05-022

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

Thomas J. Steenburgh

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

Thomas J. Steenburgh
Harvard Business School
Soldiers Field
Boston, MA 02163

tsteenburgh@hbs.edu
Phone: 617-495-6056
Fax: 617-496-5853

31 May 2006

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

Abstract

This article identifies a property of several standard discrete-choice models that amounts to an implicit assumption about <u>individual</u> choice behavior. This property, which I call the Invariant Proportion of Substitution (IPS), implies that the proportion of growth in expected own-good choice that an individual consumer draws from a given competing alternative is the same no matter which own-good attribute is improved. The IPS and Independence from Irrelevant Alternatives (IIA) properties are similar. But models that relax IIA, such as generalized extreme value (GEV) and covariance probit models, do not necessarily also relax IPS. Some models that do relax IPS are discussed.

Keywords: Discrete-Choice Models, Econometric Models

1. Introduction

This article identifies a property of several standard discrete-choice models that amounts to an implicit assumption about individual choice behavior. This property, which I call the Invariant Proportion of Substitution (IPS), implies that the proportion of growth in expected own-good choice that an individual consumer draws from a given competing alternative is the same no matter which own-good attribute is improved. The IPS and Independence from Irrelevant Alternatives (IIA) properties are similar. But models that relax IIA, such as generalized extreme value (GEV) and covariance probit models, do not necessarily also relax IPS.

The following example illustrates the IPS property. Let's say that an individual consumer faces a choice among three laptop computers with the following attributes:

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

Laptop B is the target alternative and it has moderate speed and moderate weight. Laptop A is the lightest alternative, but it runs at the slowest speed. Laptop C is the fastest alternative, but it has the greatest weight.

We want to understand how the individual behaves in response to improvements in the target good. Discrete-choice models that possess the IPS property allow the individual to be more responsive to improvements in one attribute, say a weight reduction, than to improvements in another, a processor speed increase. Yet, the proportion of growth in expected own-good choice that an individual draws from a given competing alternative is assumed to be the same no matter which own-good attribute is improved. In other words, the proportion of growth drawn from the lightest laptop is the same regardless of whether the target good is made more similar by reducing its weight or more dissimilar by increasing its speed. Likewise, the proportion of growth drawn from the fastest laptop is the same no matter if the target good is made more similar by increasing its processor speed or more dissimilar by reducing its weight. Such behaviors seem counterintuitive.

For example, let's say that a small weight reduction produces twenty incremental units for Laptop B in a homogeneous customer segment. Fourteen of these units (70%) are drawn from the lightest alternative and six (30%) are drawn from the fastest. The IPS property implies that the same proportions must hold no matter which attribute is improved. Thus, if a small processor speed increase produces ten incremental units for Laptop B, seven of these units (70%) are drawn from the lightest alternative and three (30%) are drawn from the fastest. While this pattern of substitution is possible, it is not necessarily something that we want to impose by assumption. If anything, we might expect the proportion of growth drawn from the lightest laptop to be greater if Laptop B is made lighter instead of faster. Likewise, we might expect that the proportion of growth drawn from the fastest laptop to be greater when Laptop B is made faster instead of lighter.

Allowing for differences in consumers' tastes is important, but it does not resolve the problem of how an individual consumer would behave. Suppose that there are two types of customers, Salespeople and Scientists. Both Salespeople and Scientists prefer laptops that weigh less and that run faster. All the same, Salespeople are more responsive than Scientists to weight reductions, and Scientists are more responsive than Salespeople to processor speed increases.

Suppose the previous changes in demand describe how Salespeople respond to changes in Laptop B's attributes, and Scientists respond in the opposite way. Scientists value faster processors more than Salespeople do, and the processor speed increase produces twenty incremental units in the Scientists' customer segment. Fourteen of these units are drawn from the fastest laptop and six of these units are drawn from the lightest laptop. Furthermore, Scientists value lighter weights less than Salespeople do, and the weight reduction produces only ten incremental units in the Scientists' customer segment. Seven of these units are drawn from the fastest laptop and three are drawn from the lightest. The substitution patterns of Salespeople and Scientists are as follows:

Substitution of Salespeople and Scientists under IPS

	Salespeople		Scientists	
	Laptop B Is Made Lighter	Laptop B Is Made Faster	Laptop B Is Made Lighter	Laptop B Is Made 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%)

One reason the IPS property has been overlooked might be that it does not necessarily hold in aggregate even if it does hold in every subpopulation. Aggregating the changes in demand for Salespeople and for Scientists results in the following market-level substitution patterns:

Aggregate Substitution

	Laptop B Is Made Lighter	Laptop B Is Made 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, the proportion of growth drawn from the lightest laptop is greater (57% vs. 43%) when Laptop B is made lighter instead of faster. Likewise, the proportion of growth drawn from the fastest laptop is greater (57% vs. 43%) when Laptop B is made faster instead of lighter. Thus, the aggregate substitution patterns do not possess IPS and conform to our expectations. The proportion of growth drawn from a given competing alternative is greater when the target good is made more similar to it.

