Incorporating Probabilistic Choice Rules within Random Utility Models of Brand Choice: Theory and Empirical Illustration

AMIT PAZGAL
Associate Professor of Marketing, Washington University, St. Louis *

P.B. SEETHARAMAN †
Associate Professor of Marketing, Rice University

RICHARD R. BATSELL
Associate Professor of Marketing, Rice University

November 24 2005

Abstract

Existing random utility models of brand choice behavior, in the tradition of McFadden (1974), are based on the assumption that consumers employ the decision rule of *maximum utility* to choose a brand. However, there is a rich body of empirical work in mathematical psychology that suggests that consumers choose probabilistically between brands (see, for example, Luce 1959). In this paper, we propose a random utility model of brand choice that is based on the decision rule of *probabilistic choice*, and nests the traditional random utility model as a special case. We argue that the proposed random utility model is mathematically more flexible than existing random utility models. We illustrate the empirical implementation of the proposed random utility model using real-world choice data as well as experimental choice data. We compare the empirical performance of our proposed model to the traditional random utility model. Interestingly, our proposed model outperforms the traditional random utility model (that it nests as a special case) in explaining both consumers' real-world choices and experimental choices.

Keywords: Brand Choice, Individual-Level, Random Utility, Decision Rule, Maximum Utility, Probabilistic Choice, Scanner Panel Data.

*The authors thank Fred Feinberg and Eric Bradlow for their extensive comments on earlier versions of the manuscript.

†*Address for correspondence*: Jesse H. Jones Graduate School of Management, Rice University, 6100 Main, MS 531, Houston, TX 77005. Email address: seethu@rice.edu.

1 Introduction

Demand forecasting for brands is a central issue in marketing research, with obvious implications for marketing practitioners in strategic areas such as pricing, advertising etc. Such forecasting has at its core the econometric modeling of brand choice behavior of consumers within specific categories of products and services. It is useful to model individual consumer choices, as opposed to aggregate brand sales, since consumers tend to be heterogeneous in terms of both the variables that drive their choice behavior and the relative importances of these variables in driving choice. This paper proposes a new econometric model to capture brand choice behavior of consumers.

There is an extensive literature in marketing on estimating brand choice models using individual-level purchase data. Studies over the past two decades have proposed models that are principally based on the *random utility* framework. This framework - pioneered by McFadden (1974) and Manski (1975), applied for the first time on scanner panel data by Guadagni and Little (1983), and used in hundreds of research studies thereafter - is based on the economic principle of *utility maximization*. A consumer is modeled as choosing the brand with the highest utility among those available on the purchase occasion. The consumer's utility for a brand, in turn, is assumed to depend on the consumer's underlying preference for the brand and observable marketing variables of the brand - such as price, advertising, packaging etc. - usually in a linear manner (through unknown parameters). Since these utilities are unobserved by the marketing researcher, they are assumed to be *random variables* (hence the name *random utility*). In other words, a consumer's utility for a brand is parameterized as a deterministic (up to parameters) function of observed marketing variables plus a random error term. Making a suitable assumption on the random error term enables the researcher to compute the probability that a brand has the highest utility among all the available brands for a consumer on a purchase occasion. By minimizing an appropriate loss function applied to the deviations of these modeled probabilities from the consumer's actual brand choices, the researcher estimates the parameters of the deterministic component of the consumer's random utility function. This method of calibrating brand choice models on scanner panel data has become widely adopted by marketing researchers and practitioners over the past two decades.

While the goal of random utility models is *descriptive*, they are based on a specific *decision rule* that is assumed to govern a consumer's brand choices, i.e., the rule of *maximum utility.* This rule says that the consumer will choose the brand that offers him the highest utility among all the available brands. Any difference between the consumer's actual choice and the choice predicted by the maximum utility rule could only be attributed to the random error which captures the researcher's uncertainty about the consumer's utility function. In this paper, we propose an alternative explanation for the observed differences between observed and predicted choices of consumers, in addition to the above-mentioned one. Our explanation is that the consumer's decision rule need not be one of maximum utility. We propose a decision rule, that we call the rule of *probabilistic choice*, that says that the consumer chooses probabilistically between brands according to his random utilities for them. Under this decision rule, brands that do not offer the highest utility to the consumer still have a non-zero probability (albeit smaller than that associated with the highest-utility brand) of being chosen by the consumer. This decision rule is consistent with theories of probabilistic choice proposed by Thurstone (1927), Luce (1959), Marschak (1960), Bass (1974) etc., and nests the traditional maximum utility rule as a special case. Specifically, when the suggested probabilistic choice rule dictates that the consumer chooses the highest-utility brand with probability one, and the other brands with probability zero, it reduces to the maximum utility rule.

We test the empirical value of the probabilistic choice rule by developing an estimable random utility model of brand choice that is based on such a decision rule. This brand choice model is more flexible than the traditional random utility model, that it nests as a special case. We estimate two versions[1] of our proposed model using scanner panel data on households' brand choices in a frequently purchased product category and demonstrate that the proposed model fits and predicts observed choices better than the traditional random utility model. We further demonstrate the empirical superiority of our proposed model over the traditional random utility model using experimental data on subjects' choices among wagers in a laboratory. Our paper answers the call of Louviere et al. (2002) inviting researchers to develop more general models of choice that are based on alternative accounts of the choice process. Our proposed model is also consistent with the argument in Louviere et al. (2002) that existing brand choice models are incomplete in capturing the effects of "unobserved variability "in households' choices.

The remainder of the paper is organized as follows. In section 2 we discuss the traditional

[1] One generalizes the Multinomial Logit model, while the other generalizes the Multinomial Probit model.

random utility model of brand choice. In section 3, we develop our proposed random utility model of brand choice, and then discuss some of its properties, especially in relation to those associated with the traditional random utility model. In that section we also develop our estimation procedure. Section 4 is concerned with the application of the model to the analysis of A.C. Nielsen's scanner panel data. Along with the results from the estimation, we show the empirical gains from our model in relation to the traditional random utility model. Section 5 discusses the application of the model to the analysis of experimental choice data. In the experimental choice context, we demonstrate the internal validity of our model and its continuing superior performance over the traditional random utility model. Concluding remarks are made in Section 6.

2 Traditional Random Utility Model of Brand Choice

To appreciate the random utility model of brand choice, as described in McFadden 1974, Manski 1975, and Guadagni and Little 1983, consider the typical household h $(h = 1, 2, ..., H)$ observed over $t = 1, 2, ..., n_h$ purchase occasions in a specified product category with J brands. On any given purchase occasion t, one observes an outcome variable y_{ht} which takes the value j, $j = 1, 2, ..., J$ where j identifies the brand purchased by household h at that occasion. We assume that for every purchase occasion t, one observes the price $(Price_{htj})$, display $(Disp_{htj})$ and feature $(Feat_{htj})$ covariates faced by the household for all brands in the product category. The goal is to model the outcome variable (y_{ht}) given the above information.

The random utility model is based on a latent variable u_{hjt} that denotes the utility of household h for brand j at time t. It is further assumed that this utility can be expressed as a function of the brand and the entire set of brand covariates facing the household at time t, in the following manner:

$$u_{hjt} = \alpha_{hj} + \delta_{2h} Price_{hjt} + \delta_{3h} Disp_{hjt} + \delta_{4h} Feat_{hjt} \; + \eta_{hjt}$$

where α_{hj}, $j = 1, 2, ..., J$ are brand-specific intercepts that are household-specific and

$$\delta_h = (\alpha_{h1}, \alpha_{h2}, ..., \alpha_{hJ}, \delta_{2h}, \delta_{3h}, \delta_{4h})'$$

are household-specific coefficients, and η_{hjt} is a random error such that $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})'$ has a joint distribution $f(\eta_{ht})$ with continuous support, mean zero and covariance matrix Σ.

The brand level outcome y_{ht} is now determined using the decision rule of *maximum utility*. This decision rule says that the household will choose brand j, and the researcher would therefore observe the outcome $y_{ht} = j$ if and only if the household's utility for the jth brand exceeds the household's utilities for the remaining brands. Specifically,

$$y_{ht} = j \text{ iff } u_{hjt} > \max_{k \neq j} u_{hkt}$$

It follows that the probability of the outcome $y_{ht} = j$ is then given by

$$\Pr(y_{ht} = j | \delta_h, \Sigma) = \int_{A_{h1t}} \cdots \int_{A_{hjt}} \cdots \int_{A_{hJt}} f(\eta_{ht} | X_{ht}\delta_h, \Sigma) I_{hjt} d\eta_{ht} \quad (1)$$

where I_{hjt} is an *indicator function* that takes the value 1 if $u_{hjt} > \max_{k \neq j} u_{hkt}$, and the value 0 otherwise, and

$$A_{hjt} = (-\infty, \infty)(j = 1, ..., J)$$

and

$$X_{ht} = (X'_{h1t}, ..., X'_{hJt})'$$

and

$$X_{hjt} = (Price_{hjt}, Disp_{hjt}, Feat_{hjt})'$$

We will call equation (1) the general representation of the traditional random utility model.

