05-022

The IPS Property

Thomas J. Steenburgh

The IPS Property

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

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

30 September 2004

The IPS Property

Abstract

This article introduces the invariant proportion of substitution (IPS) property. The IPS property holds when the proportion of demand that is created by substitution away from a competing alternative is the same regardless of which of the enhanced good's attributes is improved. Since this property arises from assumptions about the representative utility rather than the assumptions about the unobserved component of utility, models that break the independence of irrelevant alternatives (IIA) property, such as the nested logit and probit models, do not necessarily also break the IPS property. Some models that do break the IPS property are discussed.

Keywords: Discrete-Choice Models, Econometric Models

1. Introduction

This article identifies a property of standard discrete-choice models that amounts to an implicit and, perhaps, overly restrictive assumption about the substitution patterns of individual consumers. The *invariant proportion of substitution* (IPS) property is said to hold for discrete-choice models in which the proportion of demand that is generated by substitution away from any given competing alternative does not depend on which of the enhanced-good's attributes is improved. While similar to the independence from irrelevant alternatives (IIA) property in the sense that both properties imply a restrictive pattern of substitution among goods, the technique commonly used to eliminate the IIA property, allowing correlation across alternatives in the unobserved component of utility, does not eliminate the IPS property. As a consequence, the nested logit and the probit discrete-choice models do not possess IIA, but do possess IPS.

The following example illustrates the IPS property. Let's say that a 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

We want to understand how improvements to Laptop B's attributes, a weight reduction or a processor speed increase, will affect the researcher's belief about how the consumer will substitute one alternative for another. Discrete-choice models that possess IPS do allow a small improvement in one attribute, say a weight decrease, to increase the expected demand for Laptop B more than an improvement in the other attribute, a processor speed increase, will increase its demand. Yet, regardless of whether Laptop B is made to weigh less or to run faster, the proportion of the increase in its expected demand that is generated by substitution away from Laptop A is the same. The same is true for the proportion of the increase that is generated by substitution away from Laptop C.

For example, let's say a small reduction in its weight produces ten incremental units for Laptop B and three of these units are generated by substitution away from Laptop A. If a small increase in its processor speed produces twenty incremental units for Laptop B, the IPS property would imply that six of these would be generated by substitution away from Laptop A. While this pattern of substitution is possible, it is not something that we would like to impose by assumption. If we expect anything a priori, it would be that a greater proportion of the increase in Laptop B's expected demand will be drawn from Laptop A following a weight reduction than will be drawn following a processor speed increase, not that these proportions are equal.

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 a priori.

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 the logit, nested logit, and probit discrete-choice models all 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 logit, nested logit, and probit models are based on the following assumptions, I will show that all of 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 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

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

where *I* denotes the indicator function.

The logit, nested logit, and probit discrete-choice models arise from different assumptions about the distribution of $\boldsymbol{\varepsilon}$. Under the logit 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. Both the nested logit and the probit models introduce correlation among the elements of $\boldsymbol{\varepsilon}$ to break the IIA restriction. The random vector $\boldsymbol{\varepsilon}$ is distributed generalized extreme value under the nested logit model and is distributed multivariate normal with a full variance-covariance matrix under the probit model. The choice probabilities of the nested logit model have a closed form, but the probabilities of the probit model do not. It is important to note, however, that the choice probabilities under all three 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 regardless of which of the own-good's 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

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

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 the 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. $\theta_j = f\left(v_1, \ldots, v_J\right) \ \forall j$.

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

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

and the discrete-choice model possesses the IPS property.

<u>Proof</u>:

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'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, $\partial\theta_k / \partial v_j$, 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, 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

Example: Nested Logit

The nested logit model best 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 the technique used to eliminate the IIA property, introducing correlation across alternatives in the unobserved component of utility, does not necessarily eliminate 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

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

If the representative utility is a linear function of the attributes, as is most common, then 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

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

Two aspects of the substitution ratio are worth noting. First, of course, since the nested logit model possesses the IPS property, the substitution ratio is the same regardless of which attribute is improved. A skeptical reader can verify this fact by directly taking the derivative of θ_k with respect to x_{ja}. Second, since the probability that a consumer chooses alternative *j* conditional on choosing nest B_n must be greater than the unconditional probability of choosing alternative *j*, a greater unobserved similarity among the goods in nest B_n (higher ρ_n) implies greater substitution ratios for these goods. In other words, following an attribute improvement, a consumer is more likely to substitute away from similar competing alternatives than from dissimilar ones, all else being equal. This is the desired consequence of relaxing the IIA property.

