Nonparametric Discrete Choice Models with Unobserved Heterogeneity

Richard A. Briesch
Cox School of Business
Southern Methodist University

Pradeep K. Chintagunta
Graduate School of Business
University of Chicago

Rosa L. Matzkin
Department of Economics
Northwestern University

September 2008
August 2007

Corresponding author. Richard Briesch; Cox School of Business; Southern Methodist University; PO Box 750333; Dallas, Tx 75275-0333. Email: rbriesch@mail.cox.smu.edu, phone: 214-768-3180, fax: 214-768-4099.

Abstract

In this research, we provide a new method to estimate discrete choice models with unobserved heterogeneity that can be used with either cross-sectional or panel data. The method imposes nonparametric assumptions on the systematic subutility functions and on the distributions of the unobservable random vectors and the heterogeneity parameter. The estimators are computationally feasible and strongly consistent. We provide an empirical application of the estimator to a model of store format choice. The key insights from the empirical application are: 1) consumer response to cost and distance contains interactions and non-linear effects which implies that a model without these effects tends to bias the estimated elasticities and heterogeneity distribution and 2) the increase in likelihood for adding nonlinearities is similar to the increase in likelihood for adding heterogeneity, and this increase persists as heterogeneity is included in the model.

JEL Classification Code: C14 ; C23 ; C33 ; C35.

Keywords: Random Effects; Heterogeneity; Discrete Choice; Nonparametric

1. INTRODUCTION

Since the early work of McFadden (1974) on the development of the Conditional Logit Model for the econometric analysis of choices among a finite number of alternatives, a large number of extensions of the model have been developed. These extensions have spawned streams of literature of their own. One such stream has focused on relaxing the strict parametric structure imposed in the original model. Another stream has concentrated on relaxing the parameter homogeneity assumption across individuals. This paper contributes to both these areas of research. We introduce methods to estimate discrete choice models where all functions and distributions are nonparametric, individuals are allowed to be heterogeneous in their preferences over observable attributes, and the distribution of these preferences is also nonparametric.

As is well known in discrete choice models, each individual possesses a utility for each available alternative, and chooses the one that provides the highest utility. The utility of each alternative is the sum of a subutility of observed attributes - the systematic subutility – and an unobservable random term – the random subutility. Manski (1975) developed an econometric model of discrete choice that did not require specification of a parametric structure for the distribution of the unobservable random subutilities. This semiparametric, distribution-free method was followed by other semiparametric distribution-free methods, developed by Cosslett (1983), Manski (1985), Han (1987), Ichimura (1989), Powell, Stock and Stoker (1989), Horowitz (1992), Klein and Spady (1993), and Moon (2004), among others. More recently, Geweke and Keane (1997) and Hirano (2002), have applied mixing techniques that allow nonparametric estimation of the error term in Bayesian models. Similarly, Klein and Sherman (2002) propose a method that allows non-parametric estimation of the density as well as the parameters for ordered response models. These methods are termed semiparametric because they require a parametric structure for the systematic subutility of the observable characteristics. A second stream of literature has focused on relaxing the parametric assumption about the systematic subutility.

Matzkin (1991)'s semiparametric method accomplished this while maintaining a parametric structure for the distribution of the unobservable random subutilities. Matzkin (1992, 1993) also proposed fully nonparametric methods where neither the systematic subutility nor the distribution of the unobservable random subutility are required to posses parametric structures.

Finally, a third stream of literature has focused on incorporating consumer heterogeneity into choice models. Wansbeek, et al (2001) noted the importance of including heterogeneity in choice models to avoid finding weak relationships between explanatory variables and choice. However, they also note the difficulty of incorporating heterogeneity into nonparametric and semiparametric models. Further, Allenby and Rossi (1999) noted the importance of allowing heterogeneity in choice models to extend into the slope coefficients. Specifications that have allowed for heterogeneous systematic subutilities include those of Heckman and Willis (1977), Albright, Lerman and Manski (1977), McFadden (1978), and Hausman and Wise (1978). These papers use a particular parametric specification, i.e., a specific continuous distribution, to account for the distribution of systematic subutilities across consumers. Heckman and Singer (1984) propose estimating the parameters of the model without imposing a specific continuous distribution for this heterogeneity distribution. Ichimura and Thompson (1993) have developed an econometric model of discrete choice where the coefficients of the linear subutility have a distribution of unknown form, which can be estimated.

Recent empirical work has relaxed assumptions on the heterogeneity distributions. Lancaster (1997) allows for non-parametric identification of the distribution of the heterogeneity in Bayesian models. Taber (2000) and Park et al (2007) apply semiparametric techniques to dynamic models of choice. Briesch, Chintagunta and Matzkin (2002) allow consumer heterogeneity in the parametric part of the choice model while restricting the non-parametric function to be homogeneous. Dahl (2002) applies non-parametric techniques to transition probabilities and dynamic models. Pinkse, et al (2002) allow for heterogeneity in semiparametric models of aggregate-level choice.

The method that we develop here combines the fully nonparametric methods for estimating discrete choice models (Matzkin 1992, 1993) with a method that allows us to estimate the distribution of unobserved heterogeneity nonparametrically as well. The unobserved heterogeneity variable is included in the systematic subutility in a nonadditive way (Matzkin 1999a, 2003). We provide conditions under which the systematic subutility, the distribution of the nonadditive unobserved heterogeneity variable, and the distribution of the additive unobserved random subutility can all be nonparametrically identified and consistently estimated from individual choice data. The method can be used with either cross sectional or panel data. These results update Briesch, Chintagunta, and Matzkin (2002).

We apply the proposed methodology to study the drivers of grocery store-format choice for a panel of households. There are two main types of formats that supermarkets classify themselves into – everyday low price (EDLP) stores or high-low price (Hi-Lo) stores. The former offer fewer promotions of lower "depth" (i.e., magnitude of discounts) than the latter. The main tradeoff facing consumers is that EDLP stores are typically located farther away (longer driving distances) than Hi-Lo stores although their prices, on average, are lower than those at Hi-Lo stores leading to a lower total cost of shopping "basket" for the consumer. Since the value of time (associated with the driving distance) is heterogeneous across households, and little is known about the shape of the utility for driving distance and expenditure, we think that the proposed method is ideally suited to understanding the nature of the tradeoff between distance and expenditure facing the consumer. To decrease the well-known dimensionality problems associated with relaxing parametric structures, we use a semiparametric version of our model. In particular, only the subutility of distance to the store and cost of refilling inventory at the store are nonparametric. We allow this subutility to be heterogeneous across consumers, and provide an estimator for both, the subutilities of the different types, the distribution of types, and the additional parameters of the model. Further, we assume that the unobserved component of utility in this application is distributed according to a type-I extreme value distribution.

In the next section we describe the model. Section 3 states conditions under which the model is identified. In Section 4 we present strongly consistent estimators for the functions and distributions in the model. Section 5 provides computational details. Section 6 presents the empirical application. Section 7 concludes.

2. THE MODEL

As is usual in discrete choice models, we assume that a typical consumer must choose one of a finite number, *J*, of alternatives, and he/she chooses the one that maximizes the value of a utility function, which depends on the characteristics of the alternatives and the consumer. Each alternative *j* is characterized by a vector, z_j, of the observable attributes of the alternatives. We will assume that $z_j \equiv (x_j, r_j)$, where $r_j \in R$ and $x_j \in R^K$ $(K \geq 1)$. Each consumer is characterized by a vector, $s \in R^L$, of observable socioeconomic characteristics for the consumer. The utility of a consumer with observable socioeconomic characteristics s, for an alternative, *j*, is given by $V^*(j, s, z_j, \omega) + \varepsilon_j$, where ε_j and ω denote the values of unobservable random variables. For any given value of ω, and any j, $V^*(j,\bullet,\omega)$ is a real valued, but otherwise unknown, function. The dependence of V^* on ω allows this systematic subutility to be different for different consumers, even if the observable exogenous variables are the same for these consumers. We denote the distribution of ω by G^* and we denote the distribution of the random vector $(\varepsilon_1,\ldots,\varepsilon_J)$ by F^*.

The probability that a consumer with socioeconomic characteristics *s* will choose alternative *j* when the vector of observable attributes of the alternatives is $z \equiv (z_1,...,z_J) \equiv (x_1, r_1,..., x_J, r_J)$ is denoted by $p(j|s,z;V^*,F^*,G^*)$. Hence,

$$p(j \mid s, z; V, F, G) = \int \Pr(j \mid s, z; \omega, V, F) dG(\omega)$$

where $\Pr(j \mid s, z; \omega, V, F)$ denotes the probability that a consumer with systematic subutility $V(\cdot,\omega)$ will choose alternative *j*, when the distribution of ε is F. By the utility maximization hypothesis,

$\Pr(j \mid s, z; \omega, V, F)$

$$= \Pr\{V(j, s, x_j, r_j, \omega) + \varepsilon_j > V(k, s, x_k, r_k, \omega) + \varepsilon_k, \forall k \neq j\}$$

$$= \Pr\{\varepsilon_k - \varepsilon_j < V(j, s, x_j, r_j, \omega) - V(k, s, x_k, r_k, \omega), \forall k \neq j\}$$

which depends on the distribution F. In particular, if we let F_1^* denote the distribution of the vector $(\varepsilon_2 - \varepsilon_1, ..., \varepsilon_J - \varepsilon_1)$, then $\Pr(1 \mid s, z; \omega, V, F^*)$

$$= F_1^*(V(1 \mid s, x_1, r_1, \omega) - V(2 \mid s, x_2, r_2, \omega), ..., V(1 \mid s, x_1, r_1, \omega) - V(J \mid s, x_J, r_J, \omega),)$$

and the probability that the consumer will choose alternative one is then $p(1 \mid s, z; V, F^*, G)$

$$= \int \Pr(1 \mid s, z; V, F^*, G) dG(\omega)$$

$$= \int F_1^*(V(1 \mid s, x_1, r_1, \omega) - V(2 \mid s, x_2, r_2, \omega), ..., V(1 \mid s, x_1, r_1, \omega) - V(J \mid s, x_J, r_J, \omega)) dG(\omega)$$

For any *j*, $\Pr(j \mid s, z; \omega, V, F^*)$ can be obtained in an analogous way, letting F_j^* denote the distribution of $(\varepsilon_1 - \varepsilon_j, \ldots \varepsilon_J - \varepsilon_j)$. A particular case of the polychotomous choice model is the Binary Threshold Crossing Model, which has been used in a wide range of applications. This model can be obtained from the Polychotomous Choice Model by letting J=2, $\eta = (\varepsilon_2 - \varepsilon_1)$ and $\forall (s, x_2, r_2, \omega), V(2, s, x_2, r_2, \omega) \equiv 0$ $\forall (s, x_2, r_2, \omega)\ V(2, s, x_2, r_2, \omega) \equiv 0$. In other words, the model can be described by: $y^* = V(x, r, \omega) - \eta$, with

$$y = \begin{cases} 1 : y^* \geq 0 \\ 0 : otherwise \end{cases}$$

where y^* is unobservable. In this model, F_1^* denotes the distribution of η. Hence, for all x, r, ω, $\Pr(1 \mid x, r; \omega, V^*, F^*) = F_1^*(V^*(x, r, \omega))$ and for all x, r the probability that the consumer will choose alternative one is

$$p(1 \mid s, z; V^*, F^*, G^*) = \int \Pr(1 \mid s, z; \omega, V^*, F^*)\ dG^*(\omega) = \int F_1^*(V^*(x, r, \omega))\ dG^*(\omega)$$

3. NONPARAMETRIC IDENTIFICATION

Our objective is to develop estimators for the function V^* and the distributions F^* and G^*, without requiring that these functions and distributions belong to parametric families. It follows from the definition of the model that we can only hope to identify the distributions of the vectors $\eta_j \equiv (\varepsilon_1 - \varepsilon_j, ..., \varepsilon_J - \varepsilon_j)$ for j=1,…,J. Let F_1^* denote the distribution of η_1. Since from F_1^* we can obtain the distribution of η_j (j=2,…,J), we will deal only with the identification of F_1^*. We let $F^* = F_1^*$.

DEFINITION: *The function V^* and the distributions F^* and G^* are identified in a set* $(W\times\Gamma_F\times\Gamma_G)$ *such that $(V^*,F^*,G^*) \in (W\times\Gamma_F\times\Gamma_G)$, if* $\forall (V,F,G) \in (W \times \Gamma_F \times \Gamma_G)$ then

$$\int \Pr(j \mid s,z;V(\cdot;\omega),F)\,dG(\omega) = \int \Pr(j \mid s,z;V(\cdot;\omega),F^*)\,dG^*(\omega) \quad \text{for j=1…J, a.s.}$$

implies that $V=V^*$, $F=F^*$ and $G=G^*$. That is, (V^*,F^*,G^*) is identified in a set $(W\times\Gamma_F\times\Gamma_G)$ to which (V^*,F^*,G^*) belongs, if any triple, (V,F,G) that belongs to $(W\times\Gamma_F\times\Gamma_G)$ and is different from (V^*,F^*,G^*) generates, for at least one alternative j, and a set of (s,z) that possesses positive probability, choice probabilities p(j|s,z;V,F,G) that are different from $p(j|s,z;V^*,F^*,G)$. We next present a set of conditions that, when satisfied, guarantee that (V^*,F^*,G^*) is identified in $(W\times\Gamma_F\times\Gamma_G)$.

