WAGNER A. KAMAKURA and RAJENDRA K. SRIVASTAVA*

The authors discuss some shortcomings of probabilistic choice models which do not capture product interdependencies. When substitutional relationships are present in the marketplace, these models may lead to biased estimates for choice/market shares. A probabilistic choice model which accounts for the interdependence among choice alternatives is proposed. This model uses parsimonious parameterization and relies on parameters which are not related to specific alternatives in the choice set. The model is compared with two other probabilistic choice models in terms of goodness of fit and predictive ability.

Predicting Choice Shares Under Conditions of Brand Interdependence

The modeling of consumer choices has an important role in a variety of marketing research/management tasks including, but not limited to, product design and positioning. Typically, choice models may be used to project demand/market shares for potential product positions given existing products/brands. If choice models based on judged preferences (vector, ideal point, conjoint, hybrid; see Green, Carroll, and Goldberg 1981; Green and Srinivasan 1978; Shocker and Srinivasan 1979) are used to estimate consumer preferences, choice simulators based on the proportion of first preferences can be used to predict market shares. Alternatively, purchase probabilities can be modeled as an increasing function of preference via transformations based on Luce’s (1959) choice axiom (see Hauser and Simmie 1981; Silk and Urban 1978), the generalization of the Thurstone Case V model (see Jain, Malhotra, and Mahajan 1979), or the multivariate generalization of the bivariate logistic model (see Bock and Jones 1968; Wiley and Low 1983).

The former approach (based on proportion of first preferences) probably has merit when used to predict market shares for high-involvement products such as durable goods where incentives for customers to optimize are apparent. However, for low-involvement products where satisfying and variety-seeking behaviors may occur, customer preferences/choices are likely to shift over time because of changes in mood/taste/anticipated product use. Then, typically, multiple brands are purchased within a product class and probabilistic choice models become more relevant. The use of first-choice proportions leads to biased predictions when there is considerable variation in brand/product utilities (Wiley and Low 1983).

If the interest is to predict shares of choices (market shares), it would seem reasonable to focus on approaches developed especially for predicting choice probabilities. Most of the approaches are based on some variant of Luce’s (1959) choice axiom (see Batsell and Lodish 1981; Gensch and Recker 1979; Louviere and Hensher 1983; Louviere and Woodworth 1983; Malhotra 1984; Punj and Staelin 1978; Reibstein 1978). Though the multinomial logit (MNL) model may lead to reasonably robust approximations to aggregate choice data under certain conditions (Huber 1982; Louviere and Woodworth 1983), the approach often is criticized because of the “independence of irrelevant alternatives” (IIA) property (Currim 1982; Debreu 1960; McFadden 1976; Tversky 1972). Unfortunately, the IIA assumption may be violated in many choice situations, leading to biased predictions. We propose a choice model which accounts for interdependence among alternatives. The model uses parsimonious parameterization and relies on parameters which do not relate to the particular alternatives included in a choice set (alternative-independent parameters). We compare this model with the random coefficients probit

*Wagner A. Kamakura is Assistant Professor, School of Management, State University of New York—Buffalo. Rajendra K. Srivastava is Associate Professor of Marketing, The University of Texas at Austin.

The research was supported, in part, by the Institute for Constructive Capitalism and the Bonham Fund, The University of Texas at Austin. The opinions expressed are those of the authors and do not necessarily reflect the views of the Institute.

420

Journal of Marketing Research
Vol. XXI (November 1984), 420–34

Copyright © 2001. All Rights Reserved.
PREDICTING CHOICE SHARES

(RCP) model (Hausman and Wise 1978) and the identity probit (PROBIT) model (the probit equivalent of the MNL model) in terms of goodness of fit and predictive ability to illustrate empirically the drawbacks of the IIA assumption.

First we briefly review probabilistic choice models to suggest desirable characteristics for such models (parsimony, alternative-independent parameters, and the absence of the need for a researcher-specified covariance structure). We then introduce a proposed probit model developed on the basis of the desirable characteristics. Next we present an empirical comparison of the PROBIT and RCP models with the proposed model. Finally, we discuss the findings and examine their implications for the optimal product design process.

PROBABILISTIC CHOICE MODELS

Three classic theories of individual choice behavior provide the basis from which most probabilistic choice models were developed: Thurstone’s (1927) Law of Comparative Judgment, Luce’s (1959) Choice Axiom, and Restle’s (1961) Elimination Process. These theoretical origins provide a framework for classifying the multitude of existing models (McFadden 1981).

“Lucent” or Constant Utility Models

This category of models shares Luce’s assumption of the discriminative process as the source of randomness in individual choice behavior. Starting from an axiom about choice probabilities, Luce derived the corollary

\[ P_j = \frac{v_j}{\sum v_i} \]

where \( v_j \); \( j \in T \) is a ratio-scaled preference value or utility for alternative \( j \). The underlying assumption in equation 1 is that a constant utility value is assigned by the decision maker to each choice alternative, and choice probabilities would be proportional to these assigned constant utilities.

One controversial consequence of Luce’s Axiom 1 is the independence of irrelevant alternatives (IIA) property. Though this principle has prevailed in the probabilistic choice field, several criticisms have been posed against it on intuitive (Debreu 1960; McFadden 1976; Tversky 1972; Tversky and Russo 1969) and empirical (Becker, DeGroot, and Marschak 1963) grounds. A number of variants of Luce’s basic formulation (equation 1) have been suggested to overcome the IIA property while retaining the computational advantages and simplicity of the model. The fully competitive (McLynn 1978) and dogit (Gaudry and Dagenais 1979) models are in this category.

“Restleian” or “Tverskian” models

A different approach to the IIA dilemma was suggested by Restle (1961). Restle viewed the choice process as a sequential one, whereby individuals would partition the choice set into groups and follow an elimination process like an “elimination tournament in tennis or chess” (Restle 1961). A study by Rumelhart and Greeno (1971), in which stimuli were designed to be similar within certain groups, resulted in empirical support to Restle’s approach.

These ideas of choice as an elimination process were generalized by Tversky (1972) into his elimination by aspects (EBA) model. The choice process is viewed as a hierarchy of decisions in which the subject chooses an evaluative criterion at random, sorts out the choice alternatives not satisfying that criterion, chooses a new criterion at random, sorts out the alternatives from the previous subset which do not satisfy the new criterion, and so on until a single alternative remains. According to this approach, randomness in choices occurs as a consequence of the random selection of criteria considered to eliminate alternatives from the choice set. This process implies a disjunctive choice decision in which there are no tradeoffs between the evaluative criteria.