Still, we might not be entirely satisfied because models that possess the IPS property 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 the IPS property:

Substitution of Salespeople and Scientists Precluded by IPS

	Salespeople		Scientists	
	Laptop B Is Made Lighter	Laptop B Is Made Faster	Laptop B Is Made Lighter	Laptop B Is Made 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. But, in this case, the behavior of both Salespeople and Scientists also depends on how the target good is improved. Both Salespeople and Scientists draw a greater proportion of demand from the lightest laptop when the target good is made lighter instead of faster, and they draw a greater proportion of demand from the fastest laptop when the target good is made faster instead of lighter. An ideal model would allow both of these substitution patterns to arise.

The IPS property has a special implication for models that contain an outside good. In this case, the IPS property implies that the proportion of demand created by market expansion, substitution away from the outside good, does not depend on which of the products' attributes are improved. A product's attributes, of course, typically include not only its physical characteristics, but also its price and marketing investment levels. In the context of pharmaceutical drugs, this would imply that the proportion of own-good demand created by market expansion would be the same whether a manufacturer chooses to improve its drug by lowering the risk of fatality, lowering the price paid by consumers, increasing the physician-directed advertising levels, or increasing the consumer-directed advertising levels. It seems doubtful that a researcher would want to impose this restriction on individual consumers' choices by assumption.

The remainder of the article is organized as follows. In section two, I derive the form of the choice probabilities under fairly general assumptions about an individual consumer's utility-maximizing behavior. The choice probabilities of generalized extreme value and covariance probit discrete-choice models take this form. In section three, I show that the form of the previously derived choice probabilities implies the IPS property and discuss some of the property's implications using the nested logit model as an example. In section four, I discuss 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's 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. Since the choice probabilities of generalized extreme value and covariance probit models take this form, 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 will choose the utility-maximizing alternative, but the utility that would be derived from any of the alternatives cannot be observed. 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 $u_j > u_k \;\; \forall k \neq j$.

While 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.

$$v_j = v\left(x_j, \beta\right)$$

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 on only that alternative's attributes, not the attributes of other alternatives; this ensures consistency with economic theory.

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 $f\left(\boldsymbol{\varepsilon}\right)$ denote the joint probability density function of the random vector $\boldsymbol{\varepsilon} = \left(\varepsilon_1, \ldots, \varepsilon_J\right)$. Conditional on the consumer's tastes, the researcher's belief about whether the consumer will choose alternative *j* is described by the probability

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

where *I* denotes the indicator function.

Generalized extreme value (GEV) and covariance probit models arise from different assumptions about the distribution of $\boldsymbol{\varepsilon}$. For example, under the multinomial logit model (which is 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 distributed generalized extreme value under GEV models and is 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, θ_k depends on x_j only through $v_j \;\; \forall j,k$. As will become clear in the next section, this assumption leads to the IPS property.

3. The IPS Property

The IPS property represents one of the researcher's 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 that is 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

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

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

The substitution ratio, $\frac{-\partial\theta_k / \partial x_{ja}}{\partial\theta_j / \partial x_{ja}}$, 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 that is gained by one alternative must be drawn from another, the substitution ratios across all competing alternatives must sum to one, that is $\sum_{\forall k \neq j} \frac{-\partial\theta_k / \partial x_{ja}}{\partial\theta_j / \partial x_{ja}} = 1.$

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

1. The representative utility that an individual consumer would derive from any alternative depends on the attributes of that alternative alone. $v_j = v\left(x_j, \beta\right) \; \forall j$.