ASSUMPTION 1: *The support of $(s,x_1,r_1,\ldots,x_J,r_J,\omega)$ is a set* $\left(S \times \prod_{j=1}^J (X_j \times R_+) \times Y\right)$, *where S and X_j (j=1,…,J) are subsets of Euclidean spaces and* Y *is a subset of* R.

ASSUMPTION 2: *The random vectors $(\varepsilon_1,\ldots \varepsilon_J)$, $(s,z_1,\ldots,x_J)$ and* ω *are mutually independent.*

ASSUMPTION 3: *For all $V \in W$ and* j, *there exists a real valued, continuous function* $v(j,\cdot,\cdot,\omega)$: $\text{int}(S \times X_j) \to R$ *such that* $\forall (s,x_j,r_j,\omega) \in (S \times X_j \times R_+ \times Y)$, $V(j,s,x_j,r_j,\omega) = v(j,s,x_j,\omega) - r_j$.

ASSUMPTION 4: $\exists (\bar{s},\bar{x}_1,...,\bar{x}_J) \in \left(S \times \prod_{j=1}^J X_j\right)$ *and* $(\alpha_1,...\alpha_J) \in R^J$ *such that* $\forall j \forall \omega \forall V \in W$,

$v(j,\bar{s},\bar{x}_j,\omega) = \alpha_j$.

ASSUMPTION 5: $\exists \tilde{j}, \beta_{\tilde{j}} \in R$, *and* $\tilde{x}_{\tilde{j}} \in X_{\tilde{j}}$ *such that* $\forall s \in S, \forall \omega \in Y, \forall V \in W$, $v(\tilde{j}, s, \tilde{x}_{\tilde{j}}, \omega) = \beta_{\tilde{j}}$

ASSUMPTION 6: $\exists \quad j^* \neq \tilde{j}$ *and* $(\hat{s}, \hat{x}_{j^*}) \in (S \times X_{j^*})$ *such that* $\forall V \in W \forall \omega \;\; v(j^*, \hat{s}, \hat{x}_{j^*}, \omega) = \omega$ **.**

ASSUMPTION 6': $\exists \quad j^* \neq \tilde{j}$ **,** $(\hat{s}, \hat{x}_{j^*}, \hat{\omega}) \in (S \times X_{j^*} \times Y)$ *and* $\gamma \in R$ *such that* $\forall V \in W$

$v(j^*, \hat{s}, \hat{x}_{j^*}, \hat{\omega}) = \gamma$, and $\forall V \in W \;\; \forall \lambda \in R \;\; v(j^*, \lambda\hat{s}, \lambda\hat{x}_{j^*}, \lambda\hat{\omega}) = \lambda\gamma$.

ASSUMPTION 6": $\exists \;\; j^* \neq \tilde{j}$ **,** $(\hat{s}, \hat{x}_{j^*}) \in (S \times X_{j^*})$, $\gamma \in R$, and there exists a real valued, continuous function $m(s, x_1, x_2\omega)$ *such that* $\forall V \in W, \forall \omega \in Y$, $v(j^*, s, x, \omega) = m(s, x_1, x_2\omega)$, and $m(\hat{s}, \hat{x}_1, \hat{x}_2\omega) = \gamma$.

ASSUMPTION 7: $\forall (s, x_{j^*}) \in (S \times X_{j^*})$ (i) *either it is known that* $v(j^*, s, x_{j^*}, \cdot)$ *is strictly increasing in* *ω, for all ω∈Y and V∈ W*, or (ii) *it is known that* $v(j^*, s, x_{j^*}, \cdot)$ *is strictly decreasing in ω, for all ω∈Y and V∈ W.*

ASSUMPTION 8: $\forall \; k \neq j^*, \tilde{j}$ *either* $v(k, s, x_k, \omega)$ *is strictly increasing in ω, for all* (s, x_k), *or* $v(k, s, x_k, \omega)$ *is strictly decreasing in ω, for all* (s, x_k).

ASSUMPTION 9: $\exists j \neq \tilde{j}$ *such that* $\forall (s, x_j, x_{\tilde{j}}) \in (S \times X_j \times X_{\tilde{j}})$ *either* (i) *it is known that* $v(j, s, x_j, \cdot) - v(\tilde{j}, s, x_{\tilde{j}}, \cdot)$ *is strictly increasing in ω, for all ω∈Y and V∈ W*, or (ii) *it is known that* $v(j, s, x_j, \cdot) - v(\tilde{j}, s, x_{\tilde{j}}, \cdot)$ *is strictly decreasing in ω, for all ω∈Y and V∈ W.*

ASSUMPTION 10: G^* *is a strictly increasing, continuous distribution on* Y.

ASSUMPTION 11: *The characteristic functions corresponding to the marginal distribution functions* $F^*_{\varepsilon_{\tilde{j}} - \varepsilon_j}(j = 1..J; j \neq \tilde{j})$ *are everywhere different from 0.*

An example of a model where Assumptions 4-9 are satisfied is a binary choice where each function v(1,•) is characterized by a function m(•) and each function v(2,•) is characterized by a function h(•) such that for all s, x_1, x_2, ω: (i) $v(1,s,x_1,\omega) = m(x_1,\omega)$, (ii) $v(2,s,x_2,\omega) = h(s,x_2,\omega)$, (iii)

$m(\bar{x}_1, \omega) = 0$, (iv) $h(s, \bar{x}_2, \omega) = 0$, (v) $h(\hat{s}, \hat{x}_2, \omega) = \omega$, (vi) $h(s, x_2, \cdot)$ is strictly increasing when $x_2 \neq \bar{x}_2$, and (vii) $m(x_1, \cdot)$ is strictly decreasing when $x_1 \neq \bar{x}_1$, where $\bar{x}_1, \bar{x}_2, \hat{s},$ and $\hat{x}_2$ are given.

In this example, Assumption 4 is satisfied when $\bar{s}$ is any value and $\alpha_1 = \alpha_2 = 0$. Assumption 5 is satisfied with $\tilde{j} = 1$, $\beta_{\tilde{j}} = 0$ and $\tilde{x}_1 = \bar{x}_1$. By (v), Assumption 6 is satisfied with j^*=2. Finally, Assumption 7-9 are satisfied by (vi) and (vii). If in the above example, (v) is replaced by assumption (v') $h(\hat{s}, \hat{x}_2, \hat{\omega}) = \alpha$ and $h(\cdot,\cdot,\cdot)$ is homogenous of degree one, where, in addition to $\bar{x}_1, \bar{x}_2, \hat{s},$ and $\hat{x}_2$, $\hat{\omega}$ and $\alpha \in R$ are also given, then the model satisfies Assumptions 4, 5, 6', and 7-9.

Assumption 1 specifies the support of the observable explanatory variables and of ω. The critical requirement in this assumption is the large support condition on $(r_1,...,r_J)$. This requirement is used, together with the other requirements, to identify the distribution of $(\varepsilon_1 - \varepsilon_2,...,\varepsilon_1 - \varepsilon_J)$. Assumption 2 requires that the vectors $(\varepsilon_1, \varepsilon_2,...,\varepsilon_J)$, $(s, z_1,..., z_J)$, ω be mutually independent. That is, for all $(\varepsilon_1, \varepsilon_2,...,\varepsilon_J)$, $(s, z_1,..., z_J)$, ω,

$$f_{(\varepsilon_1,...,\varepsilon_J),((s,z_1,...,z_J),\omega}(\varepsilon_1,...,\varepsilon_J, s, z_1,..., z_J, \omega) = f_{(\varepsilon_1,...,\varepsilon_J)}(\varepsilon_1,...,\varepsilon_J) \bullet f_{(s,z_1,...,z_J)}(s, z_1,..., z_J) \bullet f_\omega(\omega)$$

where $f_{(\varepsilon_1,...,\varepsilon_J),((s,z_1,...,z_J),\omega}$ denotes the joint density of $(\varepsilon_1,...,\varepsilon_J, s, z_1,..., z_J, \omega)$, and $f_{(\varepsilon_1,...,\varepsilon_J)}(\varepsilon_1,...,\varepsilon_J)$, $f_{(s,z_1,...,z_J)}(s, z_1,..., z_J)$, and $f_\omega(\omega)$ denote the corresponding marginals. A critical implication of this is that for all $(\varepsilon_1,...,\varepsilon_J),(s, z_1,..., z_J),\omega$, the vectors $(\varepsilon_1,...,\varepsilon_J),\omega$ and $(r_1,...,r_J)$ are mutually independent conditional on $(s, z_1,..., z_J)$, i.e.,

$$f_{(\varepsilon_1,...,\varepsilon_J),(r_1,...,r_J),\omega|(s,x_1,...,x_J)}(\varepsilon_1,...,\varepsilon_J, r_1,..., r_J, \omega) = f_{(\varepsilon_1,...,\varepsilon_J),(r_1,...,r_J),\omega}(\varepsilon_1,...,\varepsilon_J, s, z_1,..., z_J, \omega)$$

Assumption 3 restricts the form of each utility function $V(j, s, x_j, r_j, \omega)$ to be additive in r_j, and the coefficient of r_j to be known. A requirement like this is necessary even when the function $V(j, s, x_j, r_j, \omega)$ is linear in variables, to compensate for the fact that the distribution of $(\varepsilon_1,...,\varepsilon_J)$ is not specified. Since the work of Lewbel (2000), regressors such as $(r_1,...,r_J)$ have been usually called "special regressors." Assumptions 4 and 5 specify the value of the functions $v(j,\cdot)$, defined in Assumption 3, at some points of the observable variables. Assumption 4 specifies the values of these $v(j,\cdot)$ functions at one point of the observable variables, for all values of ω. This guarantees that at those points of the observable variables, the choice probabilities are completely determined by the value of the distribution of $(\varepsilon_2-\varepsilon_1,...,\varepsilon_J-\varepsilon_1)$. This, together with the support condition in Assumption 1, allows us to identify this distribution. Assumption 5 specifies the value of one of the $v(j,\cdot)$ functions at one value of x_j, for all s,ω. As in the linear specification, only differences of the utility function are identified. Assumption 5 allows to recover the values of each of the $v(j,\cdot)$ functions, for $j \neq \tilde{j}$, from the difference between $v(j,\cdot)$ and $v(\tilde{j},\cdot)$. Once the $v(j,\cdot)$ functions are identified, one can use them to identify $v(\tilde{j},\cdot)$.

Assumptions 5-10 guarantee that the difference function $m(s,x_{j^*},x_{\tilde{j}},\omega) = v(j^*,s,x_{j^*},\omega)-v(\tilde{j},s,x_{\tilde{j}},\omega)$ and the distribution of ω are identified from the joint distribution of (γ,s,x) where $\gamma = m(s,x_{j^*},x_{\tilde{j}},\omega)$. Using the results in Matzkin (2003), identification can be achieved if (i) $m(s,x_{j^*},x_{\tilde{j}},\omega)$ is either strictly increasing or strictly decreasing in ω, for all $(s,x_{j^*},x_{\tilde{j}})$, (ii) ω is distributed independently of (s,x_{j^*}) conditional on all the other coordinates of the explanatory variables, and (iii) one of the following holds: (iii.a) for some $(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}})$, the value of m is known at $m(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}})$, for all ω; (iii.b) for some $(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}},\hat{\omega})$, the value of m is known at $(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}},\hat{\omega})$ and $m(s,x_{j^*},\tilde{x}_{\tilde{j}},\omega)$ is homogeneous of degree 1 along the ray that connects $(\hat{s},\hat{x}_{j^*},\hat{\omega})$ to the origin; (iii.c) for some $(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}},\hat{\omega})$, the value of m at $(\hat{s},\hat{x}_{j^*},\tilde{x}_{\tilde{j}},\hat{\omega})$ is known and m is strictly separable into a known function of ω and at least one coordinate of (s,x_{j^*}). Assumptions 6, 6', and 6'' guarantee, together with Assumption 5, that (iii.a), (iii.b), and (iii.c) are, respectively satisfied. Assumption 7 with Assumption 5 guarantees the strict monotonicity of m in ω. Assumption 2 guarantees the conditional independence between ω and (s, x_{j^*}). Hence, under our conditions, the function m and the distribution of ω are identified nonparametrically, as long as the joint distribution of (γ, s, x), where $\gamma = m(s, x_{j^*}, x_{\tilde{j}}, \omega)$, is identified.

