Quantitative Marketing and Economics, 3, 175–200, 2005.

Generalized Reverse Discrete Choice Models

SANJOG MISRA
William E. Simon School of Business, University of Rochester, Rochester, NY 14627
E-mail: misra@simon.rochester.edu

Abstract. Marketing practitioners and academics have shown a keen interest in the processes that drive consumers' choices since the early work of Guadagni and Little (1982). Over the past decade or so, a number of alternative models have been proposed, implemented and analyzed. The common behavioral assumption that underlines these models of discrete choice is random utility maximization (RUM). The RUM assumption, in its simplest form, posits that a consumer with a finite set of brands to choose from chooses the brand that gives her the maximum amount of utility. An alternative approach would be to assume that consumers choose the alternative that offers them the least disutility. Our paper proposes and tests a broad class of generalized extreme value models based on this hypothesis. We model the decision process of the consumer the assumption random disutility minimization (RDM) and derive a new class of discrete choice models based on this assumption. Our findings reveal that there are significant theoretical and econometric differences between the discrete choice models derived from a RUM framework and the RDM framework proposed in this paper. On the theoretical front we find that the class of discrete choice models based on the assumption of disutility minimization is structurally different from the models in the literature. Further, the models in this class are available in closed form and exhibit the same flexibility as the GEV models proposed by McFadden (1978). In fact, the number of parameters are identical to and have the same interpretation as those obtained via RUM based GEV models. In addition to the theoretical differences we also uncover significant empirical insights. With the computing effort and time for both models being roughly the same this new set of models offers marketing academics and researchers a viable new tool with which to investigate discrete choice behavior.

Key words. discrete choice models, brand choice models, utility maximization, disutility minimization, logit, generalized extreme value, scanner data

JEL Classification: C25, C35, M37, D12

1. Introduction

The empirical investigation of consumer choices and the determinants of such choices has been a topic of interest to researchers in marketing since the early work of Guadagni and Little (1982). Over the past decade or so, a number of alternative models have been proposed, implemented and analyzed. A large proportion of these models are variants derived from the Generalized Extreme Value framework proposed by McFadden (1978). Perhaps the most popular of these is the multinomial logit (MNL) model. The MNL's popularity has been attributed to its analytical simplicity and even though some of its properties, particularly the independence of irrelevant alternatives property (IIA), are considered limiting, it remains the model of choice.[1] In his seminal paper on generalized extreme value (GEV)

1 Pun intended.

models, McFadden (1978) introduced a broad class of models that included, among others, the conditional logit and the nested logit. The GEV choice framework is based on the multivariate GEV distribution which relies on the specification of a dependence or generating function. The choice of this function dictates the nature of the choice model that emerges. In recent years there has been a flurry of research activity that focuses on these dependence functions.[2]

Marketing applications of GEV based choice models are numerous (see e.g., Guadagni and Little, 1983; Bell and Lattin, 1998; Horsky and Nelson, 1992 for applications of the Conditional or Multinomial Logit and Fotheringham, 1988; Dubin, 1986; Kannan and Wright, 1991 for the Nested Logit). In many of these applications the basic GEV models have been modified, and in some cases extended, to investigate specific issues such as choice dynamics (Roy et al., 1996; Erdem and Keane, 1996; Seetharaman, 2003), brand and choice maps (Chintagunta, 1994; Elrod, 1987), taste heterogeneity (Jain et al., 1994; Kamakura and Russell, 1986; Gonul and Srinivasan, 1993; Rossi et al., 1996)), loss aversion (Winer, 1986; Mazumdar and Pappatla, 2000) and more recently structural models of competition (Sudhir, 2001; Kadiyali, et al., 2000).

The common behavioral assumption that underlines these models of discrete choice is that of "random utility maximization" (RUM). This assumption, in its simplest form, posits that a consumer when faced with a finite set of brands to choose from, chooses that brand which gives her the maximum amount of utility. It would seem natural then to assume that an alternative, and equally viable, approach could be constructed in which consumers choose the alternative that offers them the least cost or disutility. Economic theory suggests that the solutions to the consumption problem under either assumption would result in the same solution. This equivalence is formalized in most economics textbooks by treating cost or expenditure minimization as the "dual" of utility maximization (see e.g., Mas-Colell et al., 1995)

There have been a number of marketing studies that focus on a model of consumer behavior that involves the minimization of some form of "random disutility". Gonul and Srinivasan (1996) model the dynamics of consumer shopping behavior based on the premise that consumers minimize their discounted expenditure function. Dubin (1986) models the choice of heating systems and remarks that the deterministic component of utility "... depends principally on life-cycle or total costs." In Bell et al., (1998) a consumer chooses a store that minimizes the sum of fixed and variable costs of shopping. In all these studies the authors derive their theoretical model based on the premise of cost or disutility minimization but do not carry their theory through to an econometric implementation. Instead they negate these disutilities to obtain *pseudo*-utilities and ultimately the standard logit type probabilities. This approach, while not wrong, implicitly assumes that minimization and maximization would result in the same probability structures. In fact, marketing is not alone in making this assumption. Similar examples can be found in studies of transportation issues, political science, geography and economics that translate disutilities and costs into utilities by invoking the textbook notion of duality coupled with the implicit assumption that the maximization and minimization formulations arrive at the same set of discrete choice probabilities. As

2 See Koppelman and Sethi (2000) for a review of various discrete choice models based on the GEV framework.

this paper will show, for a number of the models used in the extant literature, the latter assumption does not always hold.

In a recent paper Anderson and De Palma (1999) showed that for errors distributed i.i.d. extreme value (Type I), the disutility minimization approach (rather than utility maximization) does not yield conditional logit probabilities. Their derivations show the existence of a related, but significantly different, consumer choice framework. They term their model the Reverse Multinomial Logit model and investigate its properties, focusing primarily on its applicability to oligopolistic competition. While Anderson and De Palma (1999) do not undertake an empirical investigation of their framework, they do mention that research in this area might yield interesting insights. Our framework generalizes their approach and we therefore term the class of models obtained Generalized Reverse Discrete Choice Models.

This paper constructs a framework that generates a class of generalized extreme value based discrete choice models where the decision process of the consumer is modeled under the assumption of random disutility minimization (RDM). We show that this class nests the Anderson and De Palma (1999) model and contains a number of other new variants. Our findings reveal that there are significant structural and econometric differences between the discrete choice models derived from the RUM framework of McFadden (1978, 1981) and RDM framework proposed in this paper.

Our analysis shows that the class of discrete choice models based on the assumption of random disutility minimization is structurally very different from the models in the extant literature. Notable among these differences is the fact that the RDM based version of the conditional logit model does not exhibit the IIA property. In spite of these differences, the models in this class are available in closed form and exhibit the same flexibility as the GEV models proposed by McFadden (1978). In fact, the number of parameters are identical to and have the same interpretation as those obtained via RUM based GEV models. In addition to the theoretical analysis we also calibrate the "reverse" analogs of popular discrete choice models on real data. Our findings reveal interesting empirical contrasts between the estimation results from the two approaches. We find that the RDM models exhibit very different elasticity patterns than their RUM counterparts. This difference is most apparent when comparing the RUM and RDM conditional logit estimates. We also find that the RDM based models require similar computing effort and time. In light of our findings, we feel that this new set of models offers marketing academics and researchers a viable new tool with which to investigate discrete choice behavior.

The rest of this paper is organized as follows: In the next section we discuss the notion of a disutility minimizing consumer. We then formalize the RUM and RDM framework and obtain the probability structures for each. We follow this with an application to data and a discussion of the results obtained. We conclude with a discussion and directions for future research.

2. The disutility minimizing consumer

Do consumers always maximize utility? We are certainly not the first to suggest that consumers may frame decisions as a minimization problem. The consumer choice literature in marketing abounds with examples where researchers have framed the consumer's decision framework as one consistent with the minimization of some form of disutility. An early example is the work of White (1978) who examines how consumers choose between three methods of payments (i.e. credit card, check or cash). He argues that the consumer minimizes the total transactions costs incurred in a choice situation and decomposes these costs into fixed, variable and qualitative components. A very similar approach is adopted by Bell et al. (1998) who propose a cost minimizing approach to the store choice problem. In their application a consumer chooses a grocery store based on the expected fixed and variable costs of shopping. (For other applications of the minimization decisions in discrete choice frameworks see Gonul and Srinivasan, 1996; Dubin, 1986; Joskow and Mishkin, 1977; Ho et al., 1998; Hobbs, 1997)

This paper relies on a very broad definition of "disutility" and it is critical that we spell out the nature, scope and implications of what we mean by this term. Before beginning that discussion, however, it is important to note that the "utility" construct used in most marketing applications of discrete choice pertains to the indirect, conditional utility function. The utility function is *conditional* because it is conditioned on other continuous household decisions being made and is *indirect* because it involves the substitution of the budget constraint into the direct utility function. In other words the "utility" function contains both the benefits derived from choosing a given options and the costs that are incurred. Our definition of "disutility" is similar in that it is a broad construct that includes all losses that a consumer might incur by choosing a specified option net of benefits derived. Note that while part of these losses may be strictly monetary (via price) they could also include other components such as time, effort and other psychological costs (e.g., switching costs) that create disutility.

