Article:

Batley, R orcid.org/0000-0002-2487-850X and Hess, S orcid.org/0000-0002-3650-2518 (2016) Testing for regularity and stochastic transitivity using the structural parameter of nested logit. Transportation Research Part B: Methodological, 93 (Part A). pp. 355-376. ISSN 0191-2615

https://doi.org/10.1016/j.trb.2016.07.018

Testing for regularity and stochastic transitivity using the structural parameter of nested logit

Richard Batley & Stephane Hess

Institute for Transport Studies, University of Leeds, UK

Address for correspondence: Richard Batley, Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, United Kingdom

Telephone: +44 (0) 113 343 1789

Fax: +44 (0) 113 343 5334

E-mail: R.P.Batley@its.leeds.ac.uk

Keywords: regularity, stochastic transitivity, Random Utility Model, nested logit, structural parameter

Abstract

We introduce regularity and stochastic transitivity as necessary and well-behaved conditions respectively, for the consistency of discrete choice preferences with the Random Utility Model (RUM). For the specific case of a three-alternative nested logit (NL) model, we synthesise these conditions in the form of a simple two-part test, and reconcile this test with the conventional zero-one bounds on the structural ('log sum') parameter within this model, i.e. $0 < \theta \leq 1$, where θ denotes the structural parameter. We show that, whilst regularity supports the lower bound of zero, moderate and strong stochastic transitivity may, for some preference orderings, give rise to a lower bound greater than zero, i.e. impose a constraint $l \leq \theta$, where $l > 0$. On the other hand, we show that neither regularity nor the stochastic transitivity conditions constrain the upper bound at one. Therefore, if the conventional zero-one bounds are imposed in model estimation, preferences which violate regularity and/or stochastic transitivity may either go undetected (if the 'true' structural parameter is less than zero) and/or be unknowingly admitted (if the 'true' lower bound is greater than zero), and preferences which comply with regularity and stochastic transitivity may be excluded (if the 'true' upper bound is greater than one). Against this background, we show that imposition of the zero-one bounds may compromise model fit, inferences of willingness-to-pay, and forecasts of choice behaviour. Finally, we show that where the 'true' structural parameter is negative (thereby violating RUM – at least when choosing the 'best' alternative), positive starting values for the structural parameter in estimation may prevent the exposure of regularity and stochastic transitivity failures.

1. Introduction

As is well-established in microeconomic consumer theory, the fundamental preference axioms of completeness, transitivity and continuity – taken together – permit the representation of an individual's complete preference ordering by a continuous real-valued order-preserving function (Debreu, 1954). An important proposition follows from Debreu; the individual is conceptualised as making consumption choices *as if* to maximise utility. This proposition, which is the cornerstone of Neo-Classical consumer theory, has been the subject of considerable interest in the behavioural economics literature. A focus of this interest has been the design and implementation of experiments that seek to elicit empirical support for (or refutation of) the axioms of completeness, transitivity and continuity – as well as other related properties of choice behaviour. Emanating from this literature, several phenomena have been identified as giving rise to violations of the fundamental axioms and, by implication, violations of utility maximisation.

The present paper is motivated by an interest in exploring analogies to the fundamental preference axioms, and their empirical verification, in the alternative domain of probabilistic discrete choice. The discrete choice context, where the individual chooses from a finite and exhaustive set of mutually-exclusive alternatives, creates difficulties for conventional Neo-Classical consumer theory. This is because the theory employs marginal concepts derived using calculus; application to discrete choice has been described as *'awkward'* (McFadden, 1981 p199), and worse still *'impossible'* (Ben-Akiva & Lerman, 1985 p44). In response to these difficulties, a bespoke version of consumer theory has evolved, centred upon the theoretical construct of the Random Utility Model (RUM)[1].

Drawing analogy with psychophysical models of judgement and choice (Fechner, 1859; Thurstone, 1927; Luce, 1959), RUM was conceived by Marschak (1960) and Block & Marschak (1960)[2] as a probabilistic representation of the Neo-Classical theory of choice. In common with the Neo-Classical theory, RUM is couched at the individual level, is based fundamentally on the notion that the individual acts *as if* to maximise utility, and (in the original 'distribution free' form of RUM proposed by B&M, at least) is entirely supported by the notion of ordinal utility. Contrasting with Neo-Classical theory, however, RUM appeals to the context of discrete choice consumption.

The present paper relates to three strands of extant literature, as follows.

1.1 Representation theorems for RUM

The literature on representation theorems has considered the necessity and sufficiency of conditions on probabilistic choice systems (PCS) giving rise to (cardinal) utility

[1] One of the reviewers of this paper pointed out that the term 'Random Utility Model' (RUM) has sometimes been interpreted differently in different disciplines, and that a tighter and more contemporary terminology is 'choice probabilities induced by strict linear orders'. See Marley & Regenwetter's (2016) recent review of deterministic and probabilistic representations of choice, which distinguished between economic (i.e. parametric) and psychological (i.e. linear order) approaches to RUM. However, since the terminology 'choice probabilities induced by strict linear orders' is not common parlance in transport, this paper will remain faithful to 'RUM', but the reviewer's point is worthy of mention.

[2] Henceforth, we will abbreviate Block & Marschak (1960) to 'B&M'.

functions (Debreu, 1959; Davidson & Marschak, 1959) and RUM. Focussing here on representation theorems for RUM, Falmagne (1978) was first to show the necessity and sufficiency of the so-called 'B&M polynomials'[3]. Some years later (and apparently ignorant of Falmagne's paper until their attention was drawn to it in the course of peer review), Barberá & Pattanaik (1986) re-stated Falmagne's theorem in terms of rankings rather than utility scales, which allows closer correspondence with the concept of ordinal utility. More recently, Fiori (2004) contributed an elegantly concise proof of Falmagne's theorem.

Mindful of its origins in the cognate discipline of psychophysics, it is interesting to observe that RUM has attracted interest from a multidisciplinary audience, spanning several core disciplines (especially economics, psychology and mathematics), as well as a raft of sectoral applications (including transport, health and the environment). McFadden (2005) presented a useful synthesis of representation theorems for RUM and, reflecting his parent discipline of economics, he characterised such theorems as addressing the 'problem of revealed stochastic preference'[4]. Within this synthesis, McFadden & Richter's (unpublished) 1970a and 1970b papers, subsequently consolidated within their 1991 paper, covered similar ground to Falmagne (1978). Reflecting back some years later, Marley (1990) described the evolution of the literature on representation theorems for RUM, and offered specific observations concerning the links between the Falmagne and McFadden/Richter bodies of work.

A distinct but related strand of literature is that dealing with representation theorems for 'parametric' versions of RUM[5]. Motivated by an interest in its practical applicability, three independent parallel teams – namely Daly & Zachary (1976, subsequently published in 1978), Williams (1977) and McFadden (1978) – proposed alternative presentations of RUM, each formalised in terms of necessary and sufficient conditions on choice probabilities and/or random utilities giving rise to choice probabilities. In this context, and drawing similarities with McFadden's 'problem of revealed stochastic preference', the probabilistic content of RUM derives from the propensity for variability in behaviour across a population of individuals, as distinct from the intra-individual variability of a single individual in B&M. This change in emphasis, together with the extended theoretical apparatus, provided the stimulus for the adoption of RUM in mainstream econometric practice (see section 1.3 to follow).

1.2 Empirical testing of theoretical properties of choice

Following from the theoretical developments outlined above, a second strand of literature has subjected the fundamental preference axioms – as well as a broader range of theoretical properties of choice – to empirical testing. In this context, the psychology and behavioural economics literatures would seem rather more developed than the discrete choice literature, but this perhaps reflects the relative infancy of the

[3] See Theorem 4 (p60) of Falmagne (1978).

[4] According to McFadden (2005), this problem poses the question: *'Are the distributions of choices observed for a population of individuals in a variety of choice situations consistent with rational choice theory, which postulates that individuals maximize preferences?'* (p245).

[5] In this regard, Regenwetter *et al* (2010) distinguished between B&M's 'distribution free' RUM and the 'parametric' RUM that arises from (1), whilst Batley (2008) distinguished between 'ordinal' RUM and 'cardinal' RUM.

latter. Following the conception of non-parametric RUM in 1960, parametric versions of RUM entered practical usage only in the late 1960s; see McFadden's 1968 (but unpublished until 1975) pioneering application to public policy analysis. Despite their different levels of maturity, the psychology, behavioural economics and discrete choice literatures show interesting parallels in terms of the phenomena which have been observed in experimental contexts (for a recent overview of this literature, see Busemeyer & Rieskamp, 2013). Of particular relevance to the present paper are three phenomena, namely 'regularity', 'transitivity' and 'invariance'. These phenomena will be formally analysed in sections 2 and 3 to follow; the present section simply introduces the intuition for each phenomenon, and briefly summarises their respective evidential positions.

Regularity: this property asserts that the probability of choosing any given alternative from an offered set should not increase if the offered set is expanded to include additional alternatives. Violations of regularity were first reported by Huber *et al* (1982), who rationalised these violations in terms of 'asymmetric dominance'. The latter phenomenon characterises situations where a binary choice set is appended by a third alternative which is similar – but materially inferior – to one of the initial pair. According to asymmetric dominance, the third alternative is rarely chosen, but its inclusion in the offered set enhances the probability of choosing the similar alternative from the initial pair. Whilst different explanations for asymmetric dominance have been advanced in the literature (e.g. Simonson, 1989; Simonson & Tversky, 1992), there is reasonable consensus that this phenomenon is prevalent in choice experiments (Heath & Chatterjee, 1995). More generally, there exists an extensive mature literature in psychology on choice and response time for so-called 'context effects', where choice is affected by the presence or absence of other alternatives. Within this literature, the similarity between alternatives has been identified as a principal context effect (e.g. Trueblood *et al*, 2015).

Transitivity: this property asserts that if alternative x is preferred to alternative y, and y to z, then x should be preferred to z. Recognising that transitivity is ostensibly a deterministic property, the RUM literature has developed various stochastic interpretations of transitivity (referred to as 'weak', 'moderate' and 'strong'). None of these variants of transitivity are necessary for RUM, although there is a close relationship between stochastic transitivity and the so-called 'triangle condition' (see section 2.2 to follow), which *is* necessary for RUM. Following the precedent of Papandreou (1957) and Davidson & Marschak (1958)[6], researchers have subjected stochastic transitivity to empirical testing, and have generally reported evidence of violations (see Rieskamp *et al* (2006) and Hougaard *et al* (2011) for overviews of this literature). However, the recent paper by Regenwetter *et al* (2011) systematically reanalysed much of this evidence; using a non-parametric statistical test of the triangle condition, they found that most individuals did not produce statistically significant violations. More recently, Cavagnaro & Davis-Stober (2014) repeated the same analysis using a slightly refined version of Regenwetter *et al*'s test; this revealed variability in stochastic transitivity properties across individuals, but essentially corroborated Regenwetter *et al*'s finding.

[6] Moscati (2007) has contributed an insightful historical overview of this literature.

Invariance: this property asserts that the relative preference between two alternatives should be invariant to the addition/subtraction of other alternatives to/from the choice set. An important feature of early discrete choice model specifications was the 'Independence from Irrelevant Alternatives' (IIA) property (Luce, 1959), which states that the ratio of any two choice probabilities is unaffected by the presence or absence of other alternatives in the choice set. Modellers initially saw IIA as an attractive property, in the sense that the choice between alternatives could be predicted without the need for data on 'external' alternatives, and this prompted widespread application of the multinomial logit (MNL) model (McFadden, 1973). Subsequently, IIA began to be seen more as a weakness rather than a strength, since it was unable to account for similarity between alternatives, which had been identified as an important determinant of choice (Tversky, 1972a; 1972b).

This prompted the development and adoption of the nested logit (NL) model (Daly & Zachary, 1976; Williams, 1977)[7], which generalises MNL such that subsets of similar alternatives are 'nested' together, not unlike Tversky & Sattath's (1979) concept of a preference tree (or PRETREE)[8]. McFadden (1978) formalised the specification of parametric RUM models through the Generalised Extreme Value (GEV) theorem. GEV gives rise to a subset of models within the RUM class (Ibáñez, 2007), which embody general patterns of correlation between alternatives, and include MNL and NL among its members. Throughout the 1980s and 1990s, MNL and NL established themselves as the primary tools of discrete choice modellers; for example Ortúzar (2001) described them as the *'…the workhorses for the empirical analysis of travel behaviour in respect of discrete choices'* (p213). This did not however deter the exploration for further generalisations of RUM, especially in terms of the flexibility of substitution patterns between discrete choice alternatives. Cross-nested logit (CNL), which is also derived from GEV, generalises NL by allowing alternatives to belong to more than one nest, potentially with different 'degrees' of membership. Although the derivation of CNL is usually credited to Vovsha (1997), the model was clearly stated in Williams (1977) and McFadden (1978). Swait's (2001) GenL model, which is motivated by a specific interest in choice set generation, restricts CNL by suppressing the different degrees of membership. Daly and Bierlaire's Recursive Nested Extreme Value (RNEV) model (Daly, 2001; Bierlaire, 2002; Daly & Bierlaire, 2006) generalises both NL and CNL, by allowing cross-nesting with an arbitrary number of levels.

1.3 The feasible range of the structural parameter in GEV

Having exposed the key role played by the invariance property within practical specifications of RUM, let us now consider the ability to test observed behaviour for consistency with RUM. In this regard, the structural parameter θ of the GEV-based models (MNL, NL, CNL and RNEV), otherwise referred to as the coefficient of the 'inclusive value' within these models, will be the focus of our interest. Conventional practice is to constrain the choice model in estimation such that the structural parameter falls within the bounds $0 < \theta \leq 1$. Informing this convention, McFadden

[7] Building upon earlier contributions by Manheim (1973), Wilson (1974) and Ben-Akiva (1974); see the historical account in Ortúzar (2001).