Assumption 10 and 11, together with Assumption 2, guarantee that the distribution of (γ, s, x), where $\gamma = m(s, x_{j^*}, x_{\tilde{j}}, \omega)$, is identified, by using results in Teicher (1961). In particular, Assumption 11 is needed to show that one can recover the characteristic function of $v(j^*, s, x_j, \omega)$ from the characteristic functions of $\varepsilon_j - \varepsilon_{j^*}$ and of $v(j^*, s, x_j, \omega) + \varepsilon_j - \varepsilon_{j^*}$. The normal and the double exponential distributions satisfy Assumption 11. The distribution whose density is $f(x) = \pi^{-1} x^{-2} (1 - \cos(x))$ does not satisfy it.

Assumptions 8 and 9 guarantee that all the functions $v(j, \cdot)$ can be identified when the distribution of ω, the distribution of $(\varepsilon_1 - \varepsilon_2, ..., \varepsilon_1 - \varepsilon_J)$, and the function $v(j^*, \cdot)$ are identified.

Using the assumptions specified above, we can prove the following theorem:

THEOREM 1: *If Assumptions 1-11, Assumptions 1-5,6',7-11 or Assumptions 1-5,6'',7-11 are satisfied, then* (V^*, F^*, G^*) *is identified in* $(W \times \Gamma_F \times \Gamma_G)$.

This theorem establishes that one can identify the distributions and functions in a discrete choice model with unobserved heterogeneity, making no assumptions about either the parametric structure of the systematic subutilities or the parametric structure of the distributions in the model. The proof of this theorem is presented in the Appendix.

Our identification result requires only one observation per individual. If multiple observations per individual were available, one could relax some of our assumptions. One such possibility would be to use the additional observations for each individual to allow the unobserved heterogeneity variable to depend on some of the explanatory variables, as in Matzkin (2004). Another, more obvious, possibility would be to allow the nonparametric functions and/or distributions to be different across periods.

4. NONPARAMETRIC ESTIMATION

Given N independent observations $\{y^i, s^i, z^i\}_{i=1}^N$ we can define the log-likelihood function:

$$L(V,F,G) = \sum_{i=1}^{N} \log \int \left[\Pr(j \mid s^i, z^i; V(\cdot;\omega), \underline{F})\right]^{y_j^i} \, dG(\omega)$$

where $F = (F_1,...,F_J)$. We then define our estimators, $\hat{V}$, $\hat{F}$ and $\hat{G}$, for V^*, F^*, and G^*, to be the functions and distributions that maximize L(V,F,G) over triples (V,F,G) that belong to a set $(W \times \Gamma_F \times \Gamma_G)$. Let d_W, d_F and d_G denote, respectively, metric functions over the sets W, Γ_F, and Γ_G. Let $d\ :(W \times \Gamma_F \times \Gamma_G) \times (W \times \Gamma_F \times \Gamma_G) \to R_+$ denote the metric defined by

$$d[(V,F,G),(V',F',G')] = d_W(V,V') + d_F(F,F') + d_G(G,G').$$

Then, the consistency of the estimators can be established under the following assumptions:

ASSUMPTION 12: *The metrics d_W and d_F are such that convergence of a sequence with respect to d_W or d_F implies pointwise convergence. The metric d_G is such that convergence of a sequence with respect to d_G implies weak convergence.*

ASSUMPTION 13: *The set $(W \times \Gamma_F \times \Gamma_G)$ is compact with respect to the metric d.*

ASSUMPTION 14: *The functions in W and Γ_F are continuous.*

An example of a set $(W \times \Gamma_F \times \Gamma_G)$ and a metric d that satisfy assumptions 12 and 13 is where the metrics d_W, d_F, and d_G are defined, for all V, V', F, F', and G, G' by

$$d_w(V,V') = \int \left|V(s,z,\omega) - V'(s,z,\omega)\right| e^{-\|s,z,\omega\|} d(s,z,\omega)$$
$$d_F(F,F') = \int \left|F(t_1,...t_{J-1}) - F'(t_1,...t_{J-1})\right| e^{-\|(t_1,...t_{J-1})\|} dt$$
and
$$d_G(G,G') = \int \left|G(t) - G'(t)\right| e^{-\|t\|} dt\ ,$$

the set W is the closure with respect to d_W of a set of pointwise bounded and α_w-Lipschitsian functions, the set Γ_F is the closure with respect to d_F of a set of α_Γ–Lipschitsian distributions, and Γ_G is the set of monotone increasing functions on R with values in [0,1]. A function $m: R^K \to R$ is α- Lipschitsian if for all x,y in R^K, $|m(x) - m(y)| \leq \alpha \|x - y\|$. In particular, sets of concave functions with uniformly bounded subgradients are α - Lipschitsian for some α (See Matzkin 1992). The following theorem is proved in the Appendix:

THEOREM 2: *Under Assumptions 1-14* $(\hat{V}, \hat{F}, \hat{G})$ *is a strongly consistent estimator of* (V^*, F^*, G^*) *with respect to the metric d. The same conclusion holds if Assumption 6' or 6'' replace Assumption 6.*

In practice, one may want to maximize the log-likelihood function over some set of parametric functions that increases with the number of observations, in such a way that it becomes dense in the set $(W \times \Gamma_F \times \Gamma_G)$ (Elbadwai, Gallant, and Souza 1983, Gallant and Nychka 1987, Gallant and Nychka 1989). Let $W^N, \Gamma_F^N,$ *and* Γ_G^N denote, respectively, such sets of parametric functions, when the number of observations is N. Let $(\tilde{V}_N, \tilde{F}_N, \tilde{G}_N)$ denote a maximizer of the log-likelihood function $L(V, F, G)$ over $(W^N \times \Gamma_F^N \times \Gamma_G^N)$. Then, we can establish the following theorem:

THEOREM 3: *Suppose that Assumptions 1-14 are satisfied, with the additional assumption that the sequence* $\{(W^N \times \Gamma_F^N \times \Gamma_G^N)\}_{N=1}^{\infty}$ *becomes dense (with respect to d) in* $(W \times \Gamma_F \times \Gamma_G)$ *as* $N \to \infty$. *Then,* $(\tilde{V}_N, \tilde{F}_N, \tilde{G}_N)$ *is a strongly consistent estimator of* (V^*, F^*, G^*) *with respect to the metric d. The same conclusion holds if Assumption 6 is replaced by Assumption 6' or 6''.*

Note that Theorem 3 holds when only some of the functions are maximized over a parametric set, which becomes dense as the number of observations increases, and the other functions are maximized over the original set.

5. COMPUTATION

In this section, we introduce a change of notation from the previous sections. Let H be the number of households in the data (H ≤ N), and N_h be the number of observed choices for household h=1..h (N_h > 0, and $N = \sum_{h=1,H} N_h$). Therefore, the key difference in this and the following sections from the previous section is that we have repeated observations for the households. The above theorems still apply as long as the assumptions are maintained; including independence of choices (see, e.g., Lindsay 1983, Heckman and Singer 1984, Dayton and McReady 1988). Note that this assumption excludes endogenous (including lagged endogenous variables), and we leave it for future research to determine the identification and consistency conditions for models with endogenous predictors.

Let $(\hat{V}, \hat{F}, \hat{G})$ denote a solution to the optimization problem:

$$\underset{(V,F,G)\in(W\times\Gamma_F\times\Gamma_G)}{Max} L(V,F,G) = \sum_{h=1}^{H} \log \int \prod_{i=1}^{N_h} \prod_{j=1}^{J} \left[\Pr(j \mid s_h^i, z_j^i; V(\cdot,\omega), F)\right]^{y_{h,j}^i} dG(\omega)$$

It is well known that, when the set Γ_G includes discrete distributions, $\hat{G}$ is a discrete distribution with at most H points of support (Lindsay 1983, Heckman and Singer 1984). Hence, the above optimization problem can be solved by finding a solution over the set of discrete distributions, G, that possess at most H points of support. We will denote the points of support of any such G by $\omega_1,...,\omega_H$, and the corresponding probabilities by $\pi_1,...,\pi_H$. Note that the value of the objective function depends on any function $V(\cdot,\omega)$ only through the values that $V(\cdot,\omega)$ attains at the finite number of observed vectors $\{(s_1^1, x_j^1, r_j)_{j=1,...,J}, ..., (s_H^{N_H}, x_j^{N_H}, r_j)_{j=1,...,J}\}$. Hence, since at a solution, $\hat{G}$ will possess at most H points of support, we will be considering the values of at most H different functions, $V(\cdot,\omega_c)\ c = 1,...,H$, i.e., we can consider for each $j\ (j = 1,...,J)$ at most H different subutilities, $V(j,\cdot;\omega_c)\ (c = 1,...,H)$. For each *i* and *c*, we will denote the value of $V(j, s_h^i, x_j^i, r_j, \omega_c)$ by $V_{j,h,c}^i$. Also, at a solution, the value of the objective function will depend on any F only through the values that F attains at the finite number of values $(V^i_{j,1,c} - V^i_{1,1,c},\ldots,V^i_{j,H,c} - V^i_{1,H,c})$, $i = 1,\ldots,N_h$, $j = 1,\ldots,J$, $h = 1,\ldots,H$, $c = 1,\ldots,H$. We then let $F^i_{j,h,c}$ denote, for each j (j=1,…,J), the value of a distribution function F_j at the vector $(V^i_{j,1,c} - V^i_{1,1,c},\ldots,V^i_{j,H,c} - V^i_{1,H,c})$. It follows that a solution, $(\hat{V},\hat{F},\hat{G})$ for the above maximization problem can be obtained by first solving the following finite dimensional optimization problem, and then interpolating between its solution:

$$\max_{\{V^i_{j,h,c}\},\{\pi_c\},\{F^i_{j,h,c}\}} \sum_{h=1}^{H} \log \sum_{c=1}^{H} \pi_c \prod_{i=1}^{N_h} \prod_{j=1}^{J} \left[F^i_{j,h,c}\right]^{y^i_{h,j}}$$

$$\textit{subject to} \qquad (\{V^i_{j,h,c}\},\{\pi_c\},\{F^i_{j,h,c}\}) \in K$$

where K is a set of a finite number of restrictions on $\{V^i_{j,h,c}\},\{\pi_c\},\{F^i_{j,h,c}\}$. The restrictions characterize the behavior of sequences $\{V^i_{j,h,c}\},\{\pi_c\},\{F^i_{j,h,c}\}$ whose values correspond to functions V in W, probability measures G in Γ_G, and distribution functions F in Γ_F. To see what is the nature of the restrictions determined by the set K, consider for example a binary choice model where $x_1 \in R_+$, $v(1,s_h,x_1,\omega) = r(s_h) + \omega\, x_1, v(2,s_h,x_2,\omega) = h(x_2,\omega)$, $r(0) = 0$, $h(0,\omega) = 0$ for all ω, $r(\cdot)$ is concave and increasing, and $h(\cdot,\cdot)$ is concave and decreasing. Then, the finite dimensional optimization problem takes the following form:

$$\max_{\{V^i_{j,h,c}\},\{\pi_c\},\{F^i_{h,c}\},\{T^i_h\},\{D^i_{h,c}\}} \sum_{h=1}^{H} \log \sum_{c=1}^{H} \pi_c \prod_{i=1}^{N_h} \left[F^i_{h,c}\right]^{y^i_{1,h}} \left[1 - F^i_{h,c}\right]^{(1-y^i_{1,h})}$$

subject to

$$(a)\ \ F^i_{h,c} < F^k_{h1,d} \qquad \textit{if} \qquad r^i_h + x^i_{1,h}\ \omega_c - h^i_{h,c} < r^k_{h1} + x^k_{1,h1}\ \omega_d - h^k_{d,h1},\ \pi_c > 0,\ \textit{and}\ \ \pi_d > 0$$

$$F^i_{h,c} = F^k_{h1,d} \qquad \textit{if} \qquad r^i_h + x^i_{1,h}\ \omega_c - h^i_{h,c} = r^k_{h1} + x^k_{1,h1}\ \omega_d - h^k_{h1,d},\ \pi_c > 0,\ \textit{and}\ \ \pi_d > 0$$

$$(b)\ \ 0 \le F^i_{h,c} \le 1,$$

$$(c)\ \ r^i_h \le r^k_{h1} + T^k_{h1} \cdot (s^i_h - s^k_{h1})$$

$$(d)\ \ h^{i}_{h,c} \le h^{k}_{h1,c} + D^{k}_{h1,c}\cdot((x^{i}_{h,2},\omega_c)-(x^{k}_{h1,2},\omega_c)) \qquad if\ \ \pi_c > 0$$

$$(e)\ \ T^{k}_{h1} \ge 0, \quad r^{N+1}=0, \ \ s^{N+1}=0,$$

$$(f)\ \ D^{k}_{h,c} \le 0, \quad h^{N+1}_{h,c}=0, \ and\ \ x^{N+1}_{h,2}=0 \qquad if\ \ \pi_c > 0$$

for $i \in \{1,...,N_h, N+1\}; k \in \{1,...N_j, N+1\};\ c,d,h,h1 = 1,...,H.$

Constraints (a) and (b) guarantee that the $F^{i}_{h,c}$ values are those of an increasing function whose values are between 0 and 1. Constraint (c) guarantees that the r^{i}_{h} values correspond to those of a concave function. Constrains (d) guarantees that the $h^{i}_{h,c}$ values correspond to those of a concave function, as well. Constraints (e) and (f) guarantee that the r^{i}_{h} and the $h^{i}_{h,c}$ values correspond, respectively, to those of a monotone increasing and a monotone decreasing function, and that the r^{i}_{h} and the $h^{i}_{h,c}$ values correspond to functions satisfying r(0)=0 and h(0,ω)=0 for all ω. Some additional constraints would typically be needed to guarantee the compactness of the sets W, Γ_F, and Γ_G and the continuity of the distributions in Γ_F.