The reader should note that in the discrete choice framework the notions of "Utility Maximization" or "Disutility Minimization" are distinguished only in the eyes of the researcher. Clearly any problem that can be constructed as utility maximization can be recast as disutility minimization. The same is true for a number of theoretical constructs used in understanding consumer behavior. For example, one might think of the utility from being loyal to a brand as the reverse of the disutility or cost of switching. In general, benefits minus costs are thought of as indirect utility while costs net of benefits may be interpreted as disutility. The question whether a given consumer actively frames a decision problem as utility maximization or otherwise is an open and unanswered question. Clearly, the extant literature suggests that there are cases where the minimization frame is more a more appropriate descriptor of the decision scenario. A related issue is if there exist conditions under which both approaches (maximization and minimization) yield the same choice models. We return to this issue in a later section.

The above discussion clearly highlights two important aspects of discrete choice decisions. First, any choice problem can adequately be represented in the form of a disutility minimization problem and second, in certain cases the cost approach may be the appropriate behavioral model. Given these basic ideas we now move to a more detailed analysis of consumer decision making. In the section that follows, we begin with the classical RUM approach to discrete choices and then move to analyzing the probability structures generated by disutility minimization.

3. Stochastic utility and disutility based choice models

3.1. The random utility maximization framework

We start by revisiting the standard random utility models of discrete choice. In this section we present only the barebones of the model and do not belabor the details. Assume that the utility a consumer h gets from choosing option j can be described by the conditional, indirect utility function as follows

$$U_{jh} = u_{jh} + \varepsilon_{jh}. \tag{1}$$

At this point we will assume that the reader understands the implications and assumptions inherent in u_{jh}. That is, the deterministic component u_{jh} includes all those factors that impact the utility that the consumer would obtain upon choosing option j from the set $\mathcal{J}$. The consumer then chooses the alternative that offers the maximum utility. This allows us to depict the choice probability for alternative i as

$$P_{ih} = \Pr(U_{ih} \geq U_{jh} : j \in \mathcal{J}), \tag{2}$$

or

$$P_{ih} = \Pr(u_{ih} + \varepsilon_{ih} \geq u_{jh} + \varepsilon_{jh} : j \in \mathcal{J}). \tag{3}$$

This expression, for the first alternative, can be represented (suppressing h)

$$P_1^u = \int_{-\infty}^{\infty} \int_{-\infty}^{u_1+\varepsilon_1-u_2} \cdots \int_{-\infty}^{u_1+\varepsilon_1-u_J} F_{1,2,\ldots,n}\, d\varepsilon_{n\ldots} d\varepsilon_2 d\varepsilon_1. \tag{4}$$

The rest of the probabilities can be computed in a similar manner. McFadden (1978) showed that if we construct a generating function $\mathcal{G}(y_1, y_2 \ldots, y_n)$ that is defined on the orthant $y_k \geq 0$, is a non-negative function that is homogenous of degree one,[3] that tends to $+\infty$ for any element approaching $+\infty$ and whose kth cross partial derivatives are nonnegative when k is odd and nonpositive when k is even, then

$$\mathcal{F} = \exp(-\mathcal{G}(y_1, \ldots, y_n)), \tag{5}$$

is a valid multivariate distribution function. In particular, the ε are said to jointly follow a multivariate generalized extreme value (GEV) with generating function $\mathcal{G}$. McFadden (1978) also showed that under this distribution we can derive the choice probabilities in closed form. For the sake of completeness and consistency we state his result in the form of a lemma below (the h subscript is suppressed).

3 Ben-Akiva and Francois (1983) show that the generating function can be homogenous of degree $\mu > 0$. In this paper we assume $\mu = 1$ to retain consistency with the models proposed by McFadden (1978, 1981). Generalizing the analysis to incorporate arbitraty values of μ is straightforward.

Lemma 1. *Let $\mathcal{G}(y_1, y_2, \ldots, y_n)$ denote the generating function of a Generalized Extreme Value distribution and let $\mathcal{G} = \frac{\partial \mathcal{G}}{\partial y_i}$. If $U_i = u_i + \varepsilon_i$ with ε distributed GEV then the probability structure consistent with random utility maximization can be depicted by*

$$P_i^u = \frac{e^{u_i}\mathcal{G}_i(e^{u_1}, e^{u_2}, \ldots, e^{u_n})}{\mathcal{G}(e^{u_1}, e^{u_2}, \ldots, e^{u_n})} \tag{6}$$

Proof: See McFadden (1978) □

The properties of the choice structure in Lemma 1 are well known and we do not repeat them here. Details are available in McFadden (1978, 1981), Ben-Akiva and Lerman (1985) and Anderson et al. (1992). Of course, the GEV is not the only multivariate distribution function used in choice models and studies using the multivariate normal (resulting in the Multinomial Probit) abound (See e.g., McCulloch and Rossi, 1994; Chintagunta, 1992.) In those cases the choice expressions can be obtained by appropriate substitutions into (4).

3.2. A random disutility based discrete choice framework

We now present the disutility minimizing approach to the consumer's problem. To begin we need to specify the minimization program faced by the consumer. The consumer's disutility can simply be stated as

$$D_{jh} = \xi(u_{jh}) + \varepsilon_{jh} \tag{7}$$

where $\frac{\partial \xi}{\partial u} < 0$. For the rest of this study we will assume that $\xi(u_{jh}) = -u_{jh} = d_{jh}$. The random disutility assumption then implies that the consumer chooses that option which burdens her with the lowest disutility. In other words the probability of household j choosing option i can be depicted as

$$P_{ih} = \Pr(D_{ih} \leq D_{ih} : j \in \mathcal{J}) \tag{8}$$

Let F_i be the derivative of F (the multivariate distribution function of ε) with respect to the ith component, F_{ij} the derivative with respect to i and j, and so on. Consequently $F_{1,2,\ldots,n}$ denotes the derivative of F with respect to $\{1, \ldots, n\}$. The probability of choosing the first alternative can then be expressed as

$$P_1^d = \int_{-\infty}^{\infty} \int_{d_1+\varepsilon_1-d_2}^{\infty} \cdots \int_{d_1+\varepsilon_1-d_J}^{\infty} F_{1,2,\ldots,n}\, d\varepsilon_{n\ldots} d\varepsilon_2 d\varepsilon_1. \tag{9}$$

In order to explore this expression further we will need some additional notation. These are contained in the definitions listed below.

D1: The set $\mathcal{M}_i$ denotes those subsets of the power set $\mathcal{P}(\mathcal{J})$ of $\mathcal{J} = \{j : j = 1, 2, \ldots, n\}$ that include the index i and have cardinality greater than one. Also let $\mathcal{M}_{\mathcal{J}} = \bigcup_{j \in \mathcal{J}} \mathcal{M}_j$.

D2: $\Psi_{\mathbf{m}}$ is the set of disutilities (d) corresponding to the indices contained in $m \in \mathcal{M}_i$, and let $\Psi^e_{\mathbf{m}}$ consist of exponentiated elements of $\Psi_{\mathbf{m}}$.
D3: $|m|$ is the cardinality of the set $m \in \mathcal{M}_i$.
D4: $\Omega_i(\Psi_{\mathbf{m}})$ is the probability of i^{th} disutility being the maximum of the disutilities defined by the elements of $\Psi_{\mathbf{m}}$; i.e. $\Omega_i(\Psi_{\mathbf{m}}) = \Pr(d_i \geq d_k : d_k \in \Psi_{\mathbf{m}})$.

We clarify the above definitions with a simple example: Imagine an index set $\mathcal{J} = \{j : j = 1 \dots 3\}$. The power set of $\mathcal{J}$ is denoted by

$$\mathcal{P}(\mathcal{J}) = \{(\emptyset), (1), (2), (3), (1, 2), (1, 3), (2, 3), (1, 2, 3)\}.$$

Now lets say we are interested in the index $i = 1$, then

$$\mathcal{M}_1 = \{(1, 2), (1, 3), (1, 2, 3)\},$$

and $m = (1, 2, 3)$ is therefore one particular subset of $\mathcal{M}_1$. Also note that in this case $|m| = 3$ and $\Psi_{\mathbf{m}} = \{d_1, d_2, d_3\}$. This allows us to express the following

$$\Omega_1(\Psi_{\mathbf{m}}) = P(d_1 \geq d_2, d_1 \geq d_3). \tag{10}$$

Using the notation described above the RDM discrete choice probabilities can be expressed fairly succinctly. We characterize these probabilities in the shape of the theorem 1.

Theorem 1. *Define $D_i = d_i + \varepsilon_i$, with d_i representing the deterministic component of disutility, then for ε distributed with multivariate distribution function $\mathcal{F}(\varepsilon)$ the values*

$$P^d_i = 1 - \sum_{m \in \mathcal{M}_1} (-1)^{|m|} \Omega_i(\Psi_m) \tag{11}$$

define a probabilistic choice model from alternatives $i \in \mathcal{J}$ which is consistent with the minimization of random disutilities.

Proof: See Appendix □

Note that the expressions in Theorem 1 are general and not dependent on any particular distribution function. If we assume the multivariate GEV distribution then explicitly solving for these probabilities gives us the generalized reverse discrete choice framework. We formalize this with the following lemma.