[8] Pre-dating Tversky & Sattath (1979), NL also resonates with Gorman's (1968) concept of a *utility tree*. In more recent work, Batley & Daly (2006) considered formal equivalence between NL and PRETREE.

(1981) remarked (but did not prove) that: *'A necessary and sufficient condition for* [NL] *to be consistent with GEV is that…the coefficient of each inclusive value…lie*[s] *in the unit interval'* (p240). In support of the zero-one bounds, McFadden presented two arguments, one rationalising the structural parameter in terms of correlation between nested alternatives, and a second based upon testable properties of binary choice data; the latter argument is considered more fully in Annex B of the present paper.

With regards to the lower bound of the structural parameter in GEV models, Train (2003) remarked that: '*A negative value of* [the structural parameter] *is inconsistent with utility maximisation and implies that improving the attributes of an alternative (such as lowering its price) can decrease the probability of it being chosen*' (p81). McFadden (1981) further remarked: *'It should be noted that, while a negative coefficient of inclusive value leads to a local failure of the GEV conditions, a coefficient of an inclusive value exceeding one will fail to satisfy GEV only for some values of the variables. Thus it is possible that an empirical fit yielding a coefficient greater than one will be consistent with GEV over the range of the data and can be combined with a second function outside the range of the data to yield a system that satisfies GEV globally. However, this chapter has not attempted to develop a test for local consistency with GEV at the observations, or for consistency with some function that satisfies GEV globally'* (p248).

With regards to the upper bound of the structural parameter in GEV models, a number of researchers have reviewed the practical convention of constraining $\theta \leq 1$, further developing McFadden's (1981) points noted above. The initial contribution in this regard was by Börsch-Supan (1990), who sought to demonstrate that, for two-level NL, $\theta \geq 1$ is consistent with RUM for some range of (but not all) values of the explanatory variables. In this way, Börsch-Supan admitted the possibility of more flexible 'local' bounds on the structural parameter, whilst complying with the conventional zero-one bounds in a 'global' sense. As evidential support, Börsch-Supan cited examples from the literature of NL models exhibiting $\theta > 1$, namely Börsch-Supan (1985), Hensher (1984) and Small & Brownstone (1982). Train (2003) cited the additional examples of Train *et al* (1987) and Lee (1999). In these cases, $\theta > 1$ was accepted by the respective authors as a valid result, and interpreted as reflecting greater substitutability between nests than within nests. Herriges & Kling (1996), building upon Koning & Ridder (1994), subsequently corrected an oversight in Börsch-Supan, and offered proof of the definitive conditions on the structural parameter for two-level NL involving nests of two, three or four alternatives. They further applied the model empirically, showing the dependence of these conditions on the marginal probabilities of choosing the nests. For the simplest case of two-level NL involving nests of two alternatives – which will be the focus of the present paper – Herriges & Kling calculated an upper bound on the structural parameter of 20 for consistency with RUM. This result was however associated with an extreme marginal probability of 0.95; for a marginal probability of 0.5, the upper bound was reduced to 2, and for lower marginal probabilities still, the permissible range showed little increase beyond 1.

In the course of a comprehensive review of NL, Carrasco & Ortúzar (2002, section 3.5) devoted particular attention to the bounds of the structural parameter, identifying some practical limitations of the Börsch-Supan's (1990) argument (and its subsequent refinements). First, for any given dataset, tree structures associated with $\theta > 1$ may well be sub-optimal in terms of explanatory fit. Second, the admission of $\theta > 1$ may contravene the requirement for a decreasing structural parameter (and by implication an increasing scale) as one moves down multi-level NL. Third, whilst the Börsch-Supan argument has theoretical credence, it has received limited support from empirical evidence. Moreover, as a preamble to their review, Carrasco & Ortúzar (2002, section 2.3) compared and contrasted the alternative derivations of NL developed by Williams (1977) and McFadden (1978), and in particular highlighted their different rationales for the structural parameter. Williams' (1977) alternative derivation of NL represents the structural parameter as the ratio of scale parameters at adjacent levels of the tree, where the scale parameters reflect the variance of the random terms at the respective levels. The implication of this derivation is that – unlike McFadden's NL – Williams' NL constrains the structural parameter to the zero-one bounds, and requires the structural parameter to increase as one moves down the tree. Furthermore, with reference to the earlier justification for GEV-based NL models exhibiting $\theta > 1$, Williams' NL in effect constrains patterns of substitution between alternatives (Williams & Senior, 1978).

1.4 The contributions of the present paper

The present paper does not seek to revisit the question of how the $0 < \theta \leq 1$ bounds relate to the definition of RUM *per se* (whether in the context of McFadden's or Williams' derivations), but instead addresses the more general question of how these bounds relate to the properties of regularity and stochastic transitivity introduced in section 1.2 above. However, as will be apparent from the summary of literature above, any interest in the $0 < \theta \leq 1$ bounds is intertwined with interests surrounding representation theorems for RUM and the invariance property. Whilst B&M showed that stochastic transitivity – unlike regularity – is unnecessary to derive RUM, our interest in this property is motivated by the proposition that any 'well-behaved' discrete choice model might be expected to exhibit stochastic transitivity.

Against this background, the present paper offers three principal contributions:

1. We will distinguish between necessary (which we represent in terms of regularity) and well-behaved (which we represent in terms of stochastic transitivity) conditions for RUM.
2. Focussing specifically upon three-alternative NL, we will synthesise these necessary and well-behaved conditions for RUM in the form of a simple two-part test.
3. Using both theory and empirics, we will reconcile the simple two-part test with conventional criteria for determining the RUM-compliance of three-alternative NL.

To these ends, the layout of the paper is as follows. Section 2 presents a formal definition of RUM, details the theoretical conditions which give rise to this definition, and arising from these conditions identifies properties which will be the subject of empirical testing. Section 3 describes, in analytical terms, the application of these tests to a special case of RUM in the form of two-level NL. In order to illustrate the practical implication of section 3, section 4 develops broadly the same example empirically, using both simulated and real data. Section 5 provides a summary and conclusion.

2. Theoretical background of RUM

2.1 Theoretical conditions underpinning RUM

Consider an individual economic agent, who is offered a finite and exhaustive set of mutually exclusive alternatives:

$$N = \{1,...,n\}$$

Let us further restrict the analysis to a feasible subset $M \subseteq N$, which we refer to as the 'choice set'. We will not concern ourselves with the specific constraints determining feasibility, but these could include factors such as budget. B&M[9] introduced two 'conditions' (their terminology) which define RUM, thus[10]:

CONDITION (P), *Rankings consistent with the Random Utility Model*: There are $n!$ numbers $p(r)$ such that for any $x \in M$ and any M, $M \subseteq N$:

$$p(r) \geq 0 \text{ and } p_M(x) = \sum\nolimits_{R_{x;M}} p(r)$$

where $p(r)$ is the probability of the ranking r; $R_{x;M}$ is the set of all rankings r on M for which x is the first among all elements of M, i.e. $R_{x;M} = \{r \,|\, r_x \leq r_y\}$ for all $y \in M$; and $p_M(x)$ is the probability of alternative x being chosen from M, where $0 \leq p_M(x) \leq 1$ and $\sum\nolimits_{x \in M} p_M(x) = 1$.

Whilst the above condition encompasses all preference orderings on the feasible set, the condition that follows considers the subset of preference orderings where a given alternative is first ranked (i.e. is chosen).

CONDITION (U)*, Random Utility Model*: There is a random vector $(U_1,...,U_n)$ unique up to an increasing monotone transformation such that for any $x \in M$ and any M, $M \subseteq N$:

$$p_M(x) = \Pr(U_x \geq U_y) \text{ for all } y \in M$$

If the random utilities are (uniformly) continuous random variables, then this implies non-coincidence, i.e. $\Pr(U_x = U_y) = 0$ for all $y \in M$. On this basis, Condition (U) effectively defines the existence of a probability space on these random utilities.

[9] This section adheres closely to B&M's seminal 1960 paper, and the reader is referred to that paper for more detailed discussion of the various definitions and conditions. However, much of the same material is covered by Fishburn's (1998) subsequent and very authoritative review.

[10] In what follows, we deploy the following notational conventions to represent choice probability, namely: $p_M(x)$ is the probability of choosing x from the offered set M; $p_{\{x,y,z\}}(x)$ is the probability of choosing x from the trinary $x,y,z \in M$; and $p(x,y)$ is the probability of choosing x from the pair $x,y \in M$.

Finally, note that B&M (Theorem 3.1, p183) showed that conditions (U) and (P) each imply the other, i.e. $(U) \leftrightarrow (P)$[11].

2.2 Theoretical conditions as testable properties of RUM

In seeking to confirm the consistency (or otherwise) of given data with RUM, it should be acknowledged that: *'...save for the choice axiom,* [models of the RUM class] *are all stated in terms of nonobservable utility functions, and so it is impossible to test them completely until we know conditions that are necessary and sufficient to characterize them in terms of preference probabilities themselves, for only these can be estimated from data'* (Luce & Suppes, 1965 pp339-340). Contemporary discrete choice modellers have tended to overlook B&M's original work on RUM, which devoted detailed attention to testable properties and defined several conditions which follow from (U) and (P). These conditions refer to various properties of binary and multinomial choice probabilities on the feasible set M, as follows.

CONDITION (e)*, Regularity*: If $L \subseteq M$, then $p_M(x) \leq p_L(x)$ for all $x \in L, M$, $L \subseteq M$

CONDITION (e_3), *Regularity for any trinary*: For any three elements $x, y, z \in M$,

$p_{\{x,y,z\}}(x) \leq p(x,y)$, or equivalently $p_{\{x,y,z\}}(x) \leq \min(p(x,y), p(x,z))$

B&M (Theorem 3.3, p185) showed that condition (P) implies condition (e), which itself implies (but is not implied by) condition (e_3), i.e. $(P) \rightarrow (e) \mapsto (e_3)$. In general, regularity is necessary but not sufficient for RUM (Marschak, 1960, p192), but in the specific case of a trinary choice set, regularity is necessary *and sufficient* for RUM to hold. Some additional conditions follow:

CONDITION (c_3), *Triangular condition*: For any three distinct elements $x, y, z \in M$

$1 \leq p(x,y) + p(y,z) + p(z,x) \leq 2$

It is well known that for binary choice probabilities involving up to five distinct alternatives, RUM holds if and only if the relevant triangle inequalities hold (Cohen & Falmagne, 1971; 1990; McFadden & Richter, 1970a; 1970b; 1991; Fishburn, 1998; Cavagnaro & Davis-Stober, 2014). B&M (Theorem 5.6, p195) and Luce & Suppes (1965, Theorem 34, p343) further showed that $(e_3) \mapsto (c_3)$.

2.3 Other theoretical conditions as testable properties

The above relations reveal that regularity is the key condition for testing the consistency of discrete choice preference data with RUM. However, Marschak (1960) described the following conditions as *'...partial results* (which) *may, however, prove useful'.*

---

[11] In what follows, we use $\rightarrow$ to denote 'implies', $\leftrightarrow$ to denote 'implies and is implied by', and $\mapsto$ to denote 'implies but is not implied by'.

CONDITION (WST), *Weak Stochastic Transitivity*: If $p(x,y) \geq \frac{1}{2}$ and $p(y,z) \geq \frac{1}{2}$, then $p(x,z) \geq \frac{1}{2}$

CONDITION (MST), *Moderate Stochastic Transitivity*: If $p(x,y) \geq \frac{1}{2}$ and $p(y,z) \geq \frac{1}{2}$, then $p(x,z) \geq \min(p(x,y), p(y,z))$

CONDITION (SST), *Strong Stochastic Transitivity*: If $p(x,y) \geq \frac{1}{2}$ and $p(y,z) \geq \frac{1}{2}$, then $p(x,z) \geq \max(p(x,y), p(y,z))$

Marschak (1960; Theorem 12, p227) showed that $(SST) \mapsto (MST) \mapsto (WST)$, whilst B&M (Theorem 5.8, p196) showed that $(SST) \mapsto (c_3)$, and Luce & Suppes (1965; Theorems 35 and 38, pp343-346) showed that $(MST) \mapsto (c_3)$.

3. An analytical example

Following from the preceding discussion, any test of regularity on a given trinary (i.e. condition (e_3)) amounts to a comparison of the binary and trinary choice probabilities and, in particular, an examination of how these probabilities deviate depending upon the presence/absence of a 'third' alternative. Regularity will certainly hold in discrete choice contexts subject to IIA, but may not hold otherwise. Since the NL model (Williams, 1977; Daly & Zachary, 1978; McFadden, 1978) seeks to relax IIA, an interesting question is whether NL complies with regularity; this question will be the focus of section 3.3.1.

Although stochastic transitivity – unlike regularity – is not a necessary condition of RUM, we hypothesise that any 'well-behaved' discrete choice model will exhibit MST or (better still) SST. We further hypothesise that, by ensuring compliance with MST and SST, modellers will yield RUMs that embody more intuitive parameter estimates (e.g. in terms of the size and sign of implied demand elasticities), and greater explanatory power. In a similar fashion to our analysis of regularity, section 3.3.2 will consider the extent to which NL complies with stochastic transitivity.

Before embarking upon these discussions of regularity and stochastic transitivity, section 3.1 will introduce an illustrative choice problem, and section 3.2 will apply NL to this problem. In considering the compliance of NL with regularity and stochastic transitivity, a key focus will be whether these conditions corroborate the conventional $0 < \theta \leq 1$ bounds on the structural parameter.

3.1 The choice problem

In testing the compliance of RUM with the fundamental preference axioms, behavioural economists have tended to adhere to B&M's original definition of RUM, which interprets U as a random ordinal variable. By contrast, discrete choice modellers have re-interpreted U as a random cardinal variable, via the following definition:

$$U_x = V_x + \varepsilon_x \text{ for all } x \in M \qquad (1)$$

where V_x are constants referred to as 'deterministic utility', and ε_x are random variables exogenous[12] of V_x with a continuous joint finite density function. This defines the class of Additive Random Utility Models (ARUM).