A solution to the original problem is obtained by interpolating the optimal values obtained from this optimization (see Matzkin 1992, 1993, 1994, 1999b for more discussion of a similar optimization problem). To describe how to obtain a solution to this maximization problem, we let

$$\tilde{L}\begin{bmatrix}(r_1^{1},...,r_1^{N_1},...,r_H^{N_H}, r^{N+1}),(T_1^{1},...,T_1^{N_1}...,T_H^{N_H}, T^{N+1}), \\ (h^{1}_{1c},...,h^{N_1}_{1c}...,h^{N_H}_{Hc}, h^{N+1}_{c}),(D^{1}_{1c},...,D^{N_1}_{1c}...,D^{N_H}_{Hc}, D^{N+1}_{c}),(\pi_1,...,\pi_H)\end{bmatrix}$$

denote the optimal value of the following maximization problem:

$$\max_{\{V^{i}_{j,h,c}\},\{\pi_c\},\{F^{i}_{h,c}\},\{T^{i}_{h}\},\{D^{i}_{h,c}\}} \ \sum_{h=1}^{H} \log \sum_{c=1}^{H} \ \pi_c \prod_{i=1}^{N_h} \ \left[F^{i}_{h,c}\right]^{y^{i}_{1,h}} \left[1-F^{i}_{h,c}\right]^{(1-y^{i}_{1,h})} \tag{1}$$

subject to:

$$(a)\ \ F^{i}_{h,c} < F^{k}_{h1,d} \qquad if \quad r^{i}_{h} + x^{i}_{h,1}\ \omega_c - h^{i}_{h,c} < r^{k}_{h1} + x^{k}_{h1,1}\ \omega_d - h^{k}_{h1,d},\ \pi_c > 0,\ and\ \ \pi_d > 0$$

$$F^{i}_{h,c} = F^{k}_{h1,d} \qquad if \quad r^{i}_{h} + x^{i}_{h,1}\ \omega_c - h^{i}_{h,c} = r^{k}_{h1} + x^{k}_{h1,1}\ \omega_d - h^{k}_{h1,d}\,,\ \pi_c > 0,\ and\ \ \pi_d > 0$$

$(b)\ \ 0 \le F^{i}_{h,c} \le 1$.

A solution to this latter problem can be obtained by using a random search over vectors $(F^{1}_{1,c},...,F^{N_H}_{H,c})_{c=1,...,H}$ that satisfy the monotonicity constraint (a) and the boundary constraint (b). Then, a solution to the full optimization problem can be obtained by using a random search over vectors $(r^{1}_{1},...r^{N_1}_{1},...,r^{N_H}_{H},r^{N+1})$, $(h^{1}_{1c},...,h^{N_1}_{1c}...,h^{N_H}_{Hc},h^{N+1}_{c})$, and $(\pi_1,...,\pi_H)$ that satisfy, respectively, constraints (c) and (e), constraints (d) and (f), and the constraints of $\pi_c \ge 0$ (c=1, .., H) and $\sum_{c=1}^{H}\pi_c = 1$.

Instead of estimating the distribution function F using (a) and (b), one could add alternative-specific random intercepts to the model and assume that ε has a known, parametric distribution. When the specified distribution for ε is smooth, this has the effect of smoothing the likelihood function. Let $\Theta = \{\beta_1,...,\beta_{SC},\pi_1,...,\pi_{SC}\}$ be the set of parametric parameters, where $\{\beta_1,...,\beta_{SC}\}$ are the parameters of the utility function; and SC is the number of discrete consumer segments with SC ≤ H; let $\omega = \{\omega_1,...,\omega_{SC}\}$ be the vector of unobserved heterogeneity, and $H = \{h^{1}_{1},...,h^{1}_{N+1},...h^{SC}_{N+1}\}$ be the set of values for the non-parametric function. The computational problem then becomes estimating (Θ,ω,H) efficiently using the likelihood function described in equation (1). Clearly, a random search over the entire parameter space is infeasible as the parametric parameters are unconstrained and the heterogeneity parameters are only constrained to be positive (from Assumption 9). Therefore, we adapted the algorithm for concave functions developed in Matzkin (1999b) and later used by Briesch, Chintagunta and Matzkin (2002) to monotone functions of the form described above. This is a random search algorithm combined with maximum likelihood estimation for the parametric parameters.

6. EMPIRICAL APPLICATION

We are interested in answering the research question of how consumers select between an Every Day Low Price (EDLP) format retailer (e.g., Wal-Mart) and a HiLo format (e.g., Kroger or many grocery stores) retailer. An EDLP retailer generally has lower price variance over time than a HiLo retailer (Tang, et al 2001). In studying store choice, the marketing literature typically assumes that consumers look in their household inventory, construct a list (written or mental) of needed items and quantities of these items, then determine which store to visit based upon the cost of refilling their inventory and the distance to the store, and potentially, interactions among them (Huff 1962, Bucklin 1971, Bell and Lattin 1998, Bell, Ho and Tang 1998). While consumers may make purchases for reasons other than replenishing inventory (e.g., "impulse" purchase), we leave the analysis of such purchases for future research.

The functional relationship between distance to store and the cost of refilling inventory on the one hand and the utility derived from going to a store is not well understood. Distance enters the indirect utility function non-linearly (i.e., natural logarithm) in Smith (2004), whereas it enters linearly in Bell and Lattin (1998). Further, Rhee and Bell (2002) find that the cost of a format is not significant in the consumer's decision to switch stores, so they conclude that the "price image" drives store choice, not cost to refill the inventory (at odds with Bell and Lattin 1998, and Bell, Ho and Tang 1998). It is also likely that the influence of these variables is heterogeneous across consumers. So how do consumers make tradeoffs between distance to the store and the price paid for the shopping basket? To answer this, we apply a semiparametric version of the method described above (details on follow) that allows us to (a) recover the appropriate functional form for the effects of inventory replenishing cost and distance on format choice; and (b) account for heterogeneity in this response function across consumers.

We specify the utility of a format as a tradeoff between a consumer's cost of refilling inventory at that format and the distance of the format as:

$$V_{h,f}^{t} = h(-S_{h,f}^{t}, -D_{h,f}, \omega) + X_{h,f}^{t}\beta_h + \varepsilon_{h,f}^{t}$$

where: $V_{h,f}^{t}$ is the value of format *f* for household *h* in period (or shopping occasion) *t*, $S_{h,f}^{t}$ is the cost of refilling household *h*'s inventory at format *f* in period *t*, $D_{h,f}$ is the distance (in minutes) from household *h* to format *f*, *h*(.) is a monotonically increasing function of its parameters, $X_{h,f}^{t}$ are other variables that need to be controlled for, ω is the unobserved heterogeneity and $\varepsilon_{h,f}^{t}$ is the error term. Note that the matrix of variables in $X_{h,f}^{t}$ can be used to provide identification of the non-parametric function $h(-S, -D, \omega)$, if required.

Time invariant format characteristics such as service quality, parking, assortment, etc. (see Bucklin 1971, Tang, et al 2001) are reflected in a household-specific intercept term for the price format. In addition to the format-specific intercept, key consumer demographic variables are also included. Finally, several researchers (e.g., Bell and Lattin, 1998; Bell, Ho, and Tang, 1998) have found that consumers prefer to go to EDLP stores when the expected basket size is large or as time between shopping trips increases (e.g., Leszczyc, et al., 2000). Our objective is to account fully for all observable sources of heterogeneity so any unobserved heterogeneity estimated from the data is to the extent possible, unobservable. Therefore, we can rewrite the utility of each format as in equation (2),

$$V_{h,f}^{t} = \beta_{0f} + EDLP * (\beta_1 TS_h^t + \beta_2 E_h + \beta_3 HS_h + \beta_4 I_h + \beta_5 CE_h) + \beta_6 L_{h,f} + h(-S_{h,f}^{t}, -D_{h,f}, \omega) + \varepsilon_{h,f}^{t} \quad (2)$$

where: *EDLP* is a binary variable set to one when the format is EDLP and zero otherwise, TS_h^t is the elapsed time (in days) from the previous shopping trip to the current shopping trip, E_h is a binary indicator set to one if the household is classified as "Elderly" (head of household is at least 65 years old), HS_h is the size of the household, I_h is the household income, CE_h is a binary indicator of whether the head of household is college educated, and L_{hf} is a format-specific loyalty term. We use the same measure of loyalty as Bell, Ho and Tang (1998) after adjusting for using store formats instead of stores:

$$L_{h,EDLP} = \left(NV^{i}_{h,EDLP} + 0.5\right) / \left(NV^{i}_{h,EDLP} + NV^{i}_{h,HiLo} + 1\right);\ L_{h,HiLo} = 1 - L_{h,EDLP}$$

where $NV^{i}_{h,EDLP}$ is the number of visits to EDLP stores by household h during an initialization period denoted by i, and $NV^{i}_{h,HiLo}$ is the number of visits to HiLo stores by household h during the same initialization period, i. The data from the initialization period temporally precede the data that we use for the estimation for each household included in our sample. In this way we are not directly using any information on our dependent variable as a model predictor.

We make the assumption that the error term, $\varepsilon^{t}_{h,f}$, has an extreme value distribution. We use a discrete model of consumer heterogeneity, where there are SC segments of consumers (or points of support). The parameter vector is allowed to be segment specific, so the utility function in equation (1) can be rewritten using segment-specific subscripts for the appropriate parameters and ω. To guarantee that Assumption 3 is satisfied, we impose the restriction that for all segments s, $\beta_{1,s} = -0.02$. This value was determined from the single segment parametric model. To guarantee that Assumptions 4, 5, and 6'' are satisfied, we impose the restrictions that (1) $h(-S^{t}_{h,f}, -D_{h,f}, \omega_s) = m(-S^{t}_{h,f}\omega_s, -D_{hf})$, (2) $m(0, -D_{N+1}) = \alpha$, and (3) $m(1, -D_{N+1}) = \gamma$, with N being the number of observations in the data, D_{N+1}, α and γ known. Assumption 4 is then satisfied when $D=D_{N+1}$ and $S=S_{N+1}=0$. The variable S is measured in differences from its average. To guarantee that Assumptions 7 and 8 are satisfied, the function $m(\cdot)$ is restricted to be monotonically increasing.

We define the probabilities for the support points, π_c, such that $\sum_{c=1}^{SC} \pi_c = 1$. Following the literature on discrete segments (Dayton and McReady 1988, Kamakura and Russell 1989), we write the mass points as $\pi_c = \exp(\lambda_c) \Big/ \left(\sum_{k=1}^{SC} \exp(\lambda_k)\right)$, with λ_1 set to zero for identification.

We next describe how the computational algorithm presented in the previous section can be modified to leverage our specific form of the nonparametric function. For expository ease, we drop the household subscript from the predictor variables and treat them as if they are independent observations. Since there are two alternatives (EDLP and Hi-Lo), we have $2*N+2$ pairs of cost and distance (note that we are assuming that the pairs are all unique. If there are repeated pairs, then we use the subset of unique pairs. Additionally, there are two constraints.) If a continuous, differentiable and constraint maintaining (i.e., does not violate any of the AH ≤ 0 constraints) approximation of $m(-\omega S^i, -D^i)$ for all $\omega>0$, is used, then the dimensionality of the random search (or the size of the vector H) can be reduced from $2*(N+1)*SC$ to $2*(N+1)$ and maximum likelihood can be used to estimate ω as well as Θ.

A natural choice for this interpolation would be a multidimensional kernel where the weight placed on observation i, in calculating the value for observation j is inversely proportional to the Euclidian distance from $\{S^j\omega_c, D^j\}$ to $\{S^i, D^i\}$, i.e., $E_{i,j,c} \equiv \left|\{\omega_c S^j, D^j\} - \{S^i, D^i\}\right|$. The problem with using a multi-dimensional kernel is that is does not preserve the shape restrictions (a stylized proof is available from the authors). Therefore, we use a single dimensional kernel to smooth the $S^j\omega_c$ values.

