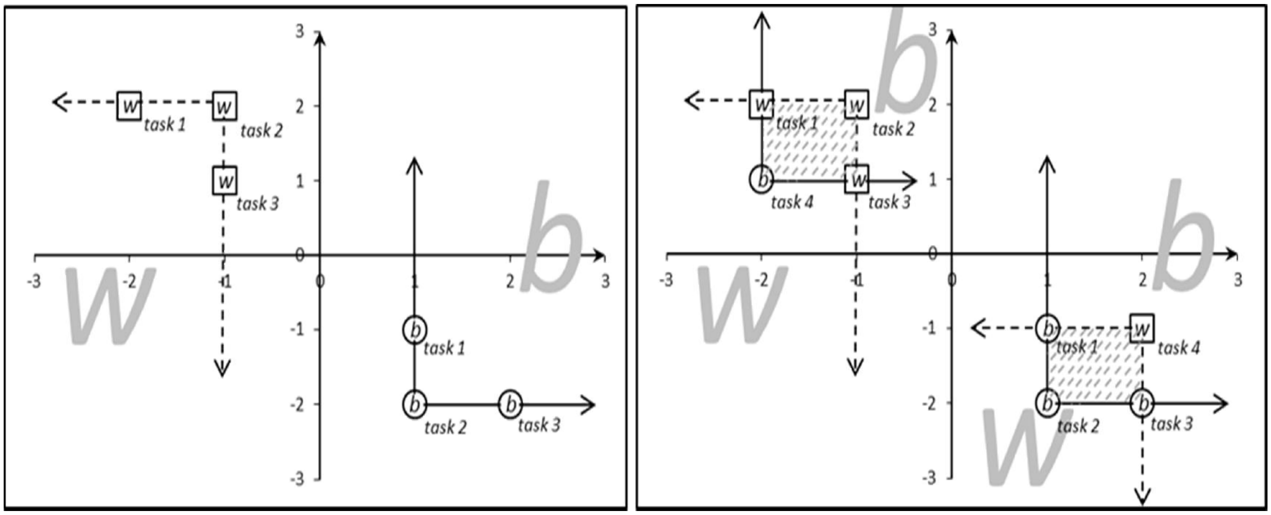
Considering Table 1, and in particular focusing on the first line of each task (i.e. *Alt.1-Alt.2*), shows the following: in all tasks Alt.1 dominates Alt.2 with respect to attribute 1 (att.1), while Alt.2 dominates Alt.1 with respect to att.2, implying a trade-off between the two alternatives. In the first case, the respondent chose an alternative that is one level better with respect to att.1, but one level worse with respect to att.2. This implies that the respondent values att.1 more than att.2. In task 2 the respondent chose the same alternative when att.2 is two levels worse. This choice is not inconsistent with the choice made in task 1. Indeed, task 2 provides analysts with more information than task 1. That is, the respondent values att.1 much more than att.2. The choice in task 3 is consistent with the preference structure implied by the first two choices tasks.

TABLE 1 An Example Of Attribute Levels And Decisions Made

	Alternative levels				Difference between attribute levels		
Task	Alt.	Att.1	Att.2	Decision	Att.1	Att.2	Decision
1	1	L_3	L_1	*Select*	*1*	*-1*	*b*
	2	L_2	L_2	*Reject*	*-1*	*1*	*w*
2	1	L_3	L_1	*Select*	*1*	*-2*	*b*
	2	L_2	L_3	*Reject*	*-1*	*2*	*w*
3	1	L_3	L_1	*Select*	*2*	*-2*	*b*
	2	L_1	L_3	*Reject*	*-2*	*2*	*w*
4	1	L_3	L_2	*Reject*	*2*	*-1*	*w*
	2	L_1	L_3	*Select*	*-2*	*1*	*b*

This is represented in Figure 1A. All points representing *Alt.1-Alt.2* fall in the lower right quadrant (4^{th} quadrant), and the points associated with *Alt.2-Alt.1* fall in the upper left (2^{nd}) quadrant. Because results are symmetric in the two quadrants, we concentrate on the 4^{th} quadrant. The 2D cone drawn by the continuous lines emanating from the point represented by the choice made in task 2 shows an area inside which all points dominate task 2. Based on the decision made in task 2, *Alt.1* was the implied better alternative (*b*). In all tasks for which the point representing *Alt.1-Alt.2* falls inside the cone, the respondent *should* choose *Alt.1*.

In task 4, however, the first alternative is two levels better and one level worse than the second one with respect to *att.1* and *att.2*, respectively. This is shown in Figure 1B. This task dominates all other tasks with respect to both attributes (i.e. it falls inside the cone). That is in task 4 *Alt.1* is superior to *Alt.2* (based on the respondent's observed preference structure from task 2), with respect to the both attributes. So, assuming transitivity, and given the knowledge inferred from the first three tasks, the respondent *should* choose *Alt.1*. However, s/he selected *Alt.2*, which is *inconsistent* with the other choices.



(A) (B)

FIGURE 1 - Dominance cones and inconsistent behaviour (A – consistent, B – inconsistent)

This is similar to how some other researchers have analysed the question of inconsistency in DCEs in the past (e.g. (Hess, et al., 2010; Rose, et al., 2013)). In order to do so, used substitution ratio benefits (between two attributes) relative to other alternatives as an indicator used to detect inconsistency across respondent choices (Hess, Rose, & Polak, 2010). In the graphical representation above, for each choice, the ratio benefit value is equal to the negative of the inverse of the slope of the line connecting the origin of the coordinate system to the point associated with that choice. Such an approach can be relatively easily performed in simpler experiments (e.g. experiments using only two attributes such as time and cost), but using the ratio benefit presents difficulties as experiments become more complex (i.e. include more attributes) (Hess, Rose, & Polak, 2010).

