Unobserved Di¤erentiation in Discrete Choice Models

Marc Rysman[¤]
Boston University

February 1, 2000

Abstract

Discrete choice models used in statistical applications typically interpret an unobservable term as the interaction of unobservable horizontal di¤erentiation and idiosyncratic consumer preferences. An implicit assumption in most such models is that all choices are equally horizontally di¤erentiated from each other. This assumption is problematic in a number of recent studies that use discrete choice frameworks to evaluate the welfare e¤ects from di¤erent numbers of goods (e.g. Berry and Waldfogel, 1999; Rysman, 2000). Researchers might think that it is possible for product space to "…ll up" and that ignoring this issue might lead to an overestimate of welfare as the number of new products increases.

This paper proposes a solution whereby the researcher estimates the decrease in value that agents receive from higher numbers of products as a result of the decreasing importance of horizontal di¤erentiation. The paper reviews previous results on how a linear random utility model (LRUM) can be mapped into an address (Hotelling) model. The paper shows how realistic assumptions on di¤erentiation in an address setting can be mapped into an LRUM.

LRUM models imply that all choices are strong gross substitutes. In order to preserve that condition in an address model, n choices must be di¤erentiated along at least n ¡ 1 dimensions. This paper proposes that utility drawn from di¤erent dimensions be weighted di¤erently. Mapping this feature into an LRUM requires weighting the utility from each choice based upon the dimension along which it is di¤erentiated from others. As researchers will typically be unwilling to make assumptions about which dimension products di¤er on, the paper discusses integrating over the di¤erent possibilities in a computationally inexpensive way that still allows the researcher to relax the assumption of symmetric di¤erentiation.

[¤] I would like to thank Dan Ackerberg and Phil Haile for helpful discussion.

1 Introduction

In typical discrete choice models used for statistical applications, two products with identical observable characteristics split the market. The reasons that some consumers choose one product and some choose the other is ascribed to an interaction between unobservable horizontal di¤erentiation and idiosyncratic consumer tastes. While not often discussed, the features of the discrete choice model have strong implications for the nature of unobserved horizontal di¤erentiation. For instance, if a third product with identical observable characteristics also enters, the three products split the market. Therefore, the discrete choice model implicitly assumes that each product is equivalently di¤erentiated from each other product. This assumption is problematic in a number of recent studies. For instance, Rysman (2000) and Berry and Waldfogel (1999) both use discrete choice frameworks to consider the welfare implications of increasing the number products in their models. If it is possible for product space to "…ll up", then these papers overestimate the welfare gains from higher numbers of products. In fact, the assumption of equivalent di¤erentiation could be misleading in any situation in which di¤erent agents face di¤erent numbers of choices. For instance, Aricidiacono (1999) studies students choosing a college after they have received acceptances. While students with more choices are surely better o¤ (controlling for quality), standard discrete choice models may overstate the amount if students with many choices have colleges which are not very di¤erentiated.

This paper proposes a way to account for these issues in discrete choice models. The key to doing so is to map the statistical discrete choice model into a discrete choice model that uses consumer address locations - a Hotelling model. Most discrete choice models of interest to econometricians are based around a linear random utility function. There is no intuitive role for horizontal di¤erentiation in such a model. However, the standard theoretical approach to horizontal differentiation is the address model, where horizontal differentiation is explicit and easy to work with. This paper uses techniques from Anderson, DePalma and Thisse (1992) (ADT) to link the linear random utility model (LRUM) to the address model. That is, this paper shows the conditions required such that both models imply the same market shares and elasticities. Standard features in the LRUM place important restrictions on the address model in order to maintain the link. Similarly, this paper develops an address model that handles horizontal differentiation in a more realistic way. The paper shows how to take the new features of the address model to the LRUM, allowing researchers to estimate a discrete choice model without the assumption of equal horizontal differentiation.

An important feature in random utility models is that all products are strong gross substitutes, i.e. the shares of all products are sensitive to the mean utility of each other product. In order to preserve that condition in an address model, n products must be differentiated along at least $n - 1$ dimensions. The standard LRUM suggests that, in the address model, each dimension is weighted equally in the consumer's utility function. This paper proposes that different dimensions should be allowed to have different weights. In other words, features in the LRUM preclude us from lifting the equal differentiation assumption by restricting the number of dimensions along which products may differentiate. Instead, we should escape the equal differentation assumption by having products differentiate into dimensions which consumers care less and less about.