Henceforth, we restrict the scope of the paper to the case where the choice set consists of the trinary[13]:

$$M = \{x, a, b\} \subseteq N$$

wherein a and b show some degree of similarity not possessed by x. For example, in the case of travel mode choice, a and b could represent two alternative bus services to a given location, whilst x could represent car. Reflecting these features, let us create a subset containing the two similar (i.e. bus) alternatives a and b:

$$L = \{a, b\} \subset M$$

The NL model arises from a special case of (1) where the random variables for all three alternatives are identically Gumbel distributed, but where ε_a and ε_b (e.g. the random variables for the bus alternatives) are correlated with each other, whilst ε_x (e.g. the random variable for the car alternative) is independent of ε_a and ε_b. The correlation between ε_a and ε_b seeks to capture the degree of similarity between a and b.

3.2 A nested logit representation of the choice problem

Following McFadden (1987), a NL representation of the aforementioned choice problem is uniquely determined by two choice probabilities, namely the marginal probability of choosing $L \subset M$:

$$p_M(L) = \frac{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]}}{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]} + e^{V_x}} \qquad (2)$$

and the conditional probability of choosing $a \in L$:

[12] This assumption of exogeneity is almost unavoidable if there is wish to apply RUM to welfare analysis; see McFadden (1995) or Batley (2014).

[13] The three-alternative choice set is but one example of real world or experimental choices. However, it is arguably the most common case of NL considered in both the literature (e.g. the widely used 'red bus-blue bus' example) and practice, and embodies important features which readily generalise to larger choice sets.

$$p_L(a) = p(a,b) = \frac{e^{\frac{V_a}{\theta}}}{e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}} \quad (3)$$

where θ continues to denote the structural parameter, and the model is implicitly upper normalised (Hensher & Greene, 2002; Carrasco & Ortúzar, 2002). Drawing reference to Carrasco & Ortúzar's critique of Börsch-Supan (1990), which was summarised in section 1.3 above, note that (2) and (3) give rise to two-level NL, thereby avoiding any complications associated with multiple levels.

In these terms, the probabilities of choosing the similar alternatives a and b are given by:

$$p_M(a) = p_M(L) \cdot p(a,b)$$

$$p_M(b) = p_M(L) \cdot (1 - p(a,b))$$

whilst the probability of choosing the dissimilar alternative x is given by:

$$p_M(x) = 1 - p_M(L)$$

This tree structure is illustrated in the top left panel of Figure 1.

FIGURE 1 ABOUT HERE

3.3 Applying the testable properties to three-alternative nested logit

Again drawing from B&M's derivation of the theoretical conditions, but now focussing on the trinary choice set (and employing the notation (P_3) to represent the application of condition (P) to this trinary set and all of its non-empty subsets), the testable properties from section 2 above are summarised in Figure 2.

FIGURE 2 ABOUT HERE

An important practical property of NL (and indeed of any parametric RUM) is that, having established a model on the complete choice set M, it readily lends itself to the derivation of choice probabilities for any reduced choice (i.e. for any subset of M). This property, which avoids the need to systematically model the full permutation of preference orderings, will be exploited in what follows. We will return to this point when introducing the empirical example in section 4.

3.3.1 Compliance with regularity

In testing compliance with the regularity condition, two general cases are of relevance.

Case 1: Intra-nest choice

For the three alternative NL choice problem under examination, regularity is satisfied if both:

$p(a,b) \geq p_M(a)$ and $p(b,a) \geq p_M(b)$

With reference to Annex A, it is trivial to show that, for intra-nest choice, compliance with regularity is guaranteed, irrespective of the value taken by the structural parameter θ.

Case 2: Inter-nest choice

In this case, regularity is satisfied if:

$p(x,a) \geq p_M(x)$, $p(a,x) \geq p_M(a)$, $p(x,b) \geq p_M(x)$, and $p(b,x) \geq p_M(b)$

For inter-nest choice, Annex A further shows that compliance with regularity will depend upon the relative magnitudes of the marginal and conditional probabilities[14]. In particular, negative values of the structural parameter are non-compliant, but values greater than one could be compliant.

Drawing together Cases 1 and 2 above, Figures 3 and 4 provide an empirical example of the choice problem under examination. Whilst Figure 3 assumes $V_a = V_b = V_x$, such that the three alternatives are deterministically indifferent, Figure 4 assumes $V_a = 9, V_b = 8, V_x = 10$, such that x is deterministically preferred to a, a to b, and x to b. In both figures, the upper and lower panels compare, for each of the inter-nest choices, the binary and multinomial choice probabilities as the structural parameter is increased from -10 to +10. With reference to (1), the structural parameter effectively represents the magnitude and interdependence of the random variables for the three alternatives. Despite the differences in deterministic preferences, both figures corroborate our theoretical proposition that regularity requires $\theta > 0$, since at negative values of the structural parameter one or more of the binary choice probabilities are less than their associated multinomial choice probabilities. Furthermore, in the case of Figure 4, regularity also gives rise to an upper bound, since $p(b,x) < p_M(b)$ where $\theta > 1.5$.

FIGURE 3 ABOUT HERE

FIGURE 4 ABOUT HERE

3.3.2 Compliance with stochastic transitivity

Relative to the discussion of regularity above, the discussion of stochastic transitivity will require rather more exposition. With reference to the general case outlined in section 2.3 above (i.e. not specific to NL), we begin by introducing the notation xyz to represent a complete set of binary stochastic preferences on the trinary choice set $\{x,y,z\}$ such that $p(x,y) \geq 1/2$, $p(y,z) \geq 1/2$ and $p(x,z) \geq 1/2$. In other words, xyz represents a preference ordering that complies with WST as a minimum (and possibly also complies with MST and SST)[15].

[14] This dependence resonates with Herriges & Kling's (1996) findings reported in section 1.

[15] As pointed out by one of the anonymous reviewers of this paper, xyz does not in general imply $p_M(x) \geq p_M(y) \geq p_M(z)$. This implication does however follow under the particular condition of 'order independence' (Luce & Suppes, 1965; Definition 9, pp411-412), a condition which characterises MNL.

Now relating this notation to the specific case of the three alternative NL under examination here, we refer to axb as the *intrinsic* preference ordering, on the grounds that the first two stochastic binary choices in the transitivity chain (i.e. the inter-nest binary choices between a and x (see top right panel of Figure 1), and between x and b (see bottom left panel of Figure 1)) will be independent of the value of the structural parameter, whilst the final stochastic 'transitive' choice (i.e. the intra-nest binary choice between a and b (see bottom right panel of Figure 1)) will be dependent on the value of the structural parameter[16].

Following the rationale outlined in Annex B, the intrinsic preference ordering allows us to infer, for given inter-nest binary choices, the upper bound on the structural parameter such that the intra-nest binary choice complies with stochastic transitivity, thus:

$$\theta \leq \frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} \tag{4}$$

where:

$$p(a,x)/p(x,a) = (1+u)$$

$$p(x,b)/p(b,x) = (1+v)$$

$$u,v \geq 0$$

and:

$$p(a,b)/p(b,a) \geq (1+w)$$

$w = \max(u,v)$ in the case of SST

$w = \min(u,v)$ in the case of MST

$w = 0$ in the case of WST

That is to say, conditional upon a being stochastically preferred to x, and x to b, (4) elicits the upper bound on θ which ensures that a is stochastically preferred to b with sufficient strength that stochastic transitivity holds for the intrinsic preference ordering axb. Since MST is associated with the minimum value of w, and SST is associated with the maximum, these two transitivity conditions may give rise to different upper bounds on the structural parameter.

Having derived (4) for the *intrinsic* preference ordering axb, let us now consider its application as a test of stochastic transitivity for *actual* preference orderings covering all possibilities. To this end, two general cases are of relevance, depending on whether

[16] This definition of the intrinsic preference ordering resonates with Herriges & Kling's (1996) comment: *'…restrictions are imposed on* [the structural parameter] *by consistency condition C.3 are expressed in terms of* [the marginal probability]*, with no cross-group terms involved'* (p37). In passing, note that since the assignment of the nested alternatives as a or b is arbitrary, we could instead adopt bxa as the intrinsic preference ordering.

the 'transitive' choice (i.e. the final binary choice of the actual preference ordering) is intra-nest (i.e. in the manner of the intrinsic preference ordering) or inter-nest[17].

Case 3: Where the 'transitive' choice is intra-nest

This case deals with *actual* preference orderings axb and bxa (again reflecting complete binary stochastic preferences on the trinary choice set, as per the notational definition at the beginning of section 3.3.2). However, having defined the discrete choice problem (section 3.1), and determined which alternatives should be nested together (section 3.2), it is arbitrary as to whether a given nested alternative is labelled a or b. The implication follows that the same bounds on the structural parameter will (in essence[18]) apply to the actual preference orderings axb and bxa. Moreover, Case 3 is in substantive terms consistent with the intrinsic preference ordering, and focussing here upon axb, we can re-state (4)[19]:

$$\theta_{axb} = \frac{\ln\left((1+u)(1+v)\right)}{\ln\left((1+w)+k_{axb}\right)} \leq \frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} = \theta_{\max} \qquad (5)$$

where k_{axb} is a non-negative constant (see Table 1 for additional working). From (5) we can infer that:

- SST entails an upper bound on the structural parameter, which we denote $\theta_{\max;SST}$, and which may be greater than one.
- MST also entails an upper bound, which we denote $\theta_{\max;MST}$, and which may itself be greater than $\theta_{\max;SST}$.
- In summary: $\theta_{axb} \leq \theta_{\max;SST} \leq \theta_{\max;MST}$.

A corollary of the above findings is that (5) will elicit the conventional upper bound of one for the structural parameter only where a and/or b is indifferent to x[20].

To give these results some intuition, note that whilst the inter-nest binary probabilities (which in Case 3 account for the first and second choices in the transitivity chain) will be independent of the value of the structural parameter, the intra-nest binary probability (which in Case 3 accounts for the third 'transitive' choice) will not. For this case, we

---

[17] In passing, it is worth remarking that we confirmed the bounds for both Cases 3 and 4, by applying the stochastic transitivity tests (B1), (B2) and (B3) to a wide range of values for both the deterministic utilities (i.e. V_x, V_a, V_b) and the structural parameter (i.e. θ), and checking correspondence with the bounds on the structural parameter arising from (5), (6) and (7).

[18] With the caveat that, having adopted a given *intrinsic* preference ordering (either axb or bxa), $w = \max(u,v)$ and $w = \min(u,v)$ for the *actual* preference ordering axb will correspond to $w = \min(u,v)$ and $w = \max(u,v)$ respectively for the *actual* preference ordering bxa.

[19] A slight qualification is that we introduce the subscript axb to the structural parameter to denote the actual preference ordering; we will adopt the same convention in the subsequent working.

[20] In this case, from (B10) it must hold that $(1+w) \leq (1+u)(1+v)$, which simplifies to $w \leq u+v+uv$. If, for example, b is indifferent to x, then $v = 0$ and the latter inequality further simplifies to $w \leq u$, consistent with SST. If all three alternatives are indifferent to each other, then $w \leq 0$, consistent with WST.

wish to discern, for given inter-nest binary probabilities greater than 0.5, any bounds on the structural parameter which ensure that the intra-nest binary choice will complete the transitivity chain. Since an increasing value of the structural parameter will amplify the probability of choosing the (deterministically) inferior alternative from the intra-nest binary, Case 3 gives rise to an upper bound on the structural parameter; at higher values of the structural parameter, the (deterministically) inferior intra-nest alternative will become sufficiently attractive that stochastic transitivity fails. Consider for example the actual preference ordering axb. If the inter-nest choices are consistent with this preference ordering (i.e. a is stochastically preferred to x, and x to b), then compliance with stochastic transitivity rests upon the intra-nest choice, in particular the strength of preference for a over b, relative to the strength of the inter-nest preferences. An increasing value of the structural parameter will gradually reduce the intra-nest probability for a over b, until an upper bound is reached where stochastic transitivity fails.

Case 4: Where the 'transitive' choice is inter-nest

Whereas Case 3 dealt with actual preference orderings that are consistent with the intrinsic preference ordering, in the sense that the 'transitive' choice is intra-nest, Case 4 deals with actual preference orderings that entail inter-nest transitivity, i.e. abx, bax, xab and xba (again reflecting complete binary stochastic preferences on the trinary choice set, as per the notational definition at the beginning of section 3.3.2). As was noted in Case 3 however, having determined which alternatives should be nested together, it is arbitrary as to which alternative is labelled a and b. In practice, therefore, we need only consider two of these four preferences orderings, where the defining feature of these preference orderings is the rank of the lone alternative x.

Case 4.1: Consider the actual preference ordering xab, where the lone alternative is first-ranked (i.e. $r_x = 1$, noting that we could instead consider xba, and (in essence[21]) derive the same bounds on the structural parameter). Reconciling xab with the odds ratios (B4a) and (B4b), we can reason that (see Table 1 for additional working), in the case of the actual preference ordering xab, it must hold that $p(a,x)/p(x,a) \le 1$, $p(x,b)/p(b,x) \ge 1$ and $p(a,b)/p(b,a) \ge 1$. The implication is that, whereas Case 3 gave rise to an *upper* bound on the structural parameter (5), the present case gives rise to the *lower* bound:

$$\theta_{xab} = \frac{\ln\left((1+u)\left((1+v)+k_{xab}\right)\right)}{\ln(1+w)} \ge \frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} = \theta_{\min;r_x=1} \qquad (6)$$

[21] In an analogous fashion to Case 3, having adopted an *intrinsic* preference ordering (either axb or bxa), $w = \max(u,v)$ and $w = \min(u,v)$ for the *actual* preference ordering xab will correspond to $w = \min(u,v)$ and $w = \max(u,v)$ respectively for the *actual* preference ordering xba.

where $\theta_{\min;r_x=1}$ denotes the lower bound of the structural parameter given that the lone alternative is ranked first, $u \le 0$, $v \ge 0$, $w = \max(u,v)$ for SST, $w = \min(u,v)$ for MST, and $k_{xab} \ge 0$.