The approach used in this paper, however, can be easily extended to more complex experiments, ones that include more attributes that can be treated by mapping the information in a higher-level coordinate system (e.g. *m*D coordinate system with *m*D cones for an experiment with *m* attributes); or additional alternatives per choice task. In this case a choice can be shown with *l-1* (*l*: number of alternatives) Mutual Dominance Relations (MDRs) between the chosen alternative and the other alternatives. Then each MDR can be treated as a binary choice and mapped on the *m*D coordinate system (For more information see (Louviere, et al., 2000)).

Finally, the approach can also be extended to experiments that include the ranking of alternatives in a stated preference setting. In this case, any given response tells us that the respondent prefers the h^{th} alternative, $h=\{1, . . ., l\}$, in the preference ranking, to the *l-h* alternative that are less preferred. This can be represented as *l-h* binary choices, and therefore all MDRs between alternatives in a task can be treated as $\frac{(l)\times(l-1)}{2}$ binary choices and mapped on the *m*D coordinate system.

3. METHODOLOGY

This section first explains the econometric structure most often used to capture inconsistency in DCEs, by accounting for heterogeneity in variance (heteroskedastic error). This approach has been employed for example by Louviere et al. (2008). It then continues to describe the use of the Dominance-based Rough Set Approach (Greco, et al., 2001) as a tool to detect inconsistent behaviour in complex experiments after implementing the transformation of the DCE response data as described in the previous section.

3.1. Econometric Structure, Heteroskedastic MNL Model

As noted in the choice modelling literature, Multinomial Logit (MNL) model parameters, β_{MNL}, are not the true estimation of effect size of attributes on choices (Train, 2003). Rather, they are confounded with scale (McFadden, 1974; Ben-Akiva, et al., 1985; Swait, et al., 1993; Train, 2003; Louviere, et al., 2008). Consider the following true underlying utility function (Louviere, et al., 2008):

$$U^*_{nj} = \beta'_{true} X_{nj} + \varepsilon^*_{nj} \tag{1}$$

Where β_{true} denotes the matrix of true underlying coefficients and $Var(\varepsilon^*_{nj}) = \sigma^2_\varepsilon$ can take any value. The standard MNL model normalizes this utility function by a factor, λ ($\lambda > 0$), so that the variance of the error term (unobserved components) equals $\pi^2/6$. Mathematically,

$$\lambda\, U^*_{nj} = \lambda\, \beta'_{true} X_{nj} + \lambda\, \varepsilon^*_{nj} \tag{2}$$

$$U^*_{nj} = \beta'_{MNL} X_{nj} + \varepsilon^*_{nj} \tag{3}$$

where $\lambda\, U^*_{nj} = U_{nj}$, $\lambda\, \beta'_{true} = \beta'_{MNL}$, $\lambda\, \varepsilon^*_{nj} = \varepsilon_{nj}$, and $Var(\lambda\varepsilon^*_{nj}) = Var(\lambda\varepsilon_{nj}) = \pi^2/6$. Therefore, $\lambda = \pi \;\; /\sigma_\varepsilon\, \sqrt{6}$. The term λ is called the scale parameter. This parameter can be used as a measure of choice (in)consistency. The higher the variance of the unobserved components (the less consistency), the smaller the λ (Louviere, et al., 2008).

Although the scale parameter is not identified when estimating models with a single data source, the relative value of scales can be identified when dealing with two or more sources of data (see, e.g., (Swait, et al., 1993; Louviere, et al., 2008)). In this study, we consider each choice task in the task sequence provided to respondents, as a source of data that can be used to identify task-specific scale parameters. We use the H-MNL as a statistical model that allows us to estimate the relative scale parameters jointly with preference parameters.

The H-MNL model is often used to combine two sources of data and estimate the relative scale, such as the case of combining stated and revealed preference data (e.g.(Hensher, et al., 1998)). In our case, we combine data from multiple choice tasks for each survey. For task number *k* in the choice task sequence, the choice probability is given by

$$P_{ni,k} = \frac{\exp(\lambda_k \beta' X_{ni})}{\sum_{j=1}^{J} \exp(\lambda_k \beta' X_{nj})}, \tag{4}$$

where one of the λ's is set to unity. Then maximum likelihood is used to estimate the additional scale parameters jointly with the preference parameters, β's.

3.2. Dominance Approach

This section summarizes the description of how the Dominance approach is applied in the Systematic Consistency Test. The interested reader can see Rezaei & Patterson (2015) for a more detailed description. The Dominance-based Rough Set Approach (DRSA) is an extension of Rough Set theory (Witlox, et al., 2004) that explicitly takes into account the preference ordering of attributes (Greco, et al., 2001). It has been applied as a data mining and knowledge discovery tool in several fields (Greco, et al., 2007; Zhai, et al., 2009; Liou, 2009; Nassiri, et al., 2012). Its prediction model is in the form of decision rules (Liou, 2009). DRSA has also been extended to Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) that allows some inconsistencies in data by defining a parameter called the "consistency level".

Following Greco et al.'s (2001) description, according to DRSA theory (Greco, et al., 2001; Dembczyński, et al., 2009; Zhai, et al., 2009; Liou, 2009), information regarding choice is represented in the form of an information table. The rows of the table refer to distinct objects (actions), while the columns refer to attributes considered. An information table is the 4-tuple information system *IS = (U, Q, V, f)*, where *U* is a finite set of objects (universe), *Q={q1, q2, . . .,qm}* is a finite set of attributes, $V=\cup_{q\in Q}Vq$ in which *Vq* is the domain of attribute *q*, and *f: U×Q→V* is a total function so that f(x,q)∈Vq for each *q*∈ Q, *x*∈ *U*, called the information function. The set *Q* is, in general, divided into set *C* of condition attributes and set *D* of decision attributes.

