Demand Models for Market Level Data.*

By
Peter Davis †
MIT Sloan

First version: October 2000
This version: October 2001.
Preliminary and Incomplete Draft

J.E.L. Classification: D0, D11, D12, C51
Keywords: Discrete- and Continuous- Choice Demand Models, Flexible Functional Forms, Product Entry and Exit.

The models resulting from the continuous- and discrete-choice demand literatures have very distinct advantages for taking to market level data. In this paper I propose new continuous and discrete choice demand models. The continuous choice models are more appropriate for disaggregated data than popular demand models such as the Translog or almost ideal demand system since they can be estimated even when products enter or exit the market during the sample period. Variation in the observed set of products can then be used to help identify substitution patterns, in a way recently made popular in the discrete choice demand literature. Using the results provided by McFadden (1981), I then propose a discrete-choice demand model that is consistent with an underlying discrete choice random utility model. I demonstrate that the model provides a flexible functional form in the sense Diewert (1974). Approaching aggregate models of discrete choice data in this way is computationally much less demanding than using a random coefficient approach since the model does not require the use of simulation estimators.

*Thanks are due to Steve Berry, Arthur Lewbel, Ariel Pakes, Scott Stern, Tom Stoker and seminar participants at MIT, Yale, Harvard and Boston University for very helpful comments and suggestions.

†Mailing address: Assistant Professor of Applied Economics, E52-449, MIT Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142. Tel: (617) 253-1277, fax: (617) 258-6855, e-mail: pjdavis@mit.edu, Web: http://web.mit.edu/pjdavis/www

1 Introduction

Estimating market level demand systems is one of the most popular activities for empirically oriented micro-economists. Two rich methodological literatures emphasizing continuous- and discrete-choice models respectively guide empirical practice. Each class of models is well refined, but stark differences between the properties of existing discrete and continuous choice models remain. These differences primarily reflect the literatures disparate historical arenas of application and suggest productive avenues for further development of these tools, avenues I explore in this paper.

To illustrate the differences between these classes of models, consider six facts. First, representative agent continuous-choice demand models are rich enough in parameters that they are flexible functional forms in the sense of Diewert (1974). In contrast, existing discrete-choice models resort to introducing unobserved consumer heterogeneity through random coefficients inorder to provide market level demand models with the ability to match the rich substitution patterns observed in most datasets.

Second, discrete choice demand models like the market level logit model can be estimated using datasets where significant product entry and exit occurs (see Berry, Levinsohn, and Pakes (1995) for a recent example.) This is not true of popular continuous-choice models like the Translog or Almost Ideal Demand System (see Christensen, Jorgenson, and Lau (1975) and Deaton and Muellbauer (1980) respectively.) As a result, existing applications of continuous-choice models are largely limited to considering substitution patterns between broad aggregates of goods (eg., food and transportation,) a level of data aggregation which eliminates product entry and exit. This substantively limits application of these techniques in many areas of both marketing and industrial organization; resorting to an analysis of aggregate data clearly limits our ability to describe the substitution patterns between the goods actually being purchased by consumers.[1]

[1] The few exceptions to this general rule have involved market level data with some very special characteristics. For example, Hausman (1994) and Ellison, Cockburn, Griliches, and Hausman (1997) each estimate variants

In addition, the discrete choice literature (again, see for example Berry, Levinsohn, and Pakes (1995),) has shown that variation in the set of choices available to consumers can provide important information about the substitutability of products (altering the set of choices available to consumers and observing how demand for each product changes, provides direct evidence on the manner in which consumers substitute between products.) This is provides a useful source of pseudo-price variation which is unused in the present generation of continuous choice models.

Third, existing continuous choice models are much simpler and faster to estimate than random coefficient discrete choice models because estimation does not require simulation over heterogeneous consumer types.[2]

Fourth, discrete choice models are usually estimated when consumer's preferences are defined over product characteristics, rather than products directly. This substantially reduces the number of parameters to be estimated. However, one can easily imagine introducing product characteristics into continuous choice demand models (see Pinkse, Slade, and Brett (1997) for a rare example.)

Fifth, existing parametric continuous choice models add an error term on to the demand (or expenditure) system in an essentially add-hoc manner. This is in contrast to the recent discrete choice literature where, since Berry (1994), the error term is included explicitly in the direct utility function. Brown and Walker (1989) show in the continuous choice demand literature that adding an error to the estimating equations will introduce correlations between regressors (prices and income) and the error term whenever the true data generating process

of the Almost Ideal Demand System proposed by Deaton and Muellbauer (1980). However, in both cases, the full demand system can be estimated only by using data from time periods when *all* goods are present in the market. Naturally, in dynamic markets with large numbers of products this is frequently not an option since products enter and exit simultaneously.

[2]The purported advantage of introducing consumer heterogeneity in the discrete choice literature is the added flexibility in substitution patterns that the model can accommodate parsimoniously. For example, McFadden and Train (1998) show that the mixed multinomial logit model can approximate arbitrary substitution patterns between goods. This heterogeneity however does introduce substantial disadvantages. In particular, estimation typically requires simulation of multi-dimensional integrals and is therefore computationally intensive, while establishing the asymptotic properties of the resulting simulation estimators requires substantially more sophisticated mathematical arguments than those required to establish standard asymptotic results (see Pakes and Pollard (1989) and McFadden (1989).)

satisfies the restrictions of choice theory, specifically slutsky symmetry.

Sixth, in contrast to the indirect utility function approach favored in the continuous choice literature, discrete choice demand systems are universally specified by using a parametric model for the *direct* utility function.[3] These two starkly different approaches persist in part because an exact equivalence between specifying an indirect and a direct utility function, provided by duality theorems for continuous choice models, is *not* always available in the discrete choice case. However, Williams (1977), Daly and Zachary (1979), and McFadden (1981) do provide a fundamental result that allows an approach to discrete choice demand modelling which is entirely analogous to the continuous choice indirect utility function approach for the subclass of direct utility functions that are additive in *some* product characteristic. While this result applies to only sub-class of discrete choice models, it does include the set of models with an additively separable unobserved product characteristic that have dominated empirical practice since they were introduced by Berry (1994) and Berry, Levinsohn, and Pakes (1995). Consequently, in section 5, I explore both the direct utility specification approach and also an indirect approach to generating discrete choice demand models.

The aim of this paper then, is to develop a discrete-choice demand model and a separate but closely related continuous-choice demand model which each have distinct advantages over the models currently in use. Specifically, each model: (i) provides a flexible functional form in the sense of Diewert (1974) (ii) can accommodate and utilize data on the entry and exit of products, (iii) is relatively simple and fast to estimate because it does not require estimation via simulation (iv) may be estimated when consumer's preferences are defined over product characteristics, rather than products directly and (v) which incorporates the error term as an integral part of the model specification, thereby avoiding the critique provided by Brown and Walker (1989).

In drawing out some common features of these two literatures, I build upon McFadden

[3]Sometimes these in fact are termed 'conditional indirect utility functions' because they are conditional on choice j but in general may have already involved maximization over a set of continuous choices. However, in the pure discrete choice context this object is literally just the direct utility function with the budget constraint substituted in for the outside good (see below.)

(1981) and Anderson, de Palma, and Thisse (1992) who emphasize that a continuum of consumers making discrete choices will in fact, under some circumstances, generate an observationally equivalent demand system as a single "representative" consumer making continuous choices. My aim is *not* to show a variant of their representative consumer result. Instead, I attempt to reconcile the currently stark differences in empirical practice and model properties between the two literatures. To do so, I develop a class of demand generating functions (I shall make this term precise shortly) and identify the *different* conditions under which these functions are (i) indirect utility functions and therefore generate continuous choice demand systems via Roy's identity and (ii) are consistent with an underlying discrete choice model.[4]

The rest of the paper is as follows. In section 2, I briefly summarize the existing approaches in the demand literature and introduce the notation used throughout the paper. In section 3 I briefly introduce the demand system generating function. In section 4, I develop a class of parametric continuous choice demand models that can be used with data where we see product entry and exit. I establish that the model is a flexible functional form in the sense of Diewert (1974). In section 5, I apply the results provided by Mcfadden (1978) for the Generalized Extreme Value (GEV) model to build a model which is a member of the class of GEV models and is capable of providing a flexible description of substitution patterns without resorting to the introduction of consumer heterogeneity to rationalize market level data. Next I use the results provided by McFadden (1981) to argue that the class of discrete choice models with an additive unobserved product characteristic, introduced by Berry (1994), can be studied using an indirect approach which is analogous to the indirect utility derivation of demand models preferred by authors in the continuous choice literature. Doing so, provides a demand system which can be shown to be consistent with an underlying discrete choice model and has desirable properties without explicit integration. In section 6, I demonstrate a practical and fast ways to estimate the respective discrete and continuous choice models. In section 7 I show how

[4]Following McFadden (1981) more directly, I also provide sufficient conditions for (a sub-class of) these models to generate demand systems that are consistent with either a distribution of consumers each making discrete choices, or a single consumer making continuous choices.

product characteristics may be introduced into the models in a way which allows cross price elasticities of demand to depend directly on the distance between products in characteristics space. In section 8, I demonstrate how to introduce consumer heterogeneity into the models and then finally conclude.

2 Previous Literatures and some Notation

I consider a class of random utility models wherein consumers are endowed with preferences and solve a utility maximization problem subject to a budget constraint

$$V(\mathrm{p}, y, \delta; \theta) = \max_{x \in X} u(x, \delta; \theta) \text{ s.t. } \mathrm{p}' x \leq y.$$

where p denotes the vector of prices, y denotes the consumers' income, and δ represents factors that affect the consumers utility that are unobserved to the econometrician. The solution to this problem is a vector of demand equations for each product, $x(\mathrm{p}, y, \delta; \theta)$. Standard duality results establish conditions on the function $V(p, y, \delta; \theta)$ which ensure that specifying a parametric functional form for the indirect utility function, and then solving for the demand system using Roy's identity, is entirely equivalent to specifying the direct utility function and budget constraint.[5] By taking this dual approach, the resulting parametric demand systems are assured to be consistent with utility maximization, at least for some subset of parameter values. Moreover, since an empirical model can be generated by writing down a polynomial in the indirect utility function's arguments, the resulting demand systems are capable of generating flexible substitution patterns while successfully avoiding the explicit solution to the non-linear direct utility maximization problem. Unfortunately, as I will show, the literature has emphasized choices for the parametric form of the indirect utility function that result in

[5]See Varian (1984) for example. For any given indirect utility function which is (i) continuous in all p $>> 0$ and $y > 0$, (ii) $v(\mathrm{p}, y)$ is non-increasing in p and non-decreasing in y, (iii) $v(\mathrm{p}, y)$ is quasi-convex in (p, y) with any one element of the vector normalized to one, and (iv) homogeneous of degree zero in (p, y) there exists a direct utility function $u(x)$ which represents the same preference ordering over goods. See for example Mas-Colell, Whinston, and Green (1995) pages 24,56, and 77.

continuous choice demand models that have extremely undesirable properties for the kinds of disaggregated data-sets increasingly available in marketing, industrial organization and many other applied fields.

Discrete choice models are easily derived from the same framework by introducing additional constraints into the utility maximization problem. Specifically, if a discrete choice must be made from the set of products indexed by $\mathcal{J} = \{1, .., J\}$ where the 1^{st} option represents the choice to consume the continuous 'outside' option, then these additional constraints can be represented by enforcing $x_j x_k = 0$ for all $j \neq k$ and $x_j \in \{0, 1\}$ where $j, k \in \mathcal{J}/\infty$ (See for example Small and Rosen (1981).) In that case, the random utility model reduces to the problem

$$\max_{j \in \mathcal{J}} v_j(y - p_j \mathbb{I}(j > 1), p_1, \delta; \theta).$$

where, $v_j \equiv u((x_1, 0, .., 0, x_j, 0, ..0), \delta; \theta)$, for $j > 1$ $v_1 \equiv u((x_1, 0, 0, ..0), \delta; \theta)$. In either case, $x_1 = \frac{y - p_j(j>1)}{p_1}$ is obtained from the budget constraint, and $x_j = 1$ if choice $j > 1$ is picked. Popular discrete choice models include the Multinomial Logit and Probit models where $v_{ij} = u_i((x_1, 0, .., 0, x_j, 0, ..0), \delta; \theta) = v_j(y - p_j \mathbb{I}(j > 1), p_1, w_j, \theta) + \epsilon_{ij}$ and $\epsilon_i = (\epsilon_{i1}, .., \epsilon_{iJ})$ is assumed to have a multivariate Type 1 extreme value or a Normal distribution respectively. In both cases, each consumer i is assumed to know her own type, ϵ_i but it is unobserved by the econometrician and assumed to be identically and independently distributed across individuals.