This equation is equivalently represented as follows.

$$\Pr(y_{ht} = j | \delta_h, \Sigma) = \int_{B_{h1t}} \cdots \int_{B_{hjt}} \cdots \int_{B_{hJt}} f(\eta_{ht} | X_{ht}\delta_h, \Sigma) d\eta_{ht} \quad (2)$$

where

$$\begin{aligned} B_{hjt} &= (-\infty, \infty) \\ B_{hkt} &= (-\infty, X_{hjt}\delta_h - X_{hkt}\delta_h + \eta_{hjt}) \; \forall \; k \neq j \end{aligned}$$

The decision rule I_{hjt} has been dropped from the integrand and the limits of integration have been changed from A_{hjt} to B_{hjt}. This is the common representation of the traditional random utility model in the literature on brand choice models (see, for example, Ben-Akiva and Lerman 1985).

Different distributional assumptions on $f(\eta_{ht})$ yield different operational versions of the traditional random utility model. For example, if one assumes that the errors $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})$ are distributed iid gumbel with an unknown scale parameter μ (and location parameter equal to zero), one obtains the *Multinomial Logit* model[2], where the probability of the outcome $y_{ht} = j$ is given by (see Ben Akiva and Lerman 1985 for details).

$$Pr_{hjt} = \frac{e^{v_{hjt}}}{\sum_{k=1}^{J} e^{v_{hkt}}} \tag{3}$$

where $v_{hjt} = \alpha_{hj} + \delta_{2h}Price_{hjt} + \delta_{3h}Disp_{hjt} + \delta_{4h}Feat_{hjt}$ stands for the deterministic component of the utility of brand j to consumer h. While it is possible to allow for more flexible distributional structures for η_{ht}, such as the multivariate normal distribution (which yields the *Multinomial Probit* model, see Daganzo 1979), all operational versions of the traditional random utility model rest on the assumption of the *maximum utility* decision rule, i.e., the consumer chooses the brand that offers him the highest utility. This assumption is in conflict with a rich body of experimental research in mathematical psychology that says that consumers choose brands in an inherently probabilistic manner (For two early papers on this subject, see Mosteller and Nogee 1951 and Edwards 1955; for an excellent review of this literature, see Luce and Suppes 1965).[3] That is, instead of selecting brands with the highest utility, consumers choose brands with choice probabilities that are (possibly smooth) functions of their underlying utilities for brands. Bass (1974) explains this by saying that there is a stochastic element in the consumer's brain that renders his brand choice process to be truly probabilistic. This implies that even if the marketing researcher perfectly observes the consumer's utility function (i.e., $\eta_{ht} = 0$), the researcher still cannot deterministically predict the consumer's brand choice on a given purchase occasion. We propose a random utility model that can reflect this by relaxing the *maximum utility* rule inherent in the traditional random utility model.

[2]For the earliest empirical application of a multi-attribute model to explain consumer choices among alternatives, see Huff (1962).

[3]Although these studies argued that consumers choose in a probabilistic manner, they did not attempt to disentangle such probabilistic decision-making on the part of consumers from the utilities themselves being random from an observational standpoint.

3 Proposed Random Utility Model of Brand Choice

To appreciate our proposed extension of the traditional random utility model, one can consider a simple example. Suppose a consumer is choosing between two brands: Coke and Pepsi. If the consumer's true utilities for the two brands (known to himself, but unknown to the marketing researcher), which are represented by u_{h1t} and u_{h2t} in the notation of the previous section, are 6 utils and 3 utils respectively. The *maximum utility* rule would say that the consumer would buy Coke with probability 1 (since it offers higher utility to the consumer than does Pepsi). The *probabilistic* choice rule would say that the consumer would buy Coke with some probability higher than 0.5 (since Coke offers the consumer higher utility than does Pepsi) but less than 1 (since Pepsi offers non-zero utility to the consumer).

Under the proposed formulation, the probability of the outcome $y_{ht} = j$ is given by

$$\Pr(y_{ht} = j|\delta_h, \Sigma) = \int_{A_{h1t}} \cdots \int_{A_{hjt}} \cdots \int_{A_{hJt}} f(\eta_{ht}|X_{ht}\delta_h, \Sigma) P_{hjt} d\eta_{ht} \tag{4}$$

where P_{hjt} is a function of $u_{hjt}, j = 1, ..., J$ that is more general than the indicator function I_{hjt}, and

$$A_{hjt} \quad = \quad (-\infty, \infty)(j = 1, ..., J)$$

We call equation (4) the general representation of our proposed random utility model. Different assumptions on P_{hjt} and the distribution $f(\eta_{ht})$ would yield different operational versions of our proposed random utility model. This model is more flexible than the traditional random utility model because one can allow P_{hjt} to take a variety of functional forms instead of restricting it to be the indicator function I_{hjt}, as explained in the earlier section. While η_{ht} captures the impact on consumers' utilities of many factors that are unobserved by the econometrician, P_{hjt} captures the inherent stochasticity in the consumer's brand choice decision.[4] Such competing drivers of stochasticity in observed choices of respondents were recognized in an early paper by Becker, DeGroot and Marschak (1963), who compared the theoretical specifications of *Luce models* and *Fechner models* (which we call *probabilistic choice* models in this paper) versus random utility

[4] P_{hjt} can also be interpreted as capturing the effects of "processing error "on the part of the consumer, see Loomes and Sugden (1995).

models (which we call *maximal utility* models in this paper). However, an explicit disentangling of the two drivers using empirical data on choice outcomes has not been attempted thus far. This paper addresses this gap in the literature by proposing and estimating a more general specification of the consumer's *decision rule* (referred to as the *individual behavior rule* by Louviere, Hensher and Swait 2000). This is the central contribution of this paper.

3.1 Operational Versions of our Proposed Model

Assuming that errors $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})$ are distributed iid gumbel with an unknown scale parameter μ (and location parameter equal to zero), we develop below various operational versions of our proposed random utility model by making different assumptions about the probability function P_{hjt}.

1. $P_{hjt} = I_{hjt}$, an *indicator function* that takes the value 1 if $u_{hjt} > \max_{k \neq j} u_{hkt}$, and the value 0 otherwise. This yields the familiar *Multinomial Logit* model (McFadden 1974).

2. $P_{hjt} = I_{hjt}$, an indicator function that takes the value 1 if u_{hjt} is at least an amount C greater than u_{hkt} for all $k \neq j$, and the value $(1/J)$ otherwise. This specification assumes that the consumer cannot recognize small utility differences, and that a threshold C must be crossed before the consumer applies the maximum utility decision rule to determine his optimal brand choice. A model in this spirit, called the *Minimum Perceived Difference* Model, has been estimated for a two-brand choice situation, using survey data, by Krishnan (1977).

3. P_{hjt} is specified exactly as in the previous case, except that the consumer is assumed to have a separate threshold parameter C_j for each brand in the product category.

4. P_{hjt} is a multinomial logistic function of $u_{hjt}, j = 1, ..., J$. This is an example of McFadden and Train's (2000) *Mixed Logit* model, where the mixture distribution is assumed to be Gumbel.

5. P_{hjt} is specified exactly as in the previous case (i.e., case 4), except that each brand's utility is associated with a different scale parameter in the multinomial logit function.

6. P_{hjt} is a multinomial probit function of $u_{hjt}, j = 1, ..., J$. This is an example of McFadden and Train's (2000) *Mixed Logit* model, where the mixture distribution is assumed to be Gaussian.

7. P_{hjt} is a binary logistic function of the difference between u_{hjt} and $Max_{k \neq j}(u_{hkt})$, where the scale parameter associated with the binary logistic function is different for different brands.

It is clear by now that one can develop a flexible range of brand choice models by making different assumptions about the consumer's probabilistic choice rule P_{hjt}. If the choice rule P_{hjt} nests the indicator function I_{hjt} as a special case, the proposed model would nest the traditional random utility model as a special case. We plot the last probability function among those listed above in Figure 1 to illustrate how it generalizes the indicator function of the traditional random utility model. From the figure, it is clear that the magnitude of $\mu > 0$ determines the steepness of the probability function, with $\mu = \infty$ representing the step function. These models capture two distinct sources of uncertainty in consumers' brand choices, one of which arises because there are variables unobserved by the researcher that drive consumers' utilities for brands (i.e., η_{hjt}), and the other arises because of the inherent stochasticity in the way consumers choose among available brands (i.e., P_{hjt}). This raises the question of whether the traditional random utility model (that is based on the *maximal utility* rule) can be modified to accommodate these two sources of uncertainty by simply changing the specification of the consumer's utility function, without changing the consumer's decision rule from *maximal utility* to *probabilistic choice*. We discuss this possibility in the next section.