3.2.1. Rough Approximation by Means of the Dominance Relationship

Let $\succeq_q$ be an outranking relation on *U* with reference to attribute $q \in Q$, so that $x \succeq_q y$ means that *x* is at least as good as *y* with respect to attribute *q*. Supposing that $\succeq_q$ is a complete pre-order, i.e., a strongly complete (for each *x*, $y \in U$, at least one of $x \succeq_q y$ and $y \succeq_q x$ is verified, and hence with respect to attribute *q, x* and *y* are always comparable) and transitive binary relation. Moreover, let $\boldsymbol{Cl} = \{Cl_t, t \in T\}$, $T = \{1, \ldots, n\}$, be a set of classes of *U*, so that each $x \in U$ belongs to one and only one class $Cl_t \in \boldsymbol{Cl}$. It is worth noting that in this paper $\boldsymbol{Cl}$ includes two classes: $T = \{b, w\}$(in which *b* and *w* stand for better and worse, respectively), so that $b > w$. Therefore, the DRSA relations (Greco, et al., 2001) can be re-defined as follows:

if ≥ is a comprehensive outranking relation on *U*, then it is supposed that

$$(x \geq Cl_b,\ y \in Cl_w,\ b > w) \Rightarrow x > y, \tag{2}$$

where $x > y$ means $x \geq y$ and not $y \geq x$.

Let's define unions of classes by a specific dominated or dominating class – these unions of classes are called upward and downward unions of classes, respectively. The upward union of class *b* is defined as:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad t = w, b\text{; and} \tag{3}$$

The downward union of classes is defined as:

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \quad t = w, b \tag{4}$$

The statement $x \in Cl_t^{\geq}$ means that "x belongs at least to class Cl_t", while $x \in Cl_t^{\leq}$ means that "x belongs at most to class Cl_t". To clarify, the union $Cl_t^{\geq}$ is the set of objects belonging to class Cl_t or a more desired class, whereas the union $Cl_t^{\leq}$ is the set of objects belonging to class Cl_t or a less desired class. It should be noted that $Cl_w^{\geq} = Cl_b^{\leq} = U$, $Cl_b^{\geq} = Cl_b$ and also $Cl_w^{\leq} = Cl_w$. Consequently, we have $Cl_b^{\geq} = U - Cl_w^{\leq}$, that is, all the objects belonging to class Cl_b or more desirable belong to class *U* minus Cl_w or less desirable and simply $Cl_b = U - Cl_w$.

It is said that object *x P*-dominates object *y* (or, *x* ***P****-dominates y*) with respect to $P \subseteq C$, denoted as $xD_P y$, if $x \succeq_q y$ for all $q \in P$, and $D_P = \bigcap_{q \in P} \succeq_P$, then the dominance relation D_P is a partial pre-order. Given $P \subseteq C$ and $x \in U$, the dominance relations ("granules of knowledge") (Greco, et al., 2001) are:

$$D_p^{+}(x) = \{y \in U : yD_P x\}, \tag{5}$$

$$D_p^{-}(x) = \{y \in U : xD_P y\} \tag{6}$$

called the *P*-dominating set (a set of knowledge dominating *x*) and the *P*-dominated set (a set of knowledge dominated by *x*), respectively.

For any set of criteria $P \subseteq C$, we say that the inclusion of object $x \in U$ to the upward union of classes $Cl_t^{\geqslant}$, for *t*= *w, b*, makes an *inconsistency* if one of the following conditions happens:

(1) x belongs to class Cl_b while being *P*-dominated by an object *y* belonging to a class Cl_w, in other words, $x \in Cl_b$ but $D_p^+(x) \cap Cl_w^{\leqslant} \neq \phi$; or

(2) x belongs to class Cl_w while it *P*-dominates an object *y* belonging to class Cl_b, in other words, $x \notin Cl_b$ but $D_p^-(x) \cap Cl_b^{\geqslant} \neq \phi$.

In this case, it is said that x belongs to Cl_b with *some ambiguity*. In contrast, if $x \in Cl_b^{\geqslant}$ and there is no inconsistency, it is said that x belongs to $Cl_b^{\geqslant}$ *without any ambiguity*. That is, all objects *P-dominating* x belong to $Cl_b^{\geqslant}$, namely, $D_p^+(x) \subseteq Cl_b^{\geqslant}$.

4. DISCRETE CHOICE EXPERIMENT DATA USED

This section provides readers with a brief description of the datasets resulting from the three DCEs used in this work.

4.1. Pedestrian Preferences With Respect to Roundabouts Data

More detail on this DCE and resulting dataset can be found in Perdomo et al.,(2014). This dataset was collected through an online survey, in Canada in 2013 for a study of pedestrian preferences with respect to roundabouts (PPRR). The study was based on an unlabelled, video-based DCE survey. Each task presented two alternative roundabouts that were characterized by the following attributes: presence/type of signs; presence/type of pedestrian crossing; number of lanes; presence of a pedestrian island; traffic volume; and traffic speed. The first two attributes included three levels and the rest of attributes included two levels. Each respondent was presented with six choice situations. The sample available for estimation contained 3005 observations collected from 501 respondents.

4.2. Neighbourhood Choice Project Data Set

The second dataset was collected using face-to-face computer-assisted surveys administered on laptops, in Montreal, Canada in 2013 and 2014 for a neighbourhood location choice study (Mostofi Darbani, et al., 2014). Each task showed two alternative neighbourhoods that were characterized by the following attributes: Dwelling type (including four types); front yard depth (2 levels); space between buildings (2 levels); average home value (3 levels); travel time to work by car (2 levels); travel time to work by transit (2 levels); and travel time to nearby shops on foot (3 levels). The final sample consisted of 2430 observations collected from 405 respondents.