“Thurstonian” or Random Utility Models

Following Thurstone’s Law of Comparative Judgment (1927), random utility models assume that when making a choice decision, individuals assign a utility value to each choice alternative from a given frequency distribution, then select the one with the highest utility. The assigned utility value can thus be decomposed into two components: a deterministic or fixed component which may be explained by the attributes of the choice alternative (and their interaction with the individual’s characteristics) and a random component.

\[ U_j = \Theta'Z + \epsilon_j \]

where:

- \( U_j \) = total utility assigned to alternative \( j \) on a given choice occasion,
- \( \Theta'Z \) = deterministic utility (a linear function of the characteristics vector \( Z \), and parameter vector \( \Theta \)), and
- \( \epsilon_j \) = random utility component for alternative \( j \).

The basic difference among various formulations of random utility models is in the form of the frequency distribution that is assumed for the random component of utility (\( \epsilon_j \))—including whether or not the interdependence between alternatives is modeled.

The multinomial logit model (MNL) assumes an independent extreme value distribution for the utilities of each choice alternative. Under this assumption equation 2 results in a final form equivalent to Luce’s choice model, equation 1 (Block and Marschak 1960; McFadden 1973). If the utilities are assumed to be independently, identically, and normally distributed, the resulting formulation is the identity probit (PROBIT) model.

Though the assumption of independence among choice alternatives may lead to reasonably robust approximations to aggregate choice data (Huber 1982; Louviere and
Woodworth (1983), Manski (1973) demonstrates that the IIA property may be violated for at least two (among others) frequently occurring sources of randomness.

1. **Omitted structure**—because omitted variables (attributes) may be shared to different extents by each pair of alternatives.

2. **Inter- and intrapersonal variation of preferences**—as utilities for choice alternatives would be a function of their attributes, the random variation of attribute weights would result in the correlation of random utility components (see Hausman and Wise 1978).

The violation of the IIA assumptions has important implications for optimal product design/positioning. The MNL and PROBIT models imply linear preferences and therefore linear isoshare lines perpendicular to the preference functions. However, the choice alternatives that are similar tend to "draw" shares from each other and are more competitive/substitutable than dissimilar alternatives. The result, as we show subsequently, is nonlinear isoshare curves. Thus, to achieve optimal positioning one must take into account not only customer preferences, but also the relative competition/substitution with existing brands. Of course, the limitations of the MNL/PROBIT models depend on the size of the random utility component in relation to the deterministic component and the extent of the interaction among alternatives. Because of this potential inadequacy, several models have been suggested to overcome the IIA restriction.

**Approaches Based on Prespecified Hierarchical Structures**

The nested multinomial logit (NMNL) model developed by Domenich and McFadden (1975) treats choice decisions as hierarchical processes, in agreement with the Restle (1961) and Tversky (1972) approaches. So, whereas the MNL model represents an interface between "Thurstonian" and "Lucean" approaches, the nested logit links "Thurstonian" and "Tverskian" approaches.

This nested logit model assumes that choice alternatives can be portrayed in a known hierarchical structure, which would also represent the decision tree followed by individuals in their sequential decision. At each level of this decision tree, the conditional probability for each "branch," given the upper-level "branch," is modeled by a logit model using the shared attributes within each branch to determine its utility. The final choice probability for a given alternative is obtained by multiplying all conditional probabilities along the sequential process.

A more general formulation than the NMNL model was suggested by McFadden (1980, 1981) with his generalized extreme value (GEV) model. The approach taken in the GEV model is similar to that of the NMNL model, and it assumes similarities within evoked subsets and independence between subsets.

The GEV and NMNL models provide an adequate methodology for explaining choice behavior if the individual follows a "Tverskian" process in making choices and if the hierarchical structure of the choice set is known. This methodology presupposes that choices are made through a sequential elimination process, and (for the applicability of these models) that the hierarchical grouping of alternatives (if it exists) is known by the researcher.

**The Multinomial Probit Model**

A more general pattern of interdependencies among choice alternatives is allowed by the multinomial probit model. Instead of constraining the interdependencies to a predetermined hierarchical pattern, this model allows for any degree of interaction among all pairs of choice alternatives.

This flexibility is attained by assuming a multivariate normal distribution for the random utilities, with a mean vector determined by the attributes of each alternative (first component in equation 2), and a covariance matrix,

\[
\Sigma = \begin{bmatrix}
\sigma_{11} & \cdots & \sigma_{1J} \\
\vdots & \ddots & \vdots \\
\sigma_{1J} & \cdots & \sigma_{JJ}
\end{bmatrix}
\]

where \(J\) is the number of choice alternatives.

This covariance structure is general enough to accommodate various patterns of interaction among choice alternatives, or some of its terms may be fixed a priori to approximate more rigid structures such as the ones required by the GEV model (see Currim 1982 for an example). However, these terms must satisfy some constraints. To represent the covariance matrix of a multivariate normal distribution, \(\Sigma\) must be positive semidefinite (a single constraint to be satisfied by \(J(J + 1)/2\) parameters). Also, the specification results in a nonparsimonious model requiring \(J(J + 1)/2\) covariance elements, in addition to the attribute weights in the deterministic component of utility. Finally, the covariance terms are specific to the choice alternatives being considered. If a new alternative is added, \((J + 1)\) new parameters referring to the new alternative must be estimated. This feature would prevent the application of the MNP model for the prediction of a new brand based on the data available on existing ones.

If the random components of utility (\(\xi_i\)) for each choice alternative are assumed to be independent of each other (e.g., a diagonal covariance matrix \(\Sigma\)), and a homoscedasticity assumption is added, the multinomial probit model reduces to the identically distributed, independent disturbances probit (IPROBIT) which, though not strictly assuming the IIA property, provides results similar to those of the MNL model.

---

1More parsimonious multinomial probit models may be defined by restricting some of the covariance terms to zero. This procedure, however, will require some prior knowledge of the interaction among choice alternatives.
PREDICTING CHOICE SHARES

Hausman and Wise Probit Model or Random Coefficients Probit (RCP)

A more parsimonious formulation of the probit model, which still accounts for the interaction among choice alternatives, is proposed by Hausman and Wise (1978). They identify random variations of preferences (attribute weights) as a major source of randomness in choice behavior. According to their formulation, the utility assigned to alternative \( j \) can be expressed by

\[
U_j = \mathbf{B}' \mathbf{Z}_j + \gamma_j
\]

where:

\( \mathbf{B} = \) vector of attribute weights, normally distributed with mean \( \Theta \) and covariance \( \Omega \),

\( \mathbf{Z}_j = \) vector of attributes for alternative \( j \), and

\( \gamma_j = \) normally distributed, independent random disturbances with zero mean and standard deviation \( \sigma_{\gamma} \).

This model reduces to a multinomial probit, with a covariance matrix of the random utilities equal to

\[
\mathbf{\Sigma} = \mathbf{Z} \mathbf{\Omega} \mathbf{Z}' + \sigma^2_{\gamma} \mathbf{I}_J
\]

where \( \mathbf{I}_J = \) identity matrix of degree \( J \).