3.2 Modified Random Utility Model

The household's latent utility u_{hjt} for brand j at time t can be assumed to be the following[5]:

$$u_{hjt} = \alpha_{hj} + \delta_{2h} Price_{hjt} + \delta_{3h} Disp_{hjt} + \delta_{4h} Feat_{hjt} + \epsilon_{1hjt} + \epsilon_{2hjt} \quad (5)$$

where ϵ_{1hjt} captures the influence of unobserved variables on the consumer's utility for brand j, while ϵ_{2hjt} captures the inherent stochasticity in the consumer's brand choice decision-making.

[5]We thank an anonymous reviewer for alerting us to this alternative conceptualization of the random utility model.

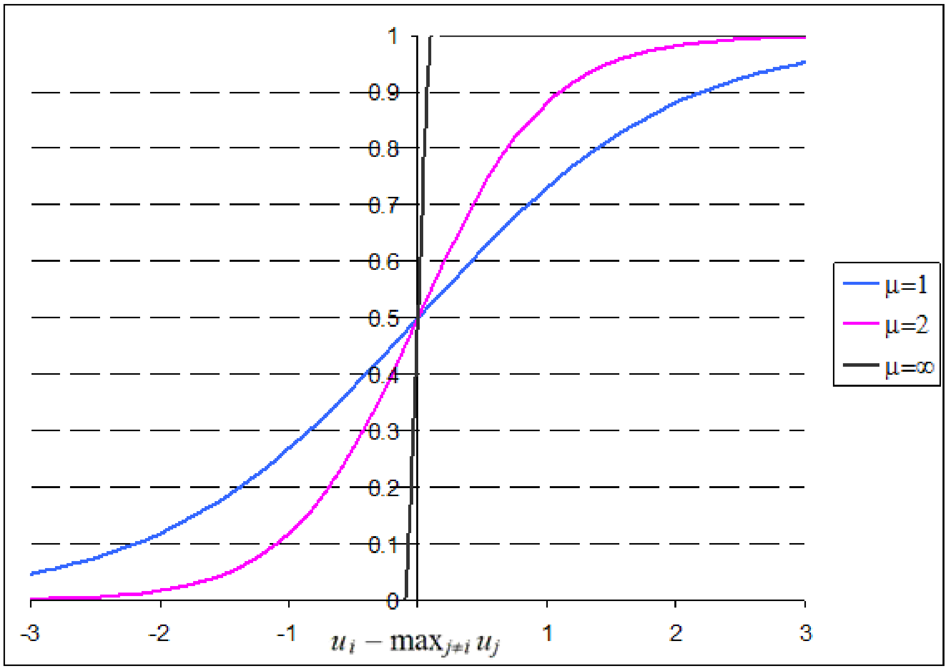

Figure 1: Probability Function

Making suitable parametric assumptions about the two error components - ϵ_{1hjt} and ϵ_{2hjt} - and invoking the maximal utility rule for the consumer's decision-making would lead to different operational versions of the modified random utility model. One could now argue that making flexible distributional assumptions on the two error components - ϵ_{1hjt} and ϵ_{2hjt} - could yield models that are more general than the traditional random utility model (that has only a single error component, η_{hjt}), while explicitly allowing for two distinct sources of uncertainty in the consumer's brand choice process (as in our proposed random utility model). This approach, called the *Mixed Logit* by McFadden and Train (2000), can, in fact, mimic operational versions 5 and 6 of our proposed model (see previous section).[6] There are at least two criticisms of this approach: 1) unless one chooses the distributions of ϵ_{1hjt} and ϵ_{2hjt} very carefully, it would be impossible to separately identify the parameters of the two error components (because they are linear and additive); and 2) this model is not as general as our proposed model. Choice probabilities generated by *any* arbitrary version of the modified random utility model can be

[6]The *Mixed Logit* model has been used in marketing and economics only to accommodate the effects of unobserved heterogeneity across households, i.e., at the household-level, and not at the choice observational-level as we argue in this section.

generated by a suitably parameterized version of our proposed model. The reason for this is that the modified random utility model is a *special case* of our proposed model when (1) the distribution of η_{hjt} in our proposed model is identical to the distribution of the sum of ϵ_{1hjt} and ϵ_{2hjt} under the modified random utility model, and (2) the consumer's decision rule in our proposed model is the *maximal utility* rule. Conversely, however, choice probabilities generated by any arbitrary version of our proposed model – under general assumptions about the consumer's decision rule P_{hjt} – cannot be mimicked by a suitably parameterized version of the modified random utility model. The reason for this is that the modified random utility model – with linear, additive structure of the household's latent utility function (see equation 5) – can only yield choice probabilities that depend on *differences* between the household's deterministic utilities, i.e., $v_{hjt} = \alpha_{hj} + \delta_{2h}Price_{hjt} + \delta_{3h}Disp_{hjt} + \delta_{4h}Feat_{hjt}$, for brands. Our proposed model, however, can generate choice probabilities that depend on *cardinal values* of the household's deterministic utilities of brands. For example, consider a household h's choice between two brands $j = 1, 2$ at time t. Choice probabilities for brand 1 that are yielded by the modified random utility model will always take the form $Pr_{h1t} = Pr[(\epsilon_{h1t} - \epsilon_{h2t}) > (v_{h2t} - v_{h1t})]$, where ϵ_{h1t} is the composite error that represents the sum of ϵ_{1h1t} and ϵ_{2h1t}. These probabilities depend on (and only on) the differences in deterministic utilities $v_{h2t} - v_{h1t}$. An estimable choice probability such as $Pr_{h1t} = \frac{v_{h1t}}{(v_{h1t} + v_{h2t})}$, cannot result from the modified random utility model. Our proposed model, however, can accommodate not only this type of choice probability but also allow for choice probabilities that are arbitrary functions of the cardinal values of the deterministic utilities, v_{h1t} and v_{h2t}.[7] In summary, a two-stage choice model, such as our proposed model, is not only statistically more flexible, but also richer in behavioral detail than a one-stage model such as the modified random utility model. For these reasons, we advocate the use of our proposed random utility model over the modified random utility model.

3.3 Estimation

The proposed random utility model of brand choice can be estimated using scanner panel data on households' brand choices within a product category. Suppose household h contributes T_h brand choice observations within the category. The household's likelihood function can then be

[7] It is useful to note that for this reason, while the *Mixed Logit* model (McFadden and Train 1986) is a special case of our proposed model, it also cannot yield all possible choice probabilities yielded by our model.

specified as follows.

$$L_h = \prod_{t=1}^{T_h} \prod_{j=1}^{J} Pr(y_{hjt} = 1)^{y_{hjt}}$$

and the sample likelihood function can be specified as follows.

$$L = \prod_{h=1}^{H} L_h$$

One can incorporate unobserved heterogeneity across households by assuming that the parameters of the brand choice model follow a semi-parametric mixture distribution across households, whose locations and probability masses are estimated using the available data (Heckman and Singer 1984). This yields the following sample likelihood function.

$$L = \prod_{h=1}^{H} \prod_{s=1}^{S} f_s L_{hs}$$

where S stands for the number of supports of the unobserved heterogeneity distribution, f_s is the probability mass associated with support s, and L_{hs} is the likelihood function of household s computed using the location of support s.

4 Empirical Illustration on Scanner Panel Data

For empirical purposes, we pick a specific operational version of our proposed random utility model. We assume that errors $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})$ are distributed iid gumbel with an unknown scale parameter μ (and location parameter equal to zero)[8], P_{hjt} is a binary logistic function of the difference between u_{hjt} and $Max_{k \neq j}(u_{hkt})$, where the scale parameter associated with the binary logistic function is different for different brands[9]. Under this operational model, the probability of the outcome $y_{ht} = j$ can be shown to be

$$\Pr(y_{ht} = j|\delta_h, \Sigma) = \int_{-\infty}^{\infty} f(\zeta_{hjt}|X_{ht}\delta_h, \Sigma) \frac{1}{1 + e^{-\mu_j (X_{hjt}\delta_h - ln \sum_k e^{X_{hkt}\delta_h} + \zeta_{hjt})}} d\zeta_{hjt}, \quad (6)$$

[8]In the appendix, we report the results for another operational version of our proposed random utility model that assumes that errors $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})$ are distributed $N(0, \Sigma)$.