2. The choice probabilities depend on the alternatives' attributes only through the representative utilities. $\theta_j = f\left(v_1,\ldots,v_J\right) \;\; \forall j$.[1]

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

$$\frac{-\partial\theta_k \big/ \partial x_{ja}}{\partial\theta_j \big/ \partial x_{ja}} = \Psi_{k/j} \quad \forall a \,,$$

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, $\partial v_k \big/ \partial x_{ja} = 0$ for $k \neq j$. Furthermore, since the choice probabilities depend on the alternatives' attributes only through the representative utilities (as opposed to, let's say, through both the representative utilities and the attributes directly), the chain rule implies

$$\frac{\partial\theta_k}{\partial x_{ja}} = \frac{\partial\theta_k}{\partial v_j}\frac{\partial v_j}{\partial x_{ja}} \;\; \forall j,k \,.$$

The first term, $\partial\theta_k \big/ \partial v_j$, 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, $\partial v_j \big/ \partial x_{ja}$, 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

[1] 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.

improved. Yet, since 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

$$\frac{\partial\theta_k / \partial x_{ja}}{\partial\theta_j / \partial x_{ja}} = \frac{\partial\theta_k / \partial v_j}{\partial\theta_j / \partial v_j}.$$

Since the ratio $\frac{\partial\theta_k / \partial v_j}{\partial\theta_j / \partial v_j}$ does not depend on *a*, the discrete-choice model possesses the IPS property. Q.E.D.

Generalized Extreme Value Models

McFadden (1978) introduced a large class of models that exhibit a wide variety of substitution patterns. These are called generalized extreme value (GEV) models[2] because the unobserved component of the individual's utility is distributed generalized extreme value, a distribution which 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 (McFadden, 1978; Daly and Zachary, 1978; and Williams, 1977) 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

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

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

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

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. Defining $\partial G / \partial y_j > 0$ for all *j*, $\partial^2 G / \left(\partial y_j \partial y_k\right) < 0$ for all *j*≠*k*, $\partial^3 G / \left(\partial y_j \partial y_k, \partial y_l\right) > 0$ for all distinct *j*, *k*, and *l*, and so on.

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

$$\frac{-\partial \theta_k / \partial y_j}{\partial \theta_j / \partial y_j} = \Psi_{k/j} \quad \forall a .$$

(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 nicely illustrates the IPS property. The choice probabilities and their derivatives take a closed form, so we can analytically determine the substitution ratio. Yet, since 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 generalized extreme value with parameter $0 \le \rho_m < 1$ 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

$$\theta_j = \theta_{B_n} \cdot \theta_{j|B_n},$$

where

$$\theta_{j|B_m} = \frac{e^{v_j/(1-\rho_m)}}{\sum_{i\in B_m} e^{v_i/(1-\rho_m)}}$$

$$\theta_{B_m} = \frac{e^{(1-\rho_m)I_m}}{\sum_{l=1}^{M} e^{(1-\rho_l)I_l}}$$

$$I_l = \ln \sum_{i\in B_l} e^{v_i/(1-\rho_l)}$$

$\theta_{j|B_m}$ is the probability of choosing alternative *j* given nest B_m is chosen, θ_{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 θ_k with respect to representative utility v_j is

$$\frac{\partial\theta_k}{\partial v_j} = \begin{cases} \dfrac{\theta_j}{1-\rho_n}\left[1-\rho_n\theta_{j|B_n} - (1-\rho_n)\theta_j\right] & \textit{for } k = j \in B_n \\ -\dfrac{\theta_k}{1-\rho_n}\left[\rho_n\theta_{j|B_n} + (1-\rho_n)\theta_j\right] & \textit{for } k \ne j \textit{ and } k, j \in B_n \\ -\theta_k\theta_j & \textit{for } k \ne j, k \in B_n, \textit{and } j \in B_m \end{cases}$$

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 only defined for $k \neq j$, is

$$\frac{-\partial\theta_k / \partial x_{ja}}{\partial\theta_j / \partial x_{ja}} = \begin{cases} \dfrac{\theta_k \left[\rho_n \theta_{j|B_n} + (1-\rho_n)\theta_j\right]}{\theta_j \left[1-\rho_n \theta_{j|B_n} - (1-\rho_n)\theta_j\right]} & \text{for } k, j \in B_n \\ \dfrac{\theta_k}{1-\rho_n \theta_{j|B_n} - (1-\rho_n)\theta_j} & \text{for } k \in B_n,\ j \in B_m \end{cases}$$

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 θ_k with respect to x_{ja}.