4.3. Shipper Preference Data

The third data set considered is from an online SP survey administered in Canada in 2005 for a study on shipper preferences with respect to carriers in the Quebec City – Windsor Corridor in Canada (Patterson, et al., 2007). Respondents were presented with 18 choice tasks, each with three unlabelled alternative carriers. The carriers were characterized by five attributes: cost of shipment; on-time reliability; damage risk; security risk; and finally, whether or not the shipment would be carried by truck only, or by truck and intermodal train. All attributes, except the last one, included three levels. The sample available for estimation contained 7,074 observations collected from 393 respondents.

5. CASE STUDIES

This section presents the findings from three case studies, using the DCE data described in the previous section, looking at the application of the SCT approach for detecting inconsistent choices in the context of learning and fatigue effects. The results of the approach are compared with the inverse of scale parameters estimate to provide a basis of comparison of the findings of the SCT approach with the more common H-MNL approach. In each case study, the following methodology was used to evaluate inconsistency of responses across choice tasks in our DCE data using the traditional scale parameter approach, and then to use the SCT approach to detect respondent inconsistent choices. First, we estimated a basic MNL with the entire dataset with Biogeme(Bierlaire, 2003). Second an H-MNL model with a different scale factor for each choice as presented in Eq. (4), was estimated. Third, the Stated Choice data were transformed as described in the section “Modelling Inconsistent Behaviour”, to derive mutual dominance relations, so that the individual decision cones could be identified. Fourth, the transformed data were used as inputs to produce decision cones for each individual. A code written in Visual Basic (available from the authors), using the SCT approach, was then used to produce decision rules for each individual and to detect inconsistencies across each individual’s choices. Finally, the findings of the two approaches are compared.

5.1. Testing for Inconsistency Using the Scale Parameter Approach

The H-MNL model was estimated for each case study. Table 2 presents the results from the H-MNL models. In each case, simple linear-in-parameters specifications of the main attributes were used to estimate the H-MNL models. In Table 2 scale estimates and their t-statistics are reported for the H-MNL models. In each case the scale of the first choice task was set to unity; hence, all scale estimates are relative to task 1. In the case of the PPRR and neighbourhood choice data, five scale parameters were estimated as respondents were asked to respond to six choice tasks. In the case of the shipper preference data, 17 scale parameters were estimated as the survey included 18 choice tasks per respondent. As expected, we reject the hypothesis that the variances of the unobserved component are the same in all choice tasks. Since the H-MNL model is nested under MNL model, we conducted likelihood-ratio tests to compare the two models in each case. The likelihood ratio statistics for the PPRR, neighbourhood choice and shipper preferences data are 7.38, 2.98 and 20.76, respectively. The critical value from the chi-square distribution at the 95% significance level, with 5 and 17 degrees of freedom, are 11.07 and 27.59, respectively. Thus, while the H-MNL model's fit is better than that of MNL model in all cases, it is not significantly statistically superior to the MNL model.

TABLE 2 Estimated Scales from H-MNL model

	PPRR data		Neighbourhood Choice Project Data		Shipper Preferences Data	
Choice task	Coeff.	*Std. err*	Coeff.	*Std. err*	*Coeff.*	*Std. err*
1	-	-	-	-	-	-
2	1.36	0.172	0.855	0.148	0.892	0.101
3	1.24	0.155	1.05	0.175	0.860	0.099
4	1.26	0.155	0.918	0.163	0.884	0.102
5	1.29	0.160	0.991	0.167	0.854	0.098
6	1.22	0.155	1.11	0.187	0.844	0.097
7					0.831	0.098
8					0.710	0.086
9					0.934	0.107
10					0.871	0.100
11					0.690	0.085
12					0.905	0.103
13					0.754	0.090
14					0.858	0.098
15					0.888	0.100
16					0.756	0.089
17					0.834	0.096
18					0.792	0.093

5.2. Testing for Inconsistency with the SCT Approach

In this section the approach presented in section 3 was employed to detect inconsistent choices. By examining individual decision rules and individual respondent choices, it was possible to identify those choices that were inconsistent with an individual's other choices, as well as to identify with which other choices the choice was inconsistent. In fact, it was possible to identify, for each choice, whether a given choice was consistent or inconsistent with all of the other choices. As such, it was also possible to identify the degree to which a given choice was consistent or inconsistent by identifying with how many other choices it was inconsistent. So, for example, supposing a choice set with two alternatives and six choice tasks, it is possible to establish whether a given choice is consistent with all, all but one, all but two, etc. other choice tasks. As a result, a choice task inconsistent with three other choice tasks is considered more inconsistent than a task inconsistent with only one other task. We then computed the average number of choices with which the first, second, third, etc. task were inconsistent. We refer to this measure as the "degree of inconsistency." Table 3 presents the degree of inconsistency for each choice task in the task sequences. As one can see, the degree of inconsistency in the shipper preference data is greater than the other cases, likely due to the greater complexity (more choice tasks and alternatives per choice task).

TABLE 3 Degree of Inconsistency Calculated

Choice task	PPRR data	Neighbourhood Choice Project Data	Shipper Preferences Data
1	0.1158	0.0197	1.226
2	0.0998	0.0148	1.567
3	0.0958	0.0049	1.371
4	0.0938	0.0099	1.389
5	0.1018	0.0148	1.552
6	0.1078	0	1.508
7			1.592
8			1.544
9			1.328
10			1.544
11			1.646
12			1.516
13			1.646
14			1.463
15			1.384
16			1.529
17			1.681
18			1.465

5.3. Comparison of the Two Approaches

In order to compare the two approaches in terms of what they imply for the detection of inconsistency as well as learning and fatigue effects, we provide a graphical analysis below. In each of the graphs we compare the degree of inconsistency with values of the inverse of the scale parameters for the corresponding choice task from the H-MNL models. We use the inverse of the scale parameter because it is more easily comparable with the degree of inconsistency than the scale parameter itself. I.e. the smaller the inverse of the scale parameter (the smaller the variance of the unobserved components), the more consistent are responses. Similarly, the smaller the degree of inconsistency, the more consistent are responses. As such, it is expected that the patterns of degree of inconsistency and the inverse of the scale parameter should be similar.