Lemma 2. *Let $\mathcal{G}(y_1, y_2, \dots, y_n)$ denote the generating function of a Generalized Extreme Value distribution and let $\mathcal{G}_i = \frac{\partial \mathcal{G}}{\partial y_i}$. If $D_i = d_i + \varepsilon_i$ with ε distributed GEV then the probability structure consistent with random disutility minimization can be depicted by*

$$P^d_i = 1 - \sum_{m \in \mathcal{M}_i} (-1)^{|m|} \left[\frac{e^{d_i} \mathcal{G}_i(\Psi_m)}{\mathcal{G}(\Psi_m)} \right] \tag{12}$$

Proof: See Appendix □

3.3. A behavioral interpretation

One quickly notices that even though the probabilities in (12) contain GEV type structures they are far more involved than the conventional GEV choice models. An expansion of (12) (as in equation A4 of the Appendix) reveals a similarity to the expansion of union set probabilities (see e.g., Theorem 2, DeGroot p.g. 39.) To understand the behavioral intuition behind this model notice that the probability of choosing option 1 out of four alternatives can be thought of as

$$\begin{aligned}
&1 - \Pr(D_1 > D_2) - \Pr(D_1 > D_3) - \Pr(D_1 > D_4) \\
&\quad + \Pr(D_1 > D_2, D_1 > D_3) + \Pr(D_1 > D_2, D_1 > D_4) \\
&\quad + \Pr(D_1 > D_3, D_1 > D_4) - \Pr(D_1 > D_2, D_1 > D_3, D_1 > D_4).
\end{aligned} \tag{13}$$

In other words, the random disutility choice probability of choosing option j can be thought of being one minus the probability of d_j being larger that all $d_{k \neq j}$. In (13), the first row simply reflects this aspect of the intuition. The second and third row simply account for any double counting that might occur. More generally, this approach arrives at the choice probabilities by simply deducting from unity the probability of being greater (in pair-wise comparisons) than other disutilities, then adding back the probability of being the maximum disutility in all triples and so on until one gets to the m-tuples. The consumer could therefore be thought of making comparisons of the focal alternative with other alternatives to ascertain if it is the minimum disutility alternative. The structure of these probabilities is very different from the RUM choice probabilities proposed by McFadden (1978). Even though it is not completely "precise" we feel compelled to point out the parallels between the RUM-RDM relation and the behavioral differences between "selecting" alternatives and "not-rejecting" them. In the RUM case the consumer evaluates the alternative and selects the best one. The RDM case is more in line with the idea that a consumer compares the options and rejects all but one. We would like to make it clear that this parallel is simply coincidental. While there might be many behavioral interpretations of the models described by Lemma 1 the reader should note that these interpretations are not based any structural or theoretical arguments. They are, however, useful in helping us understand the differences in the two approaches to discrete choice models. In the following section we derive and contrast popular RUM models and their RDM counterparts.

3.4. Illustrating the RDM and RUM GEV frameworks

Before presenting specific models, we present the general GEV probabilities under the RUM and RDM assumptions. In what follows, we use a three alternative example since it is the simplest specification in which to examine the structural differences between the two frameworks. Let the generating functions be as follows:

$$\mathcal{G}^u(y) = \mathcal{G}^u(y_1, y_2, y_3) = \mathcal{G}(e^{X_1'\beta}, e^{X_2'\beta}, e^{X_2'\beta}) \tag{14}$$

$$\mathcal{G}^d(\tilde{y}) = \mathcal{G}^d(\tilde{y}_1, \tilde{y}_2, \tilde{y}_3) = \mathcal{G}(e^{X_1'\delta}, e^{X_2'\delta}, e^{X_2'\delta}). \tag{15}$$

In the above, $\mathcal{G}^u(y)$ and $\mathcal{G}^d(\tilde{y})$ are the generating functions for the RUM and RDM specifications respectively, β and δ are parameters. We are implicitly assuming here that the utilities (and disutilities) are linear and that $d_i = \xi(u_i) = -u_i$. Since we have specified $u_i = X_i'\beta$, we can write $d_i = X_i'\delta$.[4]

Applying Lemma 1 we see that the RUM GEV probabilities are of the form:

$$P_1^u = \frac{e^{X_1'\beta}\mathcal{G}_1(e^{X_1'\beta}, e^{X_2'\beta}, e^{X_2'\beta})}{\mathcal{G}(e^{X_1'\beta}, e^{X_2'\beta}, e^{X_2'\beta})}. \tag{16}$$

$$P_2^u = \frac{e^{X_2'\beta}\mathcal{G}_2(e^{X_1'\beta}, e^{X_2'\beta}, e^{X_2'\beta})}{\mathcal{G}(e^{X_1'\beta}, e^{X_2'\beta}, e^{X_2'\beta})}. \tag{17}$$

P_3^u can be derived as $1 - P_1^u - P_2^u$. In the sequel we do not present the probability for the third alternative.

Similarly, upon applying Lemma 2 we obtain the RDM GEV probabilities,

$$P_1^d = 1 - \frac{e^{X_1'\delta}\mathcal{G}_1(e^{-X_1'\delta}, e^{X_2'\delta})}{\mathcal{G}(e^{X_1'^{\delta}}, e^{X_2'\delta})} - \frac{e^{X_1'\delta}\mathcal{G}_1(e^{X_1'\delta}, e^{X_3'\delta})}{\mathcal{G}(e^{X_1'\delta}, e^{X_3'\delta})} + \frac{e^{X_1'\delta}\mathcal{G}_1(e^{X_1'\delta}, e^{X_2'\delta}, e^{X_2'\delta})}{\mathcal{G}(e^{X_1'\delta}, e^{X_2'\delta}, e^{X_2'\delta})}. \tag{18}$$

$$P_2^d = 1 - \frac{e^{x_2'\delta}\mathcal{G}_2(e^{x_1'\delta}, e^{x_2'\delta})}{\mathcal{G}(e_1^{x'\delta}, e^{x_1'\delta})} - \frac{e^{x_2'\delta}\mathcal{G}_2(e^{x_2'\delta}, e^{-x_3'\delta})}{\mathcal{G}(e^{x_2'\delta}, e^{x_3'\delta})} + \frac{e^{x_2'\delta}\mathcal{G}_2(e^{x_1'\delta}, e^{x_2'\delta}, e^{x_2'\delta})}{\mathcal{G}(e^{x_1'\delta}, e^{x_2'\delta}e^{x_2'\delta})}. \tag{19}$$

The reader will note that the RDM probabilities are more complicated since they involve more terms in each expression. This complication, unfortunately, only increases with the addition of more alternatives. The specification does have its advantages which will become clear in later sections.

While the above GEV probabilities provide a general idea of the nature of the probability structure they do not offer the exact form of the models that might be implemented. We therefore move to a discussion of specific models.

4 Note that the signs of the estimated $\beta's$ and $\delta's$ will be the opposite of each other since RUM treats the $X's$ as utility enhancing while RDM treats it as cost enhancing. In the tables presented in the empirical application we reverse the signs of δ to facilitate comparison.

3.5. The random disutility multinomial logit model (RDMNL)

The first model we present uses the simplest and probably the most popular generating function in the GEV family. This function is described as

$$\mathcal{G}(y) = \sum_{i=1}^{n} y_i. \tag{20}$$

As earlier, in order to compare the properties of the two specifications we will use a simple three alternative choice framework. In particular assume that $\mathcal{G}(y_1, y_2, y_3) = y_1 + y_2 + y_3$. Then for the RUM framework, using Lemma 2, we obtain the familiar multinomial logit specification. Letting, $y_i = e^{u_i} = e^{X_i'\beta}$ for $i = 1 \ldots 3$ we can write this as,

$$P_1^u = \frac{e^{X_1'\beta}}{e^{X_1'\beta} + e^{X_2'\beta} + e^{X_3'\beta}} \tag{21}$$

$$P_2^u = \frac{e^{X_2'\beta}}{e^{X_1'\beta} + e^{X_2'\beta} + e^{X_3'\beta}}. \tag{22}$$

As mentioned before P_3^u can be derived as $1 - P_1^u - P_2^u$.

The derivation of the corresponding RDM probabilities are straightforward. Using Lemma 2 we obtain the RDM choice probabilities and can show that

$$P_1^d = 1 - \frac{e^{X_1'\delta}}{e^{X_1'\delta} + e^{X_2'\delta}} - \frac{e^{X_1'\delta}}{e^{X_1'\delta} + e^{X_3'\delta}} + \frac{e^{X_1'\delta}}{e^{X_1'\delta} + e^{X_2'\delta} + e^{X_3'\delta}} \tag{23}$$

$$P_2^d = 1 - \frac{e^{X_2'\delta}}{e^{X_1'\delta} + e^{X_2'\delta}} - \frac{e^{X_2'\delta}}{e^{X_1'\delta} + e^{X_3'\delta}} + \frac{e^{X_2'\delta}}{e^{X_1'\delta} + e^{X_2'\delta} + e^{X_3'\delta}}. \tag{24}$$

Collecting the terms and simplifying yields the following expressions,

$$P_1^d = \frac{(e^{X_2'\delta + X_3'\delta})(2e^{X_1'\delta} + e^{X_2'\delta} + e^{X_3'\delta})}{(e^{X_1'\delta} + e^{X_2'\delta})(e^{X_1'\delta} + e^{X_3'\delta})(e^{X_1'\delta} + e^{X_2'\delta} + e^{X_3'\delta})} \tag{25}$$

$$P_2^d = \frac{(e^{X_1'\delta + X_3'\delta})(e^{X_1'\delta} + 2e^{X_2'\delta} + e^{X_3'\delta})}{(e^{X_1'\delta} + e^{X_2'\delta})(e^{X_2'\delta} + e^{X_3'\delta})(e^{X_1'\delta} + e^{X_2'\delta} + e^{X_3'\delta})}. \tag{26}$$

We term this model the Random Disutility Multinomial Logit (RDMNL). The model presented in (23) is identical to the reverse discrete choice model proposed by Anderson and De Palma (1999). This is because the additive generating function described by (20) yields a joint density that is a product of i.i.d. extreme value variables. Thus our GEV framework, as expected, nests their model. The RDMNL choice probabilities have some interesting properties which we discuss next.