3 The Demand System Generating Function

Throughout the paper I focus on demand system generating functions of the form

$$V(p, y, \delta, \theta) \equiv d \ln H(r_1, .., r_J; \cdot) \tag{1}$$

where $H(\cdot)$ is a parametric function, θ and $d > 0$ are parameters and the arguments of the function H are (possibly parametric) functions, $r_j = \exp\{\psi_j(y, p, \delta_j)\}$. Although many functional forms are possible, for concreteness, I consider the simplest quadratic form for the function

$H(\cdot)$ as a specific example throughout the paper.

$$H(r_1, ..., r_J) \equiv \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J}(1(j=k)+b_{jk})r_j r_k = \frac{1}{2}r'(I+B)r \tag{2}$$

where B is a JxJ matrix with jk^{th} element b_{jk}, $r = (r_1, .., r_J)$, and $\psi_j()$ is a known (possibly parametric) function. Again, for concreteness, specifications of ψ that will be of particular interest include

$$\psi_j(y, p_j, p_0, \delta_j) = \frac{y - p_j}{p_0} - \delta_j, \text{ and } \quad \psi_j(y, p_j, p_0, \delta_j) = lny - lnp_j - \delta_j.$$

If the resulting demand generating function, $V()$, has the properties of an indirect utility function, applying Roy's identity provides a parametric continuous choice demand system, $x(p, y, \cdot) = -\frac{\partial V}{\partial p_j}/\frac{\partial V}{\partial y}$. In contrast, if $V(\cdot)$ has the properties of an expected additive random utility function (precisely what this means will be made explicit below) then the results provided by Williams (1977), Daly and Zachary (1979), and McFadden (1981) imply that a discrete choice demand system can be generated using the identity, $x_j(p, y, \cdot) = -\frac{\partial V}{\partial \delta_j}$ for $j = 1, .., J$. Remarkably, those authors show that for a subset of specifications, these demand systems will be identical and could therefore have been generated by either a single consumer making continuous choices or some distribution of consumers making discrete choices.

4 Consistent Continuous Choice Demand Systems

If $V(\mathrm{p}, y, \delta; \theta)$ has the properties of an indirect utility function then standard duality results imply that the demand system is easily obtained via Roy's identity.[6] In this paper, I restrict the class of functions that I consider further by requiring the indirect utility function satisfies an additional global regularity property. Specifically, if p_{-j} denotes the vector of prices excluding

[6]Recall that a function $V(\mathrm{p}, y, \cdots)$ is defined to be an indirect utility function (IU) if it is continuous at all $\mathrm{p} >> 0$, $y > 0$, non-increasing in p and non-decreasing in y, homogeneous of degree zero in (p, y), and quasi-convex in (p, y) with any one element in the vector normalized to one.

p_j, I consider only the sub-class of indirect utility functions that are globally consistent.

Definition: *An indirect utility function is globally* ***consistent*** *iff*

$$\lim_{p_j \to +\infty} V(p_j, \mathrm{p}_{-j}, y, \delta; \theta) = V(p_{-j}, y, \delta; \theta).$$

for every (y, δ, θ) *whenever* $\exists\ k \neq j$ *with* $p_k < \infty$.

Lemma 1 *Any indirect utility function which satisfies consistency and non-satiation (*$\frac{\partial V}{\partial y} \neq 0$*) generates a demand function via Roy's identity which enjoys the property that*

$$\lim_{p_j \to +\infty} x_k(p_j, p_{-j}, \cdot) = \lim_{p_j \to +\infty} \frac{\frac{\partial V}{\partial p_k}}{\frac{\partial V}{\partial y}} = \begin{cases} x_k(p_{-j}, \cdot) & \textit{for all } k \neq j \\ 0 & \textit{otherwise.} \end{cases}$$

The advantage of restricting ourselves to the class of consistent indirect utility functions to generate continuous choice demand models us that, within this class, removing a good from the choice set (which explicitly forces the level of demand to zero), is entirely equivalent to increasing its price to infinity. Surprisingly, this extremely mild and intuitive regularity condition is *not* satisfied by the vast majority of existent continuous choice models such as the Translog or the Almost Ideal Demand System (AIDS). For example, the Translog has an indirect utility function which has many terms like $\alpha_j ln p_j$ and $\beta_{jk} ln p_j ln p_k$. As a result, the Translog and AIDS also have very poor properties as the amount of price, or pseudo-price, variation is large.[7]

Following Hausman (1994), a typical response to this problem in datasets with product entry and exit has been to use disaggregated data to learn about the parameters of the model only from the period when *all* goods are observed in the market. While this approach is

[7] While the vast majority of indirect utility function specifications used to generate continuous choice demand models are *not* members of the set of consistent indirect utility functions, a very few existing demand systems are. These are generally models which have not been empirically popular. For example, the "indirect addilog" model considered by Houthakker (1960) sets $r_j(y, p_j) = \frac{(y/p_j)^{\beta_j + 1}}{\beta_j + 1}$, where β_j are parameters. Other examples of models which are *not* in the set of models with consistent indirect utility functions include the Translog Reciprocal Indirect Utiltity Function and Diewert's Reciprocal Indirect Utility Function. See for example Varian (1984) for a discussion of these models and further references.

effective (if potentially inefficient) in some markets, such as the pharmaceutical markets studied by Ellison, Cockburn, Griliches, and Hausman (1997) where generic entry is driven by loss of patent protection so all entry occurs within a very constrained period in the data, in other arena's product entry and exit occur simultaneously making even that approach largely impractical.

In Propostion 1 below, I provide a set of relatively easy to verify conditions on the functions $H(r;\theta)$ and $r(p,y,\delta)$ that are sufficient to ensure that the resulting indirect utility function is a member of the class of consistent indirect utility functions and may therefore be estimated using pseudo price variation. As I have already described, in the discrete choice literature this source of pseudo-price variation has proven extremely useful for identifying rich substitution patterns. To solidify ideas, consider the model where $H(r;B) = \frac{1}{2}r'(I+B)r$ and $r_j = \frac{y}{p_j}\exp\{-\delta_j\}$. After stating Proposition 1, I show that this particular model specification provides a flexible functional form in the sense of Diewert (1974) and it is also a member of the class of consistent indirect utility functions provided the matrix B satisfies conditions which ensure $H(r;B)$ is a convex function.

Proposition 1 *Let $V(p,y,\delta;\theta) = \frac{1}{m}\ln H(r;\theta)$ and $r_j = r_j(y,p_j,\delta_j)$ where $H(r;\theta)$ is a continuous, convex, non-decreasing, and homogeneous of degree m function of r with $\lim_{r_j\to 0} H(r;\theta) = H(r_{-j};\theta)$. If*

1. *$r_j(p_j,y,\delta_j)$ is a continuous function at all positive prices and incomes.*
2. *$r_j(p_j,y,\delta_j)$ is non-increasing in p_j, non-decreasing in y, and homogeneous degree zero in (p_j,y)*
3. *$r_j(p_j,y,\delta_j)$ is a convex function of p_j, with y normalized to 1, and*
4. *$\lim_{p_j\to\infty} r_j(y,p_j,\delta_j) = 0$*

then, $V(\mathrm{p},y,\delta;\theta) = \frac{1}{m}ln(H(r(\mathrm{p},y,\delta),\theta))$ is a consistent indirect utility function.

Proof Omitted.

An algebraic functional form for a complete system of consumer demand functions, $x(p,y,\delta,\theta)$ is said to be flexible if, at any given set of non-negative prices of commodities and income, the parameters (δ, θ) can be chosen so that the complete system of consumer demand functions, their own- and cross-price and income elasticities are capable of assuming arbitrary values at the given set of prices and commodities and income subject only to the requirements of theoretical consistency. (See Diewert (1974) or Lau (1986).) Next I show that the proposed particular functional form for the consistent continuous-choice demand model is a flexible functional forms.

Proposition 2 *Flexibility of the Continuous Choice Model. Consider the model generated from equation 1 with* $H(r) = \frac{1}{2} r'(I+B)r$*, where* $r_j = \exp\{\ln y - \ln p_j - \delta_j\}$*,* I *is the identity matrix, and* B *is a symmetric matrix of parameters. This model can match any observed vector of market shares, matrix of cross price elasticities, and vector of income elasticities provided the data generating process satisfies homogeneity, additivity, and Slutsky symmetry.*

Proof 1 *(See Appendix.)*

5 Flexible Random Utility Discrete Choice Models

The aim of this section is to develop discrete choice models that can match rich substitution patterns without requiring simulation. I divide the discussion into two cases. First, a direct utility specification approach and then an indirect approach analogous to the indirect utility function approach used in the continuous choice literature.

5.1 A Direct Utility Flexible GEV Specification

Mcfadden (1978) defines the class of generalized extreme value models by first describing a class of functions $H(r)$,

Assumption: *Generalized Extreme Value Properties*

Suppose $H(r_1, ..., r_J; \cdot)$ *has the following properties :*

1. $H(r)$ *is a non-negative, homogeneous of degree* m *function of* $(r_1, ..., r_J) \geq 0$.
2. *Suppose for any distinct* $(j_1, ..., j_k)$ *from* $\{1, ..., J\}$, $\frac{\partial^k H}{\partial r_{j_1}, ..., \partial r_{j_k}}$ *is non-negative if* k *is odd and non-positive if* k *is even.*
3. *Let* $\mathcal{J} = \{i_1, ..., i_J\}$ *and* $\mathcal{J}' = \{i'_1, ..i'_J, ..., i'_{J'}\}$. *If* $r_{i_j} = r_{i'_j}$ *for* $j = 1, ..., J$ *then* $H((r_1, .., r_J), \cdot) = H((r_1, .., r_J, 0, .., 0), \cdot)$

4. $lim_{r_j \to \infty} H(r_1, .., r_J) = +\infty$.

Mcfadden uses these properties to establish the following remarkable result:

Proposition 3 *GEV model (Very slight relaxation of Mcfadden (1978)) Suppose the utility provided by good* j *is* $u_{ij} = \overline{u}_j + \epsilon_{ij}$, *and* $(\epsilon_1, .., \epsilon_J)$ *is distributed as* $F(\epsilon_1, .., \epsilon_J) = e^{-H(e^{-\epsilon_1}, .., e^{-\epsilon_J})}$ *where* $H(r)$ *has the GEV properties described above. Then,*

$$\mathbb{P}_j = \frac{e^{\overline{u}_j} H_j(e^{\overline{u}_1}, ..., e^{\overline{u}_J})}{mH(e^{\overline{u}_1}, ..., e^{\overline{u}_J})}$$

defines a probabilistic choice model from alternatives $j = 1, .., J$, *and the expected maximum utility,*

$$V = \int_{\epsilon_1=-\infty}^{+\infty} \cdots \int_{\epsilon_J=-\infty}^{+\infty} \max_{j=1,..,J} (\overline{u}_j + \epsilon_j) f(\epsilon_1, .., \epsilon_J) d\epsilon_1 ..., d\epsilon_J$$

(with f *the density of* F*), satisfies*

$$V = \frac{1}{m} log H(e^{\overline{u}_1}, ..., e^{\overline{u}_J}) + \frac{1}{m}\gamma$$

where $\gamma = 0.5772156649...$ *is Euler's constant.*□

Proof 2 *(See Appendix.)*

These conditions provide a slight relaxation of the conditions provided by Mcfadden (1978) as sufficient for the Generalized Extreme Value Model[8] and provide the basis for all existing logit type models. For example, the standard multinomial logit model sets $H(r) = \sum_{j=1}^{J} r_j$ while the one level nested logit model with G groups assumes a specification of H that is

[8]The generalization to homogeneity of arbitrary degree m provided here analytically trivial given earlier results. Nonetheless, it does not immediately follow from Mcfadden (1978) since while a homogeneous degree m function raised to the power $1/m$ is homogeneous of degree 1, the function $H(r_1, .., r_J)^{1/m}$ will not generally satisfy the cross derivative property even if $H(r_1, .., r_J)$ does. For example, consider $H(r_1, .., r_J) = \sum_{l=1}^{J} r_l^m$. Clearly, provided $r_l \geq 0$, the first derivative property will hold for all $m \geq 0$, while all subsequent cross derivatives are zero. Now consider $H(r_1, .., r_J) = \left(\sum_{l=1}^{J} r_l^m\right)^{\frac{1}{m}}$. The first derivative is $\frac{\partial H}{\partial r_j} = \frac{1}{m}\left(\sum_{l=1}^{J} r_l^m\right)^{\frac{1}{m}-1} r_j^{m-1}$ while the second cross derivative is $\frac{\partial^2 H}{\partial r_j \partial r_k} = m\left(\frac{1}{m} - 1\right)\left(\sum_{l=1}^{J} r_l^m\right)^{\frac{1}{m}-2} r_j^{m-1} r_k^{m-1}$ which is only non-positive provided $m \leq 1$.

separable into G mutually exclusive $\{H^g\}_{g=1}^G$ functions:[9]

$$H(r;\theta) = H(H_1(r_1, ..., r_{J_1}; \theta_1), ..., H_G(r_{J_{G-1}+1}, .., r_{J_G}; \theta_G))$$

where all functions are evaluated at $r = (e^{\overline{u}_1}, .., e^{\overline{u}_J})$. It is well known that the nested models impose strong a-priori restrictions on substitution patterns and this is evidenced from the equation above since whenever $H(r) = H(H_1, .., H_g)$ is a separable function, $H_{jk} = \frac{\partial^2 H}{\partial r_j \partial r_k} = 0$ whenever j and k are not in the same nest (both arguments of at least one function H_g.)