Figures 2-4 present the degree of inconsistency and the inverse of the scale parameter for each choice task for each of the three DCEs. Degree of inconsistency is represented with dots, whereas the inverse of the scale parameter is represented with squares. Let us first consider figure 2, showing the results from the PPRR data. In this case, the first task is found to be the most inconsistent based on the degree of inconsistency as well as the inverse of the scale parameter. This makes sense from the institutional learning (Braga, et al., 2005) perspective – during the first question, respondents are just beginning to understand the survey and so it is not surprising that this is the most inconsistent. Continuing through the choice tasks, we see slightly different patterns from the degree of inconsistency and scale parameter indicator. While the inverse of the scale parameter suggests a plateau of inconsistency after the first task, the degree of inconsistency indicator suggests responses become more consistent until the 4th task suggesting a value learning effect. After that, responses become less consistent. In general the two graphs show similar increasing/decreasing trends. This is supported by the positive correlation value, +0.79, between the degree of inconsistency and inverse of scale parameter, which is significant at 0.94%. In the context of learning and fatigue effects, it looks like respondents on average are learning until the fourth task where fatigue effects become more important than the learning effects and inconsistency begins to increase. This result supports the inverted U-shaped relationship between choice consistency and choice task number, in line with the findings of Caussade et al. (2005) and Chung et al. (2010).

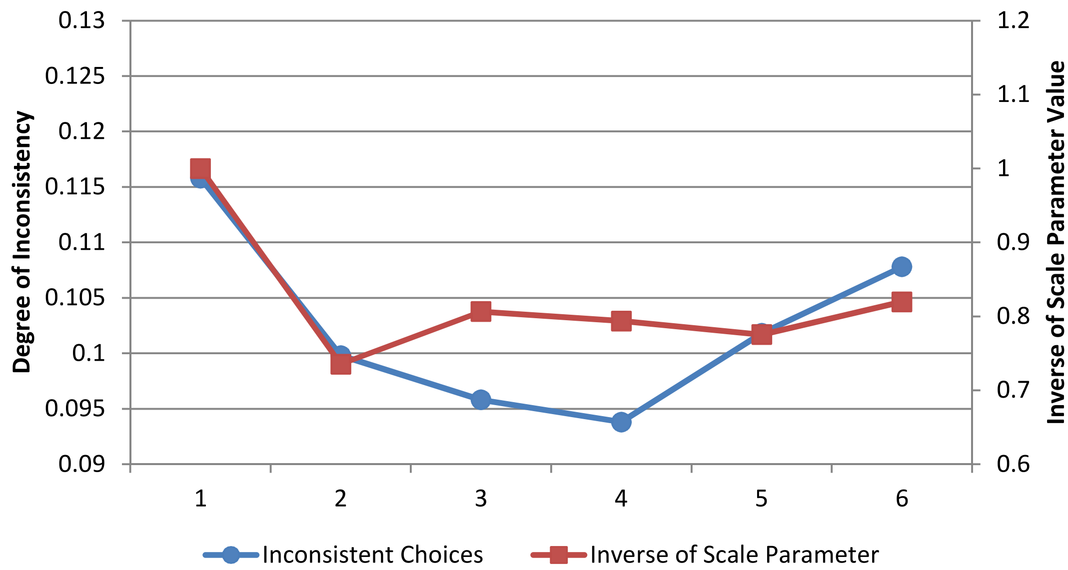


FIGURE 2 PPRR Data, Inverse of Scale Parameter & Average Degree of Inconsistency

In figure 3 the results from the two approaches generally correspond with both of them showing a general (if somewhat noisy) downward trend – i.e. reduced inconsistency with each additional task, or learning effects throughout the survey. The one marked difference between the two approaches is that for the first task. As in the PPRR data, our approach found the greatest inconsistency in the first task. At the same time, the scale parameter approach found the first task to be one of the least consistent. This results in a weak insignificant similarity between the two trends.

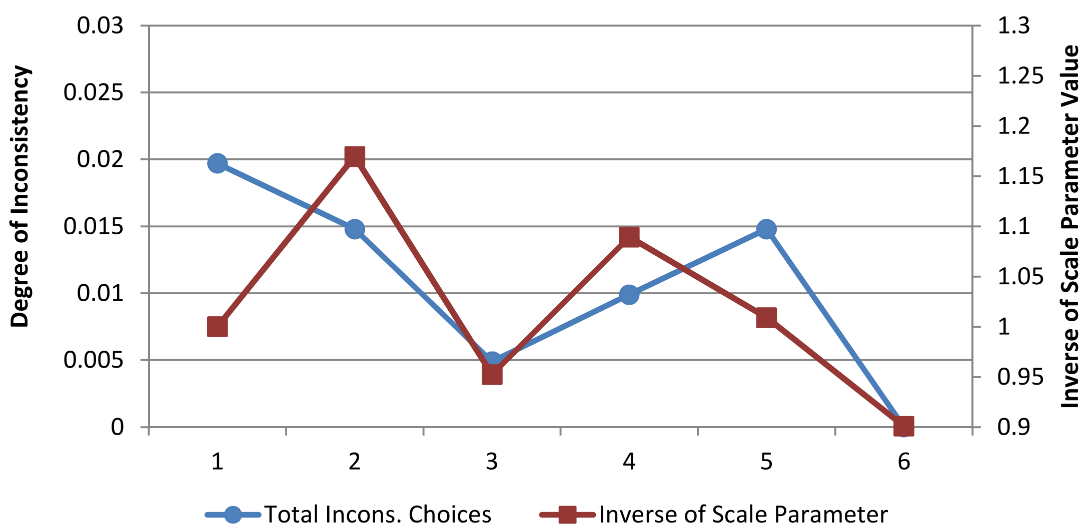


FIGURE 3 Household Location Choice Data, Inverse of Scale Parameter & Average Degree of Inconsistency