[9]The J binary logit probabilities must be scaled to sum to one.

where ζ_{hjt} is a logistic random variable with location 0 and scale 1. The basis for the above equation are the following two properties of the Gumbel distribution: (1) the maximum of $(J-1)$ independent Gumbel variates u_{hkt} with a common scale parameter and different location parameters is also distributed Gumbel with the same scale parameter, but with location parameter equal to the logarithm of the sum of the $(J-1)$ exponentiated location parameters; and (2) the difference between two gumbel variates with common scale is distributed logistic, with location equal to the difference in the location parameters of the two gumbel variates, and the same scale as the gumbel variates (see Ben-Akiva and Lerman 1985). The univariate integral in this equation does not have a closed-form, but can be evaluated using Gaussian quadrature methods. This operational version of our proposed random utility model (PRU henceforth) generalizes the widely used Multinomial Logit (MNL henceforth) model of brand choice. It would reduce to the MNL only if $\mu_j \longrightarrow \infty$ for all brands $j = 1, ..., J$. We estimate both PRU and MNL on scanner panel data in order to demonstrate the empirical gains from adopting PRU over the MNL.

We employ A.C. Nielsen's scanner panel database on household purchases in Sioux Falls, South Dakota. For our analysis, we choose the ketchup category. The dataset covers a period of 123 weeks from 1985 to 1987. Choosing households that bought only among the top eight brand sizes of ketchup (that account for 87 percent of all product sales in the ketchup category) yields 3032 households. From these households, we use only those that made at least seven purchases over the study period. This yields a sample of 529 households making a total of 5954 purchases in the category. Descriptive statistics are provided in Table 1. Among the eight brands, Heinz 32 oz has the highest market share (37.8 percent). Del Monte 32 oz has fairly high display and feature activity (25 and 29 percent respectively) compared to the other brands, as does Hunts 32 oz (30 and 14 percent respectively).

First, in order to understand the explanatory power of the proposed PRU model vis-a-vis the benchmark MNL model, we compute commonly used statistical criteria of model fit, specifically the maximized log-likelihood (LL), Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC). The fit values are reported in Table 2.[10]

It is clear that PRU outperforms MNL, in terms of both explaining the observed brand

[10] These fit values are obtained by estimating both models based on a 4-support discrete distribution for unobserved heterogeneity.

Brand	*Share*	*Price*	*Display*	*Feature*
Heinz 32 oz	0.378	0.0365	0.1649	0.1631
Control Brand 32 oz	0.176	0.0254	0.1878	0.1058
Hunts 32 oz	0.149	0.0363	0.2991	0.1371
Heinz 28 oz	0.107	0.0510	0.0534	0.0556
Del Monte 32 oz	0.060	0.0363	0.2536	0.2892
Heinz 14 oz	0.052	0.0532	0.0165	0.0506
Heinz 44 oz	0.057	0.0419	0	0.1530
Heinz 64 oz	0.021	0.0423	0	0.0969

Table 1: *Descriptive Statistics.*

FitCriterion	*PRU*	*MNL*
In-sample LL	-7662	-7757
In-sample AIC	15426	15600
In-sample SBC	15767	15887
Validation LL	-1700	-1719

Table 2: *Fit Statistics.*

choices in the product category, and validating the holdout choices[11]. We present the parameter estimates (along with their standard errors) of PRU and MNL in Tables 3 and 4 respectively.

The estimated values of μ_j are quite small in magnitude[12] (varying from 1.27 to 2.39), which indicates that consumers use “smoother ”decision rules than the maximum utility rule implied by traditional random utility models such as the MNL. Between the PRU and MNL, we find, as one would expect, that the estimated brand constants are, by and large, understated for the large share brands and overstated for the small share brands under the MNL. We also find that the marketing mix coefficients are understated in magnitude under the MNL model. In fact, when we simulate brand choice data using the PRU model, and estimate the MNL model on the simulated dataset, we find that the estimated marketing mix coefficients steadily decrease in magnitude, relative to their true counterparts, as the magnitudes of the assumed μ_j's are decreased in the simulation[13].

In order to interpret the estimated price parameters in more managerially understandable

[11] The holdout validation is performed by estimating the model using 80% of the observations, and computing the log-likelihood for the remaining 20% of the observations using the estimated parameters.

[12] Remember that as $\mu_j \longrightarrow \infty \; \forall j$, PRU reduces to MNL.

[13] The detailed simulation results are available from the authors.

Parameter	*Support*1	*Support*2	*Support*3	*Support*4
Heinz 32 oz	3.32	3.34	1.43	0.18
Control Brand 32 oz	-0.44	4.01	-3.00	-2.02
Hunts 32 oz	1.84	3.14	-1.31	1.54
Heinz 28 oz	2.13	2.14	0.28	2.04
Del Monte 32 oz	0.47	2.47	-3.18	-1.98
Heinz 14 oz	0.71	1.64	1.89	-1.28
Heinz 44 oz	0.95	0.20	1.86	-2.36
Heinz 64 oz	0	0	0	0
Price	-2.75	-3.00	0.50	0.74
Display	0.61	0.54	0.91	0.24
Feature	0.40	0.48	-0.48	-0.14
$\mu_{Heinz32}$	1.30	1.30	1.30	1.30
μ_{CB32}	1.62	1.62	1.62	1.62
$\mu_{Hunts32}$	2.20	2.20	2.20	2.20
$\mu_{Heinz28}$	1.40	1.40	1.40	1.40
μ_{DM32}	1.27	1.27	1.27	1.27
$\mu_{Heinz14}$	1.73	1.73	1.73	1.73
$\mu_{Heinz44}$	1.64	1.64	1.64	1.64
$\mu_{Heinz64}$	2.39	2.39	2.39	2.39
Probability	0.54	0.27	0.10	0.09

Table 3: *Estimation Results for PRU Model* .

terms, we convert them to price-elasticities. Given in Tables 5 and 6 are the results of this exercise (each row refers to a brand changing its price, while each column refers to a brand whose share is responding to the price change). No systematic pattern emerges when we compare the elasticities between the PRU and MNL models. Therefore, we construct clout and vulnerability measures for each brand (as discussed in Kamakura and Russell 1989), and plot these measures in Figure 2. We find that the clout and vulnerability of the largest share brand - Heinz 32 oz. - are overstated under the MNL. In contrast, for Hunts 32 oz., the clout and vulnerability measures are understated under the MNL.

Parameter	*Support1*	*Support2*	*Support3*	*Support4*
Heinz 32 oz	3.01	3.27	1.35	0.30
Control Brand 32 oz	-0.21	3.67	-2.97	-1.97
Hunts 32 oz	1.55	2.99	-1.24	1.41
Heinz 28 oz	1.98	2.31	0.46	1.70
Del Monte 32 oz	0.64	2.57	-3.15	-1.92
Heinz 14 oz	0.78	1.75	1.69	-1.16
Heinz 44 oz	0.98	0.31	1.69	-2.31
Heinz 64 oz	0	0	0	0
Price	-2.18	-2.45	0.46	0.71
Display	0.51	0.35	0.91	0.20
Feature	0.30	0.36	-0.42	-0.09
Probability	0.55	0.27	0.10	0.08

Table 4: *Estimation Results for MNL Model* .

Brand	*He32*	*CB32*	*Hu32*	*He28*	*DM32*	*He14*	*He44*	*He64*
He32	-0.83	0.46	0.94	0.62	0.75	0.25	0.61	0.64
CB32	0.17	-1.23	0.50	0.17	0.67	0.12	0.16	0.18
Hu32	0.28	0.48	-2.48	0.29	0.54	0.16	0.38	0.34
He28	0.15	0.11	0.23	-1.63	0.25	0.05	0.19	0.09
DM32	0.14	0.29	0.24	0.14	-2.71	0.07	0.15	0.15
He14	0.03	0.04	0.07	0.02	0.06	-0.77	-0.03	-0.01
He44	0.12	0.06	0.20	0.12	0.17	-0.06	-1.59	0.10
He64	0.03	0.02	0.06	0.02	0.06	-0.00	0.04	-1.88

Table 5: *Price Elasticities under the PRU Model* .

Brand	*He32*	*CB32*	*Hu32*	*He28*	*DM32*	*He14*	*He44*	*He64*
He32	-1.11	0.63	0.84	0.64	0.94	0.33	0.62	0.60
CB32	0.30	-1.32	0.47	0.17	0.71	0.15	0.12	0.14
Hu32	0.36	0.41	-2.04	0.14	0.39	0.12	0.20	0.14
He28	0.24	0.12	0.12	-1.35	0.20	0.04	0.12	0.03
DM32	0.17	0.26	0.16	0.10	-2.92	0.06	0.09	0.09
He14	0.06	0.05	0.05	0.02	0.06	-0.85	-0.04	-0.00
Hu44	0.13	0.04	0.10	0.07	0.10	-0.04	-1.37	0.05
He64	0.05	0.02	0.03	0.01	0.04	-0.00	0.02	-1.33

Table 6: *Price Elasticities under the MNL Model* .