This formulation will account for the interaction among choice alternatives even if the preferences for the different attributes are assumed to vary independently (a diagonal matrix \( \Omega \)). With this assumption the model is parsimonious, requiring two parameters (attribute weight and its standard deviation) for each attribute, in addition to the standard deviation of random disturbances \( \gamma_j \). The RCP model also has alternative-independent parameters. Consequently, no new parameter estimates are required in order to predict the choice shares for a new alternative. This model would hold, however, only if the variation of preferences within and/or across individuals were the major source of interaction among choice alternatives.

Most of the probabilistic choice models reviewed in this section focus on overcoming the IIA property. Some of these attempts are based on “corrections” of Luce’s basic model (dogit and fully competitive model). Others (NMNL and GEV) restrict the interaction among alternatives to particular subsets and require prior knowledge of the hierarchical structure of these subsets. A more promising approach seems to be the multinomial probit, which does not require any previous knowledge of the relationship among choice alternatives. However, this general formulation is nonparsimonious, with direct implications for the reliable estimation of the model. A more parsimonious RCP model, accounting for interactions among choice alternatives, was proposed by Hausman and Wise (1978), but random variation of preferences is assumed as the main source of interaction among choice alternatives.

These models were compared empirically by Currim (1982), who demonstrated the superior predictive power of the multinomial probit model despite the already mentioned limitations of this model in terms of parsimony and necessary constraints for a positive semidefinite covariance matrix. We propose a new variant of the multinomial probit model maintaining its advantages (non-specification of hierarchical structures). The model has a more parsimonious specification and alternative-independent parameters, and does not assume the random variation of “tastes” as the sole source of interdependence among choice alternatives.

A PROPOSED MULTINOMIAL PROBIT (PMNP) MODEL

The proposed model starts with the basic assumptions of “Thurstonian” models (equation 2). However, instead of specifying a full covariance matrix for the random disturbances \( \epsilon \), as done by the multinomial probit model, it expresses the covariance between two alternatives as a function of how similar they are perceived to be by the choice maker in terms of the relevant attributes. This model formulation, though similar in some respects, is different from the approach taken by Batsell (1980) in his extension of the MNL model. He uses distances to develop a measure of substitutability which is included as an “attribute” in the MNL model—i.e., is used in the “deterministic” component of utility. The proposed model utilizes distances to model the covariance structure of the random disturbances.

The notion of similarity as a cause of interaction among choice alternatives is implicit in most of the models reviewed in the preceding section, either in the form by which choice alternatives group into a hierarchical structure (NMNL, GEV) or embedded in the parameters of the model (covariance terms of the multinomial probit model). In the proposed model, the effect of similarity among choice alternatives on their interdependence is considered explicitly. Alternatives viewed as similar by the choice maker (in terms of the relevant attributes) are assumed to interact to a greater extent than ones viewed as distinct.

To measure similarity, Euclidean distances are computed over the attributes considered to estimate utilities. One also would expect alternatives which are similar in terms of important attributes (high weights in the utility function) to be more interdependent than alternatives similar on irrelevant attributes. Thus, attribute weights from the utility function are used to weight the Euclidean distances.

\[
d_{j\prime} = \left[ \sum_k \Theta_k^2 (Z_{jk} - Z_{j'k})^2 \right]^{1/2}
\]

where:

\( \Theta_k = \) weight of attribute \( k \) in the utility function,

\( Z_{jk} = \) amount of attribute \( k \) in alternative \( j \), and

\( d_{j\prime} = \) weighted Euclidean distance between alternatives \( j \) and \( j' \).

This new weighted metric also resolves a problem re-
lated to scale differences across the different attributes, as the final distances are “rescaled” to utility units regardless of the measurement units used to measure the original attributes.

What remains to be shown is how distances will result in interaction among choice alternatives and, consequently, in the correlation among the random utilities assigned to each choice alternative. Hausman and Wise (1978) suggest, “The more flexible parametrization [of the multinomial probit model] corresponds to letting data “choose” the degree of association [of random utilities] conditional on how close the observed alternatives are in the attribute space.” The proposed model constrains this “choice” to the following relationship.

\begin{equation}
    r_{ij} = Ke^{-\alpha d_{ij}}
\end{equation}

where:

\begin{align*}
    r_{ij} & = \text{correlation between the random utilities of alternatives } j \neq i \text{ and} \\
    \alpha, K & = \text{parameters to be estimated (} \alpha \geq 0; -1 \leq K \leq 1\).
\end{align*}

The last component of equation 7 \((e^{-\alpha d_{ij}})\) transforms Euclidean distances into a similarity measure between zero and one, such that alternatives infinitely apart \((d_{ij} = +\infty)\) would have a value of zero. The “decay” factor determines the rate by which similarity decreases as alternatives become farther apart in the attribute space. The \(K\)-factor transforms similarities into correlations among random utilities. Also, as in the multinomial probit model, positive correlation between two alternatives will result in a relative decrease in their choice probabilities (accounting for the substitutability among them), whereas negative correlations will increase their choice probabilities, accounting for any synergistic or complementary effect due to their similarity. This pattern is explained subsequently.

By applying equation 7 to each pair of choice alternatives, we can write the correlation matrix for the random utilities \(\epsilon_j\) as

\begin{equation}
    \Omega = \begin{bmatrix}
        Ke^{-\alpha d_{12}} & 1 \\
        Ke^{-\alpha d_{13}} & Ke^{-\alpha d_{23}} & 1 \\
        \vdots & \vdots & \vdots \\
        Ke^{-\alpha d_{1j}} & Ke^{-\alpha d_{1j}} & \ldots & \ldots
    \end{bmatrix}
\end{equation}

The covariance structure among random utilities may then be written in matrix form as

\begin{equation}
    \Sigma = \sigma \Omega \sigma
\end{equation}

where the diagonal matrix of degree \(J\),

\begin{equation}
    \sigma = \begin{bmatrix}
        \sigma_1 \\
        \sigma_2 \\
        \vdots \\
        \sigma_j
    \end{bmatrix}
\end{equation}

contains the standard deviations for the random utilities.

Equation 9 along with the definition of the total utility function in equation 2 expresses the proposed model as a special case of the multinomial probit model. The matrix \(\Omega\) determines the extent to which choice alternatives interact, based on their similarities, and some characteristics of the particular choice process \((K)\). Matrix \(\sigma\) expresses the degree of randomness in choice behavior. The larger the value of the \(\sigma_j\)'s, the smaller the proportion of variation in utility explained by the deterministic component, and thus the more random the choice behavior.

