MARKETING SCIENCE
Vol. 27, No. 2, March–April 2008, pp. 300–307
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2702 | 0300

DOI 10.1287/mksc.1070.0301

The Invariant Proportion of Substitution Property (IPS) of Discrete-Choice Models

Thomas J. Steenburgh
Harvard Business School, Boston, Massachusetts 02163, tsteenburgh@hbs.edu

This article introduces a newly discovered property of discrete-choice models, which I call the invariant proportion of substitution (IPS). Like the independence from irrelevant alternatives (IIA) property, IPS implies individual behavior that is counterintuitive in the context of choice among similar alternatives. But models that alleviate the concerns raised by IIA, such as generalized extreme value and covariance probit models, do not necessarily alleviate the concerns raised by IPS. I explore the implications of the IPS property on individual behavior in several choice contexts and discuss some models that alleviate the concerns raised by IPS.

Key words: econometric models; choice models; logit; probit; generalized extreme value; independence from irrelevant alternatives (IIA); individual choice behavior
History: This paper was received October 4, 2004, and was with the authors 27 months for 2 revisions; processed by Tulin Erdem.

1. Introduction

Discrete-choice models are widely used in marketing and economics to describe individual choice behavior. Marketing applications range from conjoint experiments, which attempt to understand consumers' preferences for product attributes, to scanner panel studies, which attempt to understand how consumers respond to elements of the marketing mix. Nevertheless, like any model of behavior, discrete-choice models have known limitations. The independence from irrelevant alternatives (IIA) property has been a longstanding concern because it implies that individuals make counterintuitive choices when facing similar alternatives. Many modeling advances have been motivated by a desire to alleviate the concerns raised by IIA and to develop new choice models that allow more realistic behavior to occur; prominent examples include the generalized extreme value (GEV) family of models, which includes the nested logit as a member, and the covariance probit model.

Given the significant effort devoted to overcoming the limitations of IIA, it is surprising to discover that GEV and covariance probit models possess another property that implies individuals exhibit similarly counterintuitive choice behavior. I call this the invariant proportion of substitution (IPS) property, and I will show that some of the well-known models that alleviate IIA, including GEV and covariance probit models, do not necessarily alleviate IPS.

Let us begin the present discussion by revisiting the counterintuitive choice behavior implied by the IIA property. Suppose an individual faces a choice between two laptop computers:

	Weight	Processor speed
Laptop A	3 lbs.	2.0 GHz
Laptop C	7 lbs.	3.4 GHz

Laptop A is the lighter alternative, but it runs at a slower speed. Laptop C is the faster alternative, but it is heavier in weight.

If a new alternative is added to the choice set, from which of the existing laptops would it draw its choice probability? Intuitively, we would expect the new alternative to draw a greater proportion of its choice probability from a given competing laptop if it is more similar to it because the alternatives are closer substitutes. In the extreme, if the new alternative is identical to Laptop A, we would expect it to draw 100% of its choice probability from Laptop A because the two alternatives are perfect substitutes. In contrast, if the new alternative is identical to Laptop C, we would expect it to draw 0% of its choice probability from Laptop A. The greater principle at work is that the similarity of the new alternative to the existing options should strongly affect the individual's substitution patterns.

The longstanding critique of models with IIA is that they prevent this type of choice behavior from occurring (Debreu 1960, Luce and Suppes 1965, Tversky 1972, McFadden 1974). Instead of allowing the individual's substitution patterns to depend on the similarity of the alternatives, the IIA property requires the new alternative's choice probability to be drawn from each of the existing laptops in proportion to their original choice probabilities. Undesirably, this implies that the proportion of the new alternative's choice probability drawn from a given competing laptop is the same no matter how the new alternative is composed. Many well-known examples illustrate this point, including Beethoven/Debussy (Debreu 1960, Tversky 1972), pony/bicycle (Luce and Suppes 1965), and red bus/blue bus (McFadden 1974).

Now, consider a second example, which illustrates the counterintuitive behavior implied by the IPS property. Suppose an individual faces a choice among the following three laptop computers:

	Weight	Processor speed
Laptop A	3 lb.	2.0 GHz
Laptop B	5 lb.	2.7 GHz
Laptop C	7 lb.	3.4 GHz

Laptops A and C are located in the same positions as before. Laptop B, the target alternative for the sake of this example, is located directly in the middle of the competing alternatives on each attribute; it has moderate speed and moderate weight.