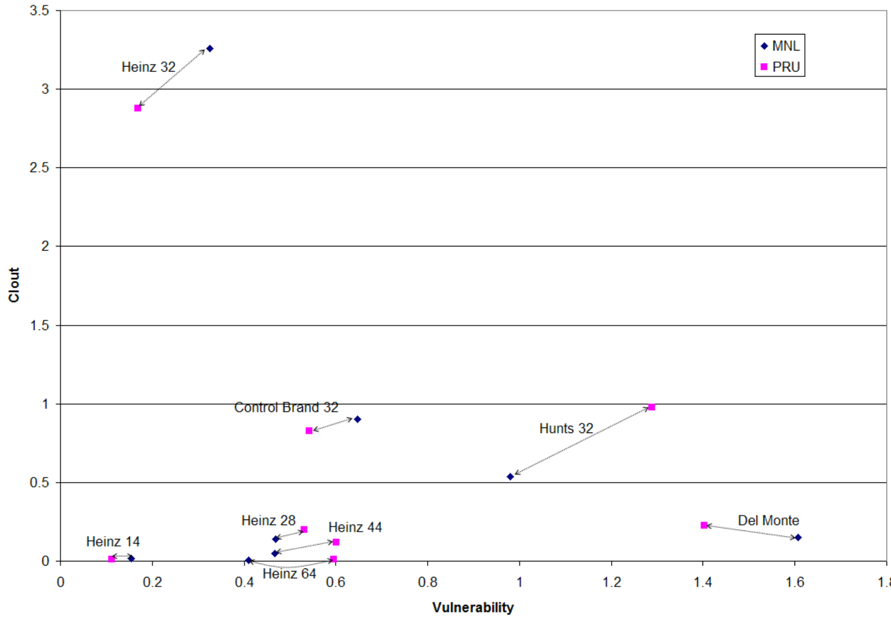

Figure 2: Clout and Vulnerability

5 Empirical Illustration on Wager Data

In order to investigate the viability of our proposed modeling framework when consumers make choices in a laboratory setting (as opposed to the real-world setting represented in scanner panel data), we use data from an experiment involving respondents choosing among wagers. Each subject is offered a choice between alternative gambles that differ in terms of risk and expected value. The wagers are constructed to vary systematically in terms of these two component attributes. The choice sets are constructed to vary systematically in terms of the number and type of wagers offered. We specify the experimental design below.

5.1 Experimental Design

A wager is a vector of six numbers. Subjects were repeatedly offered choice sets containing either two or three wagers and asked to choose one. After making his choice, the subject rolled a die. The outcome of the roll determined the amount won on that choice. For example, suppose John Doe was presented with the following choice set composed of two wagers – [24, 62, 44, 32, 36, 42] versus [51, 29, 71, 04, 18, 67] – and indicated his choice by placing an X in the blank below the first wager. If John Doe then rolled a 4, he would win 32 points toward his total winnings; if he rolled a 6, he would win 42 points toward his total winnings, etc.

13 wagers were constructed for use in the choice experiment. Each wager can be described by 2 attribute values – one indicating the expected value of the wager and calculated by taking the average of the six possible winnings, and the other indicating the risk of the wager and calculated by taking the variance of the six possible winnings. For example, the two wagers used to describe the above-mentioned choice task are characterized as (40, 9) and (40, 27) in terms of expected value and variance respectively. Other things equal, one would expect subjects to choose the wager with the highest expected value. When variances differed, however, a subject's tendency to gamble was expected to affect the choice – "risk seekers "usually choosing the wager with the highest variance, and "risk averters "usually choosing the wager with the lowest variance. The wagers were characterized by one of three expected values (40, 44, 48) and one of three variances (9, 18, 27). Using the 9 possible wagers that can be obtained by combining the three expected values and three variances, and also using 4 additional wagers that were "interior "to the expected values and variances represented in the 9 wagers[14], we constructed 106 choice sets [15](see Table 7 for a description of the 13 wagers), with 64 containing 2 wagers and 42 containing 3 wagers. Subjects were offered each of the 106 choice sets a total of 6 times (which yielded us 636 observations for each subject). Subjects chose from all 106 choice sets before any set was offered a second time, then all 106 a second time before the third offering etc. To preclude wager position bias, the order of the wagers in the choice sets was random and balanced. To preclude winning-position bias, the position of the six reward points was balanced and random. Finally, the order of the presentation of a choice set within the 106 was randomly determined.

13 subjects in the experiment received a payment for participation, their payment being determined solely on the basis of the number of reward units they won during the experiment. Because they were volunteers, the subjects do not constitute a representative sample for any population of interest. Projection to some population of interest, however, was not the purpose of the study. The method allows us to test the choice process for individual separately (since we have 636 observations for each subject, unlike in scanner panel data where we rarely had more than 10 observations from a panelist), and results are reported at the individual level.

[14]The "interior "wagers involved two expected values (42, 46) and two variances (14, 23).

[15]The descriptions of the 106 choice sets are available from the authors.

We use expected value (EV) and variance (V) as the two attributes of interest in analyzing the experimental choice data.

$Wager$	EV	V
1	40	9
2	44	9
3	48	9
4	40	18
5	44	18
6	48	18
7	40	27
8	44	27
9	48	27
10	42	14
11	46	14
12	46	23
13	42	23

Table 7: *Description of Wagers.*

5.2 Multi-Attribute Random Utility Model

Consider subject h ($h = 1, 2, ..., 13$) observed over $t = 1, 2, ..., 636$ choice tasks, where each choice task involves 3 wagers.[16] On any given choice task t, one observes an outcome variable y_{ht} which takes the value j, $j = 1, 2, 3$ where j identifies the wager chosen by subject h in that choice task. For every choice task t, one observes the expected value (EV_{hjt}) and variance (V_{hjt}) covariates faced by the subject for all wagers represented in the choice task. We label the five possible expected values as $EV_1 = 40$, $EV_2 = 44$, $EV_3 = 48$, $EV_4 = 42$, $EV_5 = 46$. We label the five possible variances as $V_1 = 9$, $V_2 = 18$, $V_3 = 27$, $V_4 = 14$, $V_5 = 23$. The goal is to model the outcome variable (y_{ht}) given the above information.

The random utility model is based on a latent variable u_{hjt} that denotes the utility of subject h for wager j in choice set t. It is further assumed that this utility can be expressed as function of the set of wager covariates facing the subject in choice set t, in the following manner:

$$u_{hjt} = \sum_{k=1}^{4} \delta_{1kh} IEV_{hjkt} + \sum_{k=1}^{4} \delta_{2kh} IV_{hjkt} + \eta_{hjt}$$

[16]The choice task involving 2 wagers can be represented by setting the consumer's utility for one wager (i.e., the missing wager) to be $-\infty$.

where

$$\delta_h = (\delta_{11h}, \delta_{12h}, \delta_{13h}, \delta_{14h}, \delta_{21h}, \delta_{22h}, \delta_{23h}, \delta_{24h})'$$

are household-specific coefficients, IEV_{hjkt} is an indicator variable that takes the value 1 if $EV_{hjt} = EV_k$ and 0 otherwise, IV_{hjkt} is an indicator variable that takes the value 1 if $V_{hjt} = V_k$ and 0 otherwise, and η_{hjt} is a random error such that $\eta_{ht} = (\eta_{h1t}, \eta_{h2t}, \eta_{h3t})'$ has a joint distribution $f(\eta_{ht})$ with continuous support, mean zero and covariance matrix Σ.

As in the scanner data example, we model the wager level outcome y_{ht} as being determined by either the decision rule of *maximum utility* or the decision rule of *probabilistic choice*. We assume $f(\eta_{ht})$ to be iid Gumbel, which renders the traditional random utility model to be the MNL.[17] Under this distributional assumption for the error terms, we make the following assumption on the consumer's probabilistic choice rule: $P_{hjt} \propto 1/(1+\mu_{hjt}[u_{hjt} - Max_{k\neq j}(u_{hkt})])$, where the scale parameter $\mu_{hjt} = \sum_{k=1}^{5} \mu_{1kh} IEV_{hjkt} + \sum_{k=1}^{5} \mu_{2kh} IV_{hjkt} > 0$ is allowed to depend on the (Expected Value, Variance) combination that represents wager j. Ten scale parameters – $\{\mu_{1kh}, k = 1, ..., 5\}$ and $\{\mu_{2kh}, k = 1, ..., 5\}$ – are explicitly estimated (instead of being assumed to be ∞ as in the traditional random utility model).[18] This decision rule assumption is made in the spirit of the probabilistic choice rule tested on scanner data.