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 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 demand generated by brand switching.

The IPS property implies that the proportion of 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. Bucklin, Gupta and Siddarth (1998) do not consciously 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 the IPS restriction on their beliefs about 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.[1]

4. Discussion

Upon reflection, it is not entirely surprising to find that the standard discrete-choice models impose the IPS restriction, or something like it, on the consumers'

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

substitution patterns. Demand systems derived from neoclassical consumer theory[2] require many more parameters to be estimated, even if the researcher is willing to impose restrictions without empirically testing their validity. Nevo (2000) discusses the dimensionality problem of these models through the following example:

> Consider, for example, a constant-elasticity or log-log, demand system, in which logarithms of quantities are linear functions of logarithms of all prices. Suppose we have 100 differentiated products; then without additional restrictions this implies estimating at least 10,000 parameters (100 demand equations, one for each product, with 100 prices in each). Even if we impose symmetry and adding up restrictions implied by economic theory, the number of parameters will be too large to estimate them. The problem becomes even harder if we want to allow for more general substitution patterns.

Yet, unless there's a free lunch, so to speak, the relatively small number of parameters required by the standard discrete-choice models must come at some expense, whether implicitly or explicitly stated. (The dimensionality seems especially small since discrete-choice models, unlike the others, account for attributes other than price.) In this light, the IPS property can be interpreted as an implicit assumption that is made to attain parsimony at the expense of flexibility.

A researcher can eliminate the IPS property by relaxing the assumptions about the consumer's representative utility. 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. Thus, the universal logit model does not possess the IPS property and allows for more flexible substitution

[2] Such models include the linear expenditure system (Stone, 1954), the Rotterdam model (Theil, 1965; Barten, 1966), the translog model (Christenson et al., 1975), and the almost ideal demand system (Deaton and Muellbauer, 1980).

patterns. Nevertheless, the model does have significant limitations. First, since the number of parameters can become very large if each of the representative utilities depends on the attributes of every alternative, the model may not be computationally feasible. Second, as McFadden (1984) points out, it is difficult to verify whether the universal logit model is consistent with economic theory.

Alternatively, idiosyncratic variation in the consumer's taste parameters can eliminate the IPS restriction. 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 eliminates the IPS property 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 attributes directly through $x_k'\boldsymbol{\tau}$, the IPS property does not hold.

It is worth pointing out that the aforementioned models, which relax assumptions about the consumer's representative utility, eliminate both the IPS and the IIA properties[3], whereas the nested logit and probit models, which relax assumptions about the unobserved component of the consumer's utility, eliminate only the IIA property. In doing so, these models allow very flexible substitution patterns to emerge, albeit at a cost of greater dimensionality. While other techniques may exist, these models provide the interested researcher with immediate solutions.

Acknowledgements

I would like to Andrew Ainslie, Steve Jordan, Elie Ofek, and Luc Wathieu for reading early drafts of this paper and providing useful comments. Kenneth Train deserves special thanks for several discussions that help broaden the ideas discussed in this article. Of course, all remaining errors are mine alone.

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

References

Barten, Anton P. (1966), “Theorie en Empirie van een Volledig Stelsel van Vraagvergelijkingen,” Doctoral Dissertation, Rotterdam: University of Rotterdam.

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

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.

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau (1975), “Transcendental Logarithmic Utility Functions,” *The American Economic Review*, 65, 3, 367-383.

Deaton, Angus and John Muellbauer (1980), “An Almost Ideal Demand System,” *The American Economic Review*, 70, 3, 312-326.

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, 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

Nevo, Aviv (2000), “A Practitioner’s Guide to Estimation of Random-Coefficient Logit Models of Demand,” *Journal of Economics and Management Strategy*, 9, 4, 513-548.

Stone, Richard (1954), “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand,” *The Economic Journal*, 64, 255, 511-527.

Theil, Henri (1965), “The Information Approach to Demand Analysis,” *Econometrica*, 33, 1, 67-87.