Recently, Bresnahan, Stern, and Trajtenberg (1997) have suggested using parametric non-separable functions to form the basis of discrete choice demand models. In that way, those authors point out that the strongest implications of a nested (separable) model structure are avoided. They propose partitioning the set of choices in multiple ways $\mathcal{J} = \mathcal{J}_1^m \bigcup ... \bigcup \mathcal{J}_{G_m}^m$, for $m = 1, .., M$ and building the non-separable function

$$H(r;\theta) = H(H^1(r_{\mathcal{J}_1^1}, .., r_{\mathcal{J}_{G_1}^1}; \theta_1), ..., H^M(r_{\mathcal{J}_1^M}, .., r_{\mathcal{J}_{G_M}^M}; \theta_M)).$$

in that way, the cross derivative H_{jk} which crucially controls estimated substitution patterns is potentially non-zero for all pairs of choices j and k.

In this sub-section, I push that idea a step further in an attempt to develop a concrete empirical strategy based on the observation, explicitly noted in Pudney (1989) but subsequently apparently abandoned, that it is possible to write specifications of discrete choice models which are of a similar form to the flexible functional form specifications used in the continuous choice literature. Specifically, I propose choosing the function H to be richly enough parameterized to ensure that the second derivatives of the function can take on any appropriate value. Perhaps

[9] The most popular of these specifications is referred to as 'the' nested multinomial logit model which assumes $H^g(r_1, .., r_{J_g}; \theta_g) = \left(\sum_{j=J_{g-1}}^{J_g} r_j^{\frac{1}{\theta_g}}\right)^{\theta_g}$ for $g = 1, .., G$. Demand models with multiple levels of nesting are easily developed. For example, a two level nested model can be generated by partitioning the set $\mathcal{J} = \mathcal{J}_{11} \bigcup ..., \mathcal{J}_{MG_M}$ with generic component set $\mathcal{J}_{mg}$ and then choosing $H(r) = H(H^1, .., H^M)$ and defining each H_m $(m = 1, .., M)$ as a separable function of G_m sub-functions $H_m(r_m) = H_m(H_{m1}, .., H_{mG_m})$ where each H_{mg_m} function depends on the mutually exclusive subset of the vector r, with indexes in $\mathcal{J}_{gm}$.

the most natural example takes $H(r) = \frac{1}{2}r'(I+B)r$ so that the second derivatives are $H_{jk} = b_{jk}$ which can clearly take on any desired value. Clearly introducing a large number of parameters reduces the degrees of freedom available in the same way that flexible functional forms do in the continuous choice literature. In section 7, I show that the introduced taste parameters may be mapped down to be parametric functions of observed product characteristics. This operation is precisely the same operation as that used in the Probit model which assumes that the covariance in tastes for two products depends directly on the observed characteristics of the products. By following the prescription of the Probit model, the GEV model avoids the numerical integration required by the Probit model.

In section 4 I defined an algebraic functional form for a complete system of consumer demand functions, $x(p, y, \delta; \theta)$ to be flexible if, at any given set of non-negative prices of commodities and income, the parameters (δ, θ) can be chosen so that the complete system of consumer demand functions, their own- and cross-price and income elasticities are capable of assuming arbitrary values at the given set of prices and commodities and income subject only to the requirements of theoretical consistency. (See Diewert (1974) or Lau (1986).) Next I show that the proposed specification of the discrete choice model is a flexible functional form.[10]

Proposition Flexibility of the discrete choice model. Consider the model $H(r) = \frac{1}{2}r(I+B)r$, where $r_j = \exp\{\ln y - \ln p_j - \delta_j\}$, B is a matrix of parameters with jk^{th} element b_{jk}. This model can match any observed vector of market shares, matrix of cross price elasticities, and vector of income elasticities provided the data generating process satisfies additivity and homogeneity of degree zero in income and prices.

Proof 3 *(See Appendix.)*

[10]I chose a specification which imposes homogeneity of degree zero in income and prices by writing the model in terms of the ratio $\frac{p_j}{y}$ instead of $\frac{y - p_j 1(j>1)}{p_1}$, which is somewhat more natural in the discrete choice setting. This formulation has the advantage that it is symmetric in all prices and hence the proofs of flexibility are simpler but introduces no evident substantive disadvantage.

5.2 An Indirect Approach to Discrete Choice Models

Consider the class of 'additive' random utility discrete choice models (ARUM). That is, any random utility model in which conditional indirect utilities have a form which is additive in some (possibly composite) characteristic.[11] That is, each consumer with individual characteristics c solves the maximization problem

$$max_{j \in \mathcal{J}} v_j(y - p_j, p_0, w_j, c, \theta) - \delta_j$$

where as before w_j denotes the product characteristics observed by the consumer, y denotes income, p_0 denotes the price of the outside alternative, c denotes a vector of this consumer's characteristics, and δ_j denotes a (possibly composite) product characteristic.[12]

Define a class of functions, $\mathcal{V}$, whose members satisfy the following expected maximum random utility ($EMRU$) properties:[13]

1. For each choice set, $\mathcal{J} = \{1, ..., J\}$, $V(\cdot)$ is a real valued function of $\delta_{\mathcal{J}} \in R^J$.
2. $V(\delta_{\mathcal{J}})$ has the additive property, that $V(\delta_{\mathcal{J}} + \theta) = V(\delta_{\mathcal{J}}) - \theta$, where θ is any real scalar and $\delta_{\mathcal{J}} + \theta$ denotes a Jx1 vector with components $\delta_j + \theta$.
3. All mixed partial derivatives of V with respect to $\delta_{\mathcal{J}}$ exist, are non-positive, and independent of the order of differentiation.
4. $\lim_{\delta_j \to -\infty} V_j(\delta) = -1$ for all $j \in \mathcal{J}$.

[11] For example, in a recent series of influential papers, Berry (1994),Berry, Levinsohn, and Pakes (1995),Berry, Levinsohn, and Pakes (1997) consider models within this class where δ_j is a linear combination of product characteristics that are observed by the econometrician, w_{1j} and a product characteristic that is unobserved by the econometrician.$\delta_j = w'_{1j}\beta + \xi_j$.

[12] Note in particular that this class of additive random utility models includes most of those used in applied work, including the generalized extreme value (GEV) class of models with an additively separable characteristic such as those considered by Berry (1994), Berry, Levinsohn, and Pakes (1995), and Berry, Levinsohn, and Pakes (1997). In that case, $v_j(y - p_j, p_0, w_j, c, \theta) = v_j(y - p_j, p_0, w_j, c_1, \theta) + \epsilon_j$ and the vector of individual characteristics $\epsilon_{\mathcal{J}} \equiv (e_1, ..., e_J)$ is a component of $c = (c_1, \epsilon)$ and has a distribution across individuals which is a member of the GEV class. Other components of c_1 may be random coefficients.

[13] McFadden (1981) calls a superset of these properties the Social Surplus (SS) properties since he is interested primarily in understanding the aggregation conditions required to interpret the Expected maximum utility function as an indirect utility function for a single representative consumer. Since one of my primary purposes is to generate flexible demand systems for discrete choice situations, that need not correspond to a single representative consumer making continuous choices, I want to separate the ideas of an underlying discrete choice random utility model and the existence of a representative consumer. Thus, I prefer the name EMRU.

5. Suppose $\mathcal{J} = \{i_1, ..., i_J\}$ and $\mathcal{J}' = \{i'_1, ..., i'_J, ..., i_J\}$ satisfy $\delta_{i_k} = \delta_{i'_k}$ for $k = 1, .., J'$. Then $V(\delta_{\mathcal{J}'}) = V(\delta_{\mathcal{J}'}, +\infty, ..., +\infty)$.

The theorem provided by Williams (1977), Daly and Zachary (1979), and McFadden (1981) (henceforth denoted the WDZM theorem and a version of which is provided as Theorem 1 below) demonstrates that any additive random utility model generates an expected maximum utility function $V(\delta, \cdot) = E_{\mathbf{v}|\cdot}[max_j\{v_j(z, y - p_j, p_1, w_j, s, c, \theta) - \delta_j\}]$ that is a member of the $EMRU$ class of functions. Moreover, their remarkable result demonstrates that the converse is also true. Any member of the class of $EMRU$ functions could be generated from an additive random utility discrete choice model (ARUM).

The primary practical implication of the WDZM theorem is that demand systems that are consistent with an underlying distribution of consumers who are each making a discrete choice can be generated by specifying parametric functional forms for the expected maximum utility function directly. In particular, the class of expected maximum utility functions described by Equation (2) is in the ERUM class of functions, for a large set of values of the parameters (δ, B). I have already shown that this particular specifications of the expected utility function can generate parametric demand systems that are capable of generating *arbitrary* substitution patterns between goods and arbitrary income elasticities of demand. Whenever it is a member of the EMRU class of functions, it could be generated by an underlying distribution of consumer types each of whom make a discrete choice from the set of available options. Given an additive RUM, the Generalized Extreme Value distribution of consumer types provides a direct utility and distribution specification for a large subset of the EMRU class of functions.

Theorem 1 *(WDZM) Consider the additive random utility model (ARUM),*

$$\max_{j \in \mathcal{J}} v_j - \delta_j$$

where the dependence of v_j on the vector $(w_j, y - p_j, p_1, c, \mathcal{J})$ is left implicit for notational simplicity. Suppose that $\mathbf{v}_{\mathcal{J}}$ is distributed in the population with conditional cumulative distribution function, $F(\mathbf{v}|w, y, \mathrm{p}, c)$ and density $f(\mathbf{v}|w, y, \mathrm{p}, c)$. Then this ARUM generates a system of choice probabilities, $Pr\{j|w, y, \mathrm{p}, c, \mathcal{J}\}$, which are non-negative, sum to one, and depend only on $(w, y, \mathrm{p}, c, \mathcal{J})$ through $\mathbf{v}$. Define

$$V(\delta) = E \max_{j \in \mathcal{J}} v_j - \delta_j \tag{3}$$

where expectations are taken with respect to $\mathbf{v}_{\mathcal{J}}$*. Then, provided* $F(\mathbf{v};\cdot)$ *has a first moment,* V *exists and satisfies the properties EMRU. Moreover,*

$$Pr\{j|w, y, \mathrm{p}, c, \mathcal{J}\} = -\frac{\partial V(\cdot)}{\partial \delta_j}. \tag{4}$$

Converse: *Suppose that* $V(\delta; w, \mathrm{p}, y, c)$ *is any function with the EMRU properties. Then Equation (4) defines a probability choice system. Further, there exists an ARUM form such that* $V(\delta)$ *could be generated by Equation (3).*

Thus a valid method of generating demand systems which are consistent with an underlying discrete choice model of demand is to follow the approach typically preferred in the classical literature: specify a flexible form for the function $V(\cdot)$ such that it satisfies EMRU.