A key feature of the RDMNL is that it does not exhibit the Independence from Irrelevant Alternatives (IIA) property. This is apparent from the probability structure denoted by (23). In particular notice that $\frac{P_1^d}{P_2^d}$ is not independent of d_3. This allows one to use the RDMNL as a possible alternative specification and test for IIA. It should also be pointed out that the set of parameters one obtains, via estimation, from either model are equivalent in their information content. In other words, since we have defined $u_i = X_1'\beta$ and correspondingly $d_i = X_i'\beta$ it is always true that dim $(\beta) = \dim(\delta)$, i.e. there are an identical number of parameters. Further the impact that particular elements of δ and β have on the choice probabilities are of comparable magnitude (while opposite in sign). While it might turn out that $\hat{\delta} \neq -\hat{\beta}$, theory suggests that their impacts should be similar. This equivalence, as will be apparent later, is true for all pairs of matching RUM/RDM models in our analysis.

3.6. The Random Disutility Nested Logit (RDNL)

We now turn our attention to another popular generating function. This generating function yields the nested logit model, which is often seen as a simpler alternative to the econometrically cumbersome multinomial probit model. The generating function is assumed to be of the form

$$\mathcal{G}(y) = \sum_{k=1}^{K} \left(\sum_{i \in \mathcal{C}_k} y_i^{\frac{1}{1-\sigma_k}} \right)^{1-\sigma_k}, \tag{27}$$

where $\mathcal{C}_k$ reflects the choice set corresponding to nest k. In order for the function to be compatible with utility maximization we need $0 \leq \sigma_k \leq 1$[5] Again, for purposes of exposition we will deal with a three alternative example with alternatives 2 and 3 forming a nest. Assuming

$$\mathcal{G}(y_1, y_2, y_3) = \left(y_1 + \left[y_2^{\frac{1}{1-\sigma}} + y_3^{\frac{1}{1-\sigma}} \right]^{1-\sigma} \right), \tag{28}$$

and invoking Lemma 1, we have (using $y_i = e^{u_i} = e^{X_i'\beta}$ as before)

$$P_1^u = \frac{e^{X_1'\beta}}{e^{X_1'\beta} + \left(e^{\frac{X_2'\beta}{1-\sigma}} + e^{\frac{X_2'\beta}{1-\sigma}} \right)^{1-\sigma}} \tag{29}$$

$$P_2^u = \frac{e^{\frac{X_2'\beta}{1-\sigma}} \left(e^{\frac{X_2'\beta}{1-\sigma}} + e^{\frac{X_3'\beta}{1-\sigma}} \right)^{-\sigma}}{e^{X_1'\beta} + \left(e^{\frac{X_2'\beta}{1-\sigma}} + e^{\frac{X_3'\beta}{1-\sigma}} \right)^{1-\sigma}}. \tag{30}$$

A similar exercise using Lemma 2 yields the Random Disutility Nested Logit (RDNL)

5 Note that the restriction on σ_k follows from the requirements for a valid GEV generating function (for details see Ben-Akiva and Lerman, 1985). Hence it follows that the restriction required for RUM is also necessary and sufficient for RDM.

choice probabilities,

$$P_1^d = 1 - \frac{e^{X_1'\delta}}{e^{X_1'\delta} + e^{X_2'\delta}} - \frac{e^{X_1'\delta}}{e^{X_1'\delta} + e^{X_3'\delta}} + \frac{e^{X_1'\delta}}{e^{X_1'\delta} + \left(e^{\frac{X_2'\delta}{1-\sigma}} + e^{\frac{X_3'\delta}{1-\sigma}}\right)^{1-\sigma}} \tag{31}$$

$$P_2^d = 1 - \frac{e^{X_2'\delta}}{e^{X_1'\delta} + e^{X_2'\delta}} - \frac{e^{\frac{X_2'\delta}{1-\sigma}}}{e^{\frac{X_2'\delta}{1-\sigma}} + e^{\frac{X_3'\delta}{1-\sigma}}} + \frac{e^{X_2'\delta}\left(e^{\frac{X_2'\delta}{1-\sigma}} + e^{\frac{X_3'\delta}{1-\sigma}}\right)^{-\sigma}}{e^{X_1'\delta} + \left(e^{\frac{X_2'\delta}{1-\sigma}} + e^{\frac{X_3'\delta}{1-\sigma}}\right)^{1-\sigma}}. \tag{32}$$

Notice that as $\sigma \to 0$ the choice probabilities coincide with those in the earlier example (23). Also we would like to point out that the term "nested" may or may not apply to the RDNL since one might not be able to decompose the probabilities as one can in standard NL model. However, the interpretation and relevance of the parameters are similar in both models.

3.7. Other variants of random disutility GEV models

The RDMNL and the RDNL simply illustrate two of the many structures that are included in the RDM-GEV model space. It should be apparent that any GEV model obtained in the RUM realm has a corresponding reverse analog in the form of a RDM model. For example we can obtain reverse discrete choice versions of the Paired Combinatorial Logit (Wen and Koppelman, 1999), The Cross-Nested Logit (Vovsha, 1997) and the Generalized Nested Logit (Koppelman and Wen, 2001; Swait, 2000).

In a recent paper, Swait(2003) proposes an interesting extension of the GEV framework. He shows out that an additive function of GEV generating functions is a valid generating function itself. Coupled with the recent work of Karlstrom (2000) this vastly enriches the gamut of GEV models from which the researcher can choose an appropriate generating function. The fact that each of these generating functions has both a disutility and utility based discrete choice model, that are distinct in form and structure from each other, greatly expands the option set available to the applied marketing researcher.

4. When does utility maximization = disutility minimization?

Before concluding our theoretical analysis of reverse discrete choice models we revisit an issue raised earlier in the paper. When do the random utility approach and the random disutility approach yield the same probability structure? For the probability structures under disutility minimization and utility maximization to be identical we need each corresponding probability to be equal. In other words we need $P_i^u = P_i^d$ for all $i \in \mathcal{J}$. Assume, as earlier, that utilities are simply negated disutilities (or vice versa). That is we can write $(u_i = -d_i : i \in \mathcal{J})$, then it has to be that the answer lies in the structure of the density that underlies the random components. Clearly if we could replace each ε_i with $-\varepsilon_i$ we would have the two probability structures being identical. This in turn implies that the densities $f(\varepsilon)$ and $f(-\varepsilon)$ have be identical or in other words the error density has to satisfy a weak form of symmetry.

A general class of densities that satisfy this property is the set of spherically symmetric distributions (SSD). This family of multivariate densities can be defined using a stochastic representation (Fang et al., 1990)

$$\mathbf{x} \overset{d}{=} r \cdot \mathbf{v}, \tag{33}$$

where $\mathbf{v}$ is a uniform random vector distributed on the unit hypersphere, r is a positive random variable independent of $\mathbf{v}$ and "$\overset{d}{=}$" signifies equivalence in distribution. This class includes most well known distributions such as the Multivariate Normal, Multivariate t and the Multivariate Cauchy. We now state the following theorem.

Theorem 2. *If* $\{u_i = d_i : i \in J\}$ *and if* $\varepsilon \overset{d}{=} (r \cdot \mathbf{v})$ *then* $P_i^d = P_i^u$ *for all* $i \in J$.

Proof: See Appendix. □

Since the multivariate normal density falls with the *SSD* class, all varieties of the Multinomial Probit will result in identical choice structures for RUM and RDM. While Anderson and de Palma (1999) recognized this equivalence for the I.I.D. multivariate normal assumption the above theorem formally extends it to all spherically symmetric distributions. This theorem also indirectly shows that the RDM and RUM frameworks arise primarily on account of the skewness of the Extreme Value density. For a more detailed discussion of this issue see Anderson and De Palma (1999).

5. Application to brand choice models

5.1. *Data*

In this section we focus on the empirical implementation of some of the models described earlier. We estimate and compare the standard multinomial logit and nested logit to their random disutility versions using datasets that are publicly available and have been used in earlier studies. In particular we use the yogurt and cracker datasets used by Jain and Chintagunta (1994). Summary statistics of the datasets are reported in Table 1.

5.2. *Model specification and methodology*

Since the primary purpose of this section is the empirical comparison of the two choice paradigms we estimate simple but popular implementations of the models seen in the literature and their random disutility counterparts. While we refrain from using more complicated models since our choice of specification might confound the comparison, future research might look at more intricate specifications. In what follows we present the results from four particular models. These models are described below.