One may achieve further parsimony in the proposed formulation by adding a homoscedasticity assumption, setting \(\sigma_1 = \sigma_2 = \ldots \sigma_j = \sigma\). This assumption also reduces the model to parameters which are not related directly to the alternatives in the choice set. In contrast to the MNP model, no additional parameter would be needed if a new alternative were included in the choice set. This feature may be useful for the positioning of new products. Choice shares for a new brand may be predicted on the basis of only its features.

A simple illustration on a two-attribute space helps explain the effect of the interaction factor \(K\). For this illustration, let us assume three choice alternatives \(A, B,\) and \(C\) with coordinates \(Z_{1A} = 3.5, Z_{1B} = 0.5, Z_{1C} = 2.0, Z_{2B} = Z_{2C} = 4.0\). Let us assume a simple formulation of the proposed model, with the decay factor \(\alpha = 1.0\) and a utility function,\(^2\) for alternative \(j (j = A, B, C)\).

\begin{equation}
    U_j = \Theta Z_{ij} + (1 - \Theta) Z_{j2} + \epsilon_j
\end{equation}

If the value of the attribute weight \(\Theta\) is assumed to be 0.7, all alternatives \((A, B, C)\) would have the same fixed utility \(\bar{U} = 2.6\). Thus, the IIA assumption would result in equal probabilities (1/3) for each one. However, the proposed model would predict different values, depending on the level of interaction among \(A, B, C\) captured by the interaction factor \(K\) (see Figure 1). A value of \(K = 0\) implies a complete independence among alternatives, resulting in equal choice probabilities (1/3) for \(A, B,\) and \(C\). The substitutability between identical alternatives \(B\) and \(C\) increases as \(K\) approaches one. At that point the two alternatives are perfectly substitutable, resulting in a .50 choice probability for \(A\). A negative \(K\) implies some synergistic interaction among similar alternatives such that the identical alternatives \(B\) and \(C\) "help" each other, decreasing the chances for \(A\).

**ESTIMATION OF THE PROPOSED MODEL**

Until recently, application of multinomial probit models has been restricted by computational constraints in the estimation of parameters. Though Bock and Jones (1968) developed an analytical form for evaluating choice

\(^2\)As in any probit model, there is an indeterminacy among the attribute weights and the parameters of the covariance matrix. The indeterminacy is eliminated either by setting one of the variance terms equal to one or by constraining the attribute weights.
probabilities under the assumptions of the multinomial probit model, that form is restricted to a maximum of three choice alternatives. More recently, two other approaches, applicable to more than three choice alternatives, have been developed. Lerman and Manski (1977) suggested a Monté Carlo approach through which choices would be simulated on the basis of the multinomial probit structure, and choice probabilities computed in terms of choice shares. The procedure used in our study is the numerical approximation to the maximum of several normally distributed variables developed by Clark (1961) and suggested by Daganzo, Bouthelier, and Sheffi (1977) for the computation of choice probabilities for the multinomial probit model. Though this approximation may result in biased choice probabilities when some utilities are correlated negatively, it has proven to be very accurate otherwise in comparison with the Monté Carlo method (Daganzo 1979). 3

A major problem of using these methods in the estimation of multinomial probit models is the warranty of a global maximum in the likelihood function. Because there is no analytical relation between attributes and choice probabilities, the likelihood function has no closed form and thus it is not possible to check analytically for a global maximum.

For the proposed model, in particular, a simulation experiment was conducted in which "true" parameters were used to generate choice data, then parameter estimates were obtained through maximum likelihood estimation. The algorithm used to estimate the parameters of the model is related closely to the CHOMP program developed by Daganzo and Schoenfeld (1978), with Clark's numerical approximation. The new algorithm uses the same definition of the log-likelihood (objective) function (based on Clark's numerical approximation) but another optimization routine, developed by Lasdon, Warren, and Ratner (1982), to search for the parameter estimates. A brief description of the algorithm is provided in the Appendix.

For the simulation experiment, we used a simple case with two attributes and five choice alternatives, assuming the deterministic utility function in equation 2. The five choice alternatives were located at (6,1), (2,5), (1,5), (4,5), (6,5), and (6,1.5) in the two-dimensional space. These alternatives were chosen to allow some pairs to be similar and other pairs relatively dissimilar. For the covariance matrix of random utilities, the scaling factor was fixed to one and "true" values assumed for \( \sigma \) and \( K \). Figure 2 is a contour plot of the likelihood function for \( \sigma \) and \( K \) (\( \Theta \) set at 0.7) computed from 100 choices generated with \( \sigma = 7 \), \( K = 0.7 \), and \( \Theta = 0.7 \). The reader may notice that the maximum of the likelihood function agrees with the parameters originally used to generate the data.

---

3When some utilities are correlated negatively and others positively (i.e., some product pairs are complements, others substitutes, respectively), it is possible to specify two interaction factors \( K_1 \) and \( K_2 \). However, this specification may result in the instability of parameter estimates because the presence of both positive and negative terms may lead to a nonpositive semidefinite covariance matrix.
Table 1

<table>
<thead>
<tr>
<th>&quot;True&quot; K</th>
<th>θ</th>
<th>σ²</th>
<th>K-factor</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.691(-0.09)</td>
<td>6.679(-.321)</td>
<td>-1.030(-0.030)</td>
<td>-157.7</td>
</tr>
<tr>
<td>0</td>
<td>0.692(-0.008)</td>
<td>6.686(-.123)</td>
<td>-0.806(-0.060)</td>
<td>-157.0</td>
</tr>
<tr>
<td>0.6</td>
<td>0.695(-0.005)</td>
<td>6.932(-.077)</td>
<td>-0.635(-0.035)</td>
<td>-157.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.697(-0.003)</td>
<td>6.970(-.033)</td>
<td>-0.416(-0.016)</td>
<td>-157.5</td>
</tr>
<tr>
<td>0.2</td>
<td>0.690(-0.010)</td>
<td>6.890(-.110)</td>
<td>-0.259(-0.059)</td>
<td>-157.7</td>
</tr>
<tr>
<td>0.0</td>
<td>0.697(-0.003)</td>
<td>6.977(-.023)</td>
<td>-0.106(-0.016)</td>
<td>-157.9</td>
</tr>
<tr>
<td>0.2</td>
<td>0.695(-0.005)</td>
<td>6.967(-.033)</td>
<td>-0.040(-0.020)</td>
<td>-158.0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.697(-0.003)</td>
<td>6.982(-.018)</td>
<td>-0.008(-0.012)</td>
<td>-158.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.695(-0.005)</td>
<td>6.955(-.045)</td>
<td>-0.004(-0.006)</td>
<td>-158.0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.698(-0.002)</td>
<td>7.003(-.003)</td>
<td>-0.002(-0.002)</td>
<td>-157.9</td>
</tr>
<tr>
<td>1.0</td>
<td>0.701(-0.001)</td>
<td>6.992(-.008)</td>
<td>1.000(0.000)</td>
<td>-157.4</td>
</tr>
</tbody>
</table>

"True" values: θ = 0.7, σ = 7.0. Values in parentheses are the differences between estimates and "true" values.