While establishing that a function is in the EMRU class of functions is sufficient for it to be consistent with an additive random utility model, the cross derivative condition that is required to establish the existence of a density of consumer types is often non-trivial to establish. Thus, following Mcfadden (1978), in Proposition 4 I provide a general set of sufficient conditions for a class of functions of particular interest that are typically much easier for the researcher to verify. These sufficient conditions on the function $H(r)$ are a strict subset of the GEV model assumptions, at the cost of the additivity assumption on the function $r_j = exp\{\phi_j(p_j, y) - \delta_j\}$.

Proposition 4 *A Set of Sufficient Conditions for ERUM*

Suppose $r_j = e^{-\delta_j}$*,* $j = 1, .., J$ *and* $H(r_1, ..., r_J; \cdot)$ *has the following properties :*

1. $H(r)$ *is a non-negative, homogeneous of degree* m *function of* $(r_1, ..., r_J) \geq 0$*.*
2. *Suppose for any distinct* $(j_1, ..., j_k)$ *from* $\{1, ..., J\}$*,* $\frac{\partial^k H}{\partial r_{j_1}, ..., \partial r_{j_k}}$ *is non-negative if* k *is odd and non-positive if* k *is even.*
3. *Let* $\mathcal{J} = \{i_1, ..., i_J\}$ *and* $\mathcal{J}' = \{i'_1, ..i'_J, ..., i'_{J'}\}$*. If* $r_{i_j}(\delta_{i_j}) = r_{i'_j}(\delta_{i'_j})$ *for* $j = 1, ..., J$ *then* $H((r_1, .., r_J), \cdot) = H((r_1, .., r_J, 0, .., 0), \cdot)$

Then, $V(\delta, \cdot) = \frac{1}{m} lnH(r(\delta))$ *is in the class of ERUM functions. Furthermore, if* $V_i(\cdot) = \frac{1}{m_i}\ln H_i(\cdot)$*,* $i = 1, .., I$ *each satisfy these three conditions, then* $V(\cdot) = \sum_{i=1}^{I} V_i(\cdot)$ *is the in the class of ERUM functions.*

Proof 4 *(See Appendix.)*

The last component of this proposition establishes that all random coefficient MNL models with an additively separable and common component to utility (such as the unobserved product attribute) generate expected utility functions that are within the set of ERUM functions. Since any member of this class of functions corresponds to a discrete choice model with some distribution of consumer attributes, the only issue is how to best pick an approximation to a sufficiently flexible function in the class of ERUM functions. The popular class of Mixed GEV models studied by Berry (1994) and Berry, Levinsohn, and Pakes (1997) generate a particular member of the ERUM class of functions and it is instructive to review and contrast the chosen approximation to the one suggested here.

Specifically, consider I individuals, each with a multinomial logit preferences, $H_i(r) = \sum_{j=1}^{J} r_{ij}$ where $r_{ij} = \exp\{\overline{u}_j(p_j, w_j, \delta_j, \alpha_i)\}$. Mixed MNL models with an additive unobserved product characteristic generate a flexible member of the class of ERUM functions by constructing

$$\begin{aligned} V(\delta) &= \sum_{i=1}^{I} V_i(\cdot) = \ln\left(\prod_{i=1}^{I} H_i(\cdot)\right) \\ &= \ln\left(\prod_{i=1}^{I}\left(\sum_{j=1}^{J} \exp\{\overline{u}_j(p_j, w_j, \delta_j, \alpha_i)\}\right)\right) \end{aligned}$$

which clearly involves using an approximation which is a polynomial of degree I in $\exp\{\overline{u}_j\}$ and hence a polynomial of degree I in functions of the product characteristics and prices $\exp\{\omega_j\}...\exp\{\omega_k\}$. While such an approximation is clearly capable of generating very flexible models,[14] the computational costs of doing so are substantial. Specifying models that generate the $H()$ functions by adding successively high order terms in an additive way provides a simpler method to generate successively flexible models, one which involves a direct generalization of the existing popular MNL and Nested MNL models. These models are particularly likely to be useful whenever the researcher is particularly interested in determining elasticities with respect to a particular characteristic such as price.[15] Moreover, as I have already shown, a

[14] McFadden and Train (1998) provide a much stronger flexibility result for the full mixed multinomial logit model than the result provided here. Again however, the random coefficient approach has substantial computational disadvantages.

[15] As an aside it is interesting to note that the random coefficient multinomial logit model has proven empir-

second order model is sufficient to guarantee the model can match substitution patterns at a point in price and income space the desirable property.

Using this proposition, it is easy to establish the conditions required for the specifications provided in Equation (2) to correspond to an additive random utility model.

Proposition 5 *Let* $V(\Psi) = \frac{1}{2} lnH(r)$ *and* $r_j = e^{\phi_j(y,p_j,\cdot)-\delta_j}$. *If* $H(r) = \frac{1}{2}r'(I+B)r$ *and (i)* $r'(I+B)r \geq 0$ *so* $H(\cdot)$ *is non-negative, (ii)* $\frac{\partial H}{\partial r_j} = r_j + \sum_{l=1}^{J} b_{lj}r_l \geq 0$ *for all* j *(iii)* $b_{jk} \leq 0$ *for all* $j \neq k$*, then* $V(\delta)$ *is in the class of EMRU functions.*

Proof 5 *See Appendix*

Finally, notice that if $V()$ also satisfies the indirect utility properties, then $V(\cdot)$ is a social indirect utility function for the set of consumers of type c. Thus, for consumers of type c, there exists a direct utility function which represents the preferences of the community of people with characteristics, c.

Proposition 6 *If* $V(\mathrm{p}, y, \delta) = lnH(r)$ *with* $r_j = \exp\{\psi_j(y, p_j, p_0, \delta_j)\}$ *is a member of the class of functions defined in EMRU and is also an indirect utility function, then* V *is a social indirect utility function if* $\psi_j(y, p_j, p_0, \delta) = \frac{y-p_j}{p_0} - \delta_j$ *for all* $j = 1, .., J$. *Ie., the demand system resulting from applying WDZM's identity is the same as the demand system that results from applying Roy's identity to* $V()$.

Proof 6 *In general, the demand systems corresponding to discrete choice behavior and continuous choice behavior will not be the same. However, the demand systems created via Roy's identity or WDZM's theorem are identical provided* $\frac{\partial V}{\partial p_j} = \frac{\partial V}{\partial y}\frac{\partial V}{\partial \delta_j}$ *for all* $j \in \mathcal{J}$ *and the shares of the outside goods match. Under the conditions in the proposition for all* $j \geq 1$,

$$\begin{aligned}
\frac{\partial V}{\partial p_j} &= \frac{\partial V}{\partial \psi_j}\frac{\partial \psi_j}{\partial p_j} = \frac{-1}{p_0}\frac{\partial V}{\partial \psi_j}, \ \textit{while} \\
\frac{\partial V}{\partial \delta_j} &= \frac{\partial V}{\partial \psi_j}\frac{\partial \psi_j}{\partial \delta_j} = -\frac{\partial V}{\partial \psi_j} \\
\frac{\partial V}{\partial y} &= \frac{1}{p_0}\sum_{k=1}^{J}\frac{\partial lnH}{\partial r_k}\frac{\partial r_k}{\partial \psi_k} = \frac{1}{p_0}\sum_{k=1}^{J}\frac{\partial lnH}{\partial r_k}r_k = \frac{1}{p_0}
\end{aligned}$$

where the final equality follows since H *is linearly homogeneous in* r.

One detail remains to be established, that the shares of the outside goods match. The result above establishes that market shares for all inside goods are equal. However, there is

ically relatively poor at determining reasonable price elasticities which authors have determined 'reasonable.' Typically estimated own price elasticities are very low.

an important distinction between the discrete choice and continuous choice models in when the outside good is consumed. That is, in the discrete case, some amount of the outside good is consumed whichever inside good is chosen. When choice j is selected, $\frac{y-p_j}{p_0}$ is spent on the outside good. Thus, in a discrete choice model, total demand for the outside good is $x^{discrete}_{outside} = \sum_{j=1}^{J} \frac{y-p_j}{p_0} x_j = \frac{y}{p_0} - \sum_{j=1}^{J} \frac{p_j}{p_0} x^{discrete}_j$. Since we have already established that $x^{discrete}_j = x^{continuous}_j$ for all $j \geq 1$, the outside market shares are equal since this expression also determines the share of the outside good chosen in the continuous choice model via the budget constraint. $\square$

6 Estimation

In this section I provide a convext program interpretation of the contraction mapping algorithm proposed by Berry, Levinsohn, and Pakes (1995). I then demonstrate that an entirely analogous approach is available for the continuous choice model. I argue that such an approach has substantial advantages over the traditional approach of the ad-hoc placement of error terms onto the market or budget share equations; placing unobservables directly into the indirect utility function ensures that the model is internally consistent and thereby avoids the fundamental critique of the ad-hoc approach to introducing unobservables provided by Brown and Walker (1989).

6.1 Discrete Choice Models

Consider the program $\max_\delta -V(\delta) - s'\delta$. If $V(\delta)$ is strictly convex in δ, the objective function is strictly concave in delta.[16] In particular, sufficient conditions for $V(\delta) = \frac{1}{2} ln\left(\frac{1}{2} r'(I+B)r\right)$ to be strictly convex is that $(I+B)$ is positive definite. A dominant diagonal argument provides sufficient condition for that provided $b_{jj} > \sum_{k \neq j} b_{jk}$ which will be true whenever $b_{jj} > 0$ and $b_{jk} \leq 0$ for all $j \neq k$, with some element strict for each j. Thus, subject to minor regularity conditions,[17] the program has a unique solution, δ^* which satisfies the first order conditions

[16] By additivity, $\sum_{j=1}^{J} s_j(\delta) = \sum_{j=1}^{J} -\frac{\partial V}{\partial \delta_j} = 1$, so that $\sum_{j=1}^{J} \frac{\partial^2 V}{\partial \delta_j \partial \delta_k} = 0$. Thus, $\frac{\partial^2 V}{\partial \delta_k^2} = -\sum_{j \neq k}^{J} \frac{\partial^2 V}{\partial \delta_j \partial \delta_k}$. If $\frac{\partial^2 V}{\partial \delta_j \partial \delta_k} \leq 0$, for each $k \neq j$, then the matrix of second derivatives of V w.r.t. δ satisfies a positive dominant diagonal condition and hence V is convex in δ. If for each k there exists some $j \neq k$ such that the mixed partial derivative is strictly negative, then V is strictly convex in δ.

[17] In particular, the solution to this problem must lie in a sufficiently large bounded subset of $\mathbb{R}^J$. Provided the observed shares are strictly in the interior, of $[0,1]^J$, the solution to the above maximization problem must

$$s_j = -\frac{\partial V(\delta)}{\partial \delta_j}, \quad j = 1, ..., J$$

Thus, provided $V(\delta)$ is strictly convex in δ, there a unique value of δ which sets the model's predicted market share equal to the vector of observed market shares. The value of δ can clearly quickly be obtained using any convex programming algorithm. Following Berry (1994) and Berry, Levinsohn, and Pakes (1995), suppose $\delta_j = w_{1j}'\beta + \xi_j$ where ξ_j represents an unobserved product characteristic and w_{1j} is a vector of observed product characteristics of good j. In that case, a generalized method of moment estimator for the parameters of the model can therefore be based on the set of moment conditions

$$E[\xi_j(\theta_o)|Z_j] = 0.$$

Finally I substantially generalize the 'contraction mapping' result provided in Berry, Levinsohn, and Pakes (1995) to show that there is a simple way to compute the solution δ^* of the maximization problem defined above. The conditions required for their contraction argument to apply are not satisfied in general here, or specifically by the flexible discrete choice model I propose. However, in the proposition below I show that a very simple iteration procedure is very generally guaranteed to converge provided we begin the iteration from a carefully chosen spot. Specifically, one need only start from a low value of δ and iterate on the element by element inverse mapping as described in the proposition below.