TABLE 1 ABOUT HERE

It should be qualified that, given the relations inherent within (6), the lower bound for MST will in principle be negative (more specifically, the numerator of the lower bound in (6) will be positive, but the denominator will be negative). In practice however, a negative structural parameter will violate regularity (section 3.3.1), and it therefore makes sense to impose a lower bound of zero for MST.

Case 4.2: Consider the actual preference ordering abx, where the lone alternative is third-ranked (i.e. $r_x = 3$, noting that bax will (in essence[22]) yield the same bounds on the structural parameter). Following an analogous line of reasoning to Case 4.1, it must in this case hold that $p(a,x)/p(x,a) \ge 1$, $p(x,b)/p(b,x) \le 1$ and $p(a,b)/p(b,a) \ge 1$, thereby giving rise to the lower bound:

$$\theta_{abx} = \frac{\ln\left(\left((1+u)+k_{abx}\right)(1+v)\right)}{\ln(1+w)} \ge \frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} = \theta_{\min;r_x=3} \tag{7}$$

where $\theta_{\min;r_x=3}$ denotes the lower bound of the structural parameter given that the lone alternative is ranked third, $u \ge 0$, $v \le 0$, $w = \max(u,v)$ for SST, $w = \min(u,v)$ for MST, and $k_{abx} \ge 0$. In practice, the lower bound for MST will again be zero.

Moreover (6) and (7) provoke the following inferences:

- In summary: $0 < \theta_{\min;r_x=3} \le \theta_{abx}$, and $0 < \theta_{\min;r_x=1} \le \theta_{xab}$
- Whether $\theta_{\min;r_x=1} < \theta_{\min;r_x=3}$, or $\theta_{\min;r_x=1} > \theta_{\min;r_x=3}$, will be an empirical issue.

To give this result some intuition, in Case 4 the intra-nest choice will be first or second in the transitivity chain, whilst the third (i.e. 'transitive' choice) will be inter-nest. As before, we wish to discern, for given inter-nest probabilities, any bounds on the structural parameter such that the intra-nest probability (which unlike Case 3 will not be the third 'transitive' choice) is consistent with the transitivity chain. Consider for example the actual preference ordering abx. If the inter-nest choices are consistent with this preference (i.e. a is stochastically preferred to x, and b to x), then compliance with stochastic transitivity rests upon the intra-nest choice, in particular the strength of preference for a over b, relative to the strength of the inter-nest preferences. As the value of the structural parameter increases, the probability of choosing a over b will decrease, whilst the probability of choosing a over x will

[22] In an analogous fashion to Case 3, having adopted an *intrinsic* preference ordering (either axb or bxa), $w = \max(u,v)$ and $w = \min(u,v)$ for the *actual* preference ordering abx will correspond to $w = \min(u,v)$ and $w = \max(u,v)$ respectively for the *actual* preference ordering bax.

remain constant, until the bound is eventually reached where $p(a,x) \geq p(a,b) \geq 0.5$ and SST is satisfied.

3.4 A simple two-part test for regularity and stochastic transitivity

Arising from the discussions of regularity (section 3.3.1) and stochastic transitivity (section 3.3.2), in relation to our NL representation (section 3.2) of a three-alternative discrete choice problem (section 3.1), the logical progression is to propose a simple two-part test, as follows.

Part I of the test considers compliance with regularity, which in the trinary case is necessary and sufficient for RUM. In principle, regularity applies to both the intra-nest and inter-nest choices, but in practice, only the latter entail a restriction on the structural parameter. More specifically, regularity implies a lower bound of zero on the structural parameter. It is important to note that, whilst excluding negative values, this condition does not guarantee that a positive value of the structural parameter will comply with regularity. From an empirical perspective, regularity will hold as the structural parameter increases through the range zero to one, and possibly in excess of one. However, a critical value will eventually be reached at which one or more choice shares approach zero or one, and regularity then fails; the specific critical value will depend on the utilities at hand.

Part II considers compliance with MST and SST, which are well-behaved conditions, but not necessary for RUM. In terms of these conditions, we must distinguish between cases where the lone alternative is first, second or third ranked, and between different forms of transitivity, namely MST and SST. Where the lone alternative is second (i.e. middle) ranked, SST and MST imply upper bounds on the structural parameter (possibly in excess of one). Where the lone alternative is first or third ranked, MST implies a lower bound of zero, whereas SST implies a lower bound greater than or equal to zero, and neither condition implies an upper bound.

To give an example, recall that Figure 4 assumes $V_x = 10, V_a = 9, V_b = 8$, such that alternative x is deterministically preferred to a, and a to b. With reference to Part I of the test, compliance with regularity requires the structural parameter to be greater than zero. Furthermore, empirical analysis of this example reveals that regularity fails as the structural parameter increases beyond 1.5, whereupon $p(b,x) < p_M(b)$. Therefore, employing a combination of theory and empirics, we can discern that – in this example – regularity will be satisfied where the structural parameter lies within the bounds $0 < \theta_{xab} \leq 1.5$, i.e. an upper bound greater than one.

With reference to Part II, application of the stochastic transitivity condition (6) reveals that, in order for alternative x to be stochastically preferred to alternative b, the structural parameter should be greater than zero for MST, and greater than or equal to 0.5 for SST. Combining the regularity and stochastic transitivity requirements, we can infer that – for this example – regularity and MST will be satisfied where $0 < \theta_{xab} \leq 1.5$, whilst regularity and SST will be satisfied where $0.5 \leq \theta_{xab} \leq 1.5$.

More generally, Table 2 summarises the simple two-part test, detailing the specific bounds on the structural parameter that apply to each possible preference ordering arising from the three-alternative choice problem.

TABLE 2 ABOUT HERE

4. An empirical example

Having reconciled the regularity and stochastic transitivity conditions with the conventional zero-one bounds on the structural parameter, the present section will consider the empirical implications of these findings, by examining the prevalence of structural parameters that fall outside the conventional bounds, and the factors that might give rise to such results. In these practical contexts, it is also appropriate to consider the applicability of the two-part test outlined in section 3 above and, where it is applicable, the ability of the test to determine the validity of structural parameters observed empirically.

Before proceeding, it is important to acknowledge that three-alternative NL could potentially be estimated using data on: i) binary choices only (using Bradley & Daly's (1997) 'trick' to normalise the scale of the different pairs); ii) trinary choices only; or iii) some mixture of binary and trinary choices. Since a key input to the two-part test is knowledge of the intrinsic preference ordering (reflecting the complete set of binary stochastic preferences on a given trinary), this would seem to favour format i). However, there is arguably an intellectual dissatisfaction in constructing a NL model of trinary choice if individuals never actually face such a choice – especially if there is analytical interest in the perceived similarity between the nested alternatives, and the perceived dissimilarity of the lone alternative. Indeed, format i) tends to be the exception rather than the rule in practical NL modelling. On the other hand, the trinary choices inherent within formats ii) and iii) might, on the face it, seem to impede the applicability of the two-part test. This is because the binary stochastic preferences inherent within these trinary choices are opaque to the analyst. However, following the rationale previously deployed by Batley & Daly (2006), the marginal (2) and conditional (3) choice probabilities of NL lend themselves to the elicitation of probabilities in reduced choice sets, by considering these as limiting cases when the utility weights of individual alternatives become zero.

For example, if $V_x = 0$ then $p_M(L) = 1$ and $p_M(a) = p(a,b)$. In support of this representation of reduced choice sets, consider the following intuition: as V_X reduces in value towards zero, and the probability of choosing x also reduces to zero, one might reasonably expect NL to 'behave' in the sense that:

$$p(a,b) = \lim[V_x \to 0]\, p_M(a)$$

This says that the probability (3) of choosing a from the reduced choice set L is the same whether x is not considered at all, or whether x is considered but its utility weight is allowed to decline to zero. Similarly, if $V_b = 0$ then $p(a,b) = 1$ and (2) gives the probability of choosing a from the reduced choice set $\{a, x\}$. Thus the probability equations for the three-alternative case yield the relevant probability equations for each of the three possible binary choices.

4.1 Empirical example using real data

Our first empirical example is based on data collected using a stated choice (SC) survey, conducted using an online panel in the United Kingdom in early 2010. For full details of the survey, see Hess *et al* (2012). The sample consisted of 387 respondents who routinely commuted by bus or rail. The respondents were each issued with ten stated choice scenarios, where each scenario involved a choice between three unlabelled journeys based on their usual mode, and where the first journey was a respondent-specific 'reference' journey that was held constant across scenarios. The three journey alternatives were described in terms of five attributes, namely travel time (in minutes), cost (£), the rate of crowding (trips out of ten), the rate of delays (trips out of ten), the average delay across delayed trips (in minutes), and the provision of a delay information text message (sms) service (three possible levels; none, charged, and free). In each scenario, the respondent was asked to choose their most preferred option as well as their least preferred option. For purposes of analysis, we combined the data on best and worst choices, yielding twenty observations per respondent (ten with the choice of the *best* alternative in each task, and ten with the choice of the *worst* alternative out of the remaining two in each task), where no differences in scale were found between best and worst choices, and where similar findings to those reported here were obtained when using only data on the best choice.

Formalising this example using the notation introduced earlier, let x be the reference journey, and let a and b be hypothetical alternatives. Overall, the choice shares were such that alternative x (which was also specified as the lone alternative in NL terms) was most preferred, followed by alternative b, and then alternative a (where the latter two alternatives were, in NL terms, nested together), i.e. $p_M(x) > p_M(b) > p_M(a)$. Two different models were estimated on this dataset, namely MNL and NL. In both models, and in line with earlier findings by Hess *et al* (2012), we used a log-transform on the fare attribute. The results are summarised in Table 3, where the estimation of the models recognised the repeated choice nature of the data in the calculation of the robust standard errors.

TABLE 3 ABOUT HERE

The MNL results show the expected signs for all key attributes, with high levels of statistical significance, along with a dislike, albeit not statistically significant, for a charged delay sms service (relative to no service). For the NL model, we first imposed a constraint on the structural parameter such that $0 < \theta \leq 1$, in line with the default option in many estimation packages – this model collapsed to a MNL structure, i.e. with $\theta = 1$. We then re-estimated the NL model without constraining the structural parameter, finding that $\theta = 1.74$. This potentially supports our propositions concerning the bounds on the structural parameter for consistency with regularity and stochastic transitivity, in the sense that the structural parameter exceeds one.

However, in order to facilitate application of the two-part test from section 3.4 – and thereby elicit a precise bound on the structural parameter for consistency with stochastic transitivity – we simplified matters by re-estimating the NL model on a restricted dataset containing only observations of the 'best' alternative from the trinary choice set (thus omitting observations of the 'worst' alternative from reduced choice sets of two alternatives). For the restricted dataset, we deployed the rationale outlined at the outset of section 4 to elicit the complete set of binary stochastic preferences associated with the trinary. This exercise identified a prevailing preference ordering of xba, and using (6) we calculated – for each and every observation in the dataset – a lower bound of 1.46 for compliance of the structural parameter with SST. Estimating NL on the reconstituted dataset, the (unconstrained) structural parameter was also found to be 1.46, thereby corroborating (6), and implying that the best-fitting model was that falling at the lower bound. That the estimated structural parameter fell at the lower bound should come as no surprise since, for given choice shares, a structural parameter greater than 1.46 would imply greater variance in the utilities of alternatives a and b, and thus a poorer fitting model.

Returning to the best-worst (i.e. unrestricted) dataset, Table 3 reports that, relative to MNL, the NL admitting $\theta > 1$ gives an improvement in log-likelihood by 83.28 units for one additional parameter (from -5,724.137 to -5,640.858), which is highly significant, giving a likelihood ratio test value of 166.56, with a critical 99% χ^2 test value of just 6.63. The improvement in fit is also reflected in the fact that the estimate for the structural parameter is significantly different from one at high levels of confidence, with a *t*-ratio against one of 8.64. It is apparent therefore that, for the present data at least, imposing the conventional $0 < \theta \leq 1$ constraint on the structural parameter leads to inferior model performance. Whilst there are of course situations where a better fitting model may be rejected on theoretical grounds, our work here allows us to determine that these higher values are in fact permissible.

Despite the improvement in fit brought by the NL with $\theta = 1.74$, it is however interesting to note that the $0 < \theta \leq 1$ constraint has little or no impact on the implied monetary valuations (i.e. ratio of the marginal utility of key attributes to the marginal utility of travel cost); again, it should be qualified that this result refers only to the present data and cannot be taken as a general outcome. Notwithstanding the usual reservations about forecasting with hypothetical data, we also conducted a simple example looking at the effect of a 10% increase in fare for the reference journey on its probability of being chosen. Recall that the reference journey was specified as the lone alternative in NL (i.e. as alternative x) and, prior to the fare increase, was first-ranked of the three alternatives. The results in Table 3 show that, relative to MNL (or indeed to any NL observing $0 < \theta \leq 1$), the estimated NL (embodying $\theta > 1$) predicts a larger decrease in the choice probability of the reference alternative. Drawing reference to the earlier discussion of substitutability between alternatives in section 1.3, this finding suggests that the imposition of an upper bound of one on the structural parameter, when this is not theoretically required or empirically supported, may yield misleading forecasts.

4.2 Empirical example using simulated data

Further to the empirical example using real data, we also conducted a larger scale simulated data exercise, using a broad range of 'true' (i.e. supposed) values for the structural parameter. The example was again based on a three-alternative choice task, where two alternatives represented rail journeys, and the third alternative represented a car journey. The alternatives were described in terms of time and cost, on the basis that the car journey was faster but more expensive than the two rail journeys. The actual attribute levels came from a D-efficient experimental design.