Mapping this feature into an LRUM requires weighting the utility from each choice based upon the dimension along which it is differentiated from others. Knowing the actual (unobserved) dimension in the address model is important because it determines how much utility in the LRUM should be adjusted. As researchers will typically be unwilling to make assumptions about this kind of information, which is unobserved by assumption, the paper discusses integrating over the different possibilities in a computationally inexpensive way. The approach still allows the researcher to estimate the proper adjustment to consumer utility due to lessened horizontal differentiation from higher number of products.[1]

The layout of the paper is as follows: Section 2 introduces the linear random utility model. Section 3 introduces the address model and provides the algorithm for linking the two models. Section 4 proposes a solution which allows the researcher to estimate a discrete choice model in which the assumption of equal differenatiation is lifted in a way that can be easily understood in terms of the underlying theoretical structure. Section 5 (to be complete) provides an example of the methodology and Section 6 (to be completed) concludes.

2 The Linear Random Utility Model (LRUM)

ADT characterize a linear random utility model (LRUM) as follows: A unit mass of agents choose 1 of n choices. Each choice is defined by quality level u_i. Each agent receives indirect utility level V_i from a given choice defined by $V_i = u_i + \epsilon_i$, where $\epsilon_1 \ldots \epsilon_n$ is a random variable drawn from the probability density $f(x)$. So the probability of an agent choosing i is:[2]

$$\hat{s}_i = \Pr\left(V_i = \max_{j=1 \ldots n} V_j\right) = \Pr\left(u_i + \epsilon_i = \max_{j=1 \ldots n} (u_i + \epsilon_i)\right) :$$

ADT show that an LRUM necessarily implies that choices are weak gross substitutes, i.e. $d\hat{s}_i / du_j \leq 0 \; i \neq j; \forall u_i; u_j$. But ADT point out that for most specific applications (mulitnomial logit, probit, nested logit), choices are actually strong gross substitutes, i.e. $ds_i / du_j < 0 \; i \neq j; \forall u_i; u_j$. I take this condition as an implication of the LRUM and develop the address model in order to satisfy the condition.

[1] The paper is written in terms of consumers choosing products. But obviously, the results are relevant to any situation in which an agent makes choices.

[2] Note that Cardell (1997) shows that the Nested Logit model can be characterized in this way for a particular distribution of ϵ.

3 The Address Model

This section sets up an address model following ADT and shows what assumptions can link it to the LRUM (in the sense of matching market shares and elasticicities with respect to u_i). The next section suggests intutitive changes to the address model that solve the problem laid out above, and imposes those changes onto the LRUM via the technology laid out in this section.

Let there be m characteristics. Each agent is characterized by a vector $\bar{z}$ of length m that describes the agent's ideal choice. Let there be n distinct products described by a quality level u_i and a location $\bar{z}_i \in \Re^m; i = 1; ...; n$. A consumer located at $\bar{z}$ who consumes product i receives utility level:

$$V_i(\bar{z}) = u_i - \tau \sum_{k=1}^{m} (z^k - z_i^k)^2 \quad i = 1 \ldots n.$$

The parameter $\tau > 0$ measures consumers' sensitivity to distance. Consumers choose the option which confers the most utility. The market space of product i is defined to be:

$$M_i = \{\bar{z} \in \Re^m; V_i(\bar{z}) \geq V_j(\bar{z}); \; j = 1 \ldots n\}.$$

Consumers are distributed in $\Re^m$ according to the probability density $g(\bar{z})$ where $g : \Re^m \to \Re^1$. The market share s_i for choice i is:

$$s_i = \int_{M_i} g(\bar{z}) d\bar{z}; \quad i = 1; \ldots ; n.$$

Note that $\sum_{i=1}^{n} s_i = 1$. The paper refers to this model as the address model. Now consider the conditions on the address model that would allow it to match the LRUM in terms of market shares and elasticicities with respect to u_i. In the LRUM, all choices are strong gross substitutes. ADT prove the following theorem for the address model:

Theorem 1 For the n variants to be strong gross substitutes, the set $\{\bar{z}_1 \ldots \bar{z}_n\}$ must contain $n - 1$ linearly independent points.