Proposition. ***Convergence to the Fixed Point*** Consider the algorithm[18]

1. Set $\delta_1 = 0$ and choose $\delta_j^0 = \overline{\overline{\delta}}$ for $j > 1$, where this global upper bound is defined by $\overline{\overline{\delta}} = \max_{j>1} \overline{\delta}_j$, where $\overline{\delta}_j$ is the solution to $s_j(\delta_j, \delta_{-j}) = s_j$ with $\delta_k = +\infty$ for all $k \neq j$.

be the same as the solution to a problem on a bounded set of $\mathbb{R}^J$ by an identical argument to that provided by Berry (1994) and Berry, Levinsohn, and Pakes (1995). Hence, if the function $V(\delta)$ is strictly convex, it must have a unique solution.

[18]This is just an 'iterated' best response algorithm or the 'Simultaneous Optimization' algorithm described in Topkis (1998) for super-modular games (see p. 191.) I am not aware of other authors who have emphasized the apparently deep connections between the super-modularity literature and Generalized Extreme Value type models. However, this connection certainly appears to suggest fruitful areas to explore particularly when searching for computationally tractable estimation algorithms within that class of econometric models.

2. Given the $(J-1\text{x}1)$ vector $\delta^k_{-1} \in [\underline{\delta}_{-1}, \overline{\delta}_{-1}]^{J-1}$ for any $k \in \mathbb{Z}^+$, define $\delta^{k+1}_{-1} = r_{-1}(\delta^k_{-1})$ where $r_{-1}(\cdot)$ is the vector of element by element inverse functions implicitly defined by $s_j = s_j(\delta_1, \delta_2, ..., r_j(\delta_{-j}, s_j), .., \delta_J)$. If $|\delta^k - \delta^{k+1}|$ is sufficiently small then stop.

3. Set $k = k+1$. Return to step (2) and continue.

This algorithm will converge to a fixed point of the $((J-1)\text{x}1)$ mapping $r_{-1}(\delta_{-1})$. By construction, the fixed point of this mapping is the vector δ^* that equates observed and predicted market shares. $\square$

Proof A straight forward application of the implicit function theorem establishes that the element by element inverse function $r_j(\delta_{-j}, s_j)$ exists provided $\frac{\partial s_j(\delta)}{\partial \delta_j} = -\frac{\partial^2 V}{\partial \delta_j^2} \neq 0$. This argument is provided in the appendix to Berry (1994). In addition, his results imply that a truncated version of the inverse function can be analyzed with $\underline{r}_{-1}(\delta, s) = \max\{r_{-1}(\delta, s), \underline{\delta}_{-1}\}$ and that $\overline{\overline{\delta}}_{-1}$ provides an upper bound such that any solution to the equations $s(\delta) = s$ must lie within the bounded set $[\underline{\delta}_{-1}, \overline{\delta}_{-1}]$.[19] Fortunately, since the original function restricted to this closed and bounded set is continuous, it has a fixed point by Brouwers theorem. By construction, this fixed point is a fixed point of the original function defined on $\mathbb{R}^{J-1}$.

Returning to the existence result for the element by element inverse function, note that existence follows for each $j = 2, .., J$, so that the $J-1$ dimensional vector inverse function also exists $r_{-1}(\delta, s) = (r_2(\delta_{-2}, s_2), .., r_J(\delta_{-J}, s_J))$. Moreover, each $r_j(\delta_{-j})$ is increasing in δ_{-j} since by the implicit function theorem, $\frac{\partial r_j}{\partial \delta_k} = -\frac{\frac{\partial s_j}{\partial \delta_k}}{\frac{\partial s_j}{\partial \delta_j}} = -\frac{\frac{\partial^2 V}{\partial \delta_j \delta_k}}{\frac{\partial^2 V}{\partial \delta_j^2}} \geq 0$. Thus, the vector function is increasing in δ_{-1}. The 'increasing' nature of the vector function $r_{-1}(\delta_{-1})$ also follows directly from the the supermodularity of the objective function which in turn follows from the sign restrictions on the mixed partial derivatives of the function V.

Next I show that the sequence δ^k_{-1} generated by $\delta^{k+1}_{-1} = r_{-1}(\delta^k_{-1})$ is a decreasing sequence. Since I have already shown that it exists in a compact set, if so we know that the sequence must converge to $\delta^*_{-1} = lim_{k\to\infty}\delta^k_{-1}$, where $r_{-1}(\delta^*_{-1}) = \delta^*_{-1}$.

That $\{\delta^k_{-1}\}^\infty_{k=0}$ provides a decreasing sequence follows immediately from the monotonicity property already demonstrated for the vector function $r_{-1}(\delta_{-1})$. More formally, the result can be shown by induction. First, δ^0_{-1} is defined at the upper bound of the feasible set of δ_{-1}'s. Then, since δ^0_{-1} is an upper bound on the feasible set it follows immediately that $\delta^1_{-1} \leq \delta^0_{-1}$. For the induction hypothesis, suppose that $\delta^k_{-1} \leq \delta^{k-1}_{-1}$ for some $k \in \mathbb{Z}^+$. Then,

[19] Note that the exact form of these upper and lower bounds just reverses the upper and lower bound results provided in Berry (1994). The switch from upper to lower occurs because δ enters the utility function with a minus sign in front in this paper for consistency with the original results due to McFadden (1981). Note also that showing existence of an inverse is equivalent to the statement that the model is able to match any vector of market shares *using only the δ parameters*. In that case, existence of an upper and lower bound follow since we can define them as the parameter values which equate the predicted market shares to a vector of 'true' market shares that are strictly below and above the actually observed vector of inside market shares respectively. Since the objective function is supermodular, the solution is monotonically decreasing in the vector of 'true' market shares and it follows that the solution with the actual vector of market shares must lie in between those two bounds.

$\delta_{-1}^{k+1} = r_{-1}(\delta_{-1}^k) \leq r_{-1}(\delta_{-1}^{k-1}) = \delta_{-1}^k$ where the inequality follows directly from $r_{-1}(\cdot)$ increasing. $\Box$.[20]

In practise, this inverse mapping algorithm works very well when taking us close to the fixed point, but subsequently takes a large number of iterations to actually converge on the fixed point. Since it uses no derivative information, this feature is not surprising. However, an algorithm comprising of a small number of steps of this inverse mapping followed by the application of Newton's method will be both robust and also fast. It will be fast because Newton's method converges quadratically when we are local to the fixed point. Define, $F(\delta_{-1}) = -V(\delta_{-1}) - \delta_{-1}s$, with corresponding first order conditions given by $f(\delta_{-1}) = -D_{\delta_{-1}}V(\delta_{-1}) - s = s(\delta) - s$ and second order conditions $H(\delta_{-1}) = -D^2_{\delta_{-1}}V(\delta_{-1})$. The Newton iteration algorithm generates the sequence from

$$\delta_{-1}^{k+1} = \delta_{-1}^k - H(\delta_{-1}^k)^{-1} f(\delta_{-1}^k).$$

Provided $H(\delta_{-1}^k) = -D^2_{\delta_{-1}}V(\delta_{-1})$ is negative definite, this algorithm will converge quadratically to the fixed point. (See Theorem 5.5.1, Judd (1999).)

6.2 Continuous Choice Models

Econometric unobservables are typically added onto the demand system in continuous choice models. This is clearly one option here, $s_j = s_j(\delta) + \xi_j$. However, Brown and Walker (1989) demonstrate that doing so introduces a correlation between all the prices and income in the demand specification and the error term whenever the data generating process satisfies slutsky symmetry.

[20] Notice that an alternative computational algorithm would start at the *lower* bound and generate an increasing sequence using the element by element inverse function. Either bound can be used as a starting value. However, the upper bound is much easier, in fact trivial, to compute. Naturally, using the notation from Berry (1994) and Berry, Levinsohn, and Pakes (1995), the lower bound will be easiest to compute. This result emerges very generally. It results immediately from the supermodularity assumptions placed directly on the class of functions, V. This fact is likely to have important implications for the analysis and computation of models such as the pure hedonic model developed in Berry and Pakes (1999) where the authors suggest an algorithm to compute the fixed point based on a homotopy technique. I leave the important task of pursuing those connections to future research. A third feature worthy of note is that monotonic variable transformations will preserve (or flip) monotonicity results. Thus, defining a new variable such as $r_j = e^{-\delta_j}$ will not fundamentally alter the convergence proofs. This fact allows both the numerical issues and computational burden associated with calculating exponents to be avoided entirely during computation of the fixed point.

An alternative is to incorporate the error term directly into the indirect utility function representing preferences. By doing so, we can avoid the problem entirely. The disadvantage is that a new computational problem, to find the value of the unobserved component of preferences that makes predicted and actual market shares identical, is introduced. However, the discrete choice literature suggests an appropriate formulation of the problem in which this computational problem reduces to that of solving a globally convex program, an easy task.

Specifically, if $V(p, y, \delta)$ is an indirect utility function, then we may invert $u = V(p, y, \delta)$ to obtain the corresponding expenditure function, $E(p, \delta, u)$. Consider the program

$$\max_{\delta} E(p, \delta, u) - s' diag\{\left(\frac{\partial \Psi_j}{\partial p_j}\right)^{-1} \frac{\partial \Psi_j}{\partial \delta_j}\}\delta$$

where s is the vector of observed market shares, and $diag\{\cdot\}$ denotes a JxJ matrix with jj^{th} element $\left(\frac{\partial \Psi_j}{\partial p_j}\right)^{-1} \frac{\partial \Psi_j}{\partial \delta_j}$.

If $\Psi_j = \phi_j(y, p_j) - \delta_j$ then the terms in the diagonal matrix are independent of δ. Thus the solution to this program satisfies the first order conditions

$$s_j = \frac{\partial \Psi_j}{\partial p_j} \left(\frac{\partial \Psi_j}{\partial \delta_j}\right)^{-1} \frac{\partial E(p, \delta, u)}{\partial \delta_j}. \tag{5}$$

However, by the chain rule, $\frac{\partial V()}{\partial p_j} = \frac{\partial V()}{\partial \Psi_j} \frac{\partial \Psi_j}{\partial p_j}$ and similarly $\frac{\partial V()}{\partial \delta_j} = \frac{\partial V()}{\partial \Psi_j} \frac{\partial \Psi_j}{\partial \delta_j}$. Thus, $\frac{\partial V}{\partial \delta_j} = \frac{\partial V}{\partial p_j} \frac{\partial \Psi_j}{\partial \delta_j} \left(\frac{\partial \Psi_j}{\partial \mathrm{p}_j}\right)^{-1}$, In addition, by the implicit function theorem $\frac{\partial E(p,u,\delta)}{\partial \delta_j} = -\left(\frac{\partial V(p,\delta,y)}{\partial y}\right)^{-1} \frac{\partial V}{\partial \delta_j}$. Substituting these into Equation (5) yields the first order conditions

$$s_j = -\left(\frac{\partial V(p, \delta, y)}{\partial y}\right)^{-1} \frac{\partial V(p, \delta, y)}{\partial p_j}.$$

Thus, the value of δ that solves the program is the value which equates observed market shares to the model's predicted market shares. For example, suppose $V(\cdot) = \frac{1}{2} lnH(\cdot)$ and $H(r) = \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} (I(j = k) + b_{jk}) \frac{y}{p_j} \frac{y}{p_k} e^{-\delta_j} e^{-\delta_k}$. Then the expenditure function is easily solved for explicitly so that

$$lnE(\mathrm{p},u) = u - \frac{1}{2}ln\left(\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J}(I(j=k)+b_{jk})\frac{e^{-\delta_j}e^{-\delta_k}}{p_jp_k}\right).$$

Provided the expenditure function is convex or concave in δ, there is a unique value of δ that solves the program and hence a unique value that equates observed and predicted market shares. Under this particular specification, $D_yV(\cdot) = \frac{1}{y}$, so $D_\delta E(\mathrm{p}, u, \delta) = -yD_\delta V(p, y, \delta)$ and E is concave in δ whenever $V(\cdot)$ is convex in δ.

Clearly, once again, estimates of the model parameters can be obtained using the set of moment conditions

$$E[\xi_j(\theta_o)|Z_j] = 0.$$

where $\xi_j = \delta_j - x_j'\beta$. In some instances, no product characteristics will be available. In that case, the only explanetary variable entering this regression would be a constant.

7 Product Characteristics

Generalizing classical demand systems, Lancaster (1966) suggests that consumers are interested in goods because of the characteristics they provide, thus he argues that a useful generalization of classical choice models provided with preferences directly over product characteristics. A technology describes the fashion in which product characteristics[21] are 'produced' from products themselves, $w = f(x)$. Thus, in a lancastrian world the consumer is assumed to solve the choice problem[22]

[21] Following the literature, I assume that all product characteristics are observed by the consumer but not necessarily by the econometrician.