The simulation was run on a loop, with the time and cost coefficients fixed at -0.025 and -0.125 respectively, but the structural parameter adjusted incrementally each time. As will be described further in the subsequent sections, the range of the structural parameter encompassed both negative and positive values. On each iteration of the loop, 10,000 choice observations were simulated, and applied to the estimation of both MNL and NL models.

For modelling purposes, the rail journeys were specified as alternatives a and b, and the car journey as alternative x; thus in NL terms, the car journey was represented as the lone alternative. Given the range of values for the 'true' structural parameter, different datasets entailed different preference orderings of alternatives a,b and x. That said, at $\theta = 1$, which represented the approximate mid-point of the range simulated, the choice shares were such that $p_M(a) > p_M(b) > p_M(x)$. Since $\theta = 1$ implies independence of the random terms of the three alternatives, order independence is justified (see footnote 15), and we can infer an underlying preference ordering (i.e. reflecting complete binary stochastic preferences on the trinary choice set) of abx. As detailed in Table 2, given this preference ordering, regularity and MST require the structural parameter to be greater than zero, but do not imply an upper bound.

4.2.1 Negative 'true' values of the structural parameter

The motivation for the first part of this analysis is somewhat different from the preceding analysis of real data, in that we are interested in the implications that arise if the 'true' structural parameter is negative (i.e. in principle, violating regularity and stochastic transitivity), but the analyst restricts the structural parameter to the conventional $0 < \theta \leq 1$ range (i.e. in practice, 'forcing' compliance with regularity and stochastic transitivity). We simulated data with values for the structural parameter ranging from -1 to -0.07, finding that values closer to zero than -0.07 led to estimation failures. For each of the 94 datasets simulated on this basis, we estimated MNL as well as NL without any constraint on the structural parameter. To reiterate our motivation here, we wish to determine whether, if the data underpinning the models embodies violations of regularity and stochastic transitivity, the estimated NL could expose these violations.

FIGURE 5 ABOUT HERE

With reference to the top right panel of Figure 5, we find that when using a positive starting value for the structural parameter in NL, the model is unable to recover the negative sign of the parameter used to generate the data – with one exception where the 'true' value is -0.99. By contrast, when using a negative starting value for the structural parameter, the 'true' value is retrieved from the data. We confirmed this result for a range of different starting values, and using all standard NL estimation packages as well as purpose-written code. These findings point to difficulties in retrieving the 'true' value for the structural parameter when this is negative, with a seeming inability of the estimation to cross from positive to negative space. More worryingly, it was not the case that the estimate for the structural parameter tended towards zero, which might have been suggestive of a 'true' negative value; on the contrary, the estimate became positive and significantly different from zero!

The remaining panels of Figure 5 present summary plots of goodness of fit, willingness-to-pay (WTP) and cost elasticity across the range of 'true' negative values of the structural parameter, distinguishing between MNL and unconstrained NL. We should note that, across the range of negative structural parameters considered, we observed preference reversals whereby the multinomial choice probability for the lone alternative exceeded one of the binary choice probabilities; this phenomenon might be rationalised as a violation of regularity or stochastic transitivity, or a violation of both.

Given a negative starting value for the structural parameter, the log-likelihood of NL is always superior to that of MNL, and increasingly so as the structural parameter approaches zero. The same outcome also arises when employing a positive starting value, except for the case where the 'true' structural parameter is equal to -0.99; in this case, the estimated value for the structural parameter – even with positive starting values – is close to the 'true' value. Since recovery of the 'true' value failed for -1 and -0.98, there is no clear reason why estimation was successful for -0.99. In general, a negative starting value for the structural parameter leads to the estimation of a structural parameter that is very close to the 'true' value.

Turning to inferences of WTP, the 'true' WTP in these datasets was £0.2/min across all settings, and this was recovered very accurately by the NL with negative starting values. Again with reference to Figure 5, MNL always estimates negative WTP measures, as does the NL with positive starting values as the 'true' structural parameter approaches zero; this would at least give an analyst some indication of problems in the data. Where the 'true' structural parameter is more negative, however, the NL with positive starting values greatly underestimates WTP.

Finally, looking at the implied cost elasticity for the lone alternative (car), we can see from Figure 5 that (with the exception of a single iteration of the simulation) this is recovered accurately by the NL with negative starting values. With positive starting values, however, the elasticity is (with the exception of a single iteration, once again) underestimated; this bias is more pronounced in the NL with positive starting values than in MNL. Moreover, the improvement in fit of NL relative to MNL and the compliance of the structural parameter with the conventional zero-one bounds might lead the analyst to (unwittingly) adopt a model that produces greater bias in its forecasts.

4.2.2 Positive 'true' values of the structural parameter

We also simulated datasets with 'true' values of the structural parameter within the range +1 to +3. In contrast to section 4.2.1, here we are interested in the implications that arise if the 'true' structural parameter is in excess of one, but the analyst restricts the structural parameter to the conventional $0 < \theta \leq 1$ range.

As noted earlier in section 4.2, given a structural parameter $\theta = 1$, the simulated data exhibited the preference ordering abx; in this case, Table 2 advises us that, in theoretical terms, regularity and MST entail a lower bound of zero for the structural parameter, but no upper bound. In empirical terms, we would expect regularity and stochastic transitivity to eventually fail as the structural parameter increases beyond some critical value and the choice shares become extreme (e.g. as in Figure 4); however, no such failures were observed across the range $1 \le \theta \le 3$.

FIGURE 6 ABOUT HERE

For each 'true' value of θ, we estimated MNL alongside the unconstrained NL (note that MNL produces the same results as NL at $\theta = 1$, but different results where $\theta > 1$). The results for the 201 models estimated on this basis are summarised in Figure 6. Not surprisingly, the results show that as the 'true' value of the structural parameter exceeds one (and especially beyond 1.3), the NL model achieves substantial gains in log-likelihood over the MNL model, and is able to closely recover the 'true' structural parameter and WTP (where the latter continues to be £0.2/min). MNL on the other hand overestimates WTP and underestimates the cost elasticity as the structural parameter increases, and these biases increase as the 'true' value of the structural parameter increases. This once again suggests that the imposition of overly-restrictive constraints on the structural parameter can bias the results.

As an aside, we also developed a counterpart to the analysis from section 4.2.1, by estimating NL with a negative starting value for the structural parameter for the present context where the 'true' values were positive. In many cases, the estimation either failed to converge or converged to negative values, whilst the 'true' positive value was recovered only in occasional cases. This suggests that, in an analogous fashion to section 4.2.1 where the 'true' structural parameter was negative, using the correct sign for the starting value is important to the estimation routine.

5. Summary and Conclusions

Drawing upon the early RUM literature by Marschak (1960) and Block & Marschak (1960), this paper introduced regularity and stochastic transitivity as necessary and well-behaved conditions respectively, for the consistency of discrete choice preferences with the Random Utility Model (RUM). A particular contribution of the paper was to combine the regularity and stochastic transitivity conditions in the form of a simple two-part test, and to illustrate the application of this test for a three-alternative discrete choice problem (i.e. treating the nests of NL as reduced choice sets).

With regards to regularity, we showed that any failures will be associated with inter-nest choices (i.e. preference reversals in relation to the lone alternative), and that the prevalence of such failures will be determined by the magnitude of the structural parameter (reflecting the degree of similarity between nested alternatives) in combination with the binary and trinary probabilities. More specifically, we found that regularity implies positivity of the structural parameter in NL, but no upper bound.

With regards to stochastic transitivity, we showed that compliance will also be determined by the magnitude of the structural parameter, as well as by the odds ratios for the different pairs within the three-alternative choice set. Furthermore, stochastic transitivity will apply differently, depending on the rank of the lone alternative within the preference ordering. More specifically, where the lone alternative is second (i.e. middle) ranked, MST and SST imply different upper bounds on the structural parameter, possibly in excess of one. On the other hand, where the lone alternative is first or third ranked, MST and SST imply lower bounds of zero and greater than or equal to zero respectively, but no upper bound.

Drawing together our analyses of regularity and stochastic transitivity, we arrive at the following conclusions for the case of three-alternative NL:

- Whilst regularity supports the conventional lower bound of zero (i.e. $\theta > 0$) on the structural parameter, SST may, for some preference orderings, give rise to a lower bound greater than zero (i.e. requiring $\theta \geq l$, where $l > 0$).
- Neither regularity nor the stochastic transitivity conditions constrain the upper bound of the structural parameter to be one.
- Therefore, if the conventional $0 < \theta \leq 1$ bounds are imposed on model estimation, either or both of two scenarios could arise:
    - Preferences which violate regularity and/or stochastic transitivity may go undetected (e.g. where the 'true' value of the structural parameter is less than zero) or be unknowingly admitted (e.g. where SST calls for a lower bound greater than zero).
    - Preferences which comply with regularity and stochastic transitivity may be unknowingly excluded (e.g. where the 'true' structural parameter is greater than one).
- Moreover, if either of the above scenarios arises, then the imposition of $0 < \theta \leq 1$ on model estimation (as is done in some standard software) may compromise model fit, inferences of willingness-to-pay, and forecasts of choice behaviour.
- Finally, even where $0 < \theta \leq 1$ is not imposed, maximum likelihood estimation may fail to recover 'true' values of the structural parameter less than zero (i.e. fail to expose regularity and stochastic transitivity violations) unless starting values are of the correct sign. This suggests that analysts may wish to test both positive and negative starting values for the structural parameter.

Whilst the present paper has focussed upon a three-alternative choice set, it would seem reasonably straightforward in principle to apply the two-part test to larger choice sets and more complex tree structures, possibly involving multiple structural parameters. Regularity will continue to require a positive structural parameter for every constituent nest, whilst stochastic transitivity will give rise to upper or lower bounds on the structural parameter for each and every triple. Since different triples will elicit different bounds for a given nest, a pragmatic implementation of the method would be to focus upon the 'global' maximum or minimum, corresponding to Cases 3 and 4 in section 4 of the paper.

Acknowledgements

Thanks are due to three anonymous peer reviewers appointed by this journal, whose comments greatly improved the paper. Responsibility for any remaining faults or limitations rest entirely with the authors.

References

Barberá, S. & Pattanaik, P.K. (1986) Falmagne and the rationalizability of random orderings. *Econometrica*, 54 (3), pp707-715.

Batley, R. (2008) On ordinal utility, cardinal utility and random utility. *Theory & Decision*, 64 (1), pp36-63.

Batley, R. (2014) The intuition behind income effects of price changes in discrete choice models, and a simple method for measuring the compensating variation. Paper presented at the ITEA conference, June 2014, Toulouse, France.

Batley, R. & Daly, A. (2006) On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. *Journal of Mathematical Psychology*, 50 (5), pp456-467.

Ben-Akiva, M.E. (1974) Structure of passenger travel demand models. *Transportation Research Record*, 526, pp26-42.

Bierlaire, M. (2001) A theoretical analysis of the cross-nested logit model. Report RO-011218, Ecole Polytechnique Fédérale de Lausanne, Department of Mathematics.

Bierlaire, M. (2002) The Network GEV model. Paper presented at the Swiss Transport Research Conference 2002, Monte Verita.

Block, H.D. & Marschak, J. (1960) Random orderings and stochastic theories of responses. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1), pp172-217. D. Reidel, Dordrecht. Reprinted from Olkin, I., Ghuryé, S.G., Hoeffding, W., Madow, W.G. & Mann, H.B. (eds) *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp97-132. Stanford University Press, Stanford.

Börsch-Supan, A. (1985) Tenure choice and housing demand. In Stahl, K. & Struyk, R. (eds) *U.S. and German Housing Markets: Comparative Economic Analyses*, pp55-105. Urban Institute Press, Washington, DC.

Börsch-Supan, A. (1990) On the compatibility of nested multinominal logit models with utility maximization. *Journal of Econometrics,* 43*,* pp373-388.

Bradley, M.A. & Daly, A.J. (1997) Estimation of logit choice models using mixed stated preference and revealed preference information. In Stopher, P.R. & Lee-Gosselin, M. (eds) *Understanding Travel Behaviour in an Era of Change*. Pergamon, Oxford.

Busemeyer, J.R. & Rieskamp, J. (2013) Psychological research and theories on preferential choice. Working paper, Psychological and Brain Sciences, Indiana University.

Carrasco, J.A. & Ortúzar, J. de D. (2002) Review and assessment of the nested logit model. *Transport Reviews*, 22 (2), pp197-218

Cavagnaro, D.R. & Davis-Stober, C.P. (2014) Transitive in our preferences, but transitive in different ways: an analysis of choice variability. *Decision*, 1 (2), pp102-122.

Daly, A. (2001) Recursive Nested Extreme Value Model, Working Paper 559, Institute for Transport Studies, University of Leeds.

Daly, A. & Bierlaire, M. (2006) A general and operational representation of GEV models. *Transportation Research Part B*, 40 (4), pp285-305.

Daly, A.J. & Zachary, S. (1976) Improved multiple choice models. In *Proceedings of the Fourth PTRC Summer Annual Meeting*. PTRC, London.

Davidson, D. & Marschak, J. (1959) Experimental tests of a stochastic decision theory. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1), pp133-171. D. Reidel, Dordrecht. Reprinted from Churchman, C.W. & Philburn, R. (eds) *Measurement: Definitions and Theories*, pp233-269. Wiley, New York.

Debreu, G. (1954) Representation of a preference ordering by a numerical function. In Thrall, R.M., Coombs, C.H. & Davis, R.L. (eds) *Decision Processes,* pp159-165. Wiley, New York.

Debreu, G. (1958) Stochastic choice and cardinal utility. *Econometrica*, 26 (3), pp440-44.

Falmagne, J.C. (1978) A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, 18 (1), pp52-72.

Fechner, G.T. (1859) *Elemente der Psychophysik*. Breitkopf and Härtel, Leipzig.

Fiorini, S. (2003) A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology*, 48 (1), pp80-82.

Fishburn, P.C. (1998) Stochastic utility. In Barberá, S., Hammond, P.J. & Seidl, C. (eds) *Handbook of Utility Theory*. Kluwer, Dordrecht, pp273-318.