How would the individual substitute among the laptops if the target alternative is improved? Intuitively, we would expect a greater proportion of the growth in target alternative's choice probability to be drawn from a given competing laptop if the alternatives become more similar. In this example, reducing Laptop B's weight makes it more similar to Laptop A, but increasing its processor speed makes it less similar. Thus, we would expect a *greater* proportion of the growth to be drawn from Laptop A (the lightest alternative) if Laptop B is made lighter as opposed to faster.

Models with IPS, however, prevent this type of choice behavior from occurring. Instead of allowing the individual's substitution patterns to depend on whether the alternatives become more similar, the IPS property requires the growth in the improved good's choice probability to be drawn from each of the competing laptops in the *same* proportion no matter which attribute is improved. Like the choice behavior implied by IIA, the choice behavior implied by IPS is counterintuitive.

Let us put some numbers to the example to make it more concrete. Suppose either reducing the target alternative's weight or increasing its processor speed leads to a twenty unit increase in demand in a homogeneous customer segment. Suppose 70% (fourteen units) of the growth is drawn from Laptop A following the weight reduction. We would expect a smaller proportion of the growth, say 30% (six units), to be drawn from Laptop A following the processor speed increase. Nevertheless, it follows from the IPS property that 70% (fourteen units) of the growth must be drawn from Laptop A in this case too.

One reason the IPS property has been overlooked may be that it does not necessarily hold in aggregate even if it does hold in every subpopulation. Consider a third example, which allows for differences in consumers' tastes. Suppose there are two types of consumers, Salespeople and Scientists. Assume both types of consumers prefer laptops that weigh less and run faster. Nevertheless, suppose Salespeople are twice as responsive as Scientists to weight reductions, and Scientists are twice as responsive as Salespeople to processor speed increases. Suppose if asked to choose between Laptops A and C, a Salesperson chooses A with probability 0.7 and C with probability 0.3, whereas a Scientist chooses A with probability 0.3 and C with probability 0.7.

Given that Salespeople are twice as responsive as Scientists to weight reductions, if reducing Laptop B's weight leads to a twenty-unit increase in demand among Salespeople, it would lead to a ten unit increase in demand among Scientists. Given that Scientists are twice as responsive as Salespeople to speed increases, if increasing Laptop B's processor speed leads to a twenty-unit increase in demand among Scientists, it would lead to a ten-unit increase in demand among Salespeople.

How would individuals substitute among the laptops if a new alternative is added to the choice set? Under the multinomial logit model, the choice probabilities imply that Salespeople would draw 70% of the increase in demand from Laptop A and 30% from Laptop C no matter which type of improvement is made to Laptop B. Scientists would draw 70% of their demand from Laptop A and 30% from Laptop C following an improvement to Laptop B. This results in the following substitution patterns for each of the subpopulations:

Substitution of Salespeople and Scientists under IPS

	Salespeople		Scientists	
	B becomes lighter	B becomes faster	B becomes lighter	B becomes faster
Laptop A	−14 units (70%)	−7 units (70%)	−3 units (30%)	−6 units (30%)
Laptop B	+20 units	+10 units	+10 units	+20 units
Laptop C	−6 units (30%)	−3 units (30%)	−7 units (70%)	−14 units (70%)

Central to the present discussion, note that the substitution patterns of neither subpopulation conform to our expectations because the proportion of demand drawn from a given competing alternative does not depend on whether the target good becomes more similar to it.

Aggregating the changes in demand for Salespeople and Scientists results in the following substitution patterns for the full population:

Aggregate substitution		
	B becomes lighter	B becomes faster
Laptop A	−17 units (57%)	−13 units (43%)
Laptop B	+30 units	+30 units
Laptop C	−13 units (43%)	−17 units (57%)

In aggregate, it appears that individuals substitute among the laptops according to our expectations. A greater proportion of demand is drawn from Laptop A, the lightest alternative, if the target alternative becomes more similar to it (57% vs. 43%). Likewise, a greater proportion of demand is drawn from Laptop C, the fastest alternative, if the target alternative becomes more similar to it (57% vs. 43%).