Proof. See ADT, Appendix 4.7.1. ■

Therefore, we must require that $m \geq n - 1$. It is easy to see why products may not be gross substitutes case in which $m = 1$ and $n = 3$. Consider 3 choices with equal quality levels and locations such that $z_1 < z_2 < z_3$. Then, $\partial s_1 / \partial u_3 = 0$.

In order to guarantee that the set $\{\mathbf{z}_1 \ldots \mathbf{z}_n\}$ forms a basis of $\Re^m$, I make the following strong assumptions about the set:

Assumption 2

$$z_i^j = \begin{cases} b & \text{if } i = j, \quad i = 1 \ldots n - 1; j = 1; \ldots; m; \\ -b & \text{otherwise} \end{cases}$$

$$z_n^j = -b \quad j = 1; \ldots; m:$$

The parameter b measures the proximity of choices. Because choices are differentiated across only their first $n - 1$ characteristics, we can restrict attention to just those characteristics, and ignore n through m. From now on, I assume $m = n - 1$: Now we can characterize market shares and establish the link between the LRUM and the address model.

The set of consumers who are indifferent between i and n is a hyperplane orthogonal to the jth axis at:

$$\hat{z}^j \equiv \frac{u_n - u_j}{4b\tau} \tag{1}$$

The set of consumers indifferent between i and j is:

$$\{\mathbf{z} \in \Re^m \mid z^i - z^j = \hat{z}^j - \hat{z}^i\} \tag{2}$$

Choice n's market share is $M_n = \{\mathbf{z} \in \Re^m; z^j \leq \hat{z}^j;\ j = 1 \ldots m\}$. Similarly, choice i's market share is $M_i = \{\mathbf{z} \in \Re^m; z^i \geq \hat{z}^i \text{ and } z^j \leq \hat{z}^j + (z^i - \hat{z}^i);\ j = 1 \ldots [i] \ldots m\}$, where the notation $[\cdot]$ indicates that the enclosed element is skipped. Now, the definition of market share can be rewritten as:

$$s_i = \int_{-\infty}^{\hat{z}^1 + (z^i - \hat{z}^i)} \ldots \int_{\hat{z}^i}^{\infty} \ldots \int_{-\infty}^{\hat{z}^{n-1} + (z^i - \hat{z}^i)} g(z^1 \ldots z^{n-1}) dz^1 \ldots dz^{n-1}:$$

hi ap ro ch ak s i ea y t de er in th de iv ti es f s .

$$\frac{\partial \; i}{\partial \; j} = \frac{1}{4\tau}\int_{1}^{z^1} (z - \mathbf{e}^i) \; : \; : \int_{z^i}^{1} \int_{i}^{\mathbf{e}^j + z^i \; \mathbf{e})} \Big] \; : \; : \int_{i}^{z^n \; 1 + z^i \; \mathbf{e})}$$

$$g(\; ^1 : \; : z \; ^{-1} \; dz \; : \; : \big[d \; ^j \big] \; : \; : z^n \; ^1$$

a d:

$$\frac{^{n-} \; s_i}{u_1 \; : \; : \partial u \;] : \; : \partial \; _{n-}} = \left(\frac{1}{4\tau}\right)^{-2} \; ^1 \; _{\mathbf{e}^i} \; g \; z^1 \; : \; : \; ^{n-} \;)d \; ^i$$

$$\frac{\partial \; ^{-1} \; _i}{\partial \; _1 : \; :[\; u_i \; : \; : \partial u} = \left(\frac{1}{4\tau}\right)^{-1} (z \; : \; : z^n \; ^1) \qquad 3)$$

nt it ve y, he ar et pa es nt rs ct t a in le oi t. er ur in al bu on of he ar et ou da ie gi es he en it of on um rs t t at oi t. qu ti n 3 s t e k y t li ki g t e a dr ss od l t th LR M. e s mp y c mp te he ef -h nd id fo a g ve LR M a d p ug n t ge th di tr bu io of (· th t a lo s t e a dr ss od l t ma ch he LR M. To compute the left-hand side for the LRUM, define ϕ_i to be:

$$\phi_i(u_1; \ldots [u_i] \ldots ; u_n) = \frac{\partial^{n-1} \hat{s}_i(u_1; \ldots ; u_n)}{\partial u_1 \ldots [\partial u_i] \ldots \partial u_n}$$