[22] Note that this specification is conceptually a generalization of the classical model since one possible production function is f(x)=x. However, if the number of product characteristics is smaller than the number of products, then more parsimonious demand systems will result.

$$\max_{x \in X} u(w;\theta) \text{ s.t. } w = f(x) \quad \text{and } \mathrm{p}'x \leq y$$

Substituting in the new constraints, $w = f(x)$, yields the equivalent program

$$\max_{x \in X} u(f(x);\theta) \text{ s.t. } \mathrm{p}'x \leq y.$$

Clearly, this latter program is precisely in the form of a classical choice model. Hence, without adding structure to $u(w)$ and $f(x)$, introducing product characteristics places no additional restrictions or structure on the form of the indirect utility function, $V(p,y)$. In principle, therefore, product characteristics may enter through any of the parameters of the model in an arbitrary fashion.

An additional natural assumption that does have considerable bite, is

$$lim_{p_j \to \infty} V(p_j, p_{-j}, w_j, w_{-j}) = V(p_{-j}, w_{-j}).$$

7.1 Product Characteristics in Continuous and Discrete Choice Models

Consider the model $V(p_j, p_{-j}, w_j, w_{-j}) = \frac{1}{m} lnH(p_j, p_{-j}, w_j, w_{-j})$ where $H(r) = \frac{1}{2} r'(I+B)r = \frac{1}{2}\sum_{l=1}^{J}\sum_{m=1}^{J}(1(m=l)+b_{lm})r_l r_m$ and $r_l = \exp\{\ln y - \ln p_l - \delta_l\}$, I is the identity matrix, and B is a symmetric matrix of parameters. Imposing $lim_{p_j \to \infty} V(p_j, p_{-j}, w_j, w_{-j}) = V(p_{-j}, w_{-j})$ requires that the parameter δ_j can only depend on the product characteristics of good j and b_{jk} can only depend on the product characteristics of goods j and k.

A second natural property of any specification for goods that are *substitutes* is that as the distance between any two products decreases in characteristics space, the sensitivity of demand for product j to a change in product k's price should increase.

In the particular specification of the continuous-choice model considered here,

$$\frac{\partial s_j}{\partial p_k} = -\left(\frac{\partial H}{\partial y}\right)^{-1}\left(\frac{\partial^2 H}{\partial p_j \partial p_k} - s_j \sum_{l=1}^{J} \frac{\partial^2 H}{\partial p_l \partial p_k}\right) - s_j s_k$$

where $\frac{\partial^2 H}{\partial p_j \partial p_k} = (I(j=k) + b_{jk}) r_j r_k \frac{1}{p_j p_k}$, while in the discrete choice model

$$\frac{\partial lns_j}{\partial lnp_k} = -\left[\frac{y^2}{p_j p_k} \frac{\exp\{-\delta_j\}\exp\{-\delta_k\} b_{jk}}{\frac{1}{2} r'(I+B)r} - s_k\right] - \mathbb{I}(j=k)$$

Consider a polynomial specification with $b_{jk} = \sum_{l=1}^{p} \tau_l d(w_j, w_k; \alpha)^l$, for all $j \neq k$ where $d(w_j, w_k; \alpha)$ is some metric in characteristics space, such as $d(w_j, w_k; \alpha) = \sqrt{\sum_{l=1}^{m} \alpha_l (w_{jl} - w_{kl})^2}$ with weights that are estimable parameters that sum to one $\sum_{l=1}^{m} \alpha_l = 1$, and the vector of diagonal parameters[23] $b_{kk} = w_k' \lambda$. This specification is the 'Distance Metric' specification suggested by Slade (2001).[24]

Alternatively, a specification closer to that suggested in Hausman and Wise (1978) for the Probit model is also attractive. In the Probit case, $u_{ij} = \overline{u}_j + \epsilon_{ij}$ where $\epsilon_i = (\epsilon_{i1}, .., \epsilon_{iJ}) \sim N(0, \Sigma)$ and the parameters of the taste distribution that are mapped down to product characteristics are $\Sigma = \{\sigma_{jk}\}_{j,k=1,..,J}$ where

$$\sigma_{jk}(\omega_{lk}, \omega_{lj}, \cdot) = \sum_{l=1}^{L} \sigma_{\beta_l} \omega_{lj} \omega_{lk} + \sigma_\gamma$$

where the summation is over all measured characteristics and $(\{\sigma_{\beta_l}\}_{l=1}^{L}, \sigma_\gamma)$ are parameters to be estimated. In the GEV model suggested here, the b_{jk} parameters are not interpretable directly as covariances, however they are clearly very closely related to cross elasticities of demand. To that extent, they should be directly related to the covariance between the measured characteristics of the products which suggests a parameterization of the same form $b_{jk} = \sigma_{jk}(\omega_{lk}, \omega_{lj})$.

Either approach will be able to capture the relationship between the distance between two products in characteristics space and the resulting substitution pattern between those

[23]??To show:: possible to defined b_{kk} implicitly as the solution to the vector of constraints provided by the derivatives of the budget constraint with respect to each p_k, $\sum_{j=1}^{J} p_j \frac{\partial s_j(B,\delta)}{\partial p_k} = -s_k$?

[24]One important caveat to this specification is that imposes symmetry on the B matrix. Thus, there may be better alternatives to this mapping between observed product characteristics and the parameters of the model in the discrete choice setting wherein .

goods. Notice that this specification imposes symmetry on the matrix B while simultaneously substantially reducing the number of parameters to be estimated whenever the number of product characteristics are fewer than the number of products.

8 Consumer Characteristics and Random Coefficients

If $V(\delta, c)$ is an EMRU function and is convex in δ for all consumer types, then the problem $\min_\delta \int V(\delta, c) f(c) dc + s'\delta$ is a convex problem programming problem in the vector, δ with first order conditions that equate the observed market shares equal to the predicted market shares.

Similarly, in the continuous choice model, inverting the indirect utility function to obtain the expenditure function for each consumer type, c, yields $E(\mathrm{p}, \delta, c, u)$. Then provided $E(\mathrm{p}, \delta, c, u)$ is concave in δ for each c, the problem $\max_\delta \int E(p, \delta, u, c) f(c) dc - s' diag\{\left(\frac{\partial \Psi_j}{\partial p_j}\right)^{-1} \frac{\partial \Psi_j}{\partial \delta_j}\}\delta$ is concave in δ, with first order conditions that equate the observed market shares equal to the predicted market shares.

Thus, all random coefficient versions of the discrete and continuous choice models, may also be considered using the identical methodology. In particular, it is likely that different types of consumers may have different preference metrics over characteristics space although our ability to separate these preferences from purely aggregate data is likely to be very limited.[25]

9 Conclusion

In this paper I propose a class of models and delineate the conditions which lead this class to be consistent with either (i) an underlying distribution of consumers making discrete choices or (2) a single agent making continuous choices. In doing so, my aim is to develop discrete- and

[25] There is one stark difference between the models. Namely, that properties of EMRU functions are preserved under aggregation across consumer types, while it is well known that the same is *not* true for indirect utility functions. Thus, aggregate discrete choice demand functions obtained by integrating across the distribution of income are EMRU demand functions. As McFadden (1981) notes (p. 216), all EMRU properties are preserved by addition. Hence, if $V(\delta, c)$ is an EMRU function for each consumer type, c, with resulting demand systems $s_j(c)$ for each $j \in \mathcal{J}$. Then, the probability mixture over consumer types, $V^*(\delta) = \int V^i(c) f(c) dc$ will also be a member of the EMRU set of functions with aggregate demand system, $s_j = \int s_j f(c) dc$.

continuous-choice models that have the advantageous properties of both previous literatures, applying the lessons learned in the discrete and continuous choice demand literatures to the other literature.

There are five main implications and advantages of the new models. First, continuous choice models can easily be formulated that allow for the introduction and exit of new products. This is a feature not shared by the current generation of continuous choice models which, as a result, are largely only appropriate for broad aggregates of goods. This feature is particularly attractive for market level studies in marketing and industrial organization where increasingly datasets are extremely disaggregated in nature and product introduction and exit are both extremely frequent and very informative about substitution patterns.

Second, the error term is explicitly a part of the models and therefore, in particular, the specifications are not subject to the critique provided by Brown and Walker (1989).

Third, the discrete-choice model proposed is consistent with an underlying distribution of consumers who each make a discrete choice from the set of available products. In contrast to models such as the multi-nomial logit model, I show that the proposed discrete-choice model is a flexible functional form in the sense of Diewert (1974) and as such is capable of approximating any observed pattern of income and price effects on demand.

Fourth, by avoiding simulation estimators the discrete choice models are substantially faster to compute than existing approaches. Finally, in both cases, preferences can be mapped down to product characteristics in a natural way, providing a parsimonious approach when product characteristics data is available.

10 Appendix A

Lemma 2 *If $\Phi(r_1, .., r_J)$ is a real valued convex function that is either*

1. *non-decreasing in $\underline{r}$, and $r_i(\mathrm{x})$ are convex in* x,
2. *non-increasing in $\underline{r}$, and $r_i(\mathrm{x})$ are concave in* x,

then $\Phi(r_1(\mathrm{x}), ..., r_J(\mathrm{x}))$ is non-increasing and convex in x.

Proof Choose any pair of vectors x and x′ and let $0 \leq \lambda \leq 1$.

1. If $\underline{r}(\mathrm{x})$ is convex in x, $\underline{r}(\lambda x + (1-\lambda)x') \leq \lambda \underline{r}(x) + (1-\lambda)\underline{r}(x')$. However, since Φ is non-decreasing in $\underline{r}$, $\Phi(\underline{r}(\lambda \mathrm{x} + (1-\lambda)\mathrm{x}') \leq \Phi(\lambda \underline{r}(\mathrm{x}) + (1-\lambda)\underline{r}(\mathrm{x}')) \leq \lambda \Phi(\underline{r}(\mathrm{x})) + (1-\lambda)\Phi(\underline{r}(\mathrm{x}'))$, where the latter inequality follows since Φ is convex in r.
2. If $\underline{r}(\mathrm{x})$ is concave in x $\underline{r}(\lambda x + (1-\lambda)x') \geq \lambda \underline{r}(x) + (1-\lambda)\underline{r}(x')$. However, since Φ is non-increasing in $\underline{r}$, $\Phi(\underline{r}(\lambda \mathrm{x} + (1-\lambda)\mathrm{x}') \leq \Phi(\lambda \underline{r}(\mathrm{x}) + (1-\lambda)\underline{r}(\mathrm{x}')) \leq \lambda \Phi(\underline{r}(\mathrm{x})) + (1-\lambda)\Phi(\underline{r}(\mathrm{x}'))$, where the latter inequality follows since Φ is convex in r.□

Lemma 3 *Let $V(r) = c + dlnH(r)$ and $r_j = e^{\psi_j(x)}$. If $H(r)$ is a non-negative, non-decreasing, and homogeneous degree $m > 0$ function of r, with $\frac{\partial^2 H}{\partial r_k \partial r_j} \leq 0\ \forall j \neq k$, and $\psi_j(x)$ is convex in x for all $j = 1, .., J$ then $V(x)$ is convex in x.*

Proof If each component function, $r_j(x)$, is a convex function of x, then the vector function $r(x)$ is also convex in x. By lemma 2 it suffices to establish that $V(r(x))$ is non-decreasing in r and $r(x)$ is a convex function of x. To do so, note that

$$\sum_{j=1}^{J} \frac{\partial lnH}{\partial \psi_j} = \sum_{j=1}^{J} \frac{\partial lnH}{\partial r_j} \frac{\partial r_j}{\partial \psi_j} = \sum_{j=1}^{J} \frac{\partial lnH}{\partial log r_j} = m$$

where the first equality follows from the chain rule for differentiation, the second since $\frac{\partial r_j}{\partial \psi_j} = r_j$, and the final equality follows from Eulers theorem since $H(r)$ is a homogeneous degree m function of r.