5.3 Results

We report the fit values (i.e., maximized LL) associated with both our proposed model (PRU), as well as the MNL, in Table 8. Our model outperforms the MNL – in terms of better explaining a consumer's observed brand choices – for all 13 consumers in the data. This provides strong empirical support for our proposed model. Our proposed model shows the most pronounced gain in fit over the MNL for consumers 12 and 5 (the gains being 15% and 10%, respectively, in terms of the improved LL). Taken in conjunction with the results from scanner panel data (discussed earlier), this shows that our proposed model has superior *external* and *internal* validity over the MNL. In other words, even if one could argue that the documented superiority of our proposed model over the MNL in scanner panel data could be partly due to the two-stage model being better able to accommodate the effects of residual heterogeneity across households that remains

[17] The iid normal assumption will yield the MNP. We also fit the MNP model on the wager data and found its fit to be almost identical to the MNL, which implied that the gumbel versus normal distinction was irrelevant for the data.

[18] The three wager choice probabilities are normalized to add up to one.

after employing the latent class specification of heterogeneity under the MNL, the documented superiority of our proposed model over the MNL using individual-level wager data strongly lays to rest such a claim. Our experimental results nicely vindicate the *theoretical* superiority of our proposed model, i.e., that consumers employ probabilistic choice rules while choosing among brands, over the traditional random utility model (MNL).

Consumer	*PRU*	*MNL*
Consumer 1	-333	-350
Consumer 2	-441	-447
Consumer 3	-376	-383
Consumer 4	-256	-261
Consumer 5	-216	-239
Consumer 6	-180	-191
Consumer 7	-477	-495
Consumer 8	-172	-185
Consumer 9	-434	-441
Consumer 10	-302	-322
Consumer 11	-494	-501
Consumer 12	-188	-221
Consumer 13	-390	-397

Table 8: *Fit Statistics.*

We report the estimates of the utility coefficients associated with the expected values of the wagers, i.e., δ_{1kh}, in Table 9. For all 13 consumers, the coefficients associated with $EV = 40$, $EV = 44$, and $EV = 48$, are correctly ordered, i.e., $\delta_{11h} < \delta_{12h} < \delta_{13h}$, which suggests that among wagers of equal variances, consumers prefer the wager with the highest expected value. This gives strong face validity to our estimates.

We report the estimates of the utility coefficients associated with the variances of the wagers, i.e., δ_{2kh}, in Table 10. We find that consumers 2, 4 and 8 are risk-averse (i.e., prefer wagers of low variance), consumers 1, 5, 7, 9, 10, 11, 12 are risk-seeking (i.e., prefer wagers of high variance), while consumers 3, 6 and 13 prefer an intermediate amount of risk. These estimated differences among consumers in risk-aversion are interesting, and also to be expected.

We report the estimates of the scale coefficients associated with the expected values of the wagers, i.e., μ_{1kh}, in Table 11. Interestingly, we see that none of the five expected values is uniformly associated with an estimated scale value of similar magnitude across the 13 consumers.

For example, consumer 7 is very likely to use the maximal utility rule if the most preferred option in a wager has an expected value of 46 or 48 (i.e., the two highest expected values), but less likely to use the rule (and would more probabilistically choose among the options) if the most preferred option in a wager takes one of the three smaller values, i.e., 40, 44 or 42. Consumer 8, on the other hand, uses the maximal utility rule only if the most preferred option in a wager has an expected value of 42. More interestingly, consumer 12 is found to maximally "disobey"the maximal utility rule (i.e., $\mu \to 0$) if the most preferred option in a wager has an expected value of 48 or 46. We report estimates of the scale coefficients associated with the variances of the wagers, i.e., μ_{2kh}, in Table 12. Again, as with the case of the scales associated with the expected values (in Table 11), no uniform pattern emerges across the 13 consumers. One can again readily see that most of the estimated scale coefficients are indeed significantly lower than infinity (as assumed in the traditional random utility model).

Consumer	$EV = 40$	$EV = 44$	$EV = 48$
Consumer 1	-3.1211	0.6997	2.4205
Consumer 2	-1.2598	-0.0073	1.2831
Consumer 3	-1.8898	-0.0958	1.9892
Consumer 4	-1.1816	-0.0067	1.1903
Consumer 5	-3.3966	0.5701	2.8269
Consumer 6	-9.8461	1.6361	8.2099
Consumer 7	-1.1142	0.0503	1.0486
Consumer 8	-13.5535	0.0418	13.5050
Consumer 9	-2.1886	0.3191	1.8449
Consumer 10	-4.9386	1.2278	3.6699
Consumer 11	-1.1707	-0.2010	1.3471
Consumer 12	-11.2041	-6.9729	18.1965
Consumer 13	-0.9355	-0.1355	1.1525

Table 9: *Individual-Level Estimation Results: Utility Coefficients for EV.*

In order to further support our claim that probabilistic choice rules are indeed employed by subjects while choosing among wagers, we summarize the degree of observed stochasticity in consumer choices (across six repetitions of identical choice sets) in Table 13. We use three different summary measures: 1. CONSISTENCY, which captures the percentage of times a subject chooses the same wager in all six repetitions of the same choice set, 2. INCONSISTENCY, which captures the percentage of times a subject chooses all available wagers in a choice set

Consumer	$V = 9$	$V = 18$	$V = 27$
Consumer 1	-1.0832	-0.0739	1.1578
Consumer 2	0.9278	0.4307	-1.3607
Consumer 3	-0.0129	0.7299	-0.7183
Consumer 4	2.2456	0.0003	-2.2469
Consumer 5	-3.0233	0.6187	2.4047
Consumer 6	-0.0433	0.3995	-0.3530
Consumer 7	-0.6884	0.2195	0.4544
Consumer 8	6.4810	-2.0703	-4.4094
Consumer 9	-0.4702	-0.2748	0.7458
Consumer 10	-1.7786	0.4637	1.2915
Consumer 11	-0.0782	-0.2826	0.3699
Consumer 12	-5.4344	2.0001	3.3810
Consumer 13	0.4841	0.6375	-1.1330

Table 10: *Individual-Level Estimation Results: Utility Coefficients for V.*

across the six repetitions of the choice set, and 3. HHI, also called the Herfindahl index, which captures the sum of squares of the fraction of times a subject chooses each wager in a choice set, averaged across the 106 choice sets in the study (maximal consistency in choices would imply an HHI value of 1, while minimal consistency would imply an HHI value of 0.33 in a choice set with 3 wagers, and 0.5 in a choice set with 2 wagers). Based on the three measures, it is clear that consumers make different wager choices under identical circumstances (i.e., when offered the same choice set at six different points of time) in a controlled laboratory setting. This offers compelling evidence of the inherent stochasticity in consumers' choice decisions, estimated on the experimental choice data. This explains why we are able to uncover two layers of uncertainty (i.e., random utility errors, plus probabilistic choice rule) in this choice situation.

6 Conclusions

In this paper, we introduce the notion of probabilistic choice rules within a random utility model of households' brand choice behavior. We uncover empirical evidence supporting the existence of one such probabilistic rule using scanner panel data on consumers' brand choices in the ketchup category. We demonstrate the implications of ignoring such probabilistic choice rules on the estimated price-elasticity measures.

To supplement the external validity of our proposed model – since we uncover two layers of

Consumer	$EV = 40$	$EV = 44$	$EV = 48$	$EV = 42$	$EV = 46$
Consumer 1	0.4733	1.9238	52.1842	19.6246	0.7248
Consumer 2	0.6903	1.6446	126.0594	13.0205	1.0850
Consumer 3	2.0379	2.1844	0.3678	3.8442	0.1358
Consumer 4	32.6598	2.4381	0.4462	1.1505	0.6423
Consumer 5	0.3481	20.7519	17.9172	0.5981	4.6365
Consumer 6	9.4977	30.4488	0.0933	110.0376	0.4108
Consumer 7	1.1455	0.5938	41.2150	0.5955	16.2930
Consumer 8	0.5739	0.7005	1.4952	414.1679	0.1282
Consumer 9	1.1020	1.7254	0.4141	0.9390	0.6549
Consumer 10	0.1250	0.4837	768.1335	994.8476	0.3873
Consumer 11	0.8486	0.7228	22.6834	1.4271	0.2087
Consumer 12	7.9591	2.3978	0.00004	146.8848	0.00001
Consumer 13	19.5950	28.9636	0.0392	20.9525	0.2612

Table 11: *Individual-Level Estimation Results: Scale Coefficients for EV.*

choice uncertainty: (1) due to unobserved variables, and (2) due to the inherent stochasticity in consumers' choices, using real world scanner panel data, – we test the internal validity of our proposed model using experimental choice outcomes where subjects choose among wagers. We uncover compelling evidence in favor of our proposed two-stage model of consumer choice in the experimental context as well.