Gorman, W. (1968) Conditions for Additive Separability. *Econometrica*, 36 (3/4), pp605-609.

Heath, T. B. & Chatterjee, S. (1995) Asymmetric decoy effects on lower quality versus higher quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, 22 (3), pp268-284.

Hensher, D. A. (1984) Full information maximum likelihood estimation of a nested logit mode-choice model. Working paper no. 13, Dimensions of automobile demand project, Macquire University, North Ryde.

Hensher, D.A. & Greene, W. (2002) Specification and estimation of the nested logit model: Alternative normalisations. *Transportation Research Part B: Methodological*, 36 (1), pp1-17.

Herriges, J.A. & Kling, C. L. (1996) Testing the consistency of nested logit models with utility maximisation. *Economics Letters*, 50 (1), pp33–39.

Hess, S., Stathopoulos, A. & Daly, A.J. (2012) Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39 (3), pp565-591.

Hougaard, J.L, Tjur, T. & Østerdal, L.P. (2006) Testing preference axioms in discrete choice experiments: a reappraisal. Discussion Paper 06-11, Department of Economics, University of Copenhagen.

Huber, J, Payne, J. W. & Pluto, C. (1982) Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9 (1), pp90-98.

Ibáñez, J.N. (2007) On the compatibility of probabilistic choice systems with random utility maximisation. ITS Working Paper 2007-X23. Institute for Transport Studies, University of Leeds.

Kahneman, D. & Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2), pp263-291.

Koning, R. H. & Ridder, G. (1994) On the compatibility of nested logit models with utility maximization: a comment. *Journal of Econometrics*, 63 (2), pp389-396.

Lee, B. (1999) Calling patterns and usage of residential toll service under self-selecting tariffs. *Journal of Regulatory Economics*, 16 (1), pp45–82.

Luce, R.D. (1959) *Individual Choice Behaviour: A Theoretical Analysis*. John Wiley, New York.

Luce, R.D. & Suppes, P. (1965) Preference, utility and subjective probability. In Luce, R.D., Bush, R.R. & Galanter, E. (eds) *Handbook of Mathematical Psychology*, Volume III, pp249-410. John Wiley & Sons, New York.

McFadden, D. (1968) The revealed preferences of a government bureaucracy. Department of Economics, University of California, Berkeley, California, unpublished.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (ed) *Frontiers in Econometrics*, pp105-142. Academic Press, New York.

McFadden, D. (1975) The revealed preferences of a government bureaucracy: theory. *The Bell Journal of Economics and Management Science*, 6 (2), pp401-416.

McFadden, D. (1978) Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F. & Weibull, J. (eds) *Spatial Interaction Theory and Residential Location*, pp75-96. North-Holland, Amsterdam.

McFadden, D. (1981) Econometric models of probabilistic choice. In Manski, C. & D. McFadden (ed) *Structural Analysis of Discrete Data: With Econometric Applications*, pp198-272. The MIT Press, Cambridge, Massachusetts.

McFadden, D.L. (1995) Computing willingness-to-pay in Random Utility Models. Working paper, Department of Economics, University of California, Berkeley.

McFadden, D.L. (2005) Revealed Preference: a synthesis. *Economic Theory*, 26 (2), pp245-264.

McFadden, D. & Richter, M. K. (1970a) Revealed stochastic preference. Unpublished manuscript, Department of Economics, University of California, Berkeley.

McFadden, D. & Richter, M. K. (1970b) Stochastic rationality and revealed stochastic preference. Unpublished manuscript, Department of Economics, University of California, Berkeley.

McFadden, D.L. & Richter, M.K. (1991) Stochastic rationality and revealed stochastic preference. In Chipman, J., McFadden, D.L. & Richter, M.K. (eds) *Preferences, Uncertainty, and Rationality*, pp161-186. Westview Press, Boulder, CO.

Manheim, M.L. (1973) Practical implications of some fundamental properties of travel demand models. *Highway Research Record*, 422, pp21-38.

Marley, A.A.J. (1990) A historical and contemporary perspective on random scale representations of choice probabilities and reaction times in the context of Cohen and Falmagne's (1990, Journal of Mathematical Psychology, 34) results. *Journal of Mathematical Psychology*, 34 (1), pp81-87.

Marley A. A. J. & Regenwetter, M. (2016) Choice, preference, and utility: probabilistic and deterministic representations. In Batchelder, W. Colonius, H., Dzhafarov, E. & Myung, J. (eds) *New Handbook of Mathematical Psychology, Volume 1: Measurement and Methodology*. Cambridge University, Cambridge.

Marschak, J. (1960) Binary choice constraints and random utility indicators. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1), pp218-239. D. Reidel, Dordrecht. Reprinted from Arrow, K.J., Karlin, S. & Suppes, P. (eds) *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford.

Moscati, I. (2007) Early experiments in consumer demand theory: 1930-1970. *History of Political Economy*, 39 (3), pp359-401.

Ortúzar, J. de Dios (2001) On the development of the nested logit model. *Transportation Research Part B*, 35 (2), pp213-216.

Papandreou, A. G. (1957) A test of a stochastic theory of choice. *University of California Publications in Economics*, 16, pp1-18.

Regenwetter, M. Dana, J. & Davis-Stober, C. (2010) Testing transitivity of preferences on two-alternatives forced choice data. *Frontiers in Psychology*, 148, pp1-15.

Regenwetter, M., Dana, J. & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118, pp42-46.

Rieskamp, J., Busemeyer, J. R. & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, 44 (3), pp631-661.

Simonson, I. (1989) Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16 (2), pp158-174.

Simonson, I. & Tversky, A. (1992) Choice in context: tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29 (3), pp281-295.

Small, K.A. & Brownstone, D. (1982) Efficient estimation of nested logit models: an application to trip timing. Economic research program research memorandum no. 296, Princeton University, Princeton, NJ.

Suppes, P., Krantz, D.H., Luce, R.D. & Tversky, A. (1989) *Foundations of Measurement, Volume II: Geometrical, Threshold, and Probabilistic Representations*. Academic Press, San Diego, California.

Thurstone, L. (1927) A law of comparative judgement. *Psychological Review*, 34 (4), pp273-286.

Train, K., McFadden, D. & Ben-Akiva, M. (1987) The demand for local telephone service: A fully discrete model of residential calling patterns and service choice. *Rand Journal of Economics*, 18 (1), pp109-123.

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

Trueblood, J. S, Brown, S. D. & Heathcote, A. (2015) The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, 122 (4), pp848-853.

Tversky, A. (1972a) Choice by elimination. *Journal of Mathematical Psychology*, 9, pp341-367.

Tversky, A. (1972b) Elimination by aspects: a theory of choice. *Psychological Review*, 79 (4), pp281-299.

Tversky, A. & Sattath, S. (1979) Preference trees. *Psychological Review*, 86 (6), pp542-573.

Vovsha, P. (1997) Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record,* 1607**,** pp6-15.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9 (3), pp285-344.

Williams, C.W.L. & Senior, M.L. (1978) Accessibility, spatial interaction and the spatial benefit analysis of land use – transportation plans. In Karlqvist, A., Lundqvist, L., Snickars, F. & Weibull, J. (eds) *Spatial Interaction Theory and Residential Location*, pp253-287. North-Holland, Amsterdam.

Wilson, A.G. (1974) *Urban and Regional Models in Geography and Planning*. Wiley, Chichester.

Table 1: Additional working behind equations (5), (6) and (7)

Stochastic preference ordering	*axb*	*xab*	*abx*
$p(a,x)/p(x,a)=$	$(1+u)$	$(1+u)$	$(1+u)+k_{abx}$
$p(x,b)/p(b,x)=$	$(1+v)$	$(1+v)+k_{xab}$	$(1+v)$
$p(a,b)/p(b,a)=$	$(1+w)+k_{axb}$	$(1+w)$	$(1+w)$
$\frac{p(a,x)}{p(x,a)}\cdot\frac{p(x,b)}{p(b,x)}=$	$(1+u)(1+v)$	$(1+u)((1+v)+k_{xab})$	$((1+u)+k_{abx})(1+v)$
$p(a,b)/p(b,a)=$	$((1+u)(1+v))^{\frac{1}{\theta_{axb}}}$	$((1+u)((1+v)+k_{xab}))^{\frac{1}{\theta_{xab}}}$	$(((1+u)+k_{abx})(1+v))^{\frac{1}{\theta_{abx}}}$
	$\theta_{axb}=\frac{\ln((1+u)(1+v))}{\ln((1+w)+k_{axb})}$	$\theta_{xab}=\frac{\ln((1+u)((1+v)+k_{xab}))}{\ln(1+w)}$	$\theta_{abx}=\frac{\ln(((1+u)+k_{abx})(1+v))}{\ln(1+w)}$
	where $u,v\geq 0$, $k_{axb}\geq 0$	where $u\leq 0, v\geq 0$, $k_{xab}\geq 0$	where $u\geq 0, v\leq 0$, $k_{abx}\geq 0$

Note: the constants $k_{axb}, k_{xab}, k_{abx}$ impose WST on the relevant odds ratio for each stochastic preference ordering (i.e. this is analogous to the inequality in (B4c) for the ordering axb).

Table 2: Summary of the two-part test

	PART I: NECESSARY		**PART II: WELL-BEHAVED**	
Stochastic preference ordering	**Regularity**		**Stochastic transitivity**	
	Intra-nest	Inter-nest	SST	MST
abx	n/a	$0 < \theta$	$0 < \theta_{\min;r_x=3} \leq \theta_{abx}$	$0 < \theta_{abx}$
axb			$\theta_{axb} \leq \theta_{\max;SST}$	$\theta_{axb} \leq \theta_{\max;SST} \leq \theta_{\max;MST}$
bax			$0 < \theta_{\min;r_x=3} \leq \theta_{bax}$	$0 < \theta_{bax}$
bxa			$\theta_{bxa} \leq \theta_{\max;SST}$	$\theta_{bxa} \leq \theta_{\max;SST} \leq \theta_{\max;MST}$
xab			$0 < \theta_{\min;r_x=1} \leq \theta_{xab}$	$0 < \theta_{xab}$
xba			$0 < \theta_{\min;r_x=1} \leq \theta_{xba}$	$0 < \theta_{xba}$

Table 3: Estimation results on stated choice data

	MNL		Unconstrained NL	
Individuals	387		387	
Choice tasks	3870		3870	
Observations	7740		7740	
Final LL	-5724.137		-5640.858	
par.	9		10	
adj. ρ^2	0.173		0.185	
	est.	**t-rat. (0)**	**est.**	**t-rat. (0)**
ASC1	0.4670	10.92	0.7680	12.06
ASC2	-0.0338	-0.81	-0.0984	-1.72
travel cost (log £)	-12.4000	-19.95	-16.5000	-19.16
travel time (min)	-0.0372	-9.04	-0.0499	-8.91
rate of crowding (0-1)	-0.2110	-10.64	-0.2910	-10.44
rate of delays (0-1)	-0.2520	-12.42	-0.3460	-11.89
average delay (mins)	-0.0347	-5.49	-0.0470	-5.64
charged delay sms	-0.0765	-1.26	-0.0396	-0.5
free delay sms	0.3150	6.35	0.3980	6.34
	est.	**t-rat. (1)**	**est.**	**t-rat. (1)**
θ	1	-	1.7391	8.64
Implied monetary valuations				
travel time (£/hr)	1.80		1.81	
crowding (one fewer trip out of 10)	0.17		0.18	
delays (one fewer trip out of 10)	0.20		0.21	
delays (£/hr)	1.68		1.71	
free delay sms (£)	0.25		0.24	
Effect of 10% increase in fare for reference alternative				
Average change in probability for reference alternative	-56.93%		-70.06%	

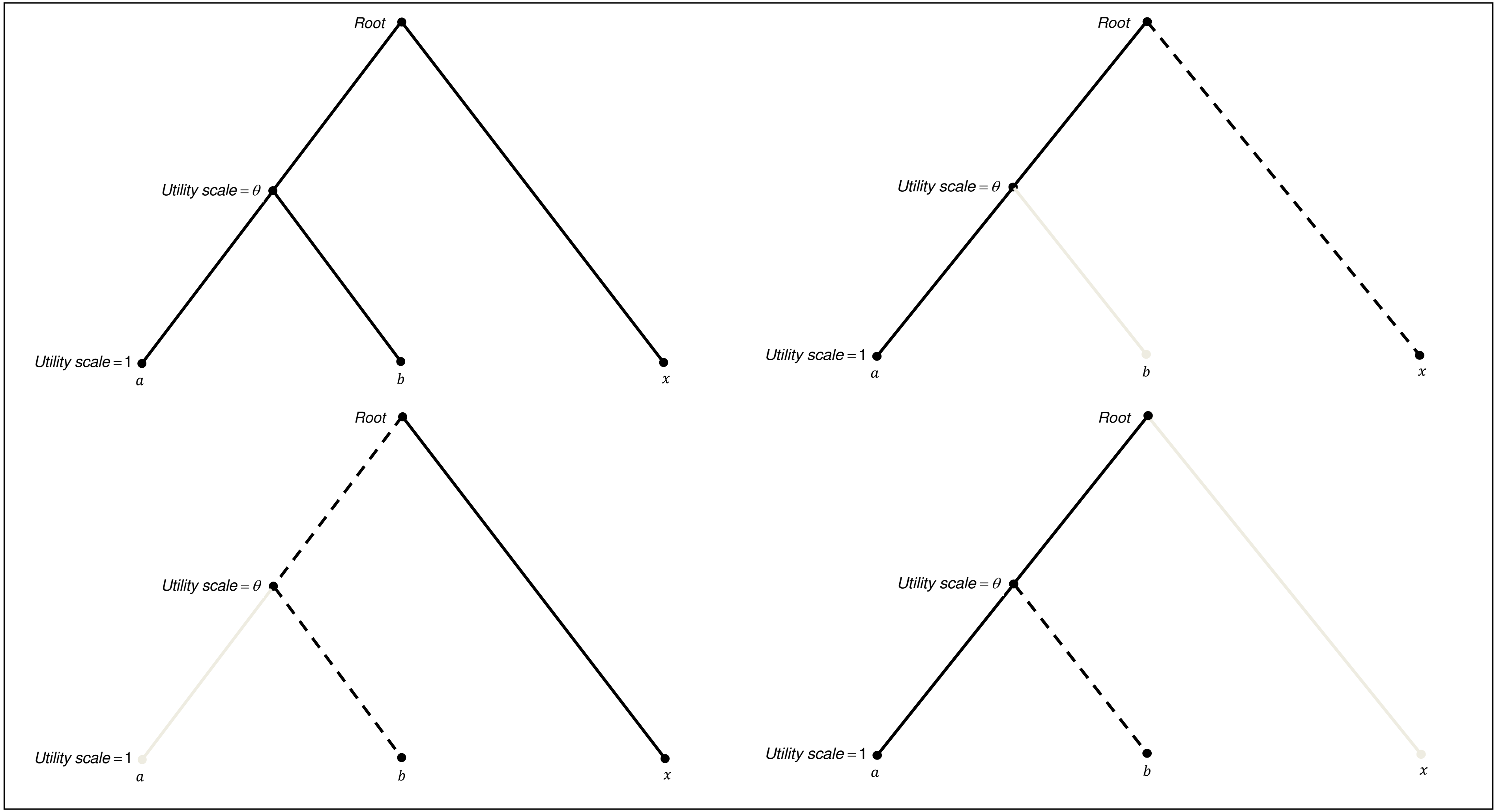

Figure 1: Tree structure for the complete trinary, together with each binary comprising the 'intrinsic' tree structure