Model	Mnemonic	Description
i.	**MNL**	Multinomial Logit
ii.	**RDMNL**	Random Disutility Multinomial Logit
iii.	**HMNL**	Heterogeneous Multinomial Logit
iv.	**HRDMNL**	Heterogeneous Random Disutility Multinomial Logit
v.	**NL**	Nested Logit
vi.	**RDNL**	Random Disutility Nested Logit

For all homogeneous parameter models the random utility and disutility specifications are as follows:

$$\begin{aligned} u_{hjt} &= \beta_{0j} + \beta_1 PRICE_{jt} + \beta_2 FEAT_{jt} + \varepsilon^u_{hjt} \\ d_{hjt} &= \delta_{0j} + \delta_1 PRICE_{jt} + \delta_2 FEAT_{jt} + \varepsilon^d_{hjt}\,. \end{aligned} \tag{34}$$

The models are of the form illustrated in (21) and (23). The usual derivation allows us to obtain the relevant likelihood functions to be maximized,

$$\mathcal{L} = \prod_{h=1}^{H}\prod_{t=1}^{T}\prod_{j=1}^{J} P^*_{hjt}(\vartheta)^{\zeta_{hjt}}\,. \tag{35}$$

Table 1. Summary statistics for Cracker and Yogurt data

Cracker data				Yogurt data			
Variable	Brand	Mean	Std. dev.	Variable	Brand	Mean	Std. Dev.
Market Share (proportion)	Sunshine	0.07260	0.25952	*Market Share* (proportion)	Yoplait	0.33914	0.47351
	Keebler	0.06865	0.25290		Dannon	0.40216	0.49043
	Nabisco	0.54435	0.49810		Weight	0.22927	0.42045
	Pvt. label	0.31440	0.46435		Hiland	0.02944	0.16906
Feature (proportion)	Sunshine	0.03767	0.19042	*Feature* (proportion)	Yoplait	0.05597	0.22991
	Keebler	0.04253	0.20182		Dannon	0.03773	0.19058
	Nabisco	0.08657	0.28125		Weight	0.03773	0.19058
	Pvt. label	0.04708	0.21185		Hiland	0.03690	0.18855
Price ($ per unit)	Sunshine	0.95703	0.13292	*Price* ($ per Oz)	Yoplait	0.10682	0.01906
	Keebler	1.12594	0.10638		Dannon	0.08163	0.01063
	Nabisco	1.07923	0.14478		Weight	0.07949	0.00774
	Pvt. label	0.68073	0.12407		Hiland	0.05363	0.00805
Observations	3292			*Observations*	2412		
households	136			*households*	100		

Here $P^* = P^u$ or P^d for the RUM and RDM models respectively, ϑ is the vector of relevant parameters and ζ_{hjt} is a choice indicator equalling one if household h chooses brand j at time t and zero otherwise. Maximizing $\mathcal{L}$ gives us estimates of the parameters of interest.

Incorporating a continuous or discrete mixing distribution, to represent heterogeneity in the parameters, is as straightforward in the RDMNL as in the standard MNL. In our analysis we restrict ourselves to incorporating heterogeneity via a continuous normal mixing density for the utility intercepts and price. Thus the utility/disutility functions are

$$\begin{aligned} u_{hjt} &= \beta_{0jh} + \beta_{1h} PRICE_{jt} + \beta_2 FEAT_{jt} + \varepsilon^u_{hjt} \\ d_{hjt} &= \delta_{0jh} + \delta_{1h} PRICE_{jt} + \delta_2 FEAT_{jt} + \varepsilon^d_{hjt} . \end{aligned} \tag{36}$$

We assume that $\beta_{0h} \overset{d}{=} \phi(\beta_0, \Sigma_{\beta 0})$ and $\beta_{1h} \overset{d}{=} \phi(\beta_1, \sigma^2_{\beta 1})$ and similarly that $\delta_{0h} \overset{d}{=} \phi(\delta_0, \Sigma_{\delta_0})$ and $\delta_{1h} \overset{d}{=} \phi(\delta_1, \sigma^2_{\delta_1})$, with ϕ denoting the normal distribution. Now, let ϑ denote the relevant (utility or disutility) parameters and Θ its corresponding 'deep' parameters (i.e. the heterogeneity mean and variance terms). This allows us to write the likelihood for the HMNL/HRDMNL as

$$\mathcal{L} = \prod_{h=1}^{H} \int_{\vartheta} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[P^*_{hjt}(\vartheta) \right]^{\zeta_{hjt}} \phi(\vartheta | \Theta) \, d\vartheta . \tag{37}$$

We use a quasi Monte-Carlo (QMC) simulation approach to approximate the integrals in the likelihood and depict the approximated log-likelihood as

$$\widehat{\ln \mathcal{L}} = \sum_{h=1}^{H} \ln \left[\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \prod_{j=1}^{J} \left(P^*_{hjt}(\vartheta^{(r)}, \Xi) \right)^{\zeta_{hjt}} \right], \tag{38}$$

where $\vartheta^{(r)}$ are draws from the relevant distributions and Ξ denote the rest of the parameters. In our implementation of QMC approximation we use randomized, scrambled Halton sequences (Bhat, 2001, 2003) rather that pseudo random numbers. Using such low discrepancy sequences greatly enhances the efficiency of the approximation and allows for more precise estimates from a smaller number of simulation runs. For simplicity, our estimation procedure assumes that the covariance matrix is a positive definite matrix with zero off-diagonal terms (i.e. covariances are assumed to be zero).

In the estimation of the NL and RDNL models we use identical nests. Since our interest is restricted to model performance, the choice of the 'right' nest is not an issue and we therefore do not permute over multiple nests. Having said that, however, we must point out that our choice of nests is fairly intuitive. For example in the Cracker data we frame the nest as {Private Label, National Label} with three brands in the {National Label} branch. The nests for the three datasets are described by the structures below.

Yogurt: {Major Brands, Minor Brands}
{Major Brands} ⇒ {Yoplait, Dannon}
{Minor Brands} ⇒ {Hiland, Weight Watcher}

Cracker: {National Label, Private Label}
{National Label} ⇒ {Keebler, Nabisco, Sunshine}
{Private Label} ⇒ {Store Brand}

The probabilities for the models are as in (29) and (31). As with the MNL and RDMNL we use maximum likelihood to estimate the NL and RDNL models. The derivation of the likelihood is similar to the process described earlier and we do not repeat it here.

5.3. Empirical results

Parameter estimates from our analysis are presented in Tables 2 and 3. We begin by examining the estimates and other auxiliary measures for the MNL and RDMNL models in detail.

5.3.1. Basic homogeneous brand choice models. A first glance at the results shows that the in-sample and out-of-sample fit statistics for the homogeneous models (MNL and RDMNL) do not resolve themselves into favoring one framework in particular. The parameter

Table 2. Estimation results (Yogurt data).

	MNL	RDMNL	HMNL	HRDMNL	NL	RDNL
Yoplait (s.e.)	4.4502 (0.187)	3.0255 (0.134)	4.9624 (0.3242)	4.3982 (0.5234)	4.3174 (0.1932)	2.487 (0.0919)
std.dev.			2.714 (0.2106)	2.804 (0.1986)		
Dannon	3.7156 (0.145)	2.4653 (0.093)	4.419 (0.2506)	2.9729 (0.1936)	3.6719 (0.1425)	2.033 (0.107)
std.dev.			1.9329 (0.1731)	3.0027 (0.2513)		
Weight Watcher	3.0744 (0.144)	1.9701 (0.089)	1.035 (0.3433)	1.8381 (0.2376)	2.9242 (0.1658)	1.2674 (0.0752)
std.dev.			4.6318 (0.3438)	2.3133 (0.2799)		
Price	−36.658 (2.436)	−28.290 (1.966)	−43.2823 (4.3468)	−33.6957 (5.7841)	−31.645 (3.657)	−21.9085 (2.877)
std.dev.			21.8045 (4.346)	30.9073 (3.6721)		
Feature	0.4914 (0.120)	0.4019 (0.095)	0.7962(0.1871)	0.7572 (0.1635)	0.4209 (0.116)	0.1864 (0.0841)
$(1-\sigma)$					1.5429 (0.252)[a]	1.7382 (0.271)[b]
−2LL	5313.8	5321.7	2494.2	2469.5	5310.7	5350.1
AIC	5323.8	5331.7	2512.2	2487.5	5322.7	5362.1
BIC	5352.7	5360.6	2564.3	2539.6	5357.4	5396.8
OOS[c] BIC	1090.3	1091.6	687.8	669.0	1091.2	1098.9

[a]Not consistent with RUM.
[b]Not consistent with RDM.
[c]Out of Sample.

Table 3. Estimation results (Cracker data).