The final issue to address is identification of the ω's. The segments are not uniquely defined as segments can be renumbered while still maintaining the same likelihood function. Additionally, we note while all of the ω's are identified, the estimation is computationally inefficient as we estimate the base function for ω_0=1, then interpolate for all SC segments. The computational efficiency can be improved by estimating ω` instead of ω, where $\omega' \equiv \omega / \omega_i$ for some non-zero ω_i. This constraint then implies we can set ω_1=1 and estimate SC-1 segment values.

6.1 Data

We use a multi-outlet panel dataset from Charlotte, North Carolina that covers a 104-week period between September 2002 and September 2004. Since panelists record all packaged and nonpackaged goods purchases using in-home scanning equipment, purchase records are not limited to a small sample of grocery stores; purchases made in all grocery and non-grocery stores are captured. This is important since packaged goods purchases are frequently made outside of grocery stores.

Households were included in the sample if at least 95 percent of their purchases were at the 7 stores (five supermarket, two mass merchandisers) for which we have geolocation data, and if they spent at least \$20 per month in the panel. The last criterion was used to ensure that the panelist was faithful in recording its purchases and remained in the panel for the entire 104 week period. The resulting data set had 161 families with a total of 26,540 shopping trips. The first 25% of the weeks were used as our "initialization" period to compute the various weights and other quantities described below. The final 26 weeks were used as a "hold out" sample and remaining weeks were used as the estimation sample. The estimation (holdout) sample had 13,857 (6573) shopping trips. On average, each household made a shopping trip every 4.6 days. Descriptive statistics for the households are provided in Table 1.

< Put Table 1 about here>

Consistent with Bell, Ho and Tang, (1998) we identified retailers as being either EDLP or HiLo based on their advertised pricing strategy. This resulted in three EDLP retailers and four HiLo retailers. The EDLP (HiLo) retailers had 58% (42%) of the shopping trips. To determine distance between a panelist and a store, we use the travel time (in minutes) from a panelist's zip+4 to the store's location (for privacy reasons, the panelists actual street address is not included in the data.)

We have detailed price information for 289 categories, of these categories we selected 150 categories based upon the following criteria. First, three common UPCs had to be carried by each retailer so that category price indices could be computed. Second, at least 5% of the selected households had to make at least three purchases in the category to ensure that the category is substantial. Third, the category had to account for at least 0.9% of total basket spending of the selected households. These categories together comprise more than 88% of the market basket on average (excluding fresh meat, fruit and vegetables). So, we use these categories to estimate the cost at each format. While not reported here, details of the categories are available from the authors. Table 2 shows specific statistics for the price formats with the standard deviations in parentheses. Note that the description of cost to refill inventory is provided in the next section.

< Put Table 2 about here>

6.1 Data Aggregation

Because our focus is on consumers selecting a type of store format (EDLP vs. Hi-Lo) rather than selecting a specific store, we need to aggregate our store-level data to the format level. This aggregation is done based on the proportion of a household's visits to each store during the initialization period defined previously. Specifically, let F_E be the set of stores with an EDLP price format, F_H be the set of stores using a HiLo format and D_{hs} be the distance from household *h*'s home to store s. The distance from household *h* to format *f* can be defined as $D_{h,f} = \sum_{s \in F_f} w_{h,f,s} D_{h,s}$, where $w_{h,f,s}$ is the proportion of visits to store *s* made by household *h* in the initialization period, and F_f is either F_E or F_H.

Because the cost to replenish the inventory (also called "cost") for a household, $S^t_{h,f}$ on a shopping occasion is not observed prior to the visit, we need to create a measure of cost for each trip. The general goal is to create a time-varying, household-specific index for each store. This index is then aggregated to the format level similar to distance above. Clearly, the cost at store *s* in period *t* by household *h*, denoted $S^t_{h,s}$, is the sum of the cost to the household in each category, c=1..C the household will purchase in period *t* (Bell, Ho and Tang 1999). Therefore, cost can be written as

$$E[S^t_{h,s}] = \sum_{c=1}^{C} E[p^t_{h,s,c}]E[q^t_{h,c}] \tag{3}$$

where $E[p^t_{h,s,c}]$ is the expected price of category *c* at store *s* in period *t* for household *h*, and $E[q^t_{h,c}]$ is the quantity household *h* needs to purchase in category *c* in period *t* to replenish its inventory. For price, we construct "market average" price indices for the EDLP and HiLo retailers based upon the retailers' long-run share of visits, i.e., visits over the entire estimation sample. The second component on the right hand side of equation (3) is quantity that needs to be replenished, $E[q_{h,c}^{t}]$. Because we do not observe household inventories, we need a mechanism to predict this quantity using data that we, as researchers, observe – quantities purchased on previous occasions. We use Tobit models (that account for household heterogeneity) to predict each household's expected purchase quantity. Using the arguments found in Nevo and Hendel (2002) and elsewhere, we define ($E[q_{h,c}^{t}]$) to depend upon previous quantity purchased, the amount of time since last purchase, and the interaction between these terms as:

$$E\left[q_{h,c}^{t}\right]/\overline{q}_{h,c} = \beta_{0,h} + \beta_{,1h}\left(q_{h,c}^{t-1}/\overline{q}_{h,c}\right) + \beta_{2,h}\left(d_{h,c}^{t}/\overline{d}_{hc}\right) + \beta_{3h}\left(\left(q_{h,c}^{t-1}/\overline{q}_{hc}\right)-1\right)\left(\left(d_{h,c}^{t}/\overline{d}_{hc}\right)-1\right) + \xi_{h,c}^{t} \qquad (4)$$

where $q_{h,c}^{t}$ is the quantity purchased in category *c* in period *t* by household *h*, $\overline{q}_{h,c}$ is the average quantity of category c purchased by household *h* conditional on purchase in the category, $d_{h,c}^{t}$ is the number of days since the category has been purchased, $\overline{d}_{hc}$ is the average number of days between purchases in category *c* by household *h*, and $\xi_{h,c}^{t}$ is an error term which has normal distribution and is independent between categories. In the interaction term we subtract one (which is the mean of both $q_{h,c}^{t-1}/\overline{q}_{hc}$ and $d_{h,c}^{t}/\overline{d}_{hc}$) from the quantity and time terms to allow clear definition of β_{1h} and β_{2h}: β_{1h} is the household response to quantity when the time between purchases is at the mean value, and β_{2h} represents household response to time when the average quantity was purchased on the prior occasion. The interaction term allows the household's consumption rate to be non-constant. The household coefficients in equation (4) are assumed to have a normal distribution with mean β and standard deviation of Σ, where Σ is a diagonal matrix. The key parameters of interest are lag quantity (expected sign is negative), days since last category purchase (expected sign is positive), and the interaction effect between these variables (expected sign is positive).

Although not reported here, we find that, at the 5% level, 3 categories (2%) had positive and significant coefficients for lag quantity and 71 categories (47%) had negative and significant coefficients for lag quantity. 95 (63%) categories had positive and significant coefficients for time since last category purchase and 13 (9%) had negative and significant coefficients for time since last category purchase. 46 (31%) has positive and significant coefficients for the interaction between these variables, and 0 (0%) has negative and significant coefficients for the interaction. These results provide support for the models where the coefficients with incorrect signs are, on aggregate, approximately 5% (16 of 450), which would be expected by chance.

We use the Tobit model results to predict the expected quantity required by each household in each period. We then follow the same aggregation scheme from above to obtain the cost of visiting a specific format. Given the complexity and computational intensity of this method, we leave it for future research to determine methods that allow computationally feasible simultaneous estimation of quantity and store choice. However, even in a parametric case, the problem of simultaneously estimating 150 category equations and one format choice equation is formidable (Note that we account for the estimation error in the standard errors by estimating Tobit models for each bootstrap simulation).

6.3 Results

We also estimated a parametric model for model comparison. For this model, the function *h*(.) is defined as a simple linear specification: $h(-S_{hft}, -D_{hf}, \omega_s) = \gamma_1^s S_{hft} + \gamma_2^s D_{hf}$.

6.3.1 Model Selection

Table 3 provides the maximized likelihood values and information criteria for the estimation and hold-out samples for one to four discrete segments. We calculate hits by assigning households to segments using the posterior probability of segment membership (see Kamakura and Russell 1989) then calculating the correct number of predicted choices using only that segment's parameters to which the household is assigned. We use 25 bootstrap simulations to calculate the standard deviations of the model fit statistics across all models in both the parametric and semiparametric estimation.

<Put Table 3 about here>

The Schwarz Criterion and the out-of-sample performance (log-likelihood and hit rate) indicate that the three segment model is the "best" parametric model. Interestingly, the Deviance Information Criterion (DIC) suggests that the four segment model is superior. However, we use the more conservative Schwarz Criterion to avoid over-parameterization of the model. In the DIC, the effective degrees of freedom are calculated as the difference between the likelihood of the estimation and the mean likelihood of the bootstrap simulations.

Given a three-segment parametric solution, we then estimated one through three segment solutions of the semiparametric model. The results of the estimation are provided in table 4, with the DIC indicating that the three segment model is superior. It is interesting to note that the improvement in the likelihood for using the one-segment semiparametric model versus the one-segment parametric model is similar to the improvement in the two-segment parametric model versus the one segment parametric model. Therefore, replacing a linear function with a monotone function has a similar impact on the likelihood as adding heterogeneity, and this relationship remains as more discrete segments are added. This result is consistent with the findings in Briesch, Chintagunta and Matzkin (2002).

<Put Table 4 about here>

6.3.2 Model Results

Table 5 provides the MLE parameter estimates from the three segment models for the parametric and semiparametric estimations, with the standard errors reported in parentheses. The segments were matched based upon the size of the mass coefficient (which roughly translates to the number of households in the segment). We note that this matching is arbitrary. Some of the key points are:

1. Demographic effects appear to be different across the parametric and semiparametric models. In the parametric model, Household Size is significant and positive for all segments, while it is significant (and negative) for only one segment in the semiparametric model. College

educated head of household is significant for all of the semiparametric segments, while it is significant for only one of the parametric segments.

2. Loyalty effects appear to be similar for two of three segments between the parametric and semiparametric models. However, for segment three of the parametric model, the coefficient is negative and significant, unlike the inertia effects in the semiparametric model.
3. Finally, all of the distance and cost coefficients for the parametric model have the expected negative coefficients, and these coefficients are significant at $p<0.05$ (one-tailed test).

<Put Table 5 about here>

Now, we turn our attention to heterogeneity distribution of cost and distance sensitivities. Table 6 provides the elasticity estimates for both parametric and semiparametric models as well as the percent of households assigned to each segment. First, the heterogeneity distribution is somewhat similar between the methods in an ordinal manner, i.e., most of the families are in the moderate cost sensitivity segment (segment 2), followed by the most cost sensitive segment (segment one) then the least cost sensitive segment (segment three). Second, the semiparametric model indicates a larger range of cost sensitivity elasticities with a larger proportion of households being very cost sensitive. Third, the range of distance elasticities is larger in the parametric model. Indeed, the semiparametric model indicates less heterogeneity in distance sensitivities across households. These differences suggest that there are likely big differences in the response surfaces. Accordingly, the estimated response surfaces for the variables of interest (cost and distance) are shown in Figures 1 and 2. Bootstrap simulations are used to get the semiparametric confidence intervals. The response surface for distance is plotted holding cost fixed at the mean value. Similarly, the response surface for cost is plotted holding distance fixed at the mean.

If we examine consumer response to distance (figure 1), we see that the semiparametric function is convex and decreasing. While not shown here, this convexity is pronounced at lower cost levels. This finding implies an interaction with the two variables would be required in a parametric representation of the model. We note that there are large differences in the parametric versus semiparametric functions, with the former having much larger slopes and different intercepts.

<Put figures 1 and 2 about here>

If we examine consumer response to cost (figures 2), we also find differences between the parametric and semiparametric response surfaces. While not shown here, in the semiparametric case, we see an interaction effect with distance. At large distances, the response surface is almost flat. However, there are significant non-linearities at shorter distances. The finding is consistent with the "tipping point" argument made in Bell, Ho and Tang (1998), although these results are much stronger.

Finally, we examine the demographic profiles of the segments in Table 7. The significant differences (at $p < 0.05$) between the parametric and semiparametric models are in bolder font in the table. There are three significant differences between the parametric and semiparametric results: percent of trips to EDLP retailer for segment one, average distance to selected formats in segments one and two and college education of segment one. The average expected spending is similar for the segments but the number of trips (and hence total cost) is different, similar to Bell, Ho and Tang (1998).