Still, we might not be entirely satisfied because models with IPS preclude individual behavior that seems reasonable. For example, the following substitution patterns result in the same aggregate changes in demand but would be precluded by models with the IPS property:

Substitution of Salespeople and Scientists Precluded by IPS				
	Salespeople		Scientists	
	B becomes lighter	B becomes faster	B becomes lighter	B becomes faster
Laptop A	−11.3 units (57%)	−4.3 units (43%)	−5.7 units (57%)	−8.7 units (43%)
Laptop B	+20 units	+10 units	+10 units	+20 units
Laptop C	−8.7 units (43%)	−5.7 units (57%)	−4.3 units (43%)	−11.3 units (57%)

As before, Salespeople differ from Scientists in their tastes for attributes. Salespeople are more responsive than Scientists to weight reductions, and Scientists are more responsive than Salespeople to processor speed increases. Nevertheless, the substitution patterns of both Salespeople and Scientists also depend on how similar the alternatives become. Both types of individuals draw a greater proportion of the growth in demand away from the lightest laptop when the target alternative becomes lighter as opposed to faster, and they draw a greater proportion of growth from the fastest alternative when the target alternative becomes faster as opposed to lighter. An ideal model would allow both of these substitution patterns to arise.

The implications of the IPS property are not limited to the context of choice among similar alternatives. Discrete-choice models are also used to study how marketing actions lead to market penetration. In this context, the IPS property implies that the proportion of growth due to market expansion (substitution away from the outside good) does not depend on which of the product's attributes[1] is improved. This also seems undesirable. For example, in a pharmaceutical drug market this would imply that the proportion of growth due to market expansion is the same no matter whether a drug is improved by lowering its risk of fatality, by lowering its price, or by increasing its consumer-directed advertising support.

The remainder of the article is organized as follows. In §2, I derive the form of the choice probabilities that give rise to IPS from fairly general assumptions about an individual consumer's utility-maximizing behavior; the choice probabilities of GEV and covariance probit models take this form. In §3, I show that the form of the previously derived choice probabilities implies the IPS property and use the nested logit model as an example. In §4, I conclude by discussing how a researcher can allow more flexible substitution patterns to emerge by relaxing the assumptions that lead to the IPS property.

2. The Form of the Choice Probabilities

Let us begin by deriving the form of the choice probabilities from fairly general assumptions about an individual consumer's utility-maximizing behavior. These probabilities represent the researcher's belief about which alternative a consumer will choose from a set of alternatives. The underlying goal is to determine the class of discrete-choice models that possess the IPS property. Because the choice probabilities of GEV and covariance probit models take this form, I will show that these models possess the IPS property.

Suppose a consumer faces a choice in which one alternative is to be selected from a set of J alternatives. Assume the consumer chooses the utility-maximizing alternative, but the utility that would be derived from any of the alternatives cannot be observed by the researcher. Denoting the utility derived from alternative j as u_j, the decision rule assumed to be governing the individual consumer's behavior is to choose alternative j if and only if $u_j > u_k \; \forall k \neq j$.

[1] Products' attributes are broadly defined to include pricing and marketing investment levels in addition to physical characteristics in these studies.

Although utility cannot be observed, the researcher does observe a subset of the alternatives' attributes that influence the choice being made, and the component of utility that depends on these attributes is referred to as the *representative utility*. The representative utility of a given alternative is a function of that alternative's attributes and the consumer's tastes. Let the vector x_j denote the observed attributes of alternative j, the vector β denote the consumer's tastes, the scalar v_j denote the representative utility derived from alternative j, and the function v denote the relationship between the observed attributes and the consumer's tastes

$$v_j = v(x_j, \beta).$$

In the standard case, the function v is assumed to be linear in the alternative's attributes, such that $v_j = x_j'\beta$, but this need not be true. Note that the representative utility of any alternative depends only on that alternative's attributes, not the attributes of other alternatives; this assumption is needed to ensure consistency with random utility theory (RUM).

The utility from alternative j is decomposed as $u_j = v_j + \varepsilon_j \ \forall j$, where ε_j denotes idiosyncratic factors other than the observed attributes that influence utility. These factors may be correlated across alternatives, but $\varepsilon_j \perp x_k \ \forall j, k$. Let $f(\boldsymbol{\varepsilon})$ denote the joint probability density function of the random vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_J)$. Conditional on the consumer's tastes, the researcher's belief about whether the consumer will choose alternative j is described by the probability

$$\begin{aligned} P_j &= \Pr\{\varepsilon_k - \varepsilon_j < v_j - v_k \ \forall k \neq j\} \\ &= \int_{\boldsymbol{\varepsilon}} I\{\varepsilon_k - \varepsilon_j < v_j - v_k \ \forall k \neq j\} f(\boldsymbol{\varepsilon})\, d\boldsymbol{\varepsilon}, \end{aligned}$$

where I denotes the indicator function.