Variable	MNL	RDMNL	HMNL	HRDMNL	NL	RDNL
Sushine	−0.6624 (0.090)	−0.4081 (0.058)	−0.182 (0.1776)	−0.5722 (0.1291)	0.0619 (0.1292) ns	0.0937 (0.1001)
std. dev.			2.741 (0.2029)	2.0463 (0.1484)		
Keebler	−0.1688 (0.117)	−0.0445 (0.077)	−0.1061 (0.2321)	0.1865 (0.1582)	0.4683 (0.1206)	0.4339 (0.0992)
std.dev.			2.6136 (0.1726)	2.0215 (0.1303)		
Nabisco	1.7928 (0.101)	1.3643 (0.075)	2.5005 (0.2374)	1.9151 (0.1583)	1.6428 (0.0971)	1.4321 (0.0742)
std.dev.			3.0075 (0.1847)	2.4631 (0.1434)		
Price	−3.1247 (0.209)	−2.2453 (0.154)	−2.9122 (0.3953)	−2.4114 (0.3231)	−2.5065 (0.239)	−2.1107 (0.172)
std.dev.			4.715 (0.5133)	4.2491 (0.3504)		
Feature	0.4961 (0.096)	0.3962 (0.078)	0.8703 (0.1506)	0.7437 (0.1267)	0.3028 (0.08299)	0.2538 (0.0699)
$(1-\sigma)$					1.6938 (0.2049)[d]	1.5699 (0.1650)[e]
−2LL	6695.4	6694.7	3685.4	3625.8	6674.7	6671.7
AIC	6707.4	6706.7	3705.4	3645.8	6688.7	6685.7
BIC	6744.0	6743.3	3766.4	3706.8	6731.4	6728.4
OOS[f] − BIC	1373.5	1371.2	891.8	878.7	1370.7	1369.8

[d] Not consistent with RUM.
[e] Not consistent with RDM.
[f] Out Of Sample.

estimates[6] presented in the tables are based on the complete data. For the out-of-sample fit/prediction statistics we used the following procedure. First we split the data randomly into hold-out (20%) and calibration (80%) samples. We then estimated the models on using only the calibration data and applied the estimates obtained to the hold-out sample. This gave us the relevant fit/prediction statistics. For the yogurt data the MNL model has a slightly better fit both within sample and out-of-sample than the RDMNL. This is reversed for the cracker data. In the datasets studied no two pairs of models with homogeneous parameters are significantly different. This should not be surprising since the underlying error structure is the essentially the same. We cannot, however, make a generalization at this point. Indeed, there might be datasets where one approach fits better than the other.

A closer look at the parameter estimates shows that the signs of all estimates from the RDMNL match those obtained in the MNL. This is true both for the yogurt and cracker results. We do find, however, that the estimates from the RDMNL are consistently smaller (in absolute value) than their MNL counterparts. This in itself does not imply that the effects of the marketing mix variables in the RDMNL are lower than those in the MNL since the estimates are identified only up to a scale factor. It does, however, lead us to delve deeper into more robust measures such as own and cross price elasticities. The reader will recall that the RDMNL does not exhibit IIA and hence its elasticity matrix should be significantly different from its MNL counterpart.

We derived the elasticity matrices for each model and dataset. These are presented in Tables 4 and 5. The elasticities are estimated at the mean market share and prices of

6 Note that the signs for the estimates in the random disutility models have been reversed to facilitate comparison. See footnote [4]

Table 4. (a) Yogurt data elasticities from MNL model. (b) Yogurt data elasticities from RDMNL model.

	Yoplait	Dannon	Weight	Hiland
		(a)		
Yoplait	−2.5973	1.2130	0.6730	0.0544
Dannon	1.3254	−1.7931	0.6730	0.0544
Weight	1.3254	1.2130	−2.2411	0.0544
Hiland	1.3254	1.2130	0.6730	−1.9117
		(b)		
Yoplait	−2.5221	1.0751	0.7213	0.0743
Dannon	1.1778	−1.6916	0.6679	0.0654
Weight	1.4181	1.1986	−2.3591	0.0969
Hilanrd	2.1491	1.7255	1.4255	−3.1674

Table 5. (a) Cracker data elasticities from MNL model. (b) Cracker data elasticities from RDMNL model.

	Sunshine	Keebler	Nabisco	Pvt. label
		(a)		
Sunshine	−2.7875	0.2313	1.8644	0.6669
Keebler	0.2029	−3.2869	1.8644	0.6669
Nabisco	0.2029	0.2313	−1.5079	0.6669
Pvt. label	0.2029	0.2313	1.8644	−1.4602
		(b)		
Sunshine	−3.5208	0.6857	1.8079	0.9494
Keebler	0.5983	−4.1722	1.8128	0.9534
Nabisco	0.1864	0.2142	−1.2282	0.5126
Pvt. label	0.2666	0.3068	1.3959	−1.2557

each brand. The tables indicate a number of significant differences between the MNL and RDMNL elasticity measures. Recall that the cross price share elasticity of x with respect to y in standard MNL models is

$$\varepsilon_{xy}^{u} = \frac{\partial P_x^u}{\partial p_y}\frac{p_y}{P_x^u}, \tag{39}$$

which simplifies to

$$\varepsilon_{xy}^{u} = -P_y^u p_y \beta. \tag{40}$$

Clearly ε_{xy}, is independent of x and is therefore constant, irrespective of the choice of x. This, however, is not true for the RDMNL, which has more complicated (and hence flexible) elasticity structures. A first glance at the tables reveals, as expected, that the cross price elasticities in the RDMNL matrices are not constant for each brand while they are in the MNL matrices. A second distinction is that the MNL elasticity estimates seem to be smaller (in absolute value) for low market share brands and higher for the larger share ones than their corresponding RDMNL estimates. For example Nabisco is the highest share brand in the cracker category and has a MNL own price elasticity of -1.5078. The RDMNL estimate is -1.2282. Similarly Keebler, which has the lowest market share has an own price elasticity of -3.2871 in the MNL table and -4.1720 in the RDMNL version. What the elasticity matrices show is that when the model is constrained by the IIA property the elasticities are also artificially constrained

The tables highlight the fact that the RDMNL approach results in elasticity measures that are significantly different from its MNL counterpart. As one would expect, the RDMNL cross-elasticities are not the same within each column since the IIA property has been relaxed. In the earlier literature the only way to deal with the IIA issue (which leads to the constraints on the elasticity matrix) was to enhance the model by adopting a nested approach, adding heterogeneity or by estimating a more general but econometrically cumbersome multinomial probit model. The RDMNL offers a simple yet effective way of arriving at cross price elasticities that are realistic without imposing any additional assumptions on the modelling framework. The aim of our analysis was to point out some key distinctions between the RDMNL and the MNL approaches. Our discussion above is simply a first look at these models. Future research might look at various other facets of comparison.

5.3.2. Heterogeneous parameter models. The incorporation of heterogeneity improves the fit of both models significantly and all heterogeneity parameters are significant. In both data sets the HRDMNL model outperforms the HMNL model in terms of fit and prediction creteria. This is especially intriguing since the homogeneous models exhibit no significant differences between the two frameworks. We conjecture that the relaxation of the IIA when coupled with the added flexibility of heterogeneous parameters in the random disutility framework allows for a closer approximation to the true underlying patterns of choices observed in the data.

The estimates from the heterogenous (random coefficients) versions of MNL and RDMNL seem to follow the same pattern as their homogenous counterparts. All coefficients are a bit smaller in the random disutility models versions. Again, this patern does not imply that the effect of price is lower in the random disutility model and we need to examine the elasticity matrices to make a better comparison. Since the heterogeneous MNL model does not suffer from the IIA problem at the aggregate level examining and comparing these elasticities might provide some interesting insights. Tables 6 and 7 depict the elasticities for the two datasets based on the estimates from the heterogenous models. We focus our attention on the Yogurt data (Tables 4 and 6) in the discussion that follows.

The elasticity estimates obtained via the heterogeneous specifications are very similar to those obtained by Jain et al. (1994). Further, the elasticity matrices based on the random utility and disutility versions do not exhibit major differences. This is to be expected, since the IIA property now no longer applies in the heterogeneous MNL model. What is more interesting is that the pattern of elasticites in the RDMNL model (Table 4b) is similar to those obtained in the heterogeneous framework. For example, Hiland was seen to be

Table 6. (a) Yogurt data elasticities from HMNL model. (b) Yogurt data elasticities from HRDMNL model.

	Yoplait	Dannon	Weight	Hiland
		(a)		
Yoplait	−1.8967	0.8692	0.3841	0.0723
Dannon	0.9246	−1.4045	0.4619	0.0921
Weight	0.8341	0.8817	−1.6724	0.0746
Hiland	1.4736	1.7776	0.7418	−2.4839
		(b)		
Yoplait	−1.4767	0.5439	0.3349	0.1141
Dannon	0.6745	−1.1649	0.3852	0.1251
Weight	0.8040	0.7462	−1.6936	0.1979
Hiland	1.7558	1.5471	1.2264	−2.8356

Table 7. (a) Cracker data elasticities from HMNL model. (b) Cracker data elasticities from HRDMNL model.

	Sunshine	Keebler	Nabisco	Pvt. label
		(a)		
Sunshine	−1.4519	0.0951	0.6046	0.6170
Keebler	0.0776	−1.1142	0.2620	0.4610
Nabisco	0.0601	0.0307	−0.5128	0.2575
Pvt. label	0.3144	0.2810	1.2902	−1.2043
		(b)		
Sunshine	−1.5079	0.1145	0.5542	0.6913
Keebler	0.0806	−1.0016	0.1485	0.4680
Nabisco	0.0584	0.0206	−0.5281	0.2751
Pvt. label	0.3297	0.3123	1.2093	−1.1937

the most elastic in the RDMNL framework and this is borne out in both the HMNL and HRDMNL based elasticity matrices. We also find that the own and cross elasticites are somewhat overstated when heterogeneity is ignored in the RDMNL model suggesting that while the RDMNL is more flexible at the individual household level it fails to account for the differences across households. This, in part, also explains why coefficients from the two approaches are different even with the incorporation of heterogeneity. The random utility logit models continue to exhibit from the IIA property at the individual level. In other words, if we conditional on knowing a particular household's parameters, standard logit type probabilities and elasticities will be obtained. The random disutility models, on the other hand, do not exhibit IIA at the individual household level. While both approaches estimate the same aggregate elasticity matrix they differ significantly at the disaggregate level and therefore in how these matrices are constructed.