(Note: with reference to the binary choices, the black line = 1st choice, dotted line = 2nd choice, grey line = choice unavailable)

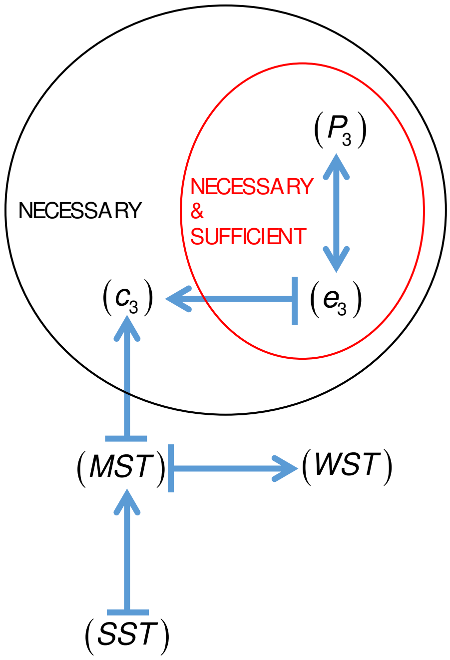

Figure 2: Relationships between properties of RUM for a trinary choice set

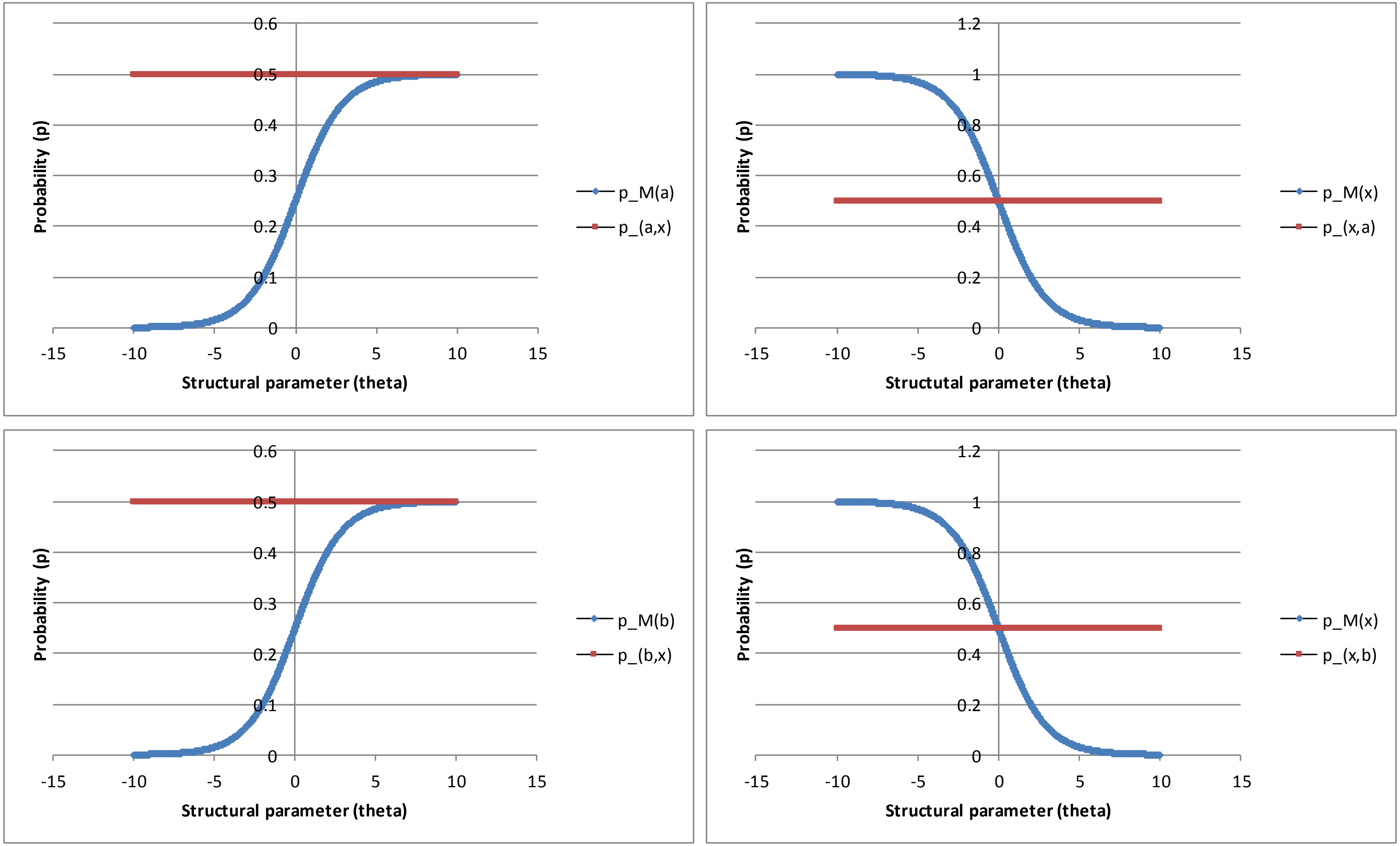

Figure 3: Regularity condition where $V_a = V_b = V_x$

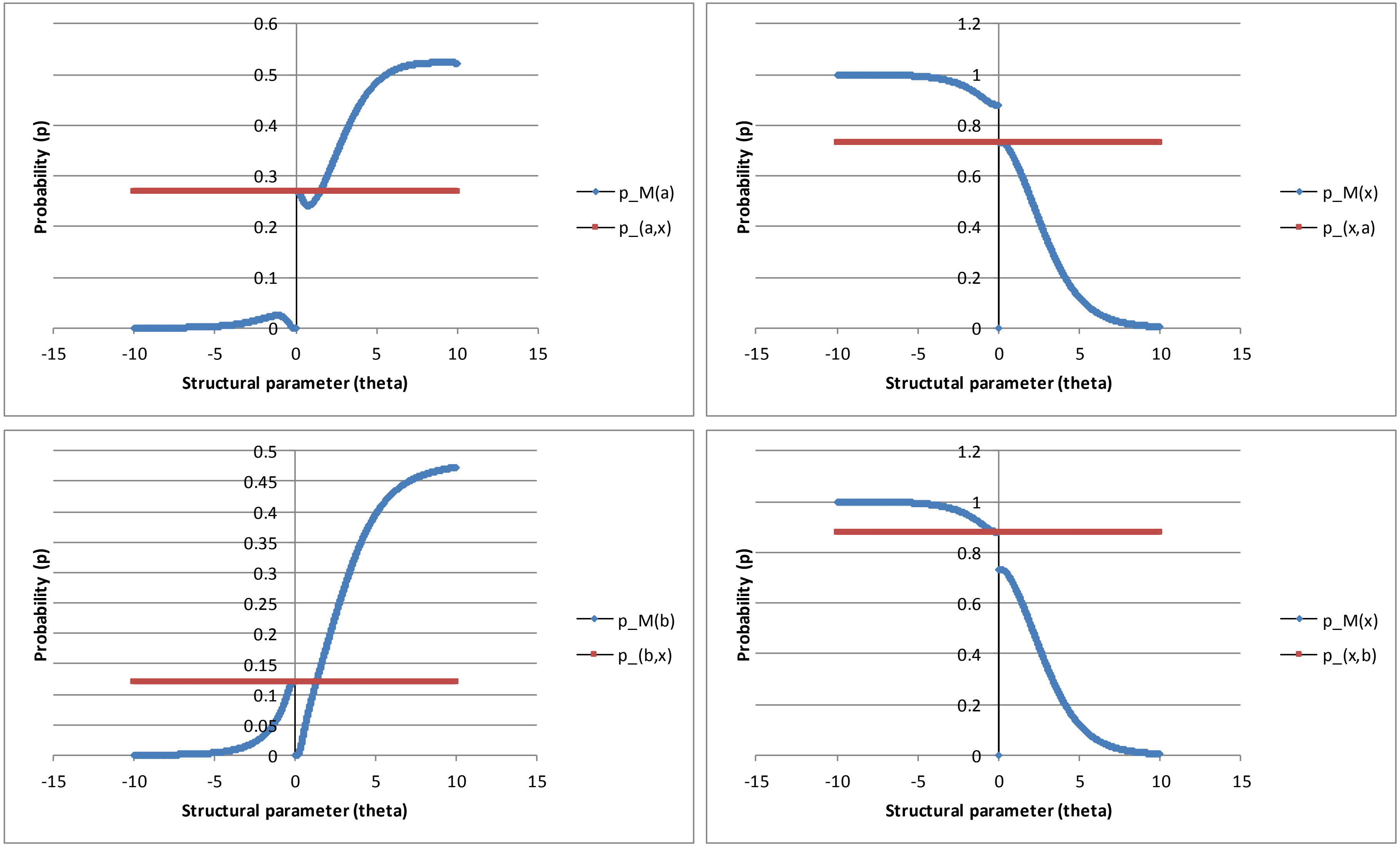

Figure 4: Regularity condition where $V_x > V_a > V_b$

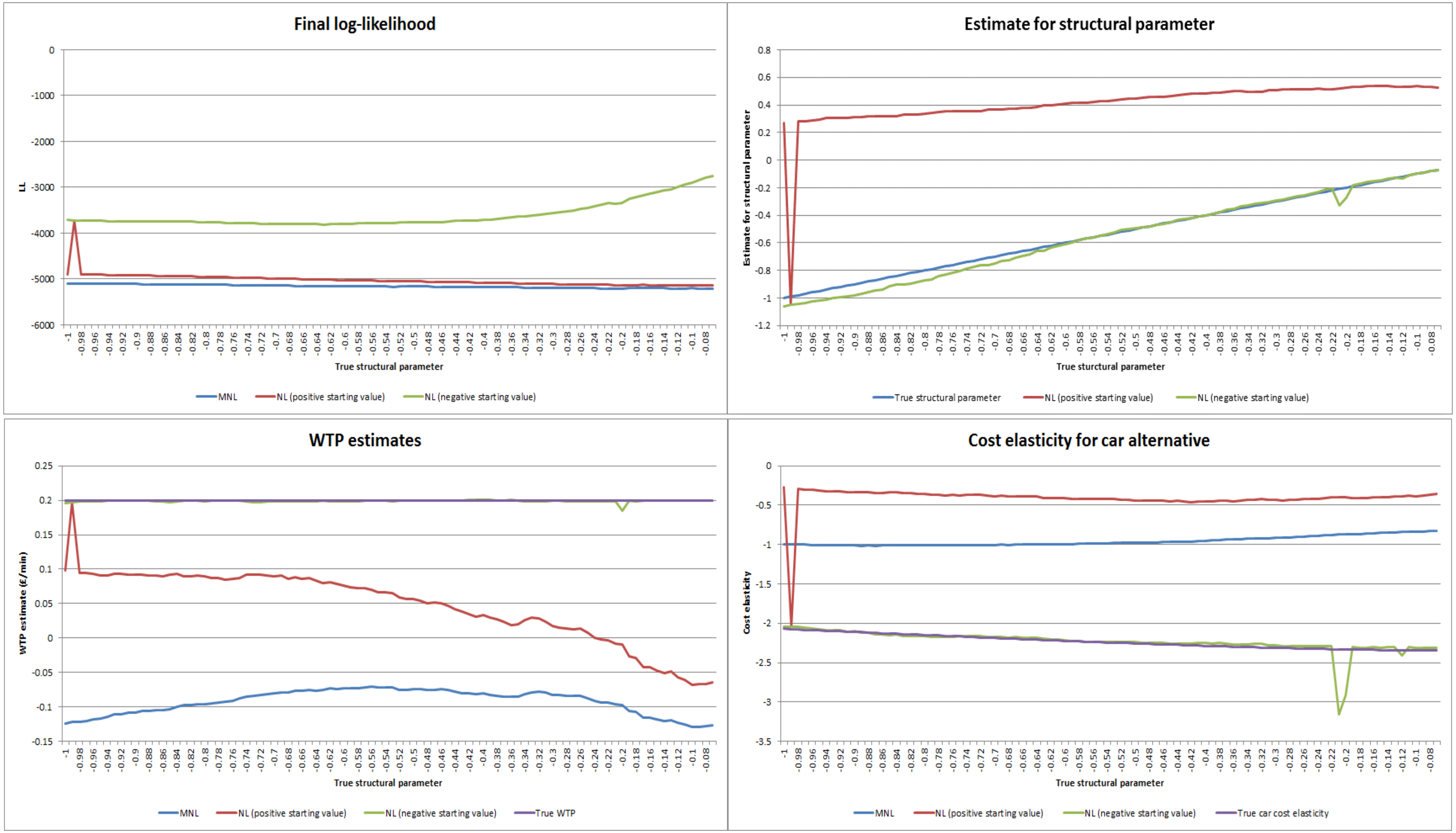

Figure 5: Estimation results on simulated data with negative structural parameter