Differentiating both sides with respect to ψ_k and rearranging yields

$$\frac{\partial^2 lnH}{\partial \psi_k^2} = -\sum_{j \neq k}^{J} \frac{\partial^2 lnH}{\partial \psi_k \partial \psi_j}$$

If $\frac{\partial^2 lnH}{\partial \psi_k \partial \psi_j} \leq 0$ for all $j \neq k$, then $\frac{\partial^2 lnH}{\partial \psi_k^2} \geq 0$ and the matrix of second derivatives of $lnH(\psi)$ has a dominant positive diagonal and is therefore convex in ψ (see Lancaster and Tismenetsky (1985), p. 373 for example.) Since $V(\psi)$ is an affine transformation of a convex function, it is convex if $lnH(\psi)$ is convex.

Thus, it suffices to establish that the conditions in the lemma ensure that $\frac{\partial^2 lnH}{\partial\psi_k\partial\psi_j} \leq 0$ for all $j \neq k$. This follows trivially, since $H, r_j, r_k, \frac{\partial H}{\partial r_j} \geq 0$ and

$$\frac{\partial^2 lnH}{\partial\psi_k\partial\psi_j} = \frac{1}{H}\frac{\partial^2 H}{\partial\psi_k\partial\psi_j} - \frac{1}{H^2}\frac{\partial H}{\partial\psi_k}\frac{\partial H}{\partial\psi_j} = \frac{1}{H}r_j r_k\left(\frac{\partial^2 H}{\partial r_k\partial r_j} - \frac{1}{H}\frac{\partial H}{\partial r_k}\frac{\partial H}{\partial r_j}\right)$$

□

Proof to proposition 4

1. Clearly if $H(r)$ is a real valued function, and $r_j = e^{\Psi_j(\delta_j)}$ is a real valued function of δ_j for $j = 1, ..., J$, then $V(\delta)$ is a real valued function of δ.

2. If $H(r)$ is homogeneous of degree m in r, then $V(\theta\delta) = \frac{1}{m}ln((e^{-\theta})^m H(\delta))$, so $V(\theta\delta) = V(\delta) - \theta$.

3. First note that the mixed cross partials with respect to δ can be written $\frac{\partial V}{\partial\delta_{1,...,\delta_k}} = (-1)^k r_1...r_k \frac{\partial V(r)}{\partial r_{1,...,\partial r_k}}$. Thus, to ensure that the mixed cross partials of $\frac{\partial V(\delta)}{\partial\delta_{1,...,\delta_k}}$ are always non-positive, the mixed cross partials of the function with respect to r must alternate in sign with even k being non-positive and odd k non-negative. Showing this requires an induction argument that is very similar to the one used by Mcfadden (1978) to characterize the Generalized Extreme Value model. Using the convention that $V_{1...,k}$ denotes $\frac{\partial V}{\partial r_1....\partial r_k}$ and $H_{1..,k} \equiv \frac{\partial H}{\partial r_1....\partial r_k}$, define, recursively, $Q_1 = H_1$ and $Q_k = \frac{\partial Q_{k-1}}{\partial r_k} - \frac{1}{H}Q_{k-1}H_k$.[26] Suppose $V_{1,...,k-1} = \frac{Q_{k-1}}{H}$. Then differentiating with respect to r_k yields $V_{1,...,k} = \left(\frac{\partial Q_{k-1}}{\partial r_k} - \frac{1}{H}Q_{k-1}H_k\right)\frac{1}{H}$. Since $V_1 = \frac{Q_1}{H}$, $V_{1,...,k} = \frac{Q_k}{H}$ for all k by induction.

   Next, I characterize the sequence Q_k. First note that Q_k is a sum of signed terms, with each term a product of cross derivatives of H of various orders. Suppose each signed term in Q_{k-1} is non-negative. Then $Q_{k-1}H_k$ is non-negative. Further, each term in $\frac{\partial Q_{k-1}}{\partial r_k}$ is non-positive, since one of the derivatives within each term has increased in order, changing from even to odd or vice versa, with a hypothesized change in sign. Hence, Q_k is non-positive. Similarly, if Q_{k-1} is non-positive then Q_k is non-negative. Since Q_1 is non-negative, the sequence of Q_k's alternates in sign with terms when k is an even number non-positive and terms with k an odd number, non-negative.

   Therefore, $\frac{\partial V}{\partial\delta_{1,...,\delta_k}} = (-1)^k r_1...r_k \frac{\partial V(r)}{\partial r_{1,...,r_k}}$ is non-positive for all k as required.

---

[26]Thus, for example,

$$V_1 = \frac{H_1}{H}$$
$$V_{12} = \frac{H_{12}}{H} - \frac{1}{H^2}H_1 H_2.$$

4. As $\delta_k \to -\infty$, the vector $(e^{\delta_k} r_1, , .., e^{\delta_k} r_J)$ converges to a vector with 1 in the k^{th} component and zeros elsewhere. Since

$$\begin{aligned}
lim_{\delta_j \to -\infty} \frac{\partial V}{\partial \delta_j} &= -lim_{\delta_j \to -\infty} \frac{r_j H_j(r, \cdot)}{mH(r, \cdot)} \\
&= -\lim_{\delta_j \to -\infty} \frac{e^{-(m-1)\delta_j} r_j H_j(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{me^{-m\delta_j} H(e^{\delta_j} r_1, .., e^{\delta_j} r_J)} \\
&= -\lim_{\delta_j \to -\infty} e^{\delta_j} r_j \frac{H_j(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{mH(e^{\delta_j} r_1, .., e^{\delta_j} r_J)} \\
&= 1 \frac{H_j((0, ..., 1, 0..0), \cdot)}{mH((0, ..., 0, 1, 0...0), \cdot)} \\
&= -1
\end{aligned}$$

where the final equality follows from taking limits of Euler's equation

$$\begin{aligned}
1 = lim_{\delta_j \to -\infty} \sum_{k=1}^{J} \frac{r_k H_k(r_1, .., r_J)}{mH(r_1, .., r_J)} &= lim_{\delta_j \to -\infty} \sum_{k=1}^{J} e^{-\delta_k} \frac{e^{+\delta_j} H_k(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{mH(e^{\delta_j} r_1, ..., e^{\delta_j} r_J)} \\
&= lim_{\delta_j \to -\infty} \sum_{k=1}^{J} e^{-\delta_k + \delta_j} \frac{H_j(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{mH(e^{\delta_j} r_1, ..., e^{\delta_j} r_J)} \\
&= lim_{\delta_j \to -\infty} \frac{H_j(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{mH(e^{\delta_j} r_1, ..., e^{\delta_j} r_J)}
\end{aligned}$$

5. If $\mathcal{J} = \{i_1, ..., i_J\}$ and $\mathcal{J}' = \{i'_1, ..., i'_J, ..., i_J\}$, satisfy $r(\delta_{i_k}) = r(\delta_{i'_k})$ for $k = 1, .., J'$ then $\delta_{i_k} = \delta_{i'_k}$ and $V(\delta_{\mathcal{J}'}, +\infty, ..., +\infty) = lnH(r_{\mathcal{J}'}, 0, .., 0) = lnH(r_{\mathcal{J}'}) = V(\delta_{\mathcal{J}'})$.

Proof to proposition 5

I show that $H(r) = a'r + r^{\frac{1}{2}'} B r^{\frac{1}{2}}$ and $H(r) = r'(I + B)r$ both have the properties used in Proposition 4 and are therefore in the class of ERUM functions. Notice that in both cases, $H(r)$ is non-negative by assumption, homogeneous of degree one and two respectively by construction, and well defined for $r \geq 0$. The mixed partial derivatives of $H(r)$ clearly exist provided all the elements in a and B are finite.

1. Then $\frac{\partial H}{\partial r_j} \geq 0$ by the hypotheses in the proposition, while $\frac{\partial^2 H}{\partial r_j \partial r_k} = \frac{1}{4} \frac{b_{jk} + b_{kj}}{2} r_j^{-\frac{1}{2}} r_k^{-\frac{1}{2}}$ for all $j \neq k$. Since $r_j \geq 0$ this is non-positive provided b_{jk} is non-positive for all $j \neq k$. All higher mixed partial derivatives are clearly zero and therefore satisfy the partial derivative conditions. The third condition is trivially satisfied.

2. Then $\frac{\partial H}{\partial r_j} \geq 0$ by the hypotheses in the proposition, while $\frac{\partial^2 H}{\partial r_j \partial r_k} = \frac{b_{jk} + b_{kj}}{2}$ for all $j \neq k$. Since $r_j \geq 0$ this is non-positive provided b_{jk} is non-positive for all $j \neq k$. All higher mixed partial derivatives are clearly zero and therefore satisfy the partial derivative conditions. The third condition is trivially satisfied.□

Theorem 2 *Generalized Extreme Value Model (GEV) (Slight relaxation of Mcfadden (1978))*

Suppose $H(r_1,..,r_J)$ is a non-negative, homogeneous of degree $m > 0$ function of $(r_1,...,r_J) \geq 0$. Suppose $\lim_{r_j \to +\infty} H(r_1,..,r_J) = +\infty$ for $j = 1,..,J$. Suppose for any distinct $(i_1,...,i_k)$ from $\{1,...,J\}$, $\frac{\partial^k H}{\partial r_{i_1},...,\partial r_{i_k}}$ is non-negative if k is odd and non-positive if k is even. Then,

$$P_j = \frac{e^{\delta_j} H_j(e^{\delta_1},...,e^{\delta_J})}{mH(e^{\delta_1},...,e^{\delta_J})}$$

defines a probabilistic choice model from alternatives $j = 1,..,J$ which is consistent with utility maximization.

Proof GEV model

The steps of this proof follow those in Theorem 1 in Mcfadden (1978). However, the theorem is a mild relaxation of that theorem since while any function $H()$ of homogeneity of degree m can be transformed into a homogeneous degree one function $\tilde{H}(\cdot) = H(\cdot)^{1/m}$, the sign properties of the derivatives of $H(\cdot)$ are not generally inherited by $\tilde{H}(\cdot)$.

Consider the function $F(\epsilon_1,..,\epsilon_J) = e^{-H(e^{-\epsilon_1},..,e^{-\epsilon_J})}$.

I first prove that this is a multi-variate extreme value distribution. If $\epsilon_j \to -\infty$, then $H \to +\infty$, implying $F \to 1$. Define, recursively, $Q_1 = H_1$ and $Q_k = Q_{k-1}H_k - \frac{\partial Q_{k-1}}{\partial r_k}$. Then Q_k is a sum of signed terms, with each term a product of cross derivatives of H of various orders. Suppose each signed term in Q_{k-1} is non-negative, Then $Q_{k-1}H_k$ is non-negative. Further, each term in $\frac{\partial Q_{k-1}}{\partial r_k}$ is non-positive, since one of the derivatives in each term has increased in order, changing from even to odd or vice versa, with a hypothesized change in sign. Hence, each term in Q_k is non-negative. By induction, Q_k i non-negative for $k = 1,..,J$.

Differentiating F, $\frac{\partial F}{\partial \epsilon_1} = e^{-\epsilon_1}Q_1F$. Suppose $\frac{\partial^{k-1}F}{\partial \epsilon_1,..,\partial e_{k-1}} = e^{\epsilon_1}Q^{k-1}F$. Then $\frac{\partial^k F}{\partial \epsilon_1,..,\partial e_k} = e^{-\epsilon_1}...e^{-\epsilon_k}\{Q_{k-1}H_kF - F\frac{\partial Q_{k-1}}{\partial r_k}\} = e^{-\epsilon_1}...e^{-\epsilon_k}Q_kF$. By induction, $\frac{\partial^J F}{\partial e^{-\epsilon_1},...,\partial e^{-\epsilon_J}}Q_JF \geq 0$. Hence, F is a cumulative distribution function. When $\epsilon_j = +\infty$ for $j \neq i$, $F = exp{-a_i e^{-\epsilon_i}}$, where $a_i = G(0,..,0,1,0,...,0)$ with the 1 in the i^{th} place. This is the univariate extreme value distribution. Hence, F is a multivariate extreme value distribution.