We hope that our work spurs further research on the development of more flexible random utility models. First, it is useful to investigate alternative specifications of P_{hjt} within our proposed random utility model. Second, there are alternative decision rules such as satisficing, conjunctive, disjunctive, lexicographic, elimination by aspects etc. that may better explain observed consumer choices than both the widely used maximum utility rule and our proposed probabilistic choice rule at least in some contexts. Understanding what these contexts are, and whether different consumers use different decision rules for the same context, are both important issues for future research.

Consumer	$V = 9$	$V = 18$	$V = 27$	$V = 14$	$V = 23$
Consumer 1	1.1926	1.1240	35.4478	22.9167	0.6207
Consumer 2	1.0709	117.9277	1.1326	1.1346	12.4512
Consumer 3	0.4056	2.0616	1.9581	0.2726	1.9156
Consumer 4	0.7975	1.2575	35.4231	0.5097	1.4498
Consumer 5	1.1258	2.9459	39.0208	2.9078	0.9537
Consumer 6	0.0300	0.6982	1287.6220	115.6316	0.3909
Consumer 7	1.5329	46.5168	0.3931	0.6788	14.2945
Consumer 8	0.3674	0.2566	6.3775	0.1253	423.8330
Consumer 9	0.5253	2.3991	0.6248	0.4016	1.5315
Consumer 10	2.3904	5.2447	3.7037	221.6516	1.7362
Consumer 11	0.5798	2.0937	11.4579	0.2218	1.3423
Consumer 12	0.0574	0.0684	0.1902	0.0014	1.3189
Consumer 13	210.6822	1.5781	0.0670	1.1962	4.5751

Table 12: *Individual-Level Estimation Results: Scale Coefficients for V.*

Consumer	*CONSISTENCY*	*INCONSISTENCY*	*HHI*
Consumer 1	37.74%	32.08%	0.754
Consumer 2	16.98%	47.17%	0.647
Consumer 3	27.36%	33.02%	0.704
Consumer 4	49.06%	21.70%	0.797
Consumer 5	60.38%	15.09%	0.846
Consumer 6	66.04%	12.26%	0.849
Consumer 7	13.21%	55.66%	0.628
Consumer 8	64.15%	12.26%	0.874
Consumer 9	18.87%	45.28%	0.645
Consumer 10	45.28%	27.36%	0.782
Consumer 11	8.49%	61.32%	0.590
Consumer 12	66.98%	13.21%	0.887
Consumer 13	18.87%	47.17%	0.601

Table 13: *Summary Measures of Inherent Stochasticity in Consumer Choices.*

References

Becker, G.M., M.H. DeGroot and J. Marschak (1963). Stochastic Models of Choice Behavior, Behavioral Science, 8, 41-55.

Bass, F.M. (1974). The Theory of Stochastic Preference and Brand Switching, Journal of Marketing Research, 11, 1, 1-20.

Ben-Akiva, M. and S.R. Lerman (1985). Discrete Choice Analysis, Cambridge: MIT Press.

Daganzo, C. (1979). Multinomial Probit: The Theory and its Applications to Demand Forecasting, New York: Academic Press.

Edwards, W. (1955). The Predictions of Decisions Among Bets, Journal of Experimental Psychology, 50, 201-214.

Guadagni, P.E. and J.D.C. Little (1983). A Logit Model of Brand Choice Calibrated on Scanner Data, Marketing Science, 2, 3, 203-238.

Heckman, J.J. and B. Singer (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration, Econometrica, 52, 3, 271-320.

Huff, D.L. (1962). A Probabilistic Analysis of Consumer Spatial Behavior, in Emerging Concepts in Marketing, W.S. Decker, ed. Chicago: American Marketing Association.

Kamakura, W.A. and G.J. Russell (1989). A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, Journal of Marketing Research, 26, 4, 379-390.

Krishnan, K.S. (1977). Incorporating Thresholds of Indifference in Probabilistic Choice Models, Management Science, 23, 11, 1224-1233.

Loomes, G. and R. Sugden (1995). Incorporating a Stochastic Element into Decision Theories, European Economic Review, 39, 641-648.

Louviere, J.J., D.A. Hensher and J.D. Swait (2000). Stated Choice Methods: Analysis and Application, Cambridge: Cambridge University Press.

Louviere, J.J., D. Street, R. Carson, A. Ainslie, J.R. Deshazo, T. Cameron, D.A. Hensher, R. Kohn and T. Marley (2002). Dissecting the Random Component of Utility, Marketing Letters, 13, 3, 177-193.

Luce, R. (1959). Individual Choice Behavior: A Theoretical Analysis, New York: Wiley Press.

Luce, R. and P. Suppes (1965). Preference, Utility and Subjective Probability, in Handbook of Mathematical Psychology, Vol. 3, R. Luce, R. Bush and E. Galanter, eds., New York: Wiley Press.

Manski, C. (1975). Maximum Score Estimation of the Stochastic Utility Model of Choice, Journal of Econometrics, 3, 205-228.

Marschak, J. (1960). Binary Choice Constraints on Random Utility Indicators, in Stanford Symposium on Mathematical Methods in the Social Sciences. K. Arrow, ed., Stanford: Stanford University Press.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, in Frontiers of Econometrics. P. Zarembka, ed., New York: Academic Press.

McFadden, D. and K. Train (2000). Mixed MNL Models for Discrete Response, Journal of Applied Econometrics, 15, 5, 447-470.

Mosteller, F. and P. Nogee (1951). An Experimental Measurement of Utility, Journal of Political Economy, 59, 371-404.

Thurstone, L. (1927). A Law of Comparative Judgment, Psychological Review, 34, 273-286.

7 Appendix - I: Empirical Illustration on Scanner Panel Data Using the Multinomial Probit Framework

For empirical purposes, we pick a specific operational version of our proposed random utility model. We assume that errors $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hJt})$ are distributed $N(0, \Sigma)$, where Σ is a diagonal matrix (with its J-th element fixed at 1 for identification), and P_{hjt} is a binary logistic function of the difference between u_{hjt} and $Max_{k \neq j}(u_{hkt})$, where the scale parameter associated with the binary logistic function (i.e., μ_j is different for different brands[19]. Under this operational model, the probability of the outcome $y_{ht} = j$ can be shown to be

$$\Pr(y_{ht} = j | \delta_h, \Sigma) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_{h1t}, ..., u_{hJt} | X_{ht}\delta_h, \Sigma) \frac{1}{1 + e^{-\mu_j(u_{hjt} - u_{hkt})}} d\eta_{h1t} \cdots d\eta_{hJt} \quad (7)$$

The multivariate integral in the above equation, which is J-dimensional and does not have a closed-form, can be evaluated using Monte Carlo Simulation. This operational version of our proposed random utility model (PRU henceforth) generalizes the widely used Multinomial Probit (MNP henceforth) model of brand choice. We estimate both PRU and MNP on scanner panel data in order to demonstrate the empirical gains from adopting PRU over the MNP.

First, in order to understand the explanatory power of the proposed PRU model vis-a-vis the benchmark MNP model, we compute commonly used statistical criteria of model fit, specifically the maximized log-likelihood (LL), Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC). The fit values for the homogeneous versions of both models are reported in Table 11.

FitCriterion	*PRU*	*MNP*
In-sample LL	-9105	-9226
In-sample AIC	18260	18486
In-sample SBC	18427	18600

Table 14: *Fit Statistics.*

It is clear that PRU outperforms MNP in terms of explaining the observed brand choices in the product category. We present the parameter estimates of the homogeneous versions of PRU and MNP in Table 12. Under the PRU, the estimated values of μ_j are quite small in magnitude[20] (varying from 1.4526 to 2.7840), which indicates that consumers use "smoother "decision rules than the maximum utility rule implied by traditional random utility models such as the MNP. Between the PRU and MNP, we find that the estimated brand constants are, by and large, understated for the large share brands and overstated for the small share brands under the MNP. We also find that the marketing mix coefficients are understated in magnitude under the MNP model. To the extent that the PRU allows for the household to switch between brands in a more stochastic manner than does the MNP, it is not surprising that the effects of marketing variables on the consumers' utilities for brands turn out to be more pronounced under the PRU. Under the MNP, observed brand-switching of households that is unrelated to marketing activities

[19]The J binary logit probabilities must be scaled to add up to one.

[20]Remember that as $\mu_j \longrightarrow \infty \; \forall j$, PRU reduces to MNP.

of brands leads us to conclude that marketing variables are less important than they truly are. Regarding the variances of a consumer's brand-specific utilities, it is interesting to note that the estimated variances decrease from the MNP to the PRU for four brands, increase for two brands, and remain unchanged for two brands. This shows that for a majority of brands, once we accommodate probabilistic decision-making on the part of consumers, the onus of having to explain brand switching using unobserved variables goes down.