Let's now discuss the IPS property in the context of some recent academic findings based on the nested logit model. Bell, Chiang and Padmanabhan (1999) and Bucklin, Gupta and Siddarth (1998) use the nested logit model to study whether demand is generated from market expansion or from brand switching. (These studies also examine whether consumers increase their purchase quantities, but we'll ignore this aspect of the consumers' decision-making process for ease of discussion.) A nested logit model that includes an outside option provides a seemingly convenient way to study this problem. The changes in the choice probabilities following an improvement to marketing instrument x_{ja} can be decomposed as

$$\frac{\partial \theta_j}{\partial x_{ja}} = -\frac{\partial \theta_0}{\partial x_{ja}} - \sum_{\substack{k=1 \\ k \neq j}}^{J} \frac{\partial \theta_k}{\partial x_{ja}} ,$$

where θ_0 denotes the probability of choosing the outside good and θ_k for $k \neq 0, j$ denotes the probability of choosing a competing alternative. Thus, the first term, $-\partial \theta_0 / \partial x_{ja}$, measures the own-good demand generated by market expansion and the second term, $-\sum_{\substack{k=1 \\ k \neq j}}^{J} \partial \theta_k / \partial x_{ja}$, measures the own-good demand generated at the other brands' expense.

The IPS property implies that the proportion of own-good demand that is generated by market expansion is the same regardless of whether the retailer chooses to drop its price, to increase its feature advertising, or to include an in-store display in support of a given brand. This is an undesirably strong assumption about how changes in the marketing mix variables will affect the consumers' decisions of whether and of which brand to purchase. Presumably, Bucklin, Gupta and Siddarth (1998) would not choose to impose it because in the opening paragraph of their article they state, 'Marketing mix variables can affect these three decisions to differing degrees.' Yet, since the authors of both studies specify a nested logit model, which imposes IPS on the consumers' substitution patterns, their findings are meaningful only if all of the marketing instruments have the same impact on the consumers' decisions of whether and of which brand to buy.[3]

[3] This may also explain why both studies only report on the effects of pricing changes.

4. Discussion

Balancing the need for parsimony and flexibility is difficult, but a few modeling approaches might be useful if a researcher wants to relax the IPS property and allow more flexible substitution patterns to emerge. The universal or 'mother' logit model, developed by McFadden (1975), provides one such solution. Under the universal logit model, the representative utility of each alternative depends not only its own attributes, but on the attributes of other alternatives too. 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. Furthermore, since these terms vary across attributes, the universal logit model does not possess the IPS property.

Despite its promise as a flexible model, few examples of the universal logit exist in the literature today. Koppelman and Sethi (2000) speculate that, "This may be due to lack of consistency with utility maximization in some cases, the potential to obtain counterintuitive elasticities[4], and the complexity of search for a preferred specification." These issues aside, the universal logit provides a useful way to allow for more flexible substitution when IPS is a concern, and this solution may be particularly useful in studies that contain an outside good.

Alternatively, introducing idiosyncratic variation in the consumer's taste parameters relaxes the IPS property. We might interpret this variation as being due either to the consumer's tastes fluctuating over time, which seems to be a strong behavioral assumption, or simply to something that the researcher cannot resolve about the consumer's tastes no matter how much data are collected. This assumption relaxes IPS

[4] Ben Akiva (1974) discusses implications of the universal logit for cross-good elasticities.

because the choice probabilities no longer depend on the alternatives' attributes through the representative utilities alone.

Consider a representative utility that is linear in the taste parameters. Let the vector $\beta = \overline{\beta} + \tau$ represent the consumer's tastes, where the vector $\overline{\beta}$ is fixed and the elements of the vector τ are zero-centered, random terms. This results in choice probabilities of the form

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

Since the choice probabilities depend on the alternatives' attributes directly through the terms $x_j'\boldsymbol{\tau}$ and $x_k'\boldsymbol{\tau}$, the IPS property does not hold. This model might be used to address the similarity issue discussed in the introduction of this paper.

It is worth pointing out that the aforementioned models relax both the IPS and the IIA properties[5], whereas GEV and covariance probit models relax only IIA. In doing so, even more flexible substitution patterns are allowed to emerge, albeit at a cost of greater dimensionality. While other techniques most certainly exist, these two models provide the interested researcher with immediate paths to follow. The hope is that this article will serve as a conceptual stepping stone in the development of even more robust models.

[5] McFadden, Train, and Tye (1978) use the universal logit model to test for violations of IIA.

Acknowledgements

I would like to Andrew Ainslie, Greg Allenby, David E. Bell, John Gourville, Steve Jordan, Frank S. Koppelman, Elie Ofek, and Luc Wathieu for reading early drafts and providing useful comments. Kenneth Train deserves special thanks for several discussions that helped broaden the ideas discussed in this article. Of course, the author alone is responsible for all remaining errors.

References

Bell, David R., Jeongwen Chiang and V. Padmanabhan (1999), “The Decomposition of Promotional Response,” *Marketing Science*, 18, 4, 504-526.

Ben-Akiva, Moshe (1974), “Note on the Specification of a Logit Model with Utility Functions that Include Attributes of Competing Alternatives,” working paper, Department of Civil Engineering, MIT, Cambridge, MA.