<Put Table 7 about here>

7. CONCLUSIONS

We have presented a method to estimate discrete choice models in the presence of unobserved heterogeneity. The method imposes weak assumptions on the systematic subutility functions and on the distributions of the unobservable random vectors and the heterogeneity parameter. The estimators are computationally feasible and strongly consistent. We described how the estimator can be used to estimate a model of format choice. Key insights from the application include: (1) the parametric model provides different estimates of the heterogeneity distribution than the semiparametric model, (2) the semiparametric suggests interactions between cost and distance that change the shape of the response function, and (3) the benefit to adding semiparametric estimation is roughly equal in magnitude to adding heterogeneity to parametric models. The benefit remains as the number of segments increases.

One drawback of the method is the computational time. The amount of time required to estimate the model is proportional to how good of a starting point is used. We used the one-segment semiparametric solution as a starting point for the three segment solution, and it took approximately two weeks for the first simulation to complete on a 1 GHz personal computer (the bootstrap simulations took a much shorter amount of time as they used the three-segment solution as a starting point). As computing power becomes cheaper, this should be less of a problem.

Some variations to the model presented in the above sections are possible. For example, instead of letting $V(j,s,x_j,r_j,\omega) = v(j,s,x_j,\omega) - r_j$, where $r_j \in R_+$, one can let r_j be an *L* dimensional vector, and specify $V(j,s,x_j,r_j,\omega) = v(j,s,x_j,\omega) - \beta \cdot r_j$. Assuming that one coordinate of β equals one, it is possible to identify β as well as all the functions and distributions that were identified in the original model. Another variation is obtained by eliminating the unobservable random variables ε_j. When $V(j,s,x_j,r_j,\omega) = v(j,s,x_j,\omega) - r_j$, the distribution of ω and the functions v(j,.) are identified as well.

Future extensions of the model will deal with the case where the heterogeneity parameter ω is multidimensional and the case where the vector of observable exogenous variables $(s,z_1,\ldots,z_J)$ is not necessarily independent of either ω or $(\varepsilon_1,\ldots,\varepsilon_J)$ (Matzkin, 2005).

Acknowledgements

Rosa Matzkin gratefully acknowledges support from NSF for this research. The authors gratefully thank Information Resources, Inc., ESRI, and Edward Fox, Director of the Retailing Center at Southern Methodist University for providing the data used in this research. Pradeep Chintagunta gratefully acknowledges support from the Kilts Center in Marketing at the University of Chicago.

APPENDIX A: Theorem Proofs

PROOF OF THEOREM 1:

Let $\eta_j = (\varepsilon_1 - \varepsilon_j, ..., \varepsilon_J - \varepsilon_j)$ (j=1,…J). To recover the distribution of η_k for all k=1,…J it is enough to determine the identification of F_1^* (see Thompson 1988). So, let (t_2,…,t_J) be given. Let (r_1,…,r_J) be such that (t_2,…,t_J) = $(-r_1 + r_2 + \alpha_1 - \alpha_2, ..., -r_1 + r_J + \alpha_1 - \alpha_J)$. Then, $\forall V \in W, G \in \Gamma_G$

$$\begin{aligned} F_1^*(t_2,...,t_J) &= \int F_1^*(t_2,...,t_J)\, dG(\omega) \\ &= \int F_1^*(-r_1 + r_2 + \alpha_1 - \alpha_2, ..., -r_1 + r_J + \alpha_1 - \alpha_J)\, dG(\omega) \\ &= p(1 \mid \bar{s}, \bar{x}_1, r_1, \bar{x}_2, r_2, ..., \bar{x}_J, r_J) \end{aligned}$$

where the last equality follows from Assumption 4. It follows that F_1^* is identified, since if for some F_1^* and $(t_2,...,t_J)$, $F_1(t_2,...,t_J) \neq F_1^*(t_2,...,t_J)$ then $\forall V, V', G, G'$

$$p(1 \mid \bar{s}, \bar{x}_1, r_1, \bar{x}_2, r_2, ..., \bar{x}_J, r_J; V, F, G) \neq p(1 \mid \bar{s}, \bar{x}_1, r_1, \bar{x}_2, r_2, ..., \bar{x}_J, r_J; V', F^*, G').$$

Since both $p(1 \mid \bullet; V, F, G)$ *and* $p(1 \mid \bullet_J; V', F^*, G')$ are continuous, the inequality holds on a set of positive probability. Next, assume w.l.o.g. that the alternative, $\tilde{j}$, that satisfies Assumption 5 is $\tilde{j} = 2$, $\beta_2 = 0$, and the alternative, j^* that satisfies either Assumption 6 or 6' or 6'' is j^*=1. To show that $v^*(1,\bullet)$ and G^* are identified, we transform the polychotomous choice model into a binary choice model by letting $r_j \to \infty$ for j≥3. Let $\eta \equiv \varepsilon_2 - \varepsilon_1$, and denote the marginal distribution of $\varepsilon_2 - \varepsilon_1$ by F_η^*. Since F^* is identified, we can assume that F_η^* is known.

$\forall (s, x_1, x_3, ..., x_J, r_1, r_2) \in (S \times X_1 \times (\prod_{j=3}^J X_j) \times R_+^2)$,

$$p(1 \mid s, x_1, \tilde{x}_2, ..., x_J, r_1, r_2, ..., r_J) = \int F_\eta^*(v^*(1, s, x_1, \omega) - r_1 + r_2)\, dG^*(\omega).$$

Let $\gamma = v^*(1, s, x_1, \omega)$ and let S^* denote the distribution of γ conditional on (s,x_1). Then, $\forall (r_1, r_2) \in R_+^2$

$$p(1 \mid s, x_1, r_1, r_2) = \int F_\eta^*(v^*(1, s, x_1, \omega) - r_1 + r_2)\, dG^*(\omega)$$
$$= \int F_\eta^*(\gamma + r_1 - r_2)\, dS^*(\gamma)$$
$$= \int F_\eta^*(\gamma + r_1 - r_2)\, dS^*(\gamma)$$

It then follows by Assumption 11 and Teicher (1961) that $S^*(\cdot)$ is identified. Let $f(s,x_1)$ be the marginal pdf of (s,x_1); $f(\cdot)$ is identified since (s,x_1) is a vector of observable variables. Hence, since $S^*(\cdot)$ is the distribution of γ conditional on (s,x_1), we can identify the joint pdf, $f(\gamma, s, x_1)$, of (γ, s, x_1). By Assumptions 5-7 and 10, $\gamma = v^*(1, s, x_1, \omega)$ is a nonparametric, continuous function, that satisfies the requirements in Matzkin (2003) for identification of $\gamma = v^*(1, s, x_1, \omega)$ and the distribution of ω. Hence, $v^*(1,\cdot,\cdot,\cdot)$ *and* G^* are identified. Substituting j^* in the above argument by k, as in Assumption 8, it follows that the distribution of $\gamma^k = v^*(k, s, x_k, \omega)$ conditional on (s, x_k) is identified. Since by Assumption 8 the nonparametric function $v^*(k, s, x_k, \omega)$ is known to be either strictly increasing or strictly decreasing in ω, and the distribution of ω has already been shown to be identified, it follows by Matzkin (2003) that $v^*(k,\cdot,\cdot,\cdot)$ is identified. Finally, since for all $j \neq \tilde{j}$, $v^*(j,\cdot,\cdot,\cdot)$ is identified, it follows by Assumption 9, using similar arguments as above that $v^*(\tilde{j},\cdot,\cdot,\cdot)$ is identified. This completes the proof of Theorem 1.

PROOF OF THEOREM 2:

We show the theorem by showing that the assumptions necessary to apply the result in Kiefer and Wolfowitz (1956) are satisfied (see also Wald 1949). For any $(V, G, F) \in (W \times \Gamma_G \times \Gamma_F)$, define

$$f(y, s, z; V, G, F) = \int \prod_{j=1}^{J} p(j \mid s, s, z; V(\cdot, \omega), F)^{y_j^i}\, dG(\omega)$$

and for any $\rho > 0$, define the function $f'(y, s, z; V, G, F, \rho)$ by

$$f'(y, s, z; V, G, F, \rho) = \sup_{d[(V,G,F),(V',G',F')] < \rho} f(y, s, z; V', G', F').$$

We need to show continuity, measurability, integrability, and identification.

CONTINUITY: $\forall \{(V_k, F_k, G_k)\}_{k=1}^{\infty}, (V, F, G)$ *such that*

$\{(V_k, F_k, G_k)\}_{k=1}^{\infty} \subset (W \times \Gamma_F \times \Gamma_G), (V, F, G) \in (W \times \Gamma_F \times \Gamma_G)$ *and* $d[(V_k, F_k, G_k), (V, F, G)] \to 0$, *one has that* $\forall (y, s, z)$, *except perhaps on a set of probability 0,* $f(y, s, z; V_k, F_k, G_k) \to f(y, s, z; V, F, G)$.

MEASURABILITY: $\forall (V, F, G) \in (W \times \Gamma_F \times \Gamma_G)$ *and* $\forall \rho > 0$, $f'(y, s, z; V, G, F, \rho)$ *is a measurable function of* (y, z).

INTEGRABILITY: $\forall (V, F, G) \in (W \times \Gamma_F \times \Gamma_G)$ $\lim_{\rho \to 0} E\left[\log\left(\frac{f'(y,s,z;V,G,F,\rho)}{f(y,s,z;V^*,G^*,F^*)}\right)\right]^+ < \infty$

IDENTIFICATION: $\forall (V, F, G) \in (W \times \Gamma_F \times \Gamma_G)$ *such that* $(V, F, G) \neq (V^*, F^*, G^*)$ *there exists a set* Ω *such that* $\int_\Omega f(y, s, z; V, G, F)\, d(y, z) \neq \int_\Omega f(y, s, z; V^*, G^*, F^*)\, d(y, z)$.

To show continuity, we note that $d[(V_k, F_k, G_k), (V, F, G)] \to 0$ implies that for all $(s, z, \omega) \in (S \times \prod_{j=1}^{J} (X_j \times R_+) \times R)$, and all j, $F_{j,k}(V_{j,k}(z, s, \omega))$ converges to $F_j(V_j(z, s, \omega))$, where $V_j(s, z, \omega) = (V(j, s, z_j, \omega) - V(1, s, z_1, \omega), ..., V(j, s, z_j, \omega) - V(J, s, z_J, \omega))$, and that G_k converges weakly to G over Y. Let $h_k : Y \to R$ and $h : Y \to R$ denote, respectively, the functions $F_{j,k}(V_{j,k}(s, z, \cdot))$ and $F_j(V_j(s, z, \cdot))$. Since h is continuous, it follows from Theorem 5.5 in Billingsley (1968) that $G_k h_k^{-1}$ converges weakly to Gh^{-1}. Hence, using a change of variables, it follows that

$$\begin{aligned}
&p(j \mid s, z; V_k, F_k, G_k) \\
&= \int F_{j,k}(V_{j,k}(s, z, \omega))\, dG_k(\omega) = \int h_k(\omega)\, dG_k(\omega) = \int t\; d(G_k h_k^{-1})(t) \\
&\to \int t\; d(Gh^{-1})(t) = \int h(\omega)\, dG(\omega) = \int F_j(V_j(s, z, \omega))\, dG(\omega) \\
&= p(j \mid s, z; V, F, G)
\end{aligned}$$

where the convergence follows because t is a continuous and bounded function on the supports of $G_k h_k^{-1}$ and Gh^{-1}. Hence, it follows that $f(y, s, z; V_k, F_k, G_k) \to f(y, s, z; V, F, G)$ for all z.

To show measurability, we first note that it suffices to show that for all j, $\sup_{d[(V,G,F),(V',G',F')]<\rho} p(j\,|\,s,z;V',F',G')$ is measurable in (s,z). Now, since $(W\times\Gamma_F\times\Gamma_G)$ is a compact space, there exists a countable, dense subset of $(W\times\Gamma_F\times\Gamma_G)$. Denote this subset by $(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)$. Then,

$$\sup\left\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F')]<\rho\,,\;(V',G',F')\in(W\times\Gamma_F\times\Gamma_G)\right\}=$$

$$\sup\left\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F')]<\rho\,,\;(V',G',F')\in(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)\right\}$$

Since, suppose that the left hand side is bigger than the right hand side, then, there must exist δ>0 and $(V',G',F')\in(W\times\Gamma_F\times\Gamma_G)$ such that $\forall(V'',G'',F'')\in(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)$,

(i) $p(j\,|\,s,z;V',F',G')>\delta>p(j\,|\,s,z;V'',F'',G'')$.