The last example, shown in figure 4, shows quite different results than the other two DCEs, although the two approaches result in similar findings, with a correlation value of +0.65 that is significant at 99%. According to both methods, the first task is the most consistent and subsequent tasks tend to become less consistent, although these results are also a little noisy. It would appear that fatigue effects appear to be stronger than learning effects throughout the survey. It is unclear why the pattern of fatigue effects is so different, but could be the result of a number of things. The shipper DCE was for example the only one with three instead of two alternatives. It was also the only one that was exclusively text based. It is worth noting that this result is in line with what Savage & Waldman (2008) report and in particular that online surveys, may show weaker learning effects and much stronger fatigue effects than mail surveys. What is most important, however, is not so much the difference in the pattern of consistency between the different surveys, but the similarity of the results between the two approaches.

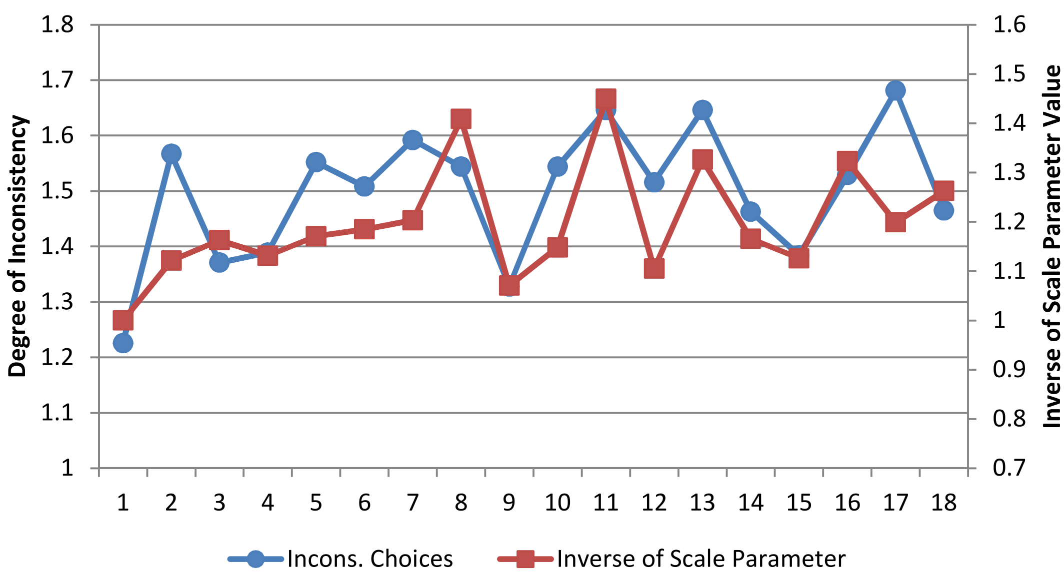


FIGURE 4 Shipper Preferences Data, Inverse of Scale Parameter & Average Degree of Inconsistency

After reviewing these results, one might argue that if the two approaches provide broadly similar results in terms of consistency of responses, than what is the advantage of the degree of consistency approach? The answer is simply that while similar results in terms of overall consistency are found, the degree of consistency approach makes it possible to identify not only which individual respondent is inconsistent, but which of their choices are consistent or inconsistent with all of their other choices. While there is debate about what to do with inconsistent responses (see e.g. (Viney, et al., 2002; Lancsar, et al., 2006; Chorus, et al., 2013)), this procedure at least makes it possible to identify inconsistent responses in a very general way and in particular in DCEs more complex than just two alternatives and two attributes.

6. CONCLUSION

This paper used the SCT approach, a systematic test to detect inconsistencies within individual respondent choices, by testing for the violation of the axiom of transitivity across choice tasks. Transitivity is a fundamental axiom in consumer choice (Samuelson, 1938), but has not often been applied in this context in previous literature (Lancsar, et al., 2006), especially in the case of complex DCEs. The paper then compared conclusions drawn about learning and fatigue effects with the most common alternative statistical approach, controlling for changes is unobservable preferences (scale heterogeneity) using the H-MNL modeling structure. Importantly, our findings suggest that the task specific scales estimated by H-MNL models, and the degree of inconsistency calculated using the SCT approach provide similar results in terms of consistency in choice tasks, as well as in conclusions about learning and fatigue effects. That is, the trends in inconsistency identified by both approaches agree with each other. The empirical analysis suggests that in the case of the PPRR data, both learning and fatigue effects were observed, while in the household location choice data we found a slight learning effect. Also, we only observed respondent fatigue in the shipper data. At the same time, the SCT approach is not only able to identify choice inconsistency, but it is able to identify which choice tasks are responsible for the inconsistency, something the traditional approach cannot.

Further investigation could use this approach to consider how the complexity of experiments influences the share of inconsistent choices, and possibly optimal complexity levels for these surveys. Similarly, the approach could be used to evaluate optimal numbers of tasks in these surveys, and many other questions in the design of DCEs.