We performed another check for the existence of a global maximum in the likelihood function. Different sets of 100 choices were simulated, combining "true" values of θ = 0.7, σ = 7.0, and K ranging from -1 to +1. Parameter estimates then obtained were obtained from each sample, starting from the same initial "guesses" of θ = 0.5, σ = 1, and K = 0. A comparison of the final estimates with the "true" parameters (see Table 1) indicates a reasonably good recovery of the underlying structure in the data. Though the two simulations illustrate the ability of the algorithm to converge to a global maximum and to recover "true" parameters, one must note that for other values of Z0 (location of choice alternatives) the likelihood function may be less well behaved.

**EMPIRICAL EVALUATION OF THE PROPOSED MODEL**

The performance of the proposed model in predicting choice probabilities was evaluated on the basis of empirical data. For this evaluation, two other probit formulations were used as a frame of reference, the random coefficients probit (RCP) and the IPROBIT models. The first was chosen because it represents a feasible formulation of the probit, still accounting for interactions among choice alternatives. The IPROBIT was used as a "proxy" of the logit model, assuming independence of choice alternatives. The advantage of this model for the comparison is that it is fully nested on the random-coefficients probit and the proposed model, allowing for likelihood-ratio tests.

**The Experiment**

For the collection of choice data, a computer decision game was developed in which subjects made repetitive choices among sets of three wagers. Though this choice setting does not represent an actual marketing application, it was chosen for several other reasons. First, it provides relatively simple means of constructing choice alternatives with varying levels of similarity, thereby affording an empirical comparison of the models in a situation where substitutability is likely to occur. Second, a gambling situation provides direct reward/penalty, improving data quality. The wagers are presented in their actual form rather than by protocols that might overemphasize attributes otherwise not considered or induce respondents to an artificial consistency in their choice patterns. Finally, game settings have been used in the choice-modeling literature. For example, Becker, DeGroot, and Marschak (1963) used an experiment involving choices among gambles to demonstrate empirically the inadequacy of the simple scalability assumption of Luce's choice model. Two-outcome gambles also were used by Tversky (1972) to test his elimination-aspects model and six-outcome gambles were used by Batell and Lodish (1981) to predict individual choice behavior.

The wagers used in the computerized decision experiments had four equiprobable outcomes. Subjects were repeatedly offered choice sets containing three wagers and asked to choose one. For example, a subject would be presented the following choice set

\((-2 \ 3 \ 7 \ 12) \ (-15 \ -2 \ 12 \ 25) \ (-18 \ -3 \ 13 \ 28)\)

\(A \ B \ C\)

and asked to select one of the three wagers, A, B, or C. The choice task was conceived as a self-paced computer game. After presenting the rules of the decision game, the computer allowed participants to play a "warmup" game. In this warmup and in the actual game, the computer randomly selected a choice set (described next) not yet played, showed the three wagers on the screen, and asked the participant to make his/her choice. Once the choice was made, the chosen wager was actually "run" and the randomly selected outcome added to or subtracted from the participant's score. The game was divided into two sessions (consisting of the same 52 randomly presented choice sets) for which the scores were computed independently. Graduate and undergraduate students were invited to play the game, with chances of winning $100 in the first session and $200 in the second. The fact that the chances of winning these cash prizes were related directly to the participant's score in each session assured a high level of involvement in the task and enhanced the validity of the measure of actual choice behavior. To maintain the "average" odds of winning as announced in the invitation, only the first 100 students replying to the invitation were allowed to participate.

The choice sets were based on three wagers (each) chosen from a set of nine "original" wagers and another set of nine wagers "similar" to the original wagers. Each of the wagers could be described by two attributes: expected value (EV) and standard deviations (STD) calculated from the four possible outcomes. The nine original wagers were based on a factorial design of EV and STD with three levels each \(EV = 1, 3, 5; STD = 5, 15, 25\). The similar wagers also were based on a \(3 \times 3\) factorial design \(EV = 0, 2, 4; STD = 10, 20, 30\). Choice sets were created by first forming pairs of the original wagers. Pairs for which all outcomes of one wager were smaller than all outcomes of another were eliminated —i.e., dominant relationships were sorted out. Next,
the triad was formed by adding a similar wager to a pair of original wagers. For example, for a pair of original wagers described by EV and STD, (1, 15) and (3, 25), one could add similar wagers (0, 10), (2, 20), or (4, 30). The result would be three triads such that there would be at least one pair of similar elements; adding the similar wager (2, 20), which is equidistant to the two original wagers, results in two pairs of similar elements. This two-step procedure led to the creation of 52 triads (choice sets) used in the experiments.

Data Analysis

The same form of deterministic utility function was defined for the three models, with the attribute weights constrained to the unitary circle,4

\[ U_i = \sqrt{1 - w^2} EV_i + wSTD_i; (\Theta_1 = \sqrt{1 - w^2}, \Theta_2 = w). \]

In addition to the attribute weights, the IPROBIT model requires estimation of the standard deviation of random utilities, \( \sigma \). The proposed model requires two more estimates: the interaction factor \( K \) and the decay parameter \( \alpha \). The random-coefficients probit requires estimation of the standard deviation for the normal distribution of each attribute weight (\( \sigma_1, \sigma_2 \)) and for the distribution of random utilities due to other unobservable factors (\( \sigma \)).

The models were compared in terms of parameter stability, goodness of fit, and predictive validity. For the parameter stability test, the sample of 5200 choices in the first session of the game across all subjects was split randomly into two subsamples of 2600 choices each. Each model then was calibrated from the two subsamples for a comparison of the estimates.

Table 2 reports the parameter estimates obtained from both subsamples for each model. As no significant differences are found in the estimates obtained from the two subsamples, one may say that the estimation algorithm provides consistent estimates under the assumptions of the three models being compared.5

---

4Some form of standardization is necessary for any probit model to eliminate the indeterminacy between the attribute weights and the covariance of random utilities.

5Because the estimates are based on repetitive choices pooled across subjects, the assumption of independence among choice occasions may be invalid within subjects. Though the estimates of the attribute weights will still be unbiased, the asymptotic standard deviations of these estimates (computed as the negative of the inverse Hessian) may not reflect the true error. Hence, the statistical tests of these estimates are questionable. This problem is common in the estimation of most logit/probit models using multiple observations per subject (as is common practice), but often is not recognized (see Louviere and Hensher 1983).
The degree of fit for each of the applications also is shown in Table 2. As the IPROBIT model is nested under the other two models, the likelihood ratios can be tested statistically. The differences in log-likelihood values between models, multiplied by two, have asymptotically a chi-square distribution with a number of degrees of freedom equal to the differences in the number of parameters specified in each model. On the basis of these statistics, one may conclude that both the proposed model and the random-coefficients probit provide a significantly better fit than the IPROBIT. Also, from the log-likelihood values one can see that the proposed model provides a better fit (higher log-likelihood) than the RCP, though this improvement in fit cannot be tested statistically.