But, $(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)$ is dense in $(W\times\Gamma_F\times\Gamma_G)$. Hence, there exists a sequence $\{(V_k,F_k,G_k)\}\subset(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)\}$ such that $d[(V_k,F_k,G_k),(V',F',G')]\to 0$. As it was shown in the proof of continuity, this implies that $p(j\,|\,s,z;V_k,F_k,G_k)\to p(j\,|\,s,z;V',F',G')$, which contradicts (i). Hence,

$$\sup\left\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F')]<\rho\,,\;(V',G',F')\in(W\times\Gamma_F\times\Gamma_G)\right\}=$$

$$\sup\left\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F')]<\rho\,,\;(V',G',F')\in(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)\right\}$$

Since $\forall(V',G',F')\in(\widetilde{W}\times\widetilde{\Gamma}_F\times\widetilde{\Gamma}_G)$, $p(j\,|\,s,z;V',F',G')$ is measurable in (s,z), it follows that

$$\sup\left\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F')]<\rho\,,\;(V',G',F')\in(W\times\Gamma_F\times\Gamma_G)\right\}$$

is measurable in (s,z).

To show integrability, we note that $\forall j\ \forall(V', G', F^{'})$, $p(j\,|\,s,z;V',F',G') \le 1$. Hence, ∀j, $\sup\{p(j\,|\,s,z;V',F',G')\,|\,d[(V,G,F),(V',G',F^{'})] < \rho\,,\ (V',G',F^{'}) \in (W \times \Gamma_F \times \Gamma_G)\} \le 1$. It follows then that $\forall \rho > 0$, $E\left[\log\left(\frac{f'(y,s,z,V,G,F,\rho)}{f(y,s,z,V^*,G^*,F^*)}\right)\right]^+ \le E\left[\log\left(\frac{1}{f(y,s,z,V^*,G^*,F^*)}\right)\right]^+$

$$= -\int_{(s,z)} \left[\sum_{j=1}^{J} p(j\,|\,s,z;V^*,F^*,G^*)\left(\log(p(j\,|\,s,z;V^*,F^*,G^*)\right)\right] dF(s,z)$$

Since the term in brackets is bounded, it follows that $E\left[\log\left(\frac{f'(y,s,z,V,G,F,\rho)}{f(y,s,z,V^*,G^*,F^*)}\right)\right]^+ < \infty$. Finally, identification follows from Theorem 1. Hence, it follows by Kiefer and Wolfowitz (1956) that the estimators are consistent.

PROOF OF THEOREM 3:

The properties shown in the proof of Theorem 2 imply that the log-likelihood function converges a.s. uniformly, over the compact set $(W \times \Gamma_F \times \Gamma_G)$, to a continuous function that is uniquely maximized over $(W \times \Gamma_F \times \Gamma_G)$ at (V^*, F^*, G^*) (see Newey and McFadden 1994). Since $\{(W^N \times \Gamma_F^N \times \Gamma_G^N)\}_{N=1}^{\infty}$ becomes dense in $(W \times \Gamma_F \times \Gamma_G)$ as the number of observations increases, it follows by Gallant and Nychka (1987, Theorem 0) that the estimators obtained by maximizing the log-likelihood over $(W^N \times \Gamma_F^N \times \Gamma_G^N)$ are strongly consistent.

REFERENCES

AINSILE, A., P.E. ROSSI (1998), "Similarities in Choice Behavior Across Product Categories," *Marketing Science*, 17 (2), 91-106.

ALBRIGHT, R.L., S.R. LERMAN, and C.F. MANSKI (1977) "Report on the Development of an Estimation Program for the Multinomial Probit Model." Report for the Federal Highway Administration. Cambridge Systematics, Inc.: Cambridge, Massachusetts.

ALLENBY, G.M. and P.E. Rossi (1999), "Marketing models of consumer heterogeneity," *Journal of Econometrics*, 89, 57-78.

BELL, D.R., and J.M. LATTIN (1998), "Shopping Behavior and Consumer Preferences for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17(1), 66-88.

BELL, D.R., T. HO, and C.S. TANG (1998), "Determining Where to Shop: Fixed and Variable Costs of Shopping," *Journal of Marketing Research*, XXXV (August), 352-369.

BOZDOGAN, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika* 52(5) 345-370.

BRIESCH, R.A., P.K. CHINTAGUNTA, and R.L. MATZKIN (2002), "Semiparametric Estimation of Brand Choice Behavior," *Journal of the American Statistical Association*, 97(460), Dec, 973-982.

BUCKLIN, L.P. (1971), "Retail Gravity Models and Consumer Choice: A Theoretical and Empirical critique," *Economic Geography*, 47 (October), 489-497.

COSSLETT, S. R. (1983), "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51(3), 765-782.

DAHL, G.B., (2002), "Mobility and the Return to Education : Testing a Roy Model with Multiple Markets," *Econometricia*, 70(6), Nov, 2367-2420.

DAYTON, C.M. and G.B. MCREADY (1988), "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Association*, 83(401), 173-178.

ELBADAWI, I, A.R. GALLANT, and G. SOUZA (1983), "An Elasticity Can Be Estimated Consistently Without A Priori Knowledge of Functional Form," Econometricia, 51, 6, 1731-1751.

FLINN, C.J. and J.J. HECKMAN (1982) "Models for the Analysis of Labor Force Dynamics," *Advances in Econometrics*, 1, 35-95.

GALLANT, A.R. and D.W. NYCHKA (1987) "Seminonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

GALLANT, A.R. and D.W. NYCHKA (1989) "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 5, 1091-1120.

GEWEKE, J. and M. KEANE (1997), "An Empirical Analysis of Male Income Dynamics in the PSID: 1968:1989," Research Department Staff Report 233, Federal Reserve Bank of Minneapolis.

HAUSMAN, J.A. and D. A. WISE (1978) “A Conditional Probit Model for Qualitative Choice Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences,” *Econometrica*, 46, 403-426.

HECKMAN, J.J. (1974) “The Effect of Day Care Programs on Women's Work Effort,” *Journal of Political Economy*.

______________ (1981a) “Statistical Models for Discrete Panel Data,” in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, M.I.T. Press.

______________ (1981b) “Heterogeneity and State Dependence,” in S. Rosen (ed.) *Studies in Labor Markets*, University of Chicago Press.

HECKMAN, J.J. and G.J. BORJAS (1980) “Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of heterogeneity and State Dependence,” *Economica*, 47, 247-283.

HECKMAN, J.J. and B. SINGER (1984) “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data,” *Econometrica*, 271-320.

HECKMAN, J.J. and C.R. TABER (1994) “Econometric Mixture Models and More General Models for Unobservables,” University of Chicago.

HECKMAN, J.J. and J. WALKER (1990a) “The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data,” *Econometrica*, 58(6), 235-275.

___________________________ (1990b) “Estimating Fecundability from Data on Waiting Times to First Conceptions,” *Journal of the American Statistical Association*, 84(408), 958-965.

HECKMAN, J.J. and R. WILLIS (1977) “A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women,” *Journal of Political Economy*.

HIRANO, K. (2002), “Semiparametric Bayesian Inference in Autoregressive Panel Data Models,” *Econometricia*, 70(2), Mar, 781-799.

HOCH, S.J., X. DREZE, and M.E. PURK (1994), “EDLP, Hi-Lo and Margin Arithmetic,” *Journal of Marketing*, 58(4), 16-27.

HOROWITZ, J.L. (1992) “A Smooth Maximum Score Estimator for the Binary Choice Model,” *Econometrica*, 60, 505-531.

HOROWITZ, J.L, and N.E. SAVIN (2001), “Binary Response Models: Logits, Probits and Semiparametrics,” *The Journal of Economic Perspectives*, 15(4), Autumn, 43-56.

HUFF, D.L. (1962), “A Probability Analysis of Consumer Spatial Behavior,” in William S. Decker (ed.), *Emerging Concepts in Marketing* (Chicago: American Marketing Association), 443-461.

ICHIMURA, H. and T.S. THOMPSON (1994) “Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution,” mimeo, University of Minnesota.

KAMAKURA, W.A. and G.J. RUSSELL (1989), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure,” Journal of Marketing Research, 26(4), Nov, 379-390.

KIEFER, J. and J. WOLFOWITZ (1956) “Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters,” *Annals of Mathematical Statistics*, 27, 887-906.

KLEIN, R.W. and R.H. SPADY (1993) "An Efficient Semiparametric Estimator for Discrete Choice Models," *Econometrica*, 61, 387-422.

KLEIN, R.W., R.P. SHERMAN (2002), "Shift Restrictions and Semiparametric Estimation in Ordered Response Models," *Econometricia*, 70(2), Mar, 663-691.

LANCASTER, T. (1979) "Econometric Methods for the Analysis of Unemployment," *Econometrica*, 47, 939-956.

_________ (1997), "Orthogonal Parameters and Panel Data," Brown University Department of Economics working paper 97-32.

LESZCZYC, P.T., A. SINAH, and H.J.P. TIMMERMANS (2000), "Consumer Store Choice Dynamics: An Analysis of Competitive Market Structure for Grocery Stores," *Journal of Retailing*, 76(3), 323-345.

LEWBEL, A. (2000) "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity and Instrumental Variables," *Journal of Econometrics,* 97, 145-177.

LINDSAY, B.G. (1983) "The Geometry of Mixture Likelihoods: A General Theory," *The Annals of Statistics*, 11(1), 86-94.

MANSKI, C. (1975) "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-228.

MATZKIN, R.L. (1991) "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 1315-1327.

______________(1992) "Nonparametric and Distribution-free Estimation of the Binary Choice and the Threshold Crossing Models," *Econometrica*, 60, 239-270.

______________(1993) "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58,137-168.

______________(1994) "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Vol. 4, by McFadden, D. and R. Engel (eds.).

______________ (1999a) "Nonparametric Estimation of Nonadditive Random Functions," mimeo, Northwestern University, presented at the 1999 Latin American Meeting of the Econometric Society.

______________(1999b) "Computation and Operational Properties of Nonparametric Shape Restricted Estimators," mimeo, Northwestern University.

______________ (2003) "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica.*71, 5, 1339-1375.

______________ (2004) "Unobservable Instruments," Mimeo, Northwestern University.

______________ (2005) "Identification in Nonparametric Simultaneous Equations," Mimeo, Northwestern University.

MOON, H. R. (2004), "Maximum score estimation of a nonstationary binary choice model," *Journal of Econometrics*, 122, 385-403.

NEVO, A. and I. HENDEL (2002), "Measuring the Implications of Sales and Consumer Stockpiling Behavior," Mimeo, Berkeley, CA: University of California.

NEWEY, W. and D. McFADDEN (1994) “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics,* Vol 4, edited by R.F. Engle and D.L. McFadden, Elsevier Science B.V.

PARK, B.U., R.C. Sickles, L. Simar (2007), “Semiparametric efficient estimation of dynamic panel data models,” *Journal of Econometrics*, 136, 281-301.

PINKSE, J., M.E. SLADE, C. BRETT (2002), “Spatial Price Competition: A Semiparametric Approach,” *Econometricia*, 70(3), May, 1111-1153.

RHEE, H., and D.R. BELL (2002), “The inter-store mobility of supermarket shoppers,” Journal of Retailing, 78, 225-237.

SMITH, Howard (2004), “Supermarket Choice and Supermarket Competition in Market Equilibrium,” *Review of Economic Studies*, 71, 235-263.

SPIEGELHALTER, D.J., N.G. BEST, B.P. CARLIN, and A. VAN DER LINDE (2002), “Bayesian measures of model complexity and fit,” Journal of Royal Statistical Society B, 64(4), 583-639.

TABER, C.R. (2000), “Semiparametric identification and heterogeneity in discrete choice dynamic programming models,” *Journal of Econometrics*, 96, 201-229.

TANG, C.S., D.R. BELL, and T. HO (2001), “Store Choice and Shopping Behavior: How Price Format Works,” California Management Review, 43(2), 56-74.

TEICHER, H. (1961) “Identifiability of Mixtures,” *Annals of Mathematical Statistics*.

THOMPSON, T.S. (1989) “Identification of Semiparametric Discrete Choice Models,” Discussion Paper No. 249, Center for Economic Research, University of Minnesota.

WALD, A. (1949) “A Note on the Consistency of the Maximum Likelihood Estimator,” *Annals of Mathematical Statistics*, 20, 595-601.

WANSBEEK, T., M. WEDEL, E. MEIJER (2001), “Comment on ‘Microeconometrics’ by J.A. Hausman,” *Journal of Econometrics*, 100, 89-91.

YACHEW, A. (1998), “Nonparametric Regression Techniques in Economics,” *Journal of Economic Literature*, 36(2 June), 699-721.

Table 1 – Descriptive statistics for the households

	Mean	Std Dev
Number of Households	161	
Average Monthly Spending	$233.9	85.4
Minimum Monthly Spending	$102.2	61.9
Number of Shopping Trips	184.2	83.4
Av Days Between Trips	4.6	1.8
Av Spending Per Trip	36.5	15.8
Elderly	13.7%	34.5%
Household Size	2.9	1.3
Income (,000)	$56.2	$25.0
College	41.0%	49.3%
Married	82.0%	38.5%

Table 2 – Price Format Statistics.