5. REFERENCES

Bateman Ian J. [et al.] Choice Set Awareness and Ordering Effects in Discrete Choice Experiments [Report] / CSERGE Working Paper EDM 08-01. - [s.l.] : CSERGE Working Paper EDM 08-01, 2008.

Ben-Akiva Moshe E and Lerman Steven R Discrete Choice Analysis, Theory and Application to Travel Demand [Book]. - Cambridge, Ma : MIT Press, 1985.

Bierlaire Michel , Axhausen Kay W and Abay George The acceptance of modal innovation: The case of Swissmetro [Conference] // 1st Swiss Transport Research Conference. - Ascona, : [s.n.], 2001. - pp. 1-15.

Bierlaire Michel BIOGEME: a free package for the estimation of discrete choice models [Conference] // 3rd Swiss Transportation Research Conference. - Ascona : [s.n.], 2003.

Błaszczyński Jerzy [et al.] jMAF - Dominance-based Rough Set Data Analysis Framework- User's guide [Journal] // Dominance-based rough set data analysis framework. - 2011.

Bradley Mark and Daly Andrew Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data [Journal] // Transportation. - 1994. - 2 : Vol. 21. - pp. 167-184.

Braga Jacinto and Starmer Chris Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis [Journal] // Environmental & Resource Economics. - 2005. - 1 : Vol. 32. - pp. 55-89.

Burge Peter and Rohr Charlene DATIV: SP Design: Proposed approach for pilot survey [Journal] // Tetra-Plan in cooperation with RAND Europe and Gallup A/S. - 2004.

Campbell Danny and Lorimer Victoria S. Accommodating attribute processing strategies in stated choice analysis: do respondents do what they say they do? [Conference] // 17th annual conference of the European Association of Environmental and Resource Economics. - Amsterdam : [s.n.], 2009. - pp. 24-27.

Caussade Sebastian [et al.] Assessing the influence of design dimensions on stated choice experiment estimates [Journal] // Transportation Research Part B: Methodological. - 2005. - 7 : Vol. 39. - pp. 621-640.

Chen Chao [et al.] Effects of task complexity and time pressure on activity-travel choices: heteroscedastic logit model and activity-travel simulator experiment [Journal] // Transportation. - 2015. - Vol. In press.

Chorus Caspar G. and Bierlaire Michel An empirical comparison of travel choice models that capture preferences for compromise alternatives [Journal] // Transportation. - 2013. - 3 : Vol. 40. - pp. 549-562.

Chung Chanjin, Boyer Tracy and Han Sungill How Many Choice Sets and Alternatives are Optimal? Consistency in Choice Experiments [Journal] // Agribusiness. - 2010. - 1 : Vol. 27. - pp. 114-125.

Czajkowski Mikolaj, Giergiczny Marek and Greene William H. Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity [Journal] // Land Economics. - 2014. - 2 : Vol. 90. - pp. 324-351.

Dembczyński Krzysztof, Greco Salvatore and Słowiński Roman Rough set approach to multiple criteria classification with imprecise evaluations and assignments [Journal] // European Journal of Operational Research. - 2009. - 2 : Vol. 198. - pp. 626-636.

Greco Salvatore, Matarazzo Benedetto and Slowinski Roman Customer satisfaction analysis based on rough set approach [Journal] // Zeitschrift für Betriebswirtschaft. - 2007. - 3 : Vol. 77. - pp. 325-339.

Greco Salvatore, Matarazzo benedetto and Slowinski Roman Rough sets theory for multi-criteria decision analysis [Journal] // European Journal of Operational Research. - 2001. - 1 : Vol. 129. - pp. 1-47.

Hensher David A. How Do Respondents Process Stated Choice Experiments? Attribute Consideration Under Varying Information Load [Journal] // Journal of Applied Econometrics. - 2006. - 6 : Vol. 21. - pp. 861-878.

Hensher David, Louviere Jordan and Swait Joffer Combining sources of preference data [Journal] // Journal of Econometrics. - 1998. - 1-2 : Vol. 89. - pp. 197-221.

Hess Stephane, Hensher David A. and Daly Andrew Not bored yet – Revisiting respondent fatigue in stated choice experiments [Journal] // Transportation Research Part A: Policy and Practice. - 2012. - 3 : Vol. 46. - pp. 626-644.

Hess Stephane, Rose John M and Polak John Non-trading, lexicographic and inconsistent behaviour in stated choice data [Journal] // Transportation Research Part D. - 2010. - 7 : Vol. 15. - pp. 405-417.

Johnson F. Reed and Desvousges William H Estimating Stated Preferences with Rated-Pair Data:Environmental, Health, and Employment Effects of Energy Programs [Journal] // Journal of Environmental Economics and Management. - 1997. - 1 : Vol. 34. - pp. 79-99.

Lancsar Emily and Louviere Jordan Deleting‘irrational’responses from discrete choice experiments: a case of investigating or imposing preferences? [Journal] // HEALTH ECONOMICS. - 2006. - 8 : Vol. 15. - pp. 797-811.

Liou James J. H. A novel decision rules approach for customer relationship management of the airline market [Journal] // Expert Systems with Applications. - April 2009. - 3 : Vol. 36. - pp. 4374-4381.

Liou James J.H. A novel decision rules approach for customer relationship management of the airline market [Journal] // Expert Systems with Applications. - 2009. - 3 : Vol. 36. - pp. 4374-4381.

Louviere Jordan J [et al.] Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? [Journal] // Journal of Consumer Research. - 2008. - 2 : Vol. 35. - pp. 360-375.

Louviere Jordan J., Hensher David A. and Swait Joffre D. Stated Choice Methods Analysis and Application [Book]. - New York : Cambridge University Press, 2000. - 1.