To compare the models in terms of their predictive power, two separate holdout tests were performed. In the first, all choice sets containing wagers with high levels of EV and STD and their respective "decoys" were held out for the prediction sample. This procedure resulted in the use of 20 of the original 52 choice sets for the calibration of the models and the remaining 32 for the predictive validity tests.

Table 3 reports the parameter estimates for each model, based on the 2000 choices in the estimation sample. These estimates were used to predict choice probabilities for the remaining 32 holdout sets under the assumptions of each model. Table 3 also shows the goodness-of-fit statistics for the estimation and prediction samples. Once again, the two models provide better fit and better predictions than the IPROBIT, demonstrating the biasing effect of the IIA assumption. Also, the proposed model shows a higher log-likelihood than the random-coefficients probit (RCP), indicating a better fit in the estimation and holdout sets. As before, because the two models are not nested, the differences cannot be tested statistically.

The differences among the three models in terms of

---

**Figure 3**

**SCATTERGRAM OF CHOICE SHARES: IPROBIT MODEL**

---

The degree of fit for each of the applications also is shown in Table 2. As the IPROBIT model is nested under the other two models, the likelihood ratios can be tested statistically. The differences in log-likelihood values between models, multiplied by two, have asymptotically a chi-square distribution with a number of degrees of freedom equal to the differences in the number of parameters specified in each model. On the basis of these statistics, one may conclude that both the proposed model and the random-coefficients probit provide a significantly better fit than the IPROBIT. Also, from the log-likelihood values one can see that the proposed model provides a better fit (higher log-likelihood) than the RCP, though this improvement in fit cannot be tested statistically.

To compare the models in terms of their predictive power, two separate holdout tests were performed. In the first, all choice sets containing wagers with high levels of EV and STD and their respective "decoys" were held out for the prediction sample. This procedure resulted in the use of 20 of the original 52 choice sets for the calibration of the models and the remaining 32 for the predictive validity tests.

Table 3 reports the parameter estimates for each model, based on the 2000 choices in the estimation sample. These estimates were used to predict choice probabilities for the remaining 32 holdout sets under the assumptions of each model. Table 3 also shows the goodness-of-fit statistics for the estimation and prediction samples. Once again, the two models provide better fit and better predictions than the IPROBIT, demonstrating the biasing effect of the IIA assumption. Also, the proposed model shows a higher log-likelihood than the random-coefficients probit (RCP), indicating a better fit in the estimation and holdout sets. As before, because the two models are not nested, the differences cannot be tested statistically.

The differences among the three models in terms of
predictive ability are seen easily in Figures 3, 4, and 5, where the predicted choice shares (computed across the 100 subjects) for each wager within the 32 holdout sets are compared with the actually observed choice shares. The first difference to be noticed is that the scattergrams for the proposed model and random-coefficients probit are more concentrated around the perfect prediction line (45°). However, the regression line between predicted and actual choice shares shows a larger deviation from the perfect prediction line (bias) for the RCP than for the proposed model. The R²'s computed around the 45° line confirm these conclusions, showing a better predictive fit for the proposed model (R² = .73), followed by the RCP (R² = .61) and IPROBIT (R² = .57).

In the second holdout test, choice sets containing wagers with intermediate levels of EV and STD were held out, resulting in a validation sample of 2200 choices and an estimation sample of 3000 choices. The tests performed on the first holdout sets were replicated and confirmed the previous results. The parameter estimates obtained from the estimation sample are listed in Table 4. The goodness-of-fit tests for the two samples are in direct agreement with the first holdout results.

The scattergrams of actual versus predicted choice shares (not shown here) indicated a better predictive ability for the proposed model, followed by the RCP and IPROBIT models (R² of .65, .52, and .51, respectively).

Segment-Level Analysis

The major difference between the proposed and random-coefficients models is the assumption about the source of randomness in choice behavior, which has direct implications for the source of interaction among choice alternatives. The RCP attributes the randomness of utilities to the variation of “tastes” within and/or between individuals, whereas the proposed model assigns this randomness to unobservable factors. According to the assumptions of the proposed model, interaction among choice alternatives would be due to the correlation of any unobservable factor (measurement errors, model misspecification, random variation of “tastes”) with the observed attributes. Thus, an ideal comparative test of these two models would be to sort out the effects of inter- and intraindividual variations, then compare the predictive ability. Unfortunately, intraindividual variations could not be estimated in our experiment because to do so would require a longer history of individual choices. Interindividual differences could be sorted out by clustering the 100 subjects according to their preferences for EV and STD estimated at the individual level (using the IPROBIT model).

Thus, a segment of 20 subjects with similar preferences for the two attributes was selected from the sample and the previously described predictive validity tests were performed. The three models were calibrated on the basis of choices among 30 choice sets, then used to predict choice shares for 22 holdout sets.

The parameter estimates in Table 5 show that the homogeneity of subjects reduced the overall randomness in choices, resulting in a decline in the estimate of σ (standard deviation of random utilities) from a value of approximately 6.0 in the aggregate analysis (see Table 2) to approximately 3.0 at the segment level. This reduction of interindividual variations also reduced the advantages of the random-coefficients and proposed models over the IPROBIT in terms of goodness of fit and predictive validity. The likelihood-ratio tests in Table 5 show
Table 4
PARAMETER ESTIMATES AND LIKELIHOOD RATIOS: SECOND PREDICTIVE TEST

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>W</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>.061(5 714)</td>
<td>5.675(14.915)</td>
</tr>
<tr>
<td>Proposed</td>
<td>.077(7.110)</td>
<td>5.680(13.670)</td>
</tr>
<tr>
<td>RCP</td>
<td>.089(5 347)</td>
<td>5.717(6.683)</td>
</tr>
</tbody>
</table>

Likelihood ratio tests

<table>
<thead>
<tr>
<th>Estimation</th>
<th>Log-likelihood</th>
<th>Chi square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>-3189.7</td>
<td>—</td>
<td>.0001</td>
</tr>
<tr>
<td>Proposed</td>
<td>-3164.0</td>
<td>73.4</td>
<td>.0001</td>
</tr>
<tr>
<td>RCP</td>
<td>-3167.4</td>
<td>70.4</td>
<td>.0001</td>
</tr>
</tbody>
</table>

We propose a relatively parsimonious model for predicting choice probabilities which accounts for the interaction among choice alternatives. The model does not require the existence or prior knowledge of a hierarchical structure of the choice set, and affords predictions for a

no significant differences among these models at the .01 level.