GEV and covariance probit models arise from different assumptions about the distribution of $\boldsymbol{\varepsilon}$. For example, under the multinomial logit model (a type of GEV model) the elements of $\boldsymbol{\varepsilon}$ are assumed to be i.i.d. extreme value across alternatives. This leads to choice probabilities with a closed form, but the substitution among alternatives is restricted by the IIA property. Other GEV models and the covariance probit model introduce correlation among the elements of $\boldsymbol{\varepsilon}$ to relax IIA. The random vector $\boldsymbol{\varepsilon}$ is the distributed generalized extreme value under GEV models and distributed multivariate normal with a full variance–covariance matrix under the covariance probit model. The choice probabilities of GEV models have a closed form, but the probabilities of the covariance probit model do not. It is important to note, however, that the choice probabilities under all of these models depend on the attributes of any alternative only through the representative utility of that alternative. In other words, P_k depends on x_j only through $v_j \ \forall j, k$. As will become clear in the next section, it is this assumption that leads to the IPS property.

3. The IPS Property

The IPS property represents one of the researcher's implicit assumptions about how an individual consumer will substitute away from competing alternatives if improvements are made to one of the available goods. It is said to hold if the proportion of demand generated by substitution away from a given competing alternative is the same no matter which own-good attribute is improved.

Definition. Let x_{ja} be attribute a of alternative j. A discrete-choice model is said to possess IPS if and only if

$$\frac{-\partial P_k/\partial x_{ja}}{\partial P_j/\partial x_{ja}} = \Psi_{k/j} \quad \forall a,$$

where $\Psi_{k/j}$ is a numerical constant for any given $k \neq j$.

The substitution ratio, $(-\partial P_k/\partial x_{ja})/(\partial P_j/\partial x_{ja})$, represents the proportion of the increase in expected demand for alternative j that is generated by substitution away from alternative k following an improvement to attribute x_{ja}. By specifying a model that possesses IPS, the researcher expresses a belief that the substitution ratio does not depend on which attribute is improved. Since demand gained by one alternative must be drawn from another, the substitution ratios across all competing alternatives must sum to one, that is, $\sum_{\forall k \neq j} (-\partial P_k/\partial x_{ja})/(\partial P_j/\partial x_{ja}) = 1$.

Proposition. *Suppose a discrete-choice model has the following characteristics*:

1. *The representative utility that an individual consumer derives from any alternative depends on the attributes of that alternative alone.* $v_j = v(x_j, \beta) \ \forall j$.
2. *The choice probabilities depend on the alternatives' attributes only through the representative utilities.* $P_j = g(v_1, \ldots, v_J) \ \forall j$.[2]

Then, the substitution ratio of alternative k into alternative j does not depend on which attribute is improved,

$$\frac{-\partial P_k/\partial x_{ja}}{\partial P_j/\partial x_{ja}} = \Psi_{k/j} \quad \forall a,$$

and the discrete-choice model possesses the IPS property.

Proof. Since the representative utility that the consumer would derive from any alternative depends on the attributes of that alternative alone, $\partial v_k/\partial x_{ja} = 0$ for $k \neq j$. Furthermore, since the choice probabilities depend on the alternatives' attributes only through the representative utilities (as opposed to, let us say, through both the representative utilities and the attributes directly), the chain rule implies

$$\frac{\partial P_k}{\partial x_{ja}} = \frac{\partial P_k}{\partial v_j} \cdot \frac{\partial v_j}{\partial x_{ja}} \quad \forall j, k.$$

[2] This would not be true, for instance, for socio-economic variables that enter the representative utility of every alternative. I thank an anonymous reviewer for suggesting this clarification.