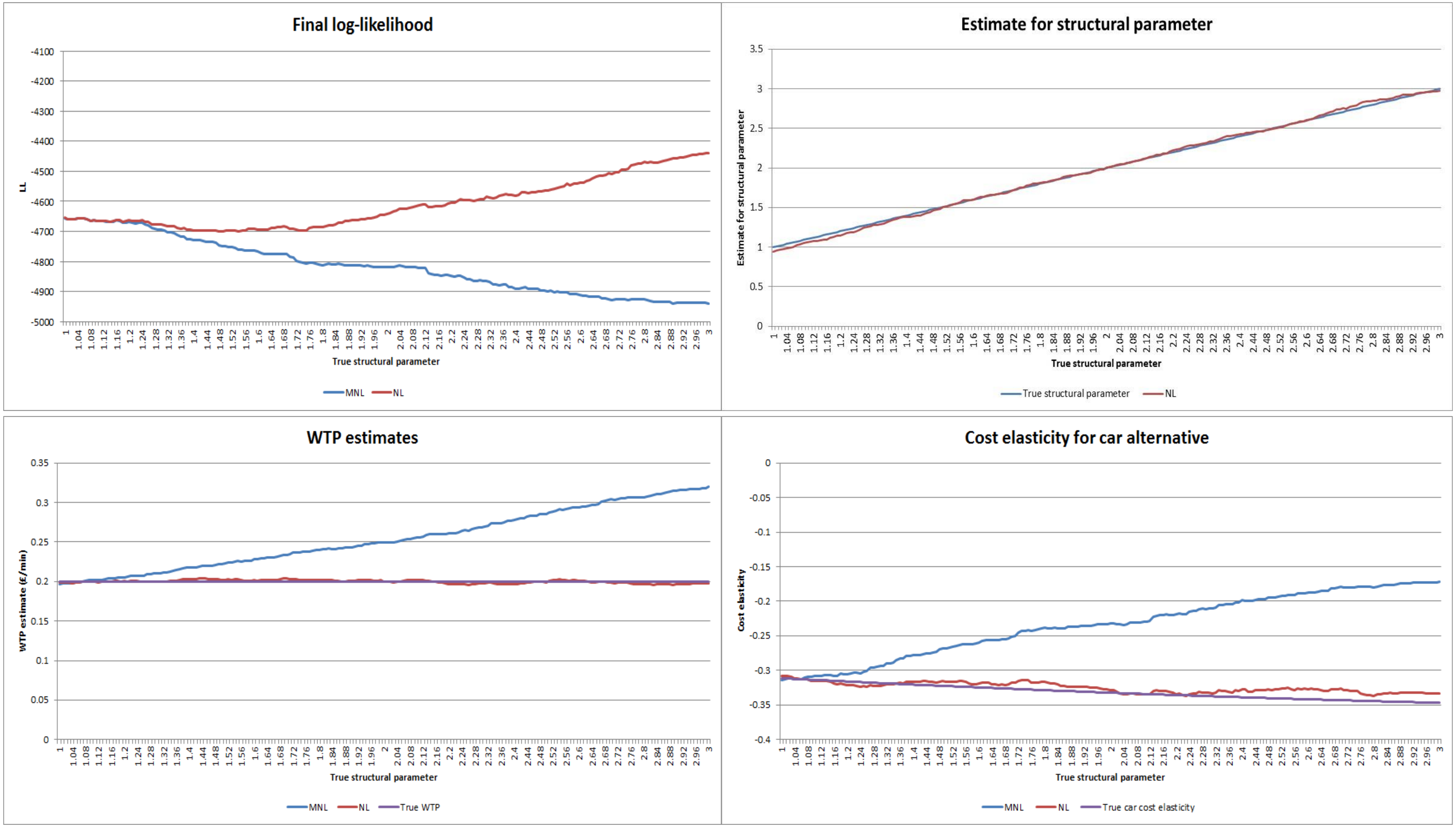

Figure 6: Estimation results on simulated data with positive structural parameter

Annex A: Analysis of regularity under Cases 1 and 2

Case 1: Intra-nest choice

In this case, regularity is satisfied if both:

$p(a,b) \geq p_M(a)$ and $p(b,a) \geq p_M(b)$

where, as defined previously, $L = \{a,b\}$ and $M = \{x,a,b\}$.

As regards the first inequality, we can substitute using (2) and (3):

$$\frac{e^{\frac{V_a}{\theta}}}{e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}} \geq \frac{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]}}{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]} + e^{V_k}} \cdot \frac{e^{\frac{V_a}{\theta}}}{e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}}$$

Simplifying, we find that:

$$1 \geq \frac{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]}}{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]} + e^{V_x}}$$

In principle, since $0 < p_M(L) \leq 1$, regularity holds regardless of the value taken by the structural parameter (with the same finding also applying to the second inequality). However, it should be acknowledged that RUM effectively gives rise to a proper continuous distribution function over the vector of random utilities, where θ embodies the utility scale that generates this distribution function. In practice, therefore, it must hold that $\theta > 0$, so as to support this notion of a distribution function.

Case 2: Inter-nest choice

In this case, regularity is satisfied if:

$p(a,x) \geq p_M(a)$, $p(b,x) \geq p_M(b)$, $p(x,a) \geq p_M(x)$ and $p(x,b) \geq p_M(x)$

For present purposes, it will suffice to consider either of the inter-nest choices; we will therefore focus on the choice between a and x (with the same conceptual issues applying analogously to the choice between b and x). Thus, substituting for the first and third inequalities above, regularity requires that:

$$\frac{e^{V_a}}{e^{V_a} + e^{V_x}} \geq \frac{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]}}{e^{\theta \ln\left[e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}\right]} + e^{V_x}} \cdot \frac{e^{\frac{V_a}{\theta}}}{e^{\frac{V_a}{\theta}} + e^{\frac{V_b}{\theta}}} \qquad \text{(A1)}$$

and:

$$\frac{e^{V_x}}{e^{V_a}+e^{V_x}} \geq \frac{e^{V_x}}{e^{\theta \ln\left[e^{\frac{V_a}{\theta}}+e^{\frac{V_b}{\theta}}\right]}+e^{V_x}} \tag{A2}$$

In contrast to the intra-nest case, compliance of (A1) and (A2) with regularity in the inter-nest case will be dependent on the value taken by the structural parameter. With reference to (2), let us abbreviate the 'log sum' construct $\ln[\cdot]=\ln\left[e^{V_a/\theta}+e^{V_b/\theta}\right]$. The role of the structural parameter within (2) is to control the utility scale of the upper level of the tree structure (i.e. pertaining to the marginal probabilities) relative to the lower level (i.e. pertaining to the conditional probabilities). In this regard, note the values taken by the (scaled) log sum for limiting values of the structural parameter, as θ approaches zero from below (Case 2.1) and above (Case 2.2):

Case 2.1: If $\theta<0$ then $\ln[\cdot]=E(\min(U_a,U_b))$, and $\theta^{-}\ln[\cdot]\rightarrow\min(V_a,V_b)$ as $\theta^{-}\rightarrow 0$.

This implies that (A1) will hold but (A2) will not hold, i.e. regularity is contravened.

Case 2.2: If $\theta>0$ then $\ln[\cdot]=E(\max(U_a,U_b))$, and $\theta^{+}\ln[\cdot]\rightarrow\max(V_a,V_b)$ as $\theta^{+}\rightarrow 0$.

This implies that (A2) will hold, but compliance with (A1) will depend upon the relative magnitudes of the marginal and conditional probabilities[23]. In particular, it is notable that values of the structural parameter in excess of one could be compliant with regularity.

[23] This dependence resonates with Herriges & Kling's (1996) findings reported in section 1.

Annex B: Analysis of stochastic transitivity under the intrinsic preference ordering

Applying the intrinsic preference ordering axb to our earlier definitions of the stochastic transitivity conditions (section 2), SST, MST and WST can be summarised, respectively:

If $p(a,x) \geq 1/2$ and $p(x,b) \geq 1/2$, then $p(a,b) \geq \max\left(p(a,x), p(x,b)\right)$ (B1)

If $p(a,x) \geq 1/2$ and $p(x,b) \geq 1/2$, then $p(a,b) \geq \min\left(p(a,x), p(x,b)\right)$ (B2)

If $p(a,x) \geq 1/2$ and $p(x,b) \geq 1/2$, then $p(a,b) \geq 1/2$ (B3)

where, as defined previously, $L = \{a,b\} \subset M$.

Following Tversky (1972a), it will prove useful to represent each of these conditions as a system of three equations, wherein each equation is defined in terms of odds ratios, as follows:

$$\frac{p(a,x)}{p(x,a)} = (1+u) \quad \text{(B4a)}$$

$$\frac{p(x,b)}{p(b,x)} = (1+v) \quad \text{(B4b)}$$

$$\frac{p(a,b)}{p(b,a)} \geq (1+w) \quad \text{(B4c)}$$

where:

$u, v \geq 0$

$w = \max(u,v)$ in the case of SST

$w = \min(u,v)$ in the case of MST

$w = 0$ in the case of WST

Now drawing reference to the example of three-alternative NL in section 3.2, if the first (B4a) and second (B4b) equations of the system hold, then we can borrow from the earlier statement of the marginal choice probability (2) to derive the identities:

$$\frac{p(a,x)}{p(x,a)} = \frac{e^{v_a}}{e^{v_x}} = (1+u) \quad \text{(B5)}$$

$$\frac{p(x,b)}{p(b,x)} = \frac{e^{v_x}}{e^{v_b}} = (1+v) \quad \text{(B6)}$$

Then combining (B5) and (B6):

$$\frac{p(a,x)}{p(x,a)} \cdot \frac{p(x,b)}{p(b,x)} = \frac{e^{V_a}}{e^{V_b}} = \frac{(1+u)}{(1+v)^{-1}} = (1+u)(1+v) \tag{B7}$$

Now relating (B7) to the conditional probability (3), it must hold that:

$$\frac{e^{\frac{V_a}{\theta}}}{e^{\frac{V_b}{\theta}}} = \left((1+u)(1+v)\right)^{\frac{1}{\theta}} \tag{B8}$$

Substituting for (B8) in the final equation of the system (B4c), we have that:

$$\frac{p(a,b)}{p(b,a)} = \left((1+u)(1+v)\right)^{\frac{1}{\theta}} \geq (1+w) \tag{B9}$$

Whereas the odds ratios for the inter-nest choices (B4a) and (B4b) are independent of the structural parameter, the odds ratio for the intra-nest choice (B9) is dependent on the structural parameter.

Rearranging (B9):

$$(1+u)(1+v) \geq (1+w)^{\theta}$$

$$(1+u)(1+v) \geq \exp\left(\theta \ln(1+w)\right) \tag{B10}$$

Then taking logarithms and rearranging again:

$$\theta \leq \frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} \tag{B11}$$

wherein the limits $1 \leq (1+u) \leq +\infty$ and $1 \leq (1+v) \leq +\infty$ must apply if a is stochastically preferred to x, and x to b.

Though not widely recognised in the literature on NL, it is worth noting that a similar identity to (B11) is reported in section 5.21 of McFadden (1981). Crucially, this generates a different result regarding the 0-1 bounds.

For the case of three-alternative NL, McFadden rationalised the structural parameter in terms of the so-called ‘trinary condition’, a condition which was originally derived by Tversky & Sattath (1979) in the context of the PRETREE model. PRETREE offers an analogy to NL, but is motivated by the behavioural paradigm of elimination-by-aspects rather than RUM; it is important to note that the trinary condition is not necessary for RUM[24].

[24] See Batley & Daly (2006) for a discussion of the correspondence between NL and PRETREE.

CONDITION (t), *Trinary Condition*: If the choice set consists of the trinary $M=\{x,a,b\}\subseteq N$, wherein a and b show some degree of similarity not possessed by x, and $p(a,b)/p(b,a)\geq 1$, then:

$$\frac{p(a,b)}{p(b,a)}\geq\frac{p(a,x)/p(x,a)}{p(b,x)/p(x,b)}\geq 1$$

Using (B5) to (B9), but adjusting (B9) to be an equality rather than an inequality, the trinary condition can be restated:

$$(1+w)\geq(1+u)(1+v)\geq 1 \qquad \text{(B12)}$$

Given this reformulation of (B9), McFadden (1981) followed the steps (B10) and (B11) as before to derive the identity:

$$\theta=\frac{\ln\left((1+u)(1+v)\right)}{\ln(1+w)} \qquad \text{(B13)}$$

The key distinction from (B11) is that (B13) embodies an equality rather than an inequality, and this gives rise to two implications:

1. Whereas (B11) derives an upper (or lower, depending on the actual – rather than intrinsic – preference ordering) bound on the structural parameter, (B13) derives a specific value of the structural parameter.
2. The identity (B12) – which embodies the trinary condition – implies that the structural parameter is constrained to be within the zero-one bounds, whereas (B11) – which embodies the stochastic transitivity condition – does not (in general) impose these specific bounds.

Mindful that neither the trinary condition nor MST/SST are necessary for RUM, the residual question would seem to be whether, in the context of NL, the trinary condition is *overly* restrictive and/or whether MST/SST are *adequately* restrictive. In other words, which condition should define the bounds of the structural parameter?