These results indicate that for this particular application, random variations of attribute weights were a major source of interaction among choice alternatives, a finding which agrees with the basic assumptions of the random-coefficients probit.

CONCLUSIONS

Table 5
PARAMETER ESTIMATES AND LIKELIHOOD RATIOS: SEGMENT ANALYSIS

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>W</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>.076(5 374)</td>
<td>2.891(11.500)</td>
</tr>
<tr>
<td>Proposed</td>
<td>.084(5 940)</td>
<td>2.914(10.335)</td>
</tr>
<tr>
<td>RCP</td>
<td>.086(5 081)</td>
<td>2.385(9.008)</td>
</tr>
</tbody>
</table>

Likelihood ratio tests

<table>
<thead>
<tr>
<th>Estimation</th>
<th>Log-likelihood</th>
<th>Chi square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>-577.7</td>
<td>—</td>
<td>.030</td>
</tr>
<tr>
<td>Proposed</td>
<td>-573.3</td>
<td>8.8</td>
<td>.025</td>
</tr>
<tr>
<td>RCP</td>
<td>-573.8</td>
<td>7.8</td>
<td>.750</td>
</tr>
</tbody>
</table>

We propose a relatively parsimonious model for predicting choice probabilities which accounts for the interaction among choice alternatives. The model does not require the existence or prior knowledge of a hierarchical structure of the choice set, and affords predictions for a

no significant differences among these models at the .01 level.

These results indicate that for this particular application, random variations of attribute weights were a major source of interaction among choice alternatives, a finding which agrees with the basic assumptions of the random-coefficients probit.

CONCLUSIONS

Table 5
PARAMETER ESTIMATES AND LIKELIHOOD RATIOS: SEGMENT ANALYSIS

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>W</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>.076(5 374)</td>
<td>2.891(11.500)</td>
</tr>
<tr>
<td>Proposed</td>
<td>.084(5 940)</td>
<td>2.914(10.335)</td>
</tr>
<tr>
<td>RCP</td>
<td>.086(5 081)</td>
<td>2.385(9.008)</td>
</tr>
</tbody>
</table>

Likelihood ratio tests

<table>
<thead>
<tr>
<th>Estimation</th>
<th>Log-likelihood</th>
<th>Chi square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>-577.7</td>
<td>—</td>
<td>.030</td>
</tr>
<tr>
<td>Proposed</td>
<td>-573.3</td>
<td>8.8</td>
<td>.025</td>
</tr>
<tr>
<td>RCP</td>
<td>-573.8</td>
<td>7.8</td>
<td>.750</td>
</tr>
</tbody>
</table>

We propose a relatively parsimonious model for predicting choice probabilities which accounts for the interaction among choice alternatives. The model does not require the existence or prior knowledge of a hierarchical structure of the choice set, and affords predictions for a

no significant differences among these models at the .01 level.

These results indicate that for this particular application, random variations of attribute weights were a major source of interaction among choice alternatives, a finding which agrees with the basic assumptions of the random-coefficients probit.

CONCLUSIONS

Table 5
PARAMETER ESTIMATES AND LIKELIHOOD RATIOS: SEGMENT ANALYSIS

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th>W</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>.076(5 374)</td>
<td>2.891(11.500)</td>
</tr>
<tr>
<td>Proposed</td>
<td>.084(5 940)</td>
<td>2.914(10.335)</td>
</tr>
<tr>
<td>RCP</td>
<td>.086(5 081)</td>
<td>2.385(9.008)</td>
</tr>
</tbody>
</table>

Likelihood ratio tests

<table>
<thead>
<tr>
<th>Estimation</th>
<th>Log-likelihood</th>
<th>Chi square</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPROBIT</td>
<td>-577.7</td>
<td>—</td>
<td>.030</td>
</tr>
<tr>
<td>Proposed</td>
<td>-573.3</td>
<td>8.8</td>
<td>.025</td>
</tr>
<tr>
<td>RCP</td>
<td>-573.8</td>
<td>7.8</td>
<td>.750</td>
</tr>
</tbody>
</table>

We propose a relatively parsimonious model for predicting choice probabilities which accounts for the interaction among choice alternatives. The model does not require the existence or prior knowledge of a hierarchical structure of the choice set, and affords predictions for a

no significant differences among these models at the .01 level.

These results indicate that for this particular application, random variations of attribute weights were a major source of interaction among choice alternatives, a finding which agrees with the basic assumptions of the random-coefficients probit.

CONCLUSIONS
new alternative without any need for new parameter estimation. The latter feature has implications not only for the usefulness of the model in predicting choices for a new brand, but also for its empirical estimation. If a model is specified in terms of alternative-dependent parameters (as in the multinomial probit), every choice decision must come from the same choice set to ensure the feasibility of the model. In other words, every subject would have to choose among the same set of alternatives, on every choice occasion.

In addition to the theoretical advantages, the proposed (PMNP) model is shown to provide better predictions than the random-coefficients probit (a parsimonious, flexible model which accounts for interactions) and the IPROBIT (which provides results similar to those of the MNL model). In two separate predictive validity tests, both the PMNP and RCP models resulted in better predictions for a holdout sample than did the IPROBIT, illustrating the effect of interdependence among choice alternatives. Better results were found also for the PMNP model than for the RCP model. However, when inter-individual differences were sorted out, no significant differences were found in the performance of the three models, indicating the random variation of “tastes” across subjects as the major source of interaction among choice alternatives in this particular choice setting (choice among wagers). One possible reason for these results is the fact that the choice alternatives were well described by the two measured attributes (EV and STD), and only a small variation was left to be explained by unobservable factors other than individual differences in preferences.

Nevertheless, the proposed (PMNP) model was able to account for the interaction due to the random variation of “tastes.” Such variation results in random components of utility directly related to the measured attributes (see Hausman and Wise 1978). This relationship was translated in the PMNP model by a large interaction factor $K$. Therefore, one would expect at least an equal predictive power for these two models if random variation of “tastes” were the sole reason for interaction among alternatives, and a better performance for the proposed (PMNP) model if other unobservable factors also affected choices. In any case, the performance of the proposed model, for the particular choice setting used in the experiment, is shown to be equal to or better than that of the random-coefficients probit at the same level of degrees of freedom. If more attributes were considered in the deterministic utility function, the random-coefficients probit model would require a larger number of parameters because it specifies at least one attribute weight and its standard deviation for each attribute.

An important characteristic of the proposed PMNP model is its flexibility in explaining choice situations with different degrees of interaction among choice alternatives. For different applications (e.g., different product categories) it would specify different values of the interaction factor $K$, representing the extent to which similarity among choice alternatives would result in interaction among them.