	EDLP	HiLo
Share of Trips	57.8%	42.2%
	35.5	12.4
Distance (minutes)	(22.6)	(10.7)
	34.3	37.2
Cost to refill inventory	(10.9)	(11.0)
Loyalty (percent of trips to format	0.57	0.43
during initialization period)	(0.30)	(0.30)

Table 3 – Estimation results for parametric models.

	Segments				
	Zero	One	Two	Three	Four
In-Sample					
-Log-Likelihood	9605	6398	6122	5904	5883
(std dev)[1]		(15)	(32)	(70)	(93)
Hits		10846	11032	11172	11167
(std dev)[1]		(14)	(47)	(58)	(72)
Hit Rate		78%	80%	81%	81%
No Parameters	0	9	19	29	39
AIC	19210	12814	12282	11866	11845
Swartz	19210	12882	12425	12085	12138
DIC		12845	12350	12123	11975
Effective DOF[2]		24	53	157	104
Out of Sample					
-Log-Likelihood	4556	3169	3010	2907	2968
(std dev)[1]		(11)	(24)	(30)	(58)
Hits		5032	5165	5219	5215
(std dev)[1]		(15)	(31)	(34)	(45)
Hit Rate		77%	79%	79%	79%

Notes: 1. Standard deviations calculated using 25 bootstrap simulations.

2. Effective Degrees of Freedom is defined as the difference between the mean of the bootstrap simulation liklihoods and the estimation sample likelihood Spiegelhalter, et al (2002) .

Table 4 – Semiparametric estimation results.

	Segments		
	One[1]	Two	Three
In-Sample			
-Log-Likelihood	6185	5816	5676
(std dev)[2]	(22)	(64)	(30)
Hits	10956	11112	11281
(std dev)[2]	(30)	(77)	(36)
Hit Rate	79%	80%	81%
No Parameters	6	14	22
DIC	12489	11896	11518
Effective DOF[3]	59	132	83
Out of Sample			
-Log-Likelihood	3224	2993	2892
(std dev)[2]	(56)	(50)	(20)
Hits	5093	5248	5270
(std dev)[2]	(24)	(33)	(26)
Hit rate	77%	80%	80%

Notes: 1. The one segment model is over-identified as the extant literature (Matzkin (1992), Briesch, Chintagunta and Matzkin (2002)) show that only one extra point and no parameter restrictions are required for identification of this model. We include this model for completeness.

2. Standard deviations calculated using 25 bootstrap simulations.

3. Effective Degrees of Freedom is defined as the difference between the mean of the bootstrap simulation liklihoods and the estimation sample likelihood Spiegelhalter, et al (2002).

Table 5 – MLE parameter estimates for three-segment model.

	Parametric			Semiparametric		
Coefficients	Segment 1	Segment 2	Segment 3	Segment 1	Segment 2	Segment 3
Mass	**1.27**	0.12	0.00	0.00	-0.26	**-1.50**
	(0.24)	(0.29)			(0.18)	(0.34)
Intercept	**-0.20**	**3.23**	**-0.74**	**1.45**	-0.01	-2.93
	(0.07)	(0.48)	(0.15)	(0.12)	(0.07)	(1.52)
Loyalty	**2.61**	**3.31**	**-0.95**	**3.21**	**3.12**	**1.90**
	(0.03)	(0.15)	(0.19)	(0.07)	(0.04)	(0.31)
Elderly	**-0.32**	**0.84**	**-4.51**	**-2.18**	**1.28**	**3.13**
	(0.05)	(0.19)	(0.29)	(0.09)	(0.09)	(1.17)
HH Size	**0.05**	**0.38**	**0.28**	0.02	**-0.14**	0.58
	(0.01)	(0.12)	(0.02)	(0.02)	(0.02)	(0.31)
Income	**0.002**	**0.008**	**0.063**	**-0.007**	**-0.003**	**0.029**
	(0.001)	(0.003)	(0.004)	(0.002)	(0.001)	(0.011)
College Educated	0.06	1.56	**-4.01**	**0.07**	**0.10**	**-0.92**
	(0.03)	(7.92)	(0.26)	(0.07)	(0.05)	(0.40)
Distance	**-0.20**	**-0.85**	**-3.82**			
	(0.01)	(0.06)	(0.21)			
Cost	**-0.95**	**-2.33**	**-0.51**			
	(0.19)	(0.82)	(0.38)			
EDLP * Days since last trip	-0.01	***-0.07***	**-0.14**	-0.013	-0.013	-0.013
	(0.01)	(0.02)	(0.02)			
Theta				1.000	**2.09**	**0.53**
					(0.01)	(0.02)

Notes: 1. One segment's mass point set to zero for identification.

2. EDLP*Days since last trip set to -0.013 for identification in semiparametric model.

3. One semiparametric theta constrained to one.

4. Segments are matched based upon mass values.

5. Bold values are significant at $p < 0.05$.

Table 6 – Cost and Distance Elasticities by Segment and Model

	Percent of Families	Cost Elasticity	Distance Elasticity
Segment 1			
Parametric	22.36	-1.19	-0.57
		(1.60)	(0.45)
Semiparametric	38.51	-2.52	-0.21
		(0.35)	(0.06)
Segment 2			
Parametric	62.11	-0.96	-0.12
		(1.36)	(0.45)
Semiparametric	52.8	-0.53	-0.22
		(0.04)	(0.10)
Segment 3			
Parametric	15.53	-0.48	-1.39
		(0.98)	(0.32)
Semiparametric	8.7	-0.22	-0.19
		(0.09)	(0.07)

Table 7 – Demographic profiles of segments.

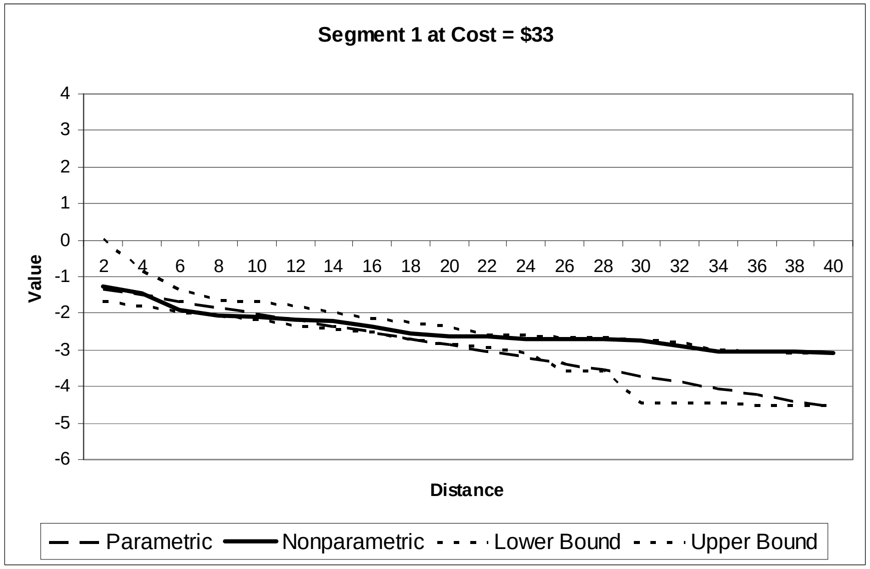
	Parametric			Semiparametric		
	Segment 1	Segment 2	Segment 3	Segment 1	Segment 2	Segment 3
Size (# families)	36	100	25	62	85	14
% Trips at EDLP	**82%** [a]	55% [b]	43% [c]	**63%** [a]	58% [a,b]	44% [b]
	(0.28)	(0.31)	(0.30)	(0.31)	(0.35)	(0.25)
Av Days between trips	5.57 [a]	5.20 [a]	4.28 [b]	4.86 [b]	5.46 [a]	4.41 [b]
	(2.20)	(1.75)	(1.79)	(2.01)	(1.73)	(1.99)
COST-Average Cost at selected format ($/10)	-0.07	0.18	0.10	0.14	0.11	0.00
	(0.80)	(0.93)	(0.71)	(0.83)	(0.85)	(1.17)
Distance to selected format/10	**3.28** [a]	***1.82*** [b]	1.38 [c]	**1.98**	***2.19***	1.85
	(2.32)	(1.07)	(0.91)	(1.44)	(1.70)	(1.20)
Income/10,000	*48.96* [b]	*59.22* [a]	*54.44* [a,b]	*50.23* [b]	*58.51* [a]	*68.39* [a]
	(19.89)	(26.10)	(25.37)	(26.83)	(21.46)	(30.72)
Family Size	2.67	2.85	3.20	2.82	2.93	2.64
	(1.17)	(1.29)	(1.63)	(1.34)	(1.33)	(1.28)
Elderly	14%	13%	16%	18%	11%	14%
	(0.35)	(0.34)	(0.37)	(0.39)	(0.31)	(0.36)
College Educated	**19%** [b]	48% [a]	44% [a]	**47%**	38%	36%
	(0.40)	(0.50)	(0.51)	(0.50)	(0.49)	(0.50)
Av. Trips per family	79.03 [b]	82.36 [b]	111.04 [a]	91.77 [a]	77.86 [b]	110.64 [a]
	(34.91)	(35.57)	(53.86)	(43.46)	(30.63)	(59.55)

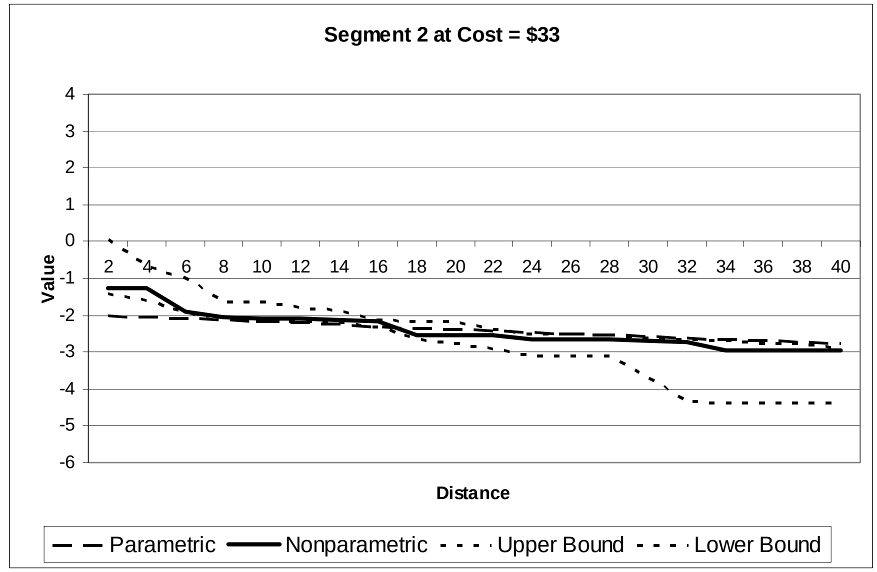
Notes: 1) Bold implies comparisons (parametric vs. semiparametric) are significant at $p < 0.05$.

2) Bold and Italic comparisons (parametric vs. semiparametric) are significant at $p<0.10$.

3) a>b>c in paired comparisons at $p < 0.10$ (within parametric or semiparametric)

FIGURE 1. Segment Distance Response Surface at Mean Cost .





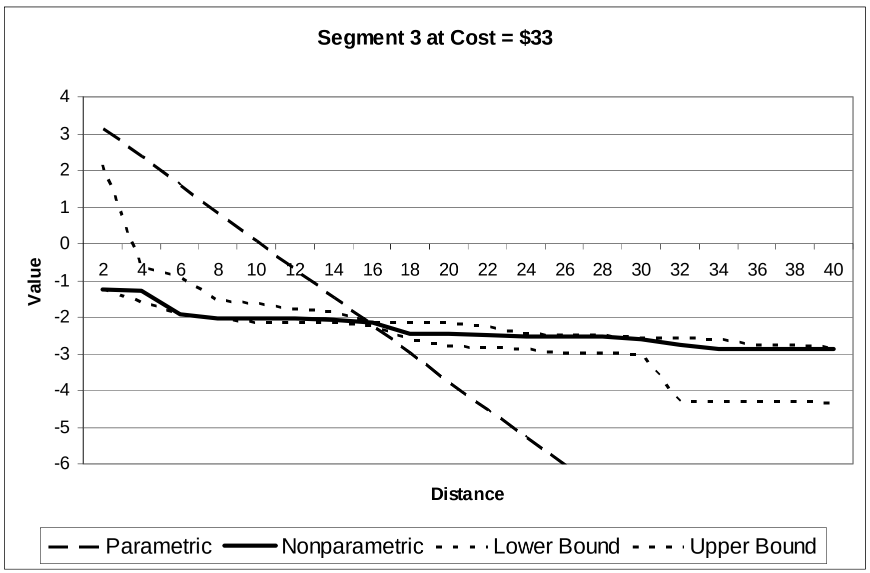


FIGURE 2. Segment Cost Response Surface at Mean Distance.