The first term, $\partial P_k/\partial v_j$, describes the rate of change in the choice probabilities for a change in representative utility v_j. This term does not depend on which attribute is improved. The second term, $\partial v_j/\partial x_{ja}$, describes the rate of change in representative utility v_j for a change in attribute x_{ja}. This term does depend on which attribute is improved. Yet, because a change in attribute x_{ja} affects every choice probability only through representative utility v_j, this term cancels out of the substitution ratio, leaving

$$\frac{-\partial P_k/\partial x_{ja}}{\partial P_j/\partial x_{ja}} = \frac{-\partial P_k/\partial v_j}{\partial P_j/\partial v_j}.$$

Since the ratio $(-\partial P_k/\partial v_j)/(\partial P_j/\partial v_j)$ does not depend on a, the discrete-choice model possesses the IPS property. Q.E.D.

GEV Models

The GEV family (McFadden 1978) represents a large class of models that allow individuals to exhibit a wide variety of substitution patterns. These models are called generalized extreme value[3] because the unobserved component of the individual's utility is a distributed generalized extreme value, a distribution that allows the unobserved utility to be correlated across alternatives. When all correlations are zero, more complex GEV models become the standard multinomial logit. The multinomial logit (McFadden 1974) and the nested logit (Ben-Akiva 1973, Williams 1977, McFadden 1978, and Daly and Zachary 1978) are the most widely used GEV models, but Koppelman and Sethi (2000) discuss a number of other models that fall into this class, including the paired combinatorial logit model (Chu 1989, Koppelman and Wen 2000), the cross-nested logit model (Vovsha 1997), the generalized nested logit model (Wen and Koppelman 2001), the generation logit model (Swait 2001), the principles of differentiation model (Bresnahan et al. 1997), and the cross-correlated model (Williams 1977). Daly and Bierlaire (2006) provide a general theoretical foundation for GEV models and propose an easy way of generating new models without a need for complicated proofs.

The choice probabilities of GEV models take the form

$$P_j = \frac{y_j}{G(y_1,\ldots,y_J)} \cdot \frac{\partial G(y_1,\ldots,y_J)}{\partial y_j},$$

where $y_j \equiv e^{v_j}$, v_j is the representative utility of good j, and the function G satisfies:

1. $G \geq 0$ for all positive values of y_j.
2. G is homogeneous of degree one.
3. $G \to \infty$ as $y_j \to \infty$ for any j.
4. The cross partial derivatives of G change signs in a particular way. Specifically, $\partial G/\partial y_j > 0$ for all j; $\partial^2 G/(\partial y_j \partial y_k) < 0$ for all $j \neq k$; $\partial^3 G/(\partial y_j \partial y_k \partial y_l) > 0$ for all distinct j, k, and l; and so on.

Given the form of the choice probabilities, the substitution ratio for GEV models is

$$\frac{-\partial P_k/\partial y_j}{\partial P_j/\partial y_j} = \Psi_{k/j} \quad \forall\, a.$$

(Proof in appendix.) Since the substitution ratio does not depend on which attribute is improved, GEV models possess the IPS property.

Example: Nested Logit

The nested logit model provides a nice illustration of the IPS property. The choice probabilities and their derivatives take a closed form, so we can analytically determine the substitution ratio. Yet, because the IIA property does not hold across all alternatives and the IPS property does, it is obvious that models that relax IIA do not necessarily also relax IPS.

Assume the nested logit model. Let the set of J alternatives be divided into M mutually exclusive nests where B_m denotes the set of alternatives in nest m. The random vector of unobserved utility $\boldsymbol{\varepsilon}$ is distributed GEV with parameter $0 \leq \rho_m < 1$ denoting the correlation among alternatives in nest m. ($\rho_m = 0$ implies no correlation.) The choice probability of alternative j in nest B_m is decomposed as

$$P_j = P_{B_m} \cdot P_{j\,|\,B_m},$$

where

$$P_{j\,|\,B_m} = \frac{e^{v_j/(1-\rho_m)}}{\sum_{i\in B_m} e^{v_i/(1-\rho_m)}},$$
$$P_{B_m} = \frac{e^{(1-\rho_m)I_m}}{\sum_{l=1}^{M} e^{(1-\rho_l)I_l}},$$
$$I_l = \ln \sum_{i\in B_l} e^{v_i/(1-\rho_l)}.$$

$P_{j\,|\,B_m}$ is the probability of choosing alternative j given nest B_m is chosen; P_{B_m} is the probability of choosing nest B_m; and I_m is the inclusive value of nest B_m.