ADT show that $\hat{s}_i(u_1; \ldots ; u_n) = \hat{s}_i(u_1 - u_i; \ldots ; u_n - u_i)$. So, we can rewrite $\phi(:)$ as:

$$\phi_i(u_1 - u_i; \ldots [u_i - u_i] \ldots ; u_n - u_i) = \frac{\partial^{n-1} \hat{s}_i(u_1 - u_i; \ldots ; u_{n-1} - u_i)}{\partial u_1 \ldots [\partial u_i] \ldots \partial u_n} \qquad (4)$$

From Equations 1 and 2, we can substitute $u_n - u_i = -4b\tau \mathbf{z}^j$ and $u_j - u_i = 4b\tau(\mathbf{z}^j - \mathbf{z}^i)$. Actually, because any agent location $\mathbf{z}$ can be reached by some combination of qualities $u_1; \ldots ; u_n$, we can use $\mathbf{z}$ instead of $\hat{\mathbf{z}}$. By equating Equation 4 and Equation 3, we can determine the distribution $g(\mathbf{z})$ that allows the address model to match the LRUM. That is, we can compute an equivalent address model for an LRUM by assuming that agents in the address model are distributed according to:

$$g(\mathbf{z}) = \left(\frac{-1}{4b\tau}\right)^{n-1} \phi_i(4b\tau(z^1 - z^i); \ldots [4b\tau(z^i - z^i)] \ldots ; 4b\tau(z^n - z^i)) \quad i = 1 \ldots n:$$

For example, write the logit model as:

$$s_i = \frac{1}{1 + \Sigma_{j=1; j \neq i} \exp[(u_j - u_i)/\mu]}$$

where $\mu > 0$ is the standard deviation of ε_i, usually normalized to 1. Then:

$$\begin{aligned} &\phi_i(u_1 - u_n; \ldots [u_i - u_n] \ldots; u_{n-1} - u_n) \\ &= \left(\frac{-1}{\mu}\right)^{n-1} (n-1)! \frac{\Pi_{j=1; j \neq i} \exp[(u_j - u_i)/\mu]}{(1 + \Sigma_{j=1; j \neq i} \exp[(u_j - u_i)/\mu])^n} \end{aligned}$$

Substituting for $u_j - u_i$ and plugging in to Equation 3 gives:

$$g(\bar{z}) = \left(\frac{4b\tau}{\mu}\right)^{n-1} (n-1)! \frac{\Pi_{j=1; j \neq i} \exp[4b\tau(z_j - z_i)/\mu]}{(1 + \Sigma_{j=1; j \neq i} \exp[4b\tau(z_j - z_i)/\mu])^n} \tag{5}$$

where $z_n = 0$. As ADT show, the right hand side is equal for all i, which is obviously necessary in order to have a coherent definition of $g(\bar{z})$. To fix intuition, consider the $n = 3$; $m = 2$ case. Figure 1(a) draws a contour map of $g(\bar{z})$ for $b = 1$; $\tau = 1$ and $\mu = 2$, assuming products are distributed according to Assumption 2. This distribution of consumers implies elasticities in the address model which would match those of the Logit model. Consider a Nested Logit model, with choices 1 and 2 in a separate nest than 3. Figures 1(b) and 1(c) show similar contour maps for successively higher degrees of within-nest correlation. ADT's system for defining the distribution of consumers mimics a Nested Logit model by putting a greater mass of consumers to the northeast, where they will prefer 1 and 2 to 3.

The graphs make it clear why Assumption 2 is not restrictive. For a different placement of choices in z space, we will just compute a different distribution of agents. And the parameters b, τ, and μ have very similar effects. Specifying choices to be farther apart is equivalent to giving agents greater disutility of travel, which is equivalent to having a smaller distribution of ε_i in the LRUM.

4 The Solution

Restrictions on the nature of horizontal differentiation in the address model are likely to be more intuitive and easier to grasp than attempting to place appropriate restrictions directly onto the LRUM. The tight link between the address model and the LRUM means that it is straightforward to map restrictions from one to the other. The link also makes it clear what cannot be done. Unequal differentiation of products cannot be modelled by restricting the number of dimensions that products expand into. Restricting the number of dimensions means restricting $m < n - 1$. In that case, the address model may imply that choices are not strong gross substitutes and the link to the LRUM will be lost.