5.3.3. Other models. The NL and RDNL yield further evidence of the pattern mentioned earlier. Interestingly, both the NL and the RDNL find that the $(1 - \sigma)$ estimate was greater than one for Cracker and Yogurt data hence rejecting the hypothesis of RUM and RDM. This could be on account of the choice of the nesting pattern chosen or because there is no nesting to begin with. Irrespective of what drives this finding, the facet important to us is the consistency of the two approaches. The findings reiterate the validity of the RDM approach as a viable discrete choice framework.

In some preliminary attempts we have had moderate success in implementing more complex GEV models such as those based on the FinMix generating function proposed by Swait (2003) and the Generalized Nested Logit model (Koppelman and Wen, 2001; Swait, 2000). In some cases we ran into a local optima problem when the starting parameters were not close to the global maxima. This was true for both the RDM and the RUM specifications and therefore has to do more with the inherent complexity of the underlying model and density function than with the choice of the decision framework. We relegate a more detailed analysis of such complex models to future research.

6. Discussion and implications

An issue that has yet to be discussed pertains to when one should choose the RDM approach over RUM and vice versa. Given the recency of the RDM approach there is not enough theoretical and/or empirical evidence to provide strict guidelines for such a choice. However, based on our experience (albeit short) with the two approaches we conjecture that the following factors will be important determinants of this choice. First, the structure of the problem being dealt with might dictate (or in the least influence) which framework to adopt. For example the Bell and Lattin (1999) store choice problem lends itself quite naturally to the RDM framework.

Second, the nature of the product/service being investigated might play a role. The transportation choice literature provides a good illustration where the individual is usually assumed to be minimizing transportation costs (since one could argue that the utility derived by reaching one's destination is the same across transportation choices). More generally, one could argue that product categories characterized by offerings that are inherently homogeneous in the benefits they provide lend themselves to the RDM approach since price may be the only variable driving decisions. Future research might be able to substantiate this argument by estimating and comparing RDM/RUM models over a number of product categories.

Third, if the choice is restricted to the MNL and the RDMNL the researcher's preferences with respect to the IIA property might play a role in determining the framework used. The reader should keep in mind that the IIA property is not necessarily "bad" and may, in fact, represent certain choice situations accurately. We note, however, that the RDM models proposed in this paper and their subsequent empirical implementations have yielded results that have significant differences from their RUM counterparts. In our opinion the foremost among these is the discrepancy between the elasticity measures generated by the two approaches. Given that the flexibility in the RDM elasticity matrix is driven, at least in part, by the relaxation of the IIA property and that the RDMNL fulfills the need for a simple choice framework that does not exhibit this property, the decision to choose between the two models might boil down to the applicability of the IIA property in the problem under investigation. This crux of this discussion suggests that managers (and researchers) should estimate choice models based on both decision frameworks and check for consistency in terms of the elasticity measures derived. The RDMNL model may also be used as a basis to develop a more formal specification test. Estimating both models (i.e. the RDMNL and the MNL) and then using the relevant Vuong (1989) statistic might be a simple way of testing for IIA. An investigation of the power of such tests, the derivation of the test statistic and subsequent implementation are avenues for future research.

To conclude this discussion we note that the RDM approach offers researchers and managers a simple yet viable framework to model choices without relying on traditional logit type models. As our results show, both homogeneous and heterogeneous imlementations of the RDM models are straightforward to estimate and perform well. There remains, of course, much more to do to ascertain the general applicability of these models which we feel future research might address.

7. Conclusion

The RDM framework has potential applicability in numerous fields apart from the scanner based brand choice applications presented in this paper. One could think using the proposed random cost based models in areas such as auction theory, conjoint analysis, theoretical and empirical IO and any other area that has uses discrete choice models. In most cases we don't expect there to be major differences in the results obtained via such substitution. However, a quick glance at the results obtained from choice models in this paper suggests that we might find new insights into processes and behavior that have till now been thought to be fairly robust. As mentioned in an earlier section, the RDM approach opens up a vast array of models that have not been tested yet. We hope that this paper spurs others to investigate this new model space.

Appendix

Preliminary definitions

Definition 1. F is a multivariate distribution function such that $F = F(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$.

Definition 2. The density function can be denoted as $F_{1,\ldots,n} = \frac{\partial^n F}{\partial\varepsilon_1 \ldots \partial\varepsilon_n}$

Definition 3. For any element $\varepsilon_j = -\infty$ the distribution function F and the marginal density function $F_{1,\ldots,j-1,j+1,\ldots,n} = \frac{\partial^{n-1} F}{\partial\varepsilon_1 \ldots \partial\varepsilon_{j-1} \partial\varepsilon_{j+1} .. \partial\varepsilon_n}$ are equal to zero. (Since the probability of any element being less than $-\infty$ is zero)

Definition 4. Using the above definitions we have

$$\begin{aligned}\int_z^\infty F_{1,\ldots,n} d\varepsilon_j &= \int_{-\infty}^\infty F_{1,\ldots,n} d\varepsilon_j - \int_{-\infty}^z F_{1,\ldots,n} d\varepsilon_j \\ &= (F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_j=\infty} - F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_n=-\infty}) \\ &\quad - (F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_n=z} - F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_n=-\infty}) \\ &= F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_j=+\infty} - F_{1,\ldots,j-1,j+1,\ldots,n}|_{\varepsilon_n=z} \end{aligned} \tag{A1}$$

Proof Theorem 1: Expanding the probability of choosing alternative 1 we have

$$\begin{aligned} P_1^d &= \int_{-\infty}^\infty \int_{d_1+\varepsilon_1-d_2}^\infty \cdots \int_{d_1+\varepsilon_1-d_n}^\infty F_{1,2,\ldots,n} d\varepsilon_n \ldots d\varepsilon_2 d\varepsilon_1 \\ &= \int_{-\infty}^\infty \int_{d_1+\varepsilon_1-d_2}^\infty \cdots \int_{d_1+\varepsilon_1-d_n-1}^\infty \\ &\quad \times (F_{1,2\ldots,n-1|\varepsilon_n=+\infty} - F_{1,2\ldots,n-1|\varepsilon_n=d_1+\varepsilon_1-d_n}) d\varepsilon_n \ldots d\varepsilon_2 d\varepsilon_1 \\ &= \int_{-\infty}^{+\infty} \begin{bmatrix} F_1(\varepsilon_1, +\infty, \ldots, +\infty) - \sum_{j=2}^n F_1(\varepsilon_1, d_1+\varepsilon_1-d_j, \varepsilon_{-j}=+\infty) \\ + \sum_{j=2}^{n-1}\sum_{k=j+1}^n F_1(\varepsilon_1, d_1+\varepsilon_1-d_j, d_1+\varepsilon_1-d_k, \varepsilon_{-j,-k}=+\infty) \\ + \cdots - (-1)^n \sum_{j_n}\cdots\sum_{j_1} F_1\big(\varepsilon_1, d_1+\varepsilon_1-d_{j_1}, \ldots, d_1+\varepsilon_1-d_{j_n}\big) \end{bmatrix} d\varepsilon_1, \end{aligned} \tag{A2}$$

and hence

$$\begin{aligned} P_1^d &= 1 - \int_{-\infty}^\infty \sum_{j=2}^n F_1(\varepsilon_1, d_1+\varepsilon_1-d_j, \varepsilon_{-j}=+\infty) d\varepsilon_1 + \cdots + \\ &\quad - \int_{-\infty}^\infty (-1)^n F_1(\varepsilon_1, d_1+\varepsilon_1-d_2, \ldots, d_1+\varepsilon_1-d_n) d\varepsilon_1. \end{aligned} \tag{A3}$$

Based on McFadden (1978) we can re-write this as (after replacing subscripts)

$$P_i^d = 1 - \begin{bmatrix} \sum_{j\neq i} \Pr(d_i+\varepsilon_i \geq d_j+\varepsilon_j) \\ - \sum_{j\neq i}\sum_{k\neq\{i,j\}} \Pr(d_i+\varepsilon_i \geq d_j+\varepsilon_j, d_i+\varepsilon_i \geq d_k+\varepsilon_k) \\ + \sum_{j\neq i}\sum_{k\neq\{i,j\}}\sum_{l\neq\{i,j,k\}} \Pr(d_i+\varepsilon_i \geq d_j+\varepsilon_j, d_i+\varepsilon_i \geq d_k \\ + \varepsilon_k, d_i+\varepsilon_i \geq d_l+\varepsilon_l)\cdots + (-1)^n \Pr(d_i+\varepsilon_i \geq d_j+\varepsilon_j : j \in \mathcal{J}) \end{bmatrix} \tag{A4}$$

and substituting in the definition of Ω obtains the expression in the theorem. □

Proof of Lemma 2: Using Lemma 1, we have (for the multivariate GEV distribution) $\Omega_i(\Psi_m) = \frac{e^{d_i}\mathcal{G}_i(\Psi_m^e)}{\mathcal{G}(\Psi_m^e)}$. The rest of the proof follows from Theorem 1 with approapriate subsitutions of Ω_i and simplification. □