The implications of this flexibility for the prediction of consumers’ reaction to new products are straightforward. The model would consider not only consumers’ preferences for the product attributes, but also the competitive effects of the other brands in the market. In predicting choice shares for a new brand, the model would not only compare attributes (as in the MNL and IPROBIT models), but also take into account possible “cannibalization” among similar brands. This important feature is illustrated in Figure 6, A and B. Suppose a choice model has been calibrated on choices among existing brands, and the objective is to determine the best position for a new brand in the market. Figure 6 shows isoshare curves for this new brand in a two-attribute map, given the existence of brands A, B, C, and D. The average utility function is represented by the vector $\hat{U}$. The isoshares predicted by the IPROBIT (Figure 6, A) suggest the location of the new brand as far as economically and technologically possible along the preference vector. In contrast, the proposed model with an interaction factor $K = 0.9$, by taking into account the substitutability (cannibalization) among similar alternatives, shows that the same choice share might be obtained at two different positions along the utility vector. For instance, position $E_1$ is considerably higher than $E_2$ in terms of deterministic utility but, given the possible cannibalization with existing brand B, it would result (according to the model) in choice shares equivalent to $E_1$. The lowest share for the new brand is indicated by L in Figure 6, B, where the brand would have direct competition from all four brands. This nonlinearity of isoshare lines under conditions of brand interdependence has obvious implications for optimal product design/positioning.

The approaches which do not involve the IIA assumption, such as the RCP, PMNP, and the generalized multinomial probit models, are recommended for estimating the shares of potential new product positions. The RCP and PMNP models, which utilize alternative-independent parameters, are particularly useful as they do not require the inclusion of all new product concepts in the data collection and model estimation phases. These models can be used to generate isoshare curves (e.g., Figure 6, B). Subsequently, given cost data, it should be feasible to scan the product space to locate new product positions in a manner similar to the procedure suggested by Hauser and Simmie (1981).

We must note, however, that further research is needed to examine whether the PMNP (and/or RCP) model parameters change substantially with new product entries. Additional problems may arise because the introduction of new brands can lead to new attributes becoming important, whereas the models would base predictions on attributes of existing brands (this problem, however, relates to most models used for product design optimization). Further, the proposed PMNP model, like other multinomial probit models, is difficult to calibrate at the individual level (because of large data requirements) and therefore would have to be estimated across individuals at the aggregate/segment level. It may be feasible to
study is actually a restricted version of a more general form. Though the simplified version was suitable to the particular choice setting used, some restrictive assumptions may be relaxed for other applications. First, if there are reasons to believe that some brands are complementary/synergistic whereas others compete with each other, two different interaction parameters, \( K_1 \) and \( K_2 \), might be specified for the two sets. Also, if choice makers are expected to have different levels of knowledge among the choice alternatives, and thus to be less certain about their preferences for some alternatives, different standard deviations of random utilities (\( \sigma \)) may be specified for each choice alternative. For example, newer brands can be expected to have larger \( \sigma \)'s. These extensions, however, use alternative-dependent parameters, requiring new parameter estimates if a new alternative is added to the set. This feature may be a drawback if one wants to predict choice shares for a new brand on the basis of the current market conditions.

Furthermore, the proposed model, as well as any random utility model, may be extended to account for time dependencies on sequential choices. For the maximum likelihood estimation of the models in our empirical analysis, an assumption was made of independence between successive choice decisions, which implies the serial independence in the utility function. However, serial dependence in the utility function or in the choice probabilities (structural dependence) may occur because of such factors as learning and variety seeking, or cumulative effects such as attribute saturation. Therefore, extensions of the proposed model may incorporate these longitudinal effects for a better understanding of consumer choice behavior (see Heckman 1981).

Finally, we acknowledge that even if representing a genuine case of human choice behavior, the experiment we used to gather empirical evidence does not replicate an actual marketing setting. Further empirical applications will be necessary to validate our results. One should not expect identical results, though, because the relative performance of these probabilistic models will depend on the major source of randomness in choice behavior and, as a consequence, on the degree of interaction among choice alternatives.

APPENDIX

ESTIMATION OF PROBIT MODELS

The algorithm used to estimate the parameters of the proposed model is related closely to the CHOMP program developed by Daganzo and Schoenfeld (1978). For any problem with more than three choice alternatives the computation of choice probabilities would require the evaluation of a cumulative multivariate normal, which is analytically intractable. Hence, these choice probabilities had to be computed through a numerical approximation proposed by Clark (1961) and first used by Daganzo and Schoenfeld (1978) for the estimation of probit models. Given a vector of multivariate normals \((U_1, U_2, \ldots, U_j)\) with mean \((u_1, u_2, \ldots, u_j)\) and covariance \(\Sigma\)
with diagonal elements equal to $\sigma^2_i$ and off-diagonal elements equal to $\sigma_{ij}r_{ij}$. Clark demonstrated that the maximum of a pair $U_i, U_j$ can be approximated by a univariate normal distribution with

$$v_1 = u_i \Phi(z_{ij}) + u_j \Phi(-z_{ij}) + \sigma_{ij} \phi(z_{ij})$$

$$v_2 = (u^2_i + \sigma^2_i) \Phi(z_{ij}) + (u^2_j + \sigma^2_j) \Phi(-z_{ij})$$

$$+ (u_i - u_j) \sigma_{ij} \phi(z_{ij})$$

where:

$\sigma_{ij}$ is the standard deviation of the difference $U_i - U_j$,

$z_{ij}$ is the standardized value of the difference $u_i - u_j$, and

$v_1, v_2$ are the first and second moments for the distribution of $U_M = \max (U_i, U_j)$.

Clark also demonstrated that the correlation of the new variable $U_M = \max (U_i, U_j)$ with a third variable of the same multivariate distribution may be approximated by

$$r(U_M, U_j) = \frac{\sigma_{ij} \phi(z_{ij}) + \sigma_{ij} \phi(-z_{ij})}{\sqrt{v_2 - v_1^2}}$$

By working with these approximations recursively, one can find the moments for the univariate distribution of the maximum of several normals.

The choice probability of one alternative can be represented as the probability that its utility value is the greatest among those of all alternatives considered by the choice maker. Hence, this probability can be computed from the distribution of the maximum of all normally distributed utility values. The accuracy of this approximation in the computation of choice probabilities for the multinomial probit model is discussed in detail by Daganzo, Bouthelier, and Sheffi (1977).

Given the attributes of the choice alternatives and the parameters of the model, the estimation algorithm uses Clark's approximation to compute choice probabilities. The computed probabilities for the actually chosen alternatives then are aggregated into a log-likelihood function. Final parameter estimates are obtained through a constrained nonlinear optimization algorithm. For our application, the generalized reduced gradient algorithm (GRG2) developed by Lasdon, Warren, and Ratner (1982) was used.

Once the final maximum likelihood estimates are obtained, their asymptotic covariance is estimated as the negative of the inverse Hessian matrix at the optimum.

REFERENCES