I propose modelling the decreasing value of horizontal differentiation by allowing different dimensions to have different travel costs. Instead of restricting the dimensions by which prodocts can differentiate, have products differentiate into dimensions which agents care less and less about. This change can be implemented by allowing τ to depend on the dimension with which it is interacting. Let $\tau(k)$ measure the disutility to travel in dimension k. Equation 3 can be rewritten as:

$$V_i(\bar{z}) = u_i - \sum_{k=1} \tau(k)(z^k - z_i^k)^2 \quad i = 1 \ldots n.$$

Under this set-up, Equation 3 becomes:

$$\frac{\partial^{n-1} s_i}{\partial u_1 \ldots [\partial u_i] \ldots \partial u_n} = \left(\frac{-1}{4b}\right)^{n-1} \frac{1}{\prod_{k=1}^{n-1} \tau(k)} g(z^1 \ldots z^{n-1}) \tag{6}$$

Now we must map the setup captured in Equation 6 into the LRUM and estimate $\tau(k)$[3]. Consider adjusting the standard deviation of ε_i. In Equation 5, a high

[3] On the other hand, one could use the original LRUM model. In this case, the algorithm for finding $g(\bar{z})$ would adjust $g(\bar{z})$ for the fact that, despite the heterogeneous travel costs, the features of the LRUM were the same. This story amounts to placing the following assumption on the LRUM model: While products may differ in the importance of their differentiation, consumers are distributed in exactly offsetting ways. This assumption is no more (or even

disutility to travel is equivalent to a low variance of ε_i. So let $\mu = \mu\lambda(j)$. Note that μ is not normally identified so $\lambda(j)$ can be identified only up to a scalar. Therefore, there is no loss of generality in assuming $\lambda(i) = 1$ for any given choice i and estimating $\lambda(j)$ $j \neq i$ as product j's variance relative to i. In this case, Equation ?? is unchanged. But the Logit model becomes:

$$\hat{s}_i = \frac{1}{1 + \Sigma_{j=1;j\neq i} \exp[(u_j/\mu\lambda(j) - u_i/\mu\lambda(i)]}$$

Now it is straightforward to find $\partial^{n-1} s_i/\partial u_1 \ldots [\partial u_i] \ldots \partial u_n$ and create a coherent definition of $g(\tilde{z})$.

This result shows that it is possible to relax the assumption of equal differentiation in a meaningful and structural way. The researcher defines some function $\lambda(j)$ where j is the index of the dimension into which product j differentiated into. Then the researcher adjusts the vertical utility measure in order to compensate for the reduced (or increased) importance of that dimension.

A major potential concern is that the researcher must know which dimension each product has differentiated into. If a choice is assigned position j, then the utility of that choice is adjusted by $\lambda(j)$. Most researchers would be unwilling to make assumptions about this "dimension indices", which are unobservable by definition. A solution is to integrate over all possibilities. If an agent faces n choices, 1 choice is selected as the "base" choice and there are $(n-2)!$ possible sequences of choices. Let $I : [1; (n-2)!] \times [1; n-1] \to [1; n-1]$ be such that $I(k; i)$ give the location of choice i in sequence k: Then $\hat{s}_i$ can be written as:

$$\hat{s}_i = \sum_{k=1}^{(n-2)!} \frac{\exp[u_i/\lambda(I(k;i))]}{1 + \sum_{j=1}^{n-1} \exp[u_j/\lambda(I(k;j))]} \frac{1}{(n-2)!}$$

where $\lambda(n)$ has been normalized to 1 and u_n has been normalized to 0. The final fraction weights each possbility equally.

less) palatable than the original assumption that all products are equally differentiated from each other in terms of both distance and importance.

The integration approach essentially applies the same adjustment to each choice, but places a bigger adjustment when there are more choices. This method gives a clear idea of how ¿̂ (¢) would be indenti…ed. Consider a Nested Logit model where the consumers may choose between an outside option and and 3 or 4 products (all in the nest). The researcher sees that in all markets, the products split market share equally and so assigns the same vertical utitliy index to each. But the share to the outside option is about the same across markets - at least the share to the outside option in the markets with 4 products is not nearly as decreased as one would expect from estimating on the markets with only 3 products. The standard Nested Logit model could not capture these features but allowing markets with 4 goods to have less unobserved horizontal di¤erentiation gives the researcher an extra degree of freedom with which to match these stylizied facts.