Proof of Theorem 2: If the random vector **x** is distributed as spherically symmetric (SSD) then x $\stackrel{d}{=}$ −x. To see this note that $\mathbf{v} \stackrel{d}{=} -\mathbf{v}$ since it is distributed on the unit sphere. And therefore

$$r\mathbf{v} \overset{d}{=} -r\mathbf{v} \overset{d}{=} -\mathbf{x} \overset{d}{=} \mathbf{x}. \tag{A5}$$

Now if $\varepsilon \overset{d}{=} SSD$, and $\{u_i = -d_i : i \in \mathcal{J}\}$ we have

$$\begin{aligned} P_i^d &\equiv \Pr(d_i + \varepsilon_i \leq d_j + \varepsilon_j : j \in \mathcal{J}) \\ &\equiv \Pr(u_i + \varepsilon_i \geq u_j + \varepsilon_j : j \in \mathcal{J}) \\ &\equiv P_i^u \end{aligned} \tag{A6}$$

for all $i \in \mathcal{J}$. □

Acknowledgments

I would like to thank the editor and referees for their suggestions and constructive comments. The paper has also benefitted from comments by participants at the Rochester Econometrics Workshop and the marketing seminar series at Cornell University, Duke University and the University of Maryland. All errors are mine and mine alone.

References

Anderson, S., A. De Palma, and J. Thisse. (1992). Discrete Choice Theory of Product Differentiation, MIT Press.

Anderson, S. and A. de Palma. (1999). “Reverse Discrete Choice Models.” Regional Science and Urban Economics 29(6), 745–764.

Bell D. and J. Lattin. (1998). “Shopping Behavior and Consumer Preference for Store Price Format: Why ‘Large Basket’ Shoppers Prefer EDLP.” Marketing Science 17(1), 66, 23.

Bell D.R, T.-H. Ho, and C.S. Tang. (1998). “Determining Where to Shop: Fixed and Variable Costs of Shipping.” Journal of Marketing Research 35 (August), 352–369.

Ben-Akiva M.E. and S.R. Lerman. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: MIT Press.

Ben-Akiva, M. and B. Francois. (1983) μ-Homogeneous Generalized Extreme Value Model, Working Paper, Dept. of Civil Engineering, Cambridge, MA, MIT.

Berry, S., J, Levinsohn, and A. Pakes. (1995). “Automobile Prices in Market Equilibrium.” Econometrica 63 (July): 841–990.

Bhat, C.R. (2003). “Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences.” Transportation Research (forthcoming).

Bhat, C.R. (2001). “Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model.” Transportation Research, 35B, 677–693.

Chintagunta, P.K. (1994). “Heterogeneous Logit Model Implications for Brand Positioning” Journal Of Marketing Research, Chicago; 31(2) 304–308.

Chintagunta, P.K. (1992). “Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments.” 11(4).

Dubin, J.A. (1986). “A Nested Logit Model of Space and Water Heat System Choice” Marketing Science, 5(2), 112–124.

DeGroot, M. (1989). “Probability and Statistics.” Reading MA: Addison Wesley Publishing.

Elrod, T. (1987). “Choice Map: Inferring a Product-Market Map from Panel Data.” Marketing Science, Linthicum; Winter 1988; 7(1) 21.

Erdem, T. and M. Keane. (1996). "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." Marketing Science 15(1).
Fang, K.T., S. Kotz, and K.W. Ng. (1990). Symmetric Multivariate and Related Distributions. Chapman and Hall, London.
Fotheringham, A.S. (1988). "Consumer Store Choice and Choice Set Definition." Marketing Science 7(3), 299–310.
Gonul, F. and K. Srinivasan. (1996). "Estimating the Impact of Consumer Expectations of Coupons on Purchase Behavior: A Dynamic Structural Model." Marketing Science 15(3), 262–279.
Guadagni, P.M. and J.D.C. Little. (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data." Marketing Science 2(3), 203–238.
Higgins, E.T., R.S. Friedman, R.E. Harlow, L.C. Idson, O.N. Ayduk, and A. Taylor (2000). "Achievement Orientations from Subjective Histories of Success: Promotion Pride Versus Prevention Pride." European Journal of Social Psychology 30, 1–23.
Higgins, E.T. (1998). "Promotion and Prevention: Regulatory Focus as a Motivational Principle." Advances in Experimental Social Psychology 30, 1–46.
Higgins, E.T. (1989). "Continuities and Discontinuities in Self-Regulatory and Self-Evaluative Processes: A Developmental Theory Relating Self and Affect." Journal of Personality 57 (2), 407–444.
Ho, T.H., C.S. Tang, and D.R. Bell. (1998). "Rational Shopping Behavior and the Option Value of Variable Pricing." Management Science 44(12, 2), 145–160
Hobbs, J.E. (1997). "Measuring the Importance of Transaction Costs in Cattle Marketing." American Journal of Agricultural Economics 79, 1083–1095.
Horsky, D. and P. Nelson. (1992). "New Brand Positioning and Pricing in an Oligopolistic Market." Marketing Science 11(2), 133–121.
Jain, D., P. Chintagunta, and N. Vilcassim. (1994). "A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data." Journal Of Business & Economic Statistics 12(3), 317.
Joskow P.L. and F.S. Mishkin. (1977). "Electric Utility Fuel Choice Behavior in the United States" International Economic Review 18(3), 719–736.
Kadiyali, V., P. Chintagunta, and N. Vilcassim. (2000). "Manufacturer-Retailer Channel Interactions and Implications for Channel Power: An Empirical Investigation of Pricing in a Local Market." Marketing Science 19(2), 127.
Kamakura W. and G. Russell. (1989). "A Probabilistic Choice Model For Market Segmentation and Elasticy Structure" Journal of Marketing Research, Chicago 26(4), 379, 12
Kannan, P.K and G. Wright. (1991). "Modeling and Testing Structured Markets: A Nested Logit Approach." Marketing Science 10(1).
Koppelman, F.S. and C.-H. Wen. (2000). "The Paired Combinatorial Logit Model: Properties, Estimation and Application." Transportation Research Part B: Methodological 34(2), 75–89.
Koppelman, F.S. and V. Sethi. (2000). "Closed Form Logit Models." In D.A. Hensher and K.J. Button (eds.), Handbook of Transport Modeling, Oxford: Pergamon Press.
Mas-Colell, A., M.D. Whinston, and J.R. Green. (1995). Microeconomic Theory, Oxford: Oxford University Press.
McFadden, D. (1978). "Modelling the Choice of Residential Location." In A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), Spatial Interaction Theory and Planning Models, 75–96, Amsterdam: North Holland.
McFadden, D. (1981). "Econometric Models of Probabilistic Choice." in C.F. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press: Cambridge, MA, 198–272.
McFadden, D. and K. Train. (1999). "Mixed MNL Models for Discrete Response." with K. Train, Journal of Applied Econometrics 15(5), 447–470, John Wiley & Sons Ltd.: New York, December.
McFadden, D. (2001). "Disaggregate Behavioral Travel Demand's RUM Side: A 30-Year Retrospective." The Leading Edge of Travel Behavior Research, David Heshner (ed.), Oxford: Pergamon Press.
McCulloch, R. and P.E. Rossi. (1994) "An Exact Likelihood Analysis of the Multinomial Probit Model." Journal of Econometrics 64, 207–240.
Mazumdar T. and P. Pappatla. (2000). "An Investigation of Reference Price Segments." Journal of Marketing Research Chicago 37(2) 246, 13.

Roy R., P. Chintagunta, and S. Haldar. (1996). "A Framework for Investigating Habits, "The Hand of the Past." and Heterogeneity in Dynamic Brand Choice." Marketing Science 15(3) 280–299.

Rossi, P.E., R.E. McCulloch, and G.M. Allenby. (1996), "The Value of Purchase History Data in Target Marketing." Marketing Science 15, 321–340.

Seetharaman, P.B. (2003), "Modeling Multiple Sources of State Dependence in Random Utility Models of Brand Choice: A Distributed Lag Approach." Marketing Science (forthcoming).

Shugan, S. (1980). "The Cost of Thinking." Journal of Consumer Research 7(2), 99–111.

Small, K. (1994). "Approximate Generalized Extreme Value Models of Discrete Choice." Journal of Econometrics, Amsterdam 62(2), 51, 32.

Sudhir, K. (2001). "Competitive Pricing Behavior in the Auto Market: A Structural Analysis." Marketing Science 20(1), 42.

Swait, J. (2003). "Flexible Covariance Structures for Categorical Dependent Variables Through Finite Mixtures of GEV Models." Journal of Business and Economic Statistics 21(1), 80–87.

Vovsha, P. (1998). "Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area, Transportation Research Record 1607, 6–15.

Vuong, Q.H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." Econometrica 57, 307–333.

Winer, R. (1986). "A Reference Price Model of Brand Choice for Frequently Purchased Products." Journal of Consumer Research, Gainesville; 13(2) 250–257.

Wen, C. and F. Koppelman. (2001). "The Generalized Nested Logit Model." Transportation Research Part B, 35(7), 627–641.