The derivative of choice probability P_k with respect to representative utility v_j is

$$\frac{\partial P_k}{\partial v_j} = \begin{cases} \dfrac{P_j}{1-\rho_n}[1-\rho_n P_{j|B_n} - (1-\rho_n)P_j] & \text{for } k = j \in B_n, \\[2ex] \dfrac{-P_k}{1-\rho_n}[\rho_n P_{j\,|\,B_n} + (1-\rho_n)P_j] & \\ \hfill \text{for } k \neq j \text{ and } k, j \in B_n, \\[2ex] -P_k P_j \quad \text{for } k \neq j,\ k \in B_n \text{ and } j \in B_m. \end{cases}$$

[3] Train (2003, Chapter 3) provides an excellent overview of GEV models.

If the representative utility is a linear function of the attributes, as is most common, then the derivative of v_j with respect to attribute x_{ja} is β_a, where β_a is the coefficient of attribute x_{ja}. This derivative allows the amount of demand that is generated by an improvement to vary across attributes, but it cancels out of the substitution ratio as previously discussed.

The substitution ratio, which is defined only for $k \neq j$, is

$$\frac{-\partial P_k/\partial x_{ja}}{\partial P_j/\partial x_{ja}} = \begin{cases} \dfrac{P_k[\rho_n P_{j\mid B_n} + (1-\rho_n)P_j]}{P_j[1-\rho_n P_{j\mid B_n} - (1-\rho_n)P_j]} \\ \quad \text{for } k, j \in B_n, \\ \dfrac{P_k}{1-\rho_n P_{j\mid B_n} - (1-\rho_n)P_j} \\ \quad \text{for } k \in B_n \text{ and } j \in B_m. \end{cases}$$

Since the substitution ratio does not depend on which attribute is improved, the nested logit model possesses the IPS property. This is to be expected, of course, because the nested logit is a GEV model, but a skeptical reader can verify this fact by directly taking the derivative of P_k with respect to x_{ja}.

4. Discussion

The implications of the IPS property on consumer behavior extend beyond the context of choice among similar alternatives. For example, many studies (c.f. Gupta 1988, Chiang 1991, Chintagunta 1993, Bucklin et al. 1998, Bell et al. 1999, van Heerde et al. 2003, Steenburgh 2007) have examined how marketing actions affect consumers' decisions of whether, which, and how much to buy. A key question in these studies is whether a marketing action entices consumers to enter the market (whether to buy) or persuades them to switch brands (which brand to buy).

Choice models are commonly used to represent this aspect of the consumers' decision. Regardless of whether the model possesses the IPS property or not, the growth in own-good choice due to a marketing action can be decomposed into

$$\frac{\partial P_j}{\partial x_{ja}} = -\frac{\partial P_0}{\partial x_{ja}} - \sum_{\substack{k=1 \\ k\neq j}}^{J} \frac{\partial P_k}{\partial x_{ja}},$$

where P_0 denotes the probability of choosing the outside good and P_k for $k \neq j, 0$ denotes the probability of choosing competing alternative k. The first term of the decomposition, $-\partial P_0/\partial x_{ja}$, measures the growth in the probability of own-good choice due to consumers entering the market, and the second term, $-\sum_{k=1;\, k\neq j}^{J} \partial P_k/\partial x_{ja}$, measures the growth stolen from other brands.

The proportion of growth due to consumers entering the market from marketing action a is measured by the substitution ratio

$$\Psi_a = \frac{-\partial P_0/\partial x_{ja}}{\partial P_j/\partial x_{ja}},$$

to which the following interpretation applies: If marketing action a generates 100 additional purchases for the target good, then $\Psi_a * 100$ of those purchases are made by consumers who would have bought nothing and $(1-\Psi_a) * 100$ are made by consumers who would have bought another brand.

The IPS property implies that the proportion of growth due to consumers entering the market is the same no matter what type of marketing action is taken. This seems to be an undesirably strong assumption to make about how consumers behave. It is well known, for example, that advertising effects can spillover to other brands; thus, we might speculate that a greater proportion of the growth in own-good choice is due to consumers entering the market when firms invest in feature advertising as opposed to when they drop price. We would need to specify a model without the IPS property, however, to empirically test whether this is true.