5 Example

To be completed with my data on Yellow Pages, and possibly other people's data as can be arranged.

6 Conclusion

To be completed.

References

[1] Anderson, S., A. de Palma and J-F Thisse, (1992), Discrete Choice Theory of Product Di¤erentiation, Cambridge: The MIT Press.

[2] Arcidiacono, P. (1999), "Option Values, College Quality, and Earnings: A Dynamic Model College and Major Choice," mimeo, Duke University.

[3] Berry, S.T. and J. Waldfogel. (1999) "Free Entry and Social Ine¢ciency in Radio Broadcsting," RAND Journal of Economics: 30(3), pp 397-420.

[4] Rysman, M. (2000), "Competition Between Networks: A Study of the Market for Yellow Pages," mimeo, Boston University.

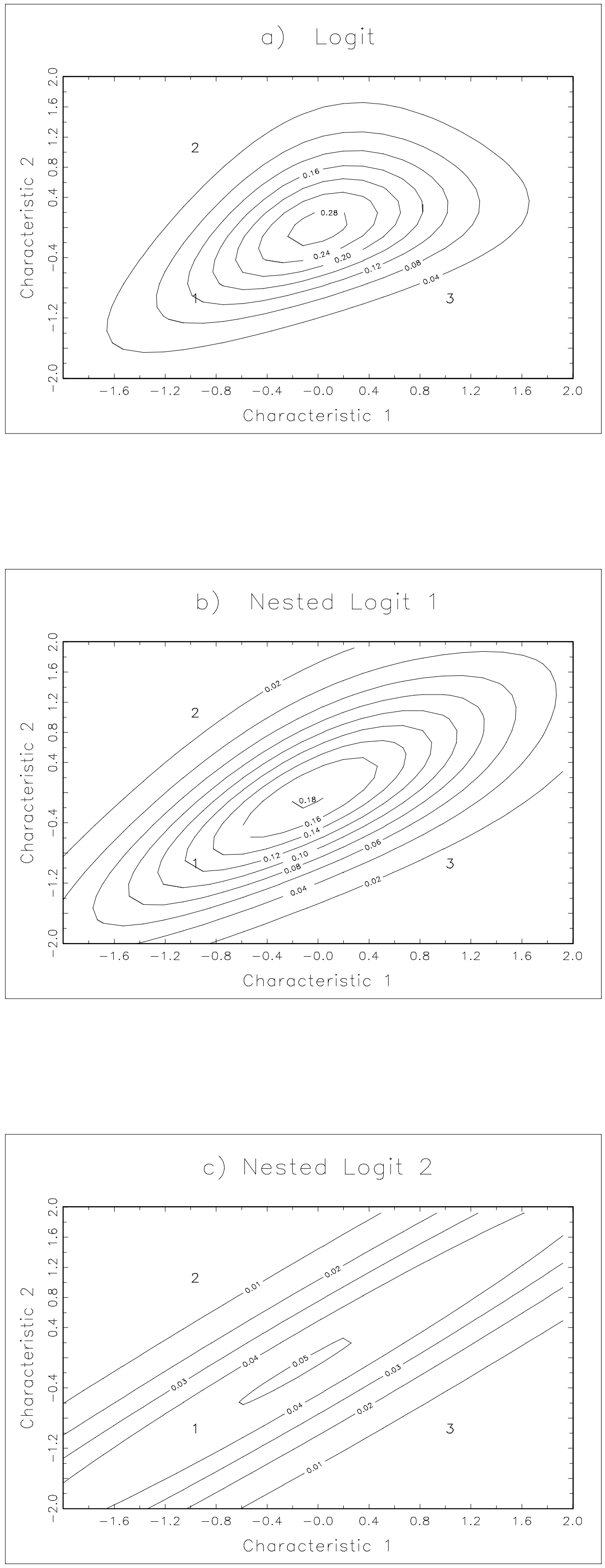

Figure 1: Consumer Distributions in the Address Model