McFadden Daniel Little Conditional logit analysis of qualitative choice behavior [Book Section] // In Frontiers in Econometrics / ed. Zarembka P.. - New York : Academic Press, 1974.

McIntosh E and Ryan M Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences [Journal] // Journal of Economic Psychology. - 2002. - Vol. 23. - pp. 367-382.

Mostofi Darbani Javad [et al.] VIDEO GAME VS. TRADITIONAL TEXT-ONLY SP SURVEY OF NEIGHBORHOOD CHOICE [Conference] // Transportation Research Board. - Washington D.C. : [s.n.], 2014. - p. 18.

Nassiri Habibollah and Rezaei Ali Air itinerary choice in a low-frequency market: A decision rule approach [Journal] // Journal of Air Transport Management. - January 2012. - 1 : Vol. 18. - pp. 34–37.

Niblett Anthony Tracking inconsistent judicial behavior [Journal] // International Review of Law and Economics. - 2013. - Vol. 34. - pp. 9-20.

Patterson Zachary, Ewing Gordon O. and Haider Murtaza Shipper Preferences Suggest Strong Mistrust of Rail Results from Stated Preference Carrier Choice Survey for Quebec City–Windsor Corridor in Canada [Journal] // Transportation Research Recoard. - December 26, 2007. - 2008. - pp. 67-74.

Perdomo Mario [et al.] Pedestrian preferences with respect to roundabouts—A video-based stated preference survey [Journal] // Accident Analysis and Prevention. - April 1, 2014. - Vol. 70. - pp. 84-91.

Rezaei Ali and Patterson Zachary Detecting, non-transitive, inconsistent responses in discrete choice experiments [Journal]. - 2015. - p. Submited to Transportation.

Rezaei Ali and Patterson Zachary Identifying Inconsistent Responses in Stated Choice Surveys Using a Dominance-Based Approach [Conference] // Transportation Research Board. - Washington, DC : [s.n.], 2015. - 15-5206.

Rose John M, Hess Stephane and Collins T Andrew What if My Model Assumptions are Wrong: The Impact of Non-standard Behaviour on Choice Model Estimation [Journal] // Journal of Transport Economics and Policy. - 2013. - 2 : Vol. 47. - pp. 245-263.

Rose John M. and Black Lain R. Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments [Journal] // Marketing Letters. - 2006. - 4 : Vol. 17. - pp. 295-310.

Rose John M., Hess Stephane and Collins Andrew T. What if my Model Assumptions are Wrong: The Impact of Non-Standard Behaviour on Choice Model Estimation [Journal] // Journal of Transport Economics and Policy. - May 2013. - 2 : Vol. 47. - pp. 245-263.

Rulleau Bénédicte Rulleau and Dachary-Bernard Jeanne Preferences, rational choices and economic valuation: Some empirical tests [Journal] // The Journal of Socio-Economics. - 2012. - 2 : Vol. 41. - pp. 198-206.

Ryan Mandy and San Miguel Fernando Revisiting the axiom of completeness in health care [Journal] // Health Economics. - 2003. - 4 : Vol. 12. - pp. 295-307.

Sælensminde Kjartan Inconsistent choices in Stated Choice data [Journal] // Transportation. - 2001. - 3 : Vol. 28. - pp. 269-296.

Sælensminde Kjartan Inconsistent choices in Stated Choice data Use of the logit scaling approach to handle resulting variance increase [Journal] // Transportation. - August 2001. - 3 : Vol. 28. - pp. 269-296.

Sælensminde Kjartan The Impact of Choice Inconsistencies in Stated Choice Studies [Journal] // Environmental and Resource Economics. - [s.l.] : Kluwer Academic, 2002. - Vol. 23. - pp. 403-420.

Samuelson Paul A A Note on the Pure Theory of Consumer's Behaviour [Journal] // Economica. - 1938. - 17 : Vol. 5. - pp. 61-71.

Savage Scott J. and Waldman Donald M. Learning and fatigue during choice experiments: a comparison of online and mail survey modes [Journal] // Journal of Applied Econometrics. - 2008. - 3 : Vol. 23. - pp. 351-371.

StataCorp Stata statistical software: release 12. - College Station : Stata Corporation, 2011.

Swait Joffre and Adamowicz Wiktor Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice [Journal] // Organizational Behavior and Human Decision Processes. - 2001. - 2 : Vol. 28. - pp. 141–167.

Swait Joffre and Adamowicz Wiktor Choice Task Complexity and Decision Strategy Selection [Report] / Dept. of Marketing, College of Business Administration. - Edmonton : University of Alberta, 1997.

Swait Joffre and Adamowicz Wiktor The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Seitching [Journal] // Journal of Consumer Research, Inc.. - 2001. - 1 : Vol. 28. - pp. 135-148.

Swait Joffre and Louviere Jordan The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models [Journal] // Journal of Marketing Research. - [s.l.] : American Marketing Association, 1993. - 3 : Vol. 30. - pp. 305-314.

Train kenneth Discrete Choice Methods with Simulation [Book]. - United Kingdom : Cambridge University Press, 2003.

Viney Rosalie, Lancsar Emily and Louviere Jordan Discrete choice experiments to measure consumer preferences for health and healthcare [Journal] // Expert Rev Pharmacoecon Outcomes Res. - 2002. - 4 : Vol. 2. - pp. 319-329.

Witlox Frank and Tindesman Hans The application of rough sets analysis in activity-based modelling. Opportunities and constraints [Journal] // Expert Systems with Applications. - 2004. - 4 : Vol. 27. - pp. 585-592.

Zhai Lain Yin, Khoo Li Pheng and Zhong Zhao Wei A dominance-based rough set approach to Kansei Engineering in product development [Journal] // Expert Systems with Applications. - 2009. - 1 : Vol. 36. - pp. 393-402.