Bresnahan, Timothy F., Scott Stern and Manuel Trajtenberg (1997), “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s,” *RAND Journal of Economics*, 28, 0, S17-S44.

Bucklin, Randolph E, Sunil Gupta, and S. Siddarth (1998), “Determining Segmentation in Sales Response Across Consumer Purchase Behaviors,” *Journal of Marketing Research*, 35, 2, 189-197.

Chu, C. (1989) “A Paired Combinatorial Logit Model for Travel Demand Analysis,” Proceedings of the Fifth World Conference on Transportation Research, Vol. 4, Ventura, California, pp. 295-309.

Daly, Andrew J. and Stanley Zachary (1978) “Improved Multiple Choice Models,” in *Determinants of Travel Choice*, David A. Hensher and Quasim Dalvi editors, Saxon House, Sussex, pp. 335-357.

Daly, Andrew and Michel Bierlaire (2006) “A General and Operational Representation of Generalised Extreme Value Models,” *Transportation Research Part B*, 40, 4, 285-305.
Barten, Anton P. (1966), “Theorie en Empirie van een Volledig Stelsel van Vraagvergelijkingen,” Doctoral Dissertation, Rotterdam: University of Rotterdam.

Koppelman, Frank S. and Chieh-Hua Wen (2000) “The Paired Combinatorial Logit Model: Properties, Estimation, and Application,” *Transportation Research Part B*, 34, 2, 75-89.

Koppelman, Frank S. and Vaneet Sethi (2000), “Closed Form Discrete Choice Models,” in *Handbook of Transportation Modeling*, David A. Hensher and Kenneth J. Button editors, Pergamon Press, Oxford, pp. 211-228.

McFadden, Daniel (1974) “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, Paul Zarembka editor, Academic Press, New York, pp. 105-142.

McFadden, Daniel (1975), “On Independence, Structure, and Simultaneity in Transportation Demand Analysis," W.P. 7511, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley, CA.

McFadden, Daniel (1978) “Modeling the Choice of Residential Location,” in *Spatial Interaction Theory and Planning Models*, Anders Karlqvist, L. Lundqvist, Folke Snickars, and J. Weibull editors, North Holland, Amsterdam, pp. 75-96.

McFadden, Daniel, Kenneth Train and William B. Tye (1978), “An Application of Diagnostic Tests for the Independence From Irrelevance Alternatives Property of the Multinomial Logit Model,” *Transportation Record Research*, 637, 39-46.

McFadden, Daniel (1984), "Econometric Analysis of Qualitative Response Models," in Z. Griliches and M. Intrilligator (eds.), Handbook of Econometrics, Vol. II, 1396-1456, Elsevier: Amsterdam, 1984

Swait, Joffre (2001) “Choice Set Generation within the Generalized Extreme Value Family of Discrete Choice Models,” *Transportation Research Part B*, 35, 7, 285-305.

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

Vovsha, Peter (1997) “The Cross-Nested Logit Model: Application to Mode Choice in the Tel-Aviv Metropolitan Area,” Transportation Research Board, 76th Annual Meeting, paper #970387.

Wen, Chieh-Hua and Frank S. Koppleman (2001), “The Generalized Nested Logit Model,” *Transportation Research Part B*, 35, 7, 627-641.

Williams, Huw (1977) “On the Formation of Travel Demand Models and the Economic Evaluation Measures of User Benefit,” *Environment and Planning A*, 9, 3, 285-344.

Technical Appendix

Under GEV models, the choice probabilities take the form:

$$\theta_k = \frac{y_k}{G\left(y_{1,\dots,}y_J\right)} \cdot \frac{\partial G\left(y_{1,\dots,}y_J\right)}{\partial y_k} \qquad \text{for } k = 1,\dots,J\,,$$

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,

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

By the chain rule, this leads to

$$\frac{\partial \theta_k}{\partial x_{ja}} = \frac{\partial \theta_k}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_{ja}} \qquad \forall k\,.$$

Thus, the substitution ratio is

$$\begin{aligned}\frac{-\partial\theta_k/\partial x_{ja}}{\partial\theta_j/\partial x_{ja}} &= \frac{-\left(\partial\theta_k/\partial y_j\right)\cdot\left(\partial y_j/\partial x_{ja}\right)}{\left(\partial\theta_j/\partial y_j\right)\cdot\left(\partial y_j/\partial x_{ja}\right)} \\ &= \frac{-\partial\theta_k/\partial y_j}{\partial\theta_j/\partial y_j}\end{aligned}$$

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