Parameter	*PRU*	*MNP*
Heinz 32 oz	2.4284	2.0194
Control Brand 32 oz	1.4110	1.3194
Hunts 32 oz	1.8108	1.5289
Heinz 28 oz	1.3466	1.4108
Del Monte 32 oz	0.7581	0.6872
Heinz 14 oz	0.7179	0.9157
Heinz 44 oz	0.9377	0.9074
Heinz 64 oz	0	0
Price	-1.2703	-0.6326
Display	0.2306	0.1656
Feature	0.1452	0.0538
${\sigma^2}_{Heinz32}$	0.5325	0.9778
${\sigma^2}_{CB32}$	0.4868	0.9832
${\sigma^2}_{Hunts32}$	0.0821	0.6704
${\sigma^2}_{Heinz28}$	2.0206	1.0098
${\sigma^2}_{DM28}$	0.9716	0.9412
${\sigma^2}_{CB32}$	1.6812	1.1981
${\sigma^2}_{Hunts32}$	1.1657	1.0541
${\sigma^2}_{Heinz28}$	1	1
$\mu_{Heinz32}$	1.8415	na
μ_{CB32}	1.7462	na
$\mu_{Hunts32}$	2.7840	na
$\mu_{Heinz28}$	1.9483	na
μ_{DM28}	1.9470	na
μ_{CB32}	1.4526	na
$\mu_{Hunts32}$	2.0588	na
$\mu_{Heinz28}$	2.0394	na

Table 15: *Estimation Results* .

8 Appendix - II: Empirical Illustration Using a Second Experiment

In order to further investigate the viability of our proposed modeling framework when consumers make choices in a laboratory setting (as opposed to the real-world setting represented in scanner panel data), we use data from an experiment involving respondents choosing among three soft drink brands – Coke, Pepsi and Sprite – offered at two possible prices – $1 and $0.5. In this experiment, consumers are shown all 57 possible subsets (i.e., 64 possible subsets minus 6 singletons and 1 null set) of the 6 alternatives (i.e., 6 brand-price combinations). For example, one of the 57 possible choice tasks for a subject is the following – [Coke $1] versus [Pepsi $0.5] – while another possible choice task is the following – – [Coke $1] versus [Pepsi $1] versus [Sprite $0.5] – etc. Each alternative in a choice task, therefore, can be described by 2 attribute values – one indicating the brand name, and the other indicating the price. Brand name being the same, one would expect a subject to choose the alternative with lower price. Prices being equal, one would expect a subject to choose the most preferred brand name. Subjects were offered each of the 57 choice sets a total of 40 times (which yielded us 2280 observations for each subject). To preclude position bias, the order of the brand names and prices in the choice sets was random and balanced. The order of the presentation of a choice set within the 57 was randomly determined. 3 subjects participated in the experiment. Because they were volunteers, the subjects do not constitute a representative sample for any population of interest. Projection to some population of interest, however, was not the purpose of the study. The method allows us to test the choice process for individual separately (since we have 2280 observations for each subject, unlike in scanner panel data where we rarely had more than 10 observations from a panelist), and results are reported at that level. We use brand name and price as the two attributes of interest while analyzing the experimental choice data.

Alternative	*BrandName*	*Price*
1	Coke	$1
2	Coke	$0.5
3	Pepsi	$1
4	Pepsi	$0.5
5	Sprite	$1
6	Sprite	$0.5

Table 16: *Description of Alternatives.*

Consider subject h ($h = 1, 2, 3$) observed over $t = 1, 2, ..., 57$ choice tasks, where each choice task involves 2, 3, 4, 5 or 6 soft drinks. For illustration purposes, let us consider a choice task involving 6 soft drinks.[21] On any given choice task t, one observes an outcome variable y_{ht} which takes the value j, $j = 1, 2, 3, 4, 5, 6$ where j identifies the soft drink chosen by subject h in that choice task. For every choice task t, one observes the brand name (Coke, Pepsi, Sprite) and price ($1, $0.5) faced by the subject for all soft drinks represented in the choice task. We label

[21] The choice task involving a smaller number of soft drinks can be represented by setting the consumer's utility for the missing alternatives to be $-\infty$.

the 3 possible brand names as $B_1 = Coke$, $B_2 = Pepsi$, $B_3 = Sprite$. The goal is to model the outcome variable (y_{ht}) given the above information.

The random utility model is based on a latent variable u_{hjt} that denotes the utility of subject h for soft drink j in choice set t. It is further assumed that this utility can be expressed as function of the set of soft drink covariates facing the subject in choice set t, in the following manner:

$$u_{hjt} = \sum_{k=1}^{2} \delta_{1kh} IB_{hjkt} + \delta_{2h} P_{hjt} + \eta_{hjt}$$

where

$$\delta_h = (\delta_{11h}, \delta_{12h}, \delta_{2h})'$$

are household-specific coefficients, IB_{hjkt} is an indicator variable that takes the value 1 if $B_{hjt} = B_k$ and 0 otherwise, P_{hjt} is the price of alternative j, and η_{hjt} is a random error such that $\eta_{ht} = (\eta_{h1t}, \eta_{h2t}, ..., \eta_{h6t})'$ has a joint distribution $f(\eta_{ht})$ with continuous support, mean zero and covariance matrix Σ.

As in the scanner data and wager examples, we model the choice outcome y_{ht} as being determined by either the decision rule of *maximum utility* or the decision rule of *probabilistic choice*. We assume $f(\eta_{ht})$ to be iid Gumbel, which renders the traditional random utility model to be the MNL. Under this distributional assumption for the error terms, we make the following assumption on the consumer's probabilistic choice rule: $P_{hjt} \propto 1/(1+\mu_{hjt}[u_{hjt} - Max_{k \neq j}(u_{hkt})])$, where the scale parameter $\mu_{hjt} = \sum_{k=1}^{3} \mu_{kh} IB_{hjkt}$ is allowed to vary by brand name. Three scale parameters – $\{\mu_{kh}, k = 1, 2, 3\}$ – are explicitly estimated (instead of being assumed to be ∞ as in the traditional random utility model).[22] This decision rule assumption is made in the spirit of the probabilistic choice rule tested on the scanner and wager datasets.

We report the fit values (i.e., maximized LL) associated with both our proposed model (PRU), as well as the MNL, in Table 17. Our model outperforms the MNL – in terms of better explaining a consumer's observed brand choices – for 2 out of 3 consumers in the data. This provides additional empirical support for our proposed model. Our proposed model shows the most pronounced gain in fit over the MNL for consumer 3 (the gains being 9% in terms of the improved LL). Taken in conjunction with the results from scanner panel data and wager data (discussed earlier), this further demonstrates that our proposed model has superior *external* and *internal* validity over the MNL.

Consumer	*PRU*	*MNL*
Consumer 1	-1057	-1095
Consumer 2	-1287	-1287
Consumer 3	-484	-534

Table 17: *Fit Statistics.*

We report the individual-level estimates in Table 17. Each consumer is found to have a different brand as their most preferred brand. Consumer 1 prefers Coke, while consumers 2 and

[22] The six alternative choice probabilities are normalized to add up to one.

3 prefer Sprite and Pepsi respectively. The price coefficient is negatively signed, as expected. In terms of the estimated scale coefficients, we find that all 3 consumers are very likely to use the maximal utility rule if the most preferred option in a choice task is Coke or Pepsi, but consumers 1 and 3 are less likely to use the rule (and would more probabilistically choose among the options) if the most preferred option in a choice task is Sprite, their least preferred brand. This makes intuitive sense and gives face validity to our estimated scale coefficients. Consistent with the findings in Table 17 – that our proposed model does not improve over the MNL – is the finding in Table 18 that consumer 2 employs the maximal utility rule regardless of which brand is the most preferred in a choice task. This finding also agrees with the fact that the descriptive statistic, CONSISTENCY, (i.e., the percentage of times the consumer picked the same alternative across the 40 replications of the same choice task) is a measly 8.77%, the smallest among the three consumers in the sample.

$Consumer$	$Consumer1$	$Consumer2$	$Consumer3$
Coke	3.3454	-0.0387	7.0415
Pepsi	1.2659	-4.7971	10.7505
Sprite	0	0	0
Price	-9.5329	-3.6337	-12.1858
μ_{Coke}	104.5497	20.0855	699.0936
μ_{Pepsi}	12.3753	5.9321	142.2760
μ_{Sprite}	0.6121	1096.6332	0.4469

Table 18: *Individual-Level Estimation Results.*