A universal logit model (McFadden 1975, Koppelman and Sethi 2000) might provide a useful solution to this problem. The universal logit allows the representative utility of each alternative to depend on the attributes of competing alternatives in addition to its own attributes. In this example, the representative utility of each brand would depend on the price and advertising levels of competing brands in addition to its own price and advertising levels. Since the terms $\partial v_k/\partial x_{ja}$ are no longer restricted to be zero, the term $\partial v_j/\partial x_{ja}$ does not cancel out of the substitution ratio; thus, the universal logit model does not possess the IPS property. This generalization is described as the "universal" or "mother" logit because it can be used to approximate all qualitative choice models to any desired degree of accuracy. It is possible to specify the model such that the substitution ratio will vary across marketing instruments.

In spite of its promise as a flexible model, the universal logit does raise significant concerns. While it may be consistent with RUM in some cases, it is not guaranteed to be consistent in all cases. Further work would need to show when the universal logit satisfies RUM requirements. Koppelman and Sethi (2000), in fact, point out that few examples of the universal logit exist in the literature today and speculate that "This may be due to lack of consistency with [random] utility maximization in some cases, the potential to obtain counterintuitive elasticities, and the complexity of search for a preferred specification (Ben Akiva 1974)."

A different modeling approach may be useful in the context of choice among similar alternatives. The trick in this case is to allow the unobserved utility of any two alternatives to become more correlated as the alternatives become more similar and therefore become closer substitutes. This can be accomplished by relaxing the assumption that the unobserved utility is statistically independent of the alternatives' attributes. For example, consider an error components model in which the unobserved utility is specified as $\varepsilon_k + x_k\tau$, where ε_k is a random variable and τ is a random vector. This results in choice probabilities of the form:

$$P_j = \int_{\boldsymbol{\varepsilon}} \int_{\boldsymbol{\tau}} I\{(\varepsilon_k + x_k'\tau) - (\varepsilon_j + x_j'\tau) < v_j - v_k \ \forall\, k \neq j\} f(\boldsymbol{\tau}, \boldsymbol{\varepsilon})\, d\boldsymbol{\tau}\, d\boldsymbol{\varepsilon}.$$

These choice probabilities depend on the alternatives' attributes not only through the individual's representative utilities v_j and v_k, but also through the terms $x_j'\tau$ and $x_k'\tau$; thus, the model does not possess the IPS property. Furthermore, as any two alternatives become more similar, their errors become more correlated. By contrast, GEV and covariance probit models assume the correlation between alternatives to be fixed no matter how the alternatives' attributes change.

Discrete-choice models are undoubtedly useful in their current form, and the implications of the IPS property need to be balanced against many other considerations when specifying a choice model. Nevertheless, the IPS property does imply counterintuitive behavior that may be important to alleviate in some research settings, and the aforementioned models provide the interested researcher with immediate paths to follow. The hope is that further investigation will lead to even more robust choice models in the future.

Acknowledgments

The author thanks Andrew Ainslie, Greg Allenby, Jill Avery, David E. Bell, John Gourville, Steve Jordan, Frank S. Koppelman, Jordan Louviere, Elie Ofek, and Luc Wathieu for providing useful comments. Kenneth Train deserves special thanks for several discussions that helped broaden the ideas presented in this article. Of course, the author alone is responsible for all remaining errors.

Appendix

Under GEV models, the choice probabilities take the form

$$P_k = \frac{y_k}{G(y_1, \ldots, y_J)} \cdot \frac{\partial G(y_1, \ldots, y_J)}{\partial y_k} \quad \text{for } k = 1, \ldots, J,$$

where $y_k \equiv e^{v_k}$ and v_k is the representative utility of alternative k.

Since the representative utility of good j is a function of the attributes of only alternative j, the variable y_j is a function of the attributes of only alternative j. Thus,

$$\frac{\partial y_k}{\partial x_{ja}} = 0 \quad \text{for } k \neq j.$$

By the chain rule, this leads to

$$\frac{\partial P_k}{\partial x_{ja}} = \frac{\partial P_k}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_{ja}} \quad \forall\, k.$$

Thus, the substitution ratio is

$$\frac{-\partial P_k/\partial x_{ja}}{\partial P_j/\partial x_{ja}} = \frac{-\partial P_k/\partial y_j}{\partial P_j/\partial y_j} \cdot \frac{\partial y_j/\partial x_{ja}}{\partial y_j/\partial x_{ja}} = \frac{-\partial P_k/\partial y_j}{\partial P_j/\partial y_j}.$$

Since the substitution ratio is the same no matter which attribute is improved, GEV models possess the IPS property. Q.E.D.

References

Bell, D. R., J. Chiang, V. Padmanabhan. 1999. The decomposition of promotional response. *Marketing Sci.* **18**(4) 504–526.

Ben-Akiva, M. 1973. The structure of travel demand models. PhD Thesis, Department of Civil Engineering, MIT Press, Cambridge, MA.

Ben-Akiva, M. 1974. Note on the specification of a logit model with utility functions that include attributes of competing alternatives. Working paper, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Bresnahan, T. F., S. Stern, M. Trajtenberg. 1997. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND J. Econom.* **28**(0) S17–S44.

Bucklin, R. E., S. Gupta, S. Siddarth. 1998. Determining segmentation in sales response across consumer purchase behaviors. *J. Marketing Res.* **35**(2) 189–197.

Chiang, J. 1991. A simultaneous approach to the whether, what, and how much to buy questions. *Marketing Sci.* **10** 297–315.

Chintagunta, P. K. 1993. Investigating purchase incidence, brand choice, and purchase quantity decisions of households. *Marketing Sci.* **12** 184–208.

Chu, C. 1989. A paired combinatorial logit model for travel demand analysis. *Proc. Fifth World Conf. Transportation Res.*, Vol. 4, Ventura, CA, 295–309.

Daly, A. J., S. Zachary. 1978. Improved multiple choice models. D. A. Hensher, Q. Dalvi, eds. *Determinants of Travel Choice*. Saxon House, Sussex, UK, 335–357.

Daly, A., M. Bierlaire. 2006. A general and operational representation of generalised extreme value models. *Transportation Res. Part B* **40**(4) 285–305.

Debreu, G. 1960. Review of R. D. Luce *Individual Choice Behavior*. *Amer. Econom. Rev.* **50** 186–188.

Gupta, S. 1988. Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* **25**(4) 342–355.

Koppelman, F. S., V. Sethi. 2000. Closed form discrete choice models. D. A. Hensher, K. J. Button, eds. *Handbook of Transportation Modeling*. Pergamon, Oxford, UK, 211–228.

Koppelman, F. S., C.-H. Wen. 2000. The paired combinatorial logit model: Properties, estimation, and application. *Transportation Res. Part B* **34**(2) 75–89.

Luce, R. D., P. Suppes. 1965. Preference, utility and subjective probability. R. D. Luce, R. R. Bush, E. Galanter, eds. *Handbook of Mathematical Psychology, III*. John Wiley and Sons, New York, 249–410.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.

McFadden, D. 1975. On independence, structure, and simultaneity in transportation demand analysis. Working Paper 7511, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, CA.

McFadden, D. 1978. Modeling the choice of residential location. A. Karlqvist, L. Lundqvist, F. Snickars, J. Weibull, eds. *Spatial Interaction Theory and Planning Models*. North Holland, Amsterdam, 75–96.

McFadden, D. 1984. Econometric analysis of qualitative response models. Z. Griliches, M. Intrilligator, eds. *Handbooks of Econometrics*, Vol. II. Elsevier, Amsterdam, 1396–1456.

McFadden, D., K. Train, W. B. Tye. 1978. An application of diagnostic tests for the independence from irrelevance alternatives property of the multinomial logit model. *Transportation Record Res.* **637** 39–46.

Steenburgh, T. J. 2007. Measuring consumer and competitive impact with elasticity decompositions. *J. Marketing Res.* **44**(4) 636–646.

Swait, J. 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Res. Part B* **35**(7) 285–305.

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

Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psych. Rev.* **79**(4) 281–299.

Van Heerde, H. J., S. Gupta, D. R. Wittink. 2003. Is 75% of the sales promotion bump due to brand switching? No, only 33% is. *J. Marketing Res.* **40**(4) 481–491.

Vovsha, P. 1997. The cross-nested logit model: Application to mode choice in the Tel-Aviv metropolitan area. *Transportation Res. Board, 76th Annual Meeting*, Transportation Research Board, Washington, D.C., Paper 970387.

Wen, C.-H., F. S. Koppelman. 2001. The generalized nested logit model. *Transportation Res. Part B* **35**(7) 627–641.

Williams, H. 1977. On the formation of travel demand models and the economic evaluation measures of user benefit. *Environment Plann. A* **9**(3) 285–344.