Suppose a population has utilities $u_i = \delta_i + \epsilon_i$, where $(\epsilon_1,..,\epsilon_J)$ is distributed as F. Then, the probability that the first alternative is selected satisfies

$$
\begin{aligned}
\pi_i &= \int_{\epsilon=-\infty}^{+\infty} F_1(\epsilon, \delta_1 - \delta_2 + \epsilon, ..., \delta_1 - \delta_J + \epsilon) d\epsilon \\
&= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} H_1(e^{-\epsilon}, e^{-\delta_1+\delta_2-\epsilon}, ..., e^{-\delta_1+\delta_J-\epsilon}) exp\{-H(e^{-\epsilon}, e^{-\delta_1+\delta_2-\epsilon}, ..., e^{-\delta_1+\delta_J-\epsilon})\} d\epsilon \\
&= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} (e^{\epsilon+\delta_1})^{-(m-1)} H_1(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J}) exp\{-(e^{\epsilon+\delta_1})^{-m} H(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})\} d\epsilon \\
&= e^{\delta_1} \int_{u=-\infty}^{+\infty} e^{-m(u)} H_1(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J}) exp\{-(e^{-m(u)} H(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})\} du \\
&= e^{\delta_1} \frac{H_1(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})}{H(e^{\delta_1}, .., e^{\delta_J})} \int_{u=-\infty}^{+\infty} e^{-m(u)+lnH(e^{\delta_1}, .., e^{\delta_J})} exp\{-(e^{-m(u)+lnH}\} du \\
&= \frac{1}{m} e^{\delta_1} \frac{H_1(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})}{H(e^{\delta_1}, .., e^{\delta_J})} \int_{u=-\infty}^{+\infty} m e^{-m(u-\frac{1}{m}lnH(e^{\delta_1}, .., e^{\delta_J}))} exp\{-(e^{-m(u-\frac{1}{m}lnH(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J}))}\} du \\
&= \frac{e^{\delta_1} H_1(e^{\delta_1}, ..., e^{\delta_J})}{m H(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})}
\end{aligned}
$$

where the second equality follows since $H()$ is homogeneous of degree m and the third follows by a change of variable $u = \epsilon + \delta_1$. Now the type 1 extreme value probability density function with parameters (θ, ξ) is $p(x) = \theta^{-1} e^{-\frac{x-\xi}{\theta}} exp\{-e^{-\frac{(x-\xi)}{\theta}}\}$ (see Johnson, Kotz, and Balakrishnan (1995), p11). Setting $\xi = \frac{1}{m} lnH(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})$ and $\theta = m^{-1}$ establishes the final equality since the area under a density function is one. Since this argument can be applied to any alternative, the theorem is proved. □.

Corollary

Under the hypotheses of Theorem 3, expected maximum utility, defined by

$$
V = \int_{\epsilon_1=-\infty}^{+\infty} \cdots \int_{\epsilon_J=-\infty}^{+\infty} \max_{j=1,..,J} (\delta_j + \epsilon_j) f(\epsilon_1, .., \epsilon_J) d\epsilon_1 ..., d\epsilon_J
$$

(with f the density of F), satisfies

$$
V = \frac{1}{m} logH(e^{\delta_1}, ..., e^{\delta_J}) + \frac{1}{m}\gamma
$$

where $\gamma = 0.5772156649...$ is Euler's constant and

$$
P_i = \frac{\partial \tilde{U}}{\partial \delta_i}
$$

Proof

The probability density function for the extreme value distribution with parameters (ξ, θ) is described by the function $f(x) = \theta^{-1} exp\{\frac{-(x-\xi)}{\theta} - e^{\frac{-(x-\xi)}{\theta}}\}$, with $\theta > 0$, and has mean $\xi + \gamma\theta$.

The integral in 10 can be partitioned into regions where each alternative has maximum utility, yielding

$$\tilde{U} = \sum_j \int_{\epsilon_j=-\infty}^{+\infty} (\delta_j + \epsilon_j) F_j(\epsilon_1, \delta_1 + \epsilon_1 - \delta_2, ..., \delta_1 + \epsilon_1 - \delta_J) d\epsilon_j.$$

Let $\xi = \frac{1}{m} lnH(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})$ and $\theta = m^{-1}$. Then,

$$F_j(\delta_j + \epsilon_j - \delta_1, \delta_j + \epsilon_j - \delta_2 ..., \delta_j + \epsilon_j - \delta_J) = \frac{H_j(e^{-\delta_j - \epsilon_j + \delta_1}, ..., e^{-\delta_j - \epsilon_j + \delta_J}) e^{-\epsilon_j}}{exp\{H(e^{-\delta_j - \epsilon_j + \delta_1}, ..., e^{-\delta_j - \epsilon_j + \delta_J})\}}$$

Making the transformation $u = \delta_j + \epsilon_j$, Equation 10 becomes

$$\begin{aligned}
\tilde{U} &= \sum_j \int_{u=-\infty}^{+\infty} uexp\{-e^{-m(u)} H(e^{\delta_1}, ..., e^{\delta_J})\} e^{-(m-1)(u)} H_j(e^{\delta_1}, ..., e^{\delta_J}) e^{-u+\delta_j} du \\
&= \sum_j m e^{\delta_j} \frac{H_j(e^{\delta_1}, ..., e^{del_J})}{H(e^{\delta_1}, ..., e^{\delta_J})} \int_{u=-\infty}^{+\infty} \frac{1}{m} u \exp\{e^{-m(u - \frac{1}{m} lnH(e^{\delta_1}, ..., e^{\delta_J}))}\} e^{-m(u - \frac{1}{m} lnH(e^{\delta_1}, ..., e^{\delta_J}))} du \\
&= E[u] = \xi + \gamma\theta = \frac{1}{m} lnH(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J}) + \frac{1}{m}\gamma. \square
\end{aligned}$$

Proposition Flexibility of the continuous choice model. Consider the model $H(r) = \frac{1}{2} r'(I + B) r$, where $r_j = \exp\{\ln y - \ln p_j - \delta_j\}$, I is the identity matrix, and B is a symmetric matrix of parameters. This model can match any observed vector of market shares, matrix of cross price elasticities, and vector of income elasticities provided the data generating process satisfies additivity and slutsky symmetry.

Proof An algebraic functional form for a complete system of consumer demand functions, $s(p, y, \theta)$ is said to be flexible if at any given set of non-negative prices of commodities and income the parameters, θ, can be chosen so that the complete system of consumer demand functions, their own- and cross-price and income elasticities are capable of assuming arbitrary values at the given set of prices and commodities and income subject only to the requirements of theoretical consistency. (See Diewert (1974) or Lau (1986).)

Here I shall take the only requirements of theoretical consistency to be additivity and slutsky symmetry. We want to show that at an arbitrary point (p^*, y^*), if we observe some s^*, $\frac{\partial lns_j^*}{\partial lnp_k}$, and $\frac{\partial lns_j^*}{\partial lny}$ that satisfy additivity and slutsky symmetry, then we can always choose the parameters of the model, $\theta = (\delta, B)$ that satisfy the following equations:

$$s_j^* = s_j(p^*, y^*, \theta) \quad j = 1, ..., J \tag{6}$$

$$\frac{\partial lns_j^*}{\partial lnp_k} = \frac{\partial lns_j(p^*, y^*, \theta)}{\partial lnp_k} \quad j, k = 1, ..., J \tag{7}$$

$$\frac{\partial lns_j^*}{\partial lny} = \frac{\partial lns_j(p^*, y^*, \theta)}{\partial lny} \quad j = 1, .., J \tag{8}$$

Without loss of generality we can choose $(p^*, y^*) = (1, .., 1)$ since the physical units of each demand equation can be chosen. At that point, additivity from the budget constraint implies that $\sum_{j=1}^{J} s_j^* = 1$, $\sum_{j=1}^{J} \frac{\partial s_j^*}{\partial p_k} = -s_k^*$ and $\sum_{j=1}^{J} \frac{\partial s_j^*}{\partial y} = 1$, while slutsky symmetry ensures that

$$\frac{\partial s_j^*}{\partial p_k} + s_k^* \frac{\partial s_j^*}{\partial y} = \frac{\partial s_k^*}{\partial p_j} + s_j^* \frac{\partial s_k^*}{\partial y}.$$

First note, that if the true values of the observed demands and elasticities satisfy additivity and slutsky symmetry, then at any value of θ which satisfies Equations (6),(7), and (8), so will the model. Thus, we can seek values of θ which satisfy the equations when the model is constrained to satisfy these theoretical consistency constraints.

Hence, in terms of the model, the additivity constraints amount to $\sum_{j=1}^{J} s_j(p^*, y^*, \theta) = 1$, $\sum_{j=1}^{J} \frac{\partial s_j(p^*,y^*,\theta)}{\partial p_k} = -s_k(p^*, y^*, \theta)$ and $\sum_{j=1}^{J} \frac{\partial s_j(p^*,y^*,\theta)}{\partial y} = 1$ or,

$$\sum_{j=1}^{J} -\frac{\partial H(p^*, y^*, \theta)}{\partial p_j} = \frac{\partial H(p^*, y^*, \theta)}{\partial y} \tag{9}$$

$$-\sum_{j=1}^{J} \frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial p_k} - \frac{\partial H(p^*, y^*, \theta)}{\partial p_k} = \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y \partial p_k} \tag{10}$$

$$-\sum_{j=1}^{J} \frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial y} = \frac{\partial H(p^*, y^*, \theta)}{\partial y} + \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y^2} \tag{11}$$

Provided $H(\cdot)$ is chosen to be homogeneous of degree zero in (p, y) the model automatically satisfies all of these additivity constraints at every value of the parameters by Eulers Theorem.

Thus, establishing flexibility reduces to finding a value of θ so that the equations

$$\begin{aligned}
\left(\frac{\partial H(p^*, y^*, \theta)}{\partial y}\right) s_j^* &= -\frac{\partial H(p^*, y^*, \theta)}{\partial p_j} \\
\left(\frac{\partial H(p^*, y^*, \theta)}{\partial y}\right) \frac{\partial s_j^*}{\partial p_k} &= -\frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial p_k} - s_j^* \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y \partial p_k} \\
\left(\frac{\partial H(p^*, y^*, \theta)}{\partial y}\right) \frac{\partial s_j^*}{\partial y} &= -\frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial y} - s_j^* \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y^2}
\end{aligned}$$

are satisfied.

First, notice that these equations are all satisfied provided $\frac{\partial H(p^*,y^*,\theta)}{\partial y} = 1$,$-\frac{\partial H(p^*,y^*,\theta)}{\partial p_j} = s_j^*$ $\frac{\partial^2 H}{\partial p_j \partial p_k} = -\frac{\partial s_j^*}{\partial p_k} - 2 s_j^* s_k^* + s_j \frac{\partial s_k^*}{\partial y}$ since, via the additivity constraints this solution ensures that $\sum_{k=1}^{J} \frac{\partial^2 H}{\partial p_j \partial p_k} = \frac{\partial s_j^*}{\partial y} - s_j^*$, while $\frac{\partial^2 H}{\partial p_k \partial y} = -\frac{\partial s_k^*}{\partial y} + 2 s_k^*$, $\sum_{k=1}^{J} \sum_{j=1}^{J} \frac{\partial^2 H}{\partial p_j \partial p_k} = 0$, and $\frac{\partial^2 H}{\partial y^2} = -\sum_{j=1}^{J} \frac{\partial^2 H}{\partial p_j \partial y} - \frac{\partial H}{\partial y} = -\left(\sum_{j=1}^{J} \left(-\frac{\partial s_j^*}{\partial y} + 2 s_j\right)\right) - 1 = -2$.

Thus, it remains only to show that we can choose (δ, B) so that the predicted shares match the observed shares and $\frac{\partial^2 H}{\partial p_j \partial p_k}$ may be set in the fashion required by this solution. For the particular $H()$ function stated in the proposition,

$$\frac{\partial H}{\partial y} = \sum_{j=1}^{J}\left(e^{-2\delta_j} + \sum_{l=1}^{J} b_{jl}e^{-\delta_l}e^{-\delta_j}\right)$$

$$s_j^* = e^{-2\delta_j} + \sum_{l=1}^{J} b_{jl}e^{-\delta_l}e^{-\delta_j}$$

$$\frac{\partial^2 H}{\partial p_j \partial p_k} = b_{jk}e^{-\delta_j}e^{-\delta_k}$$

By choosing $b_{jk}e^{-\delta_j}e^{-\delta_k} = -\frac{\partial s_j^*}{\partial p_k} - 2s_j^* s_k^* + s_j \frac{\partial s_k^*}{\partial y}$ and $e^{-2\delta_j} = 2s_j^* - \frac{\partial s_j^*}{\partial y}$ it is easy to verify that each of these constraints are satisfied since $\sum_{l=1}^{J} b_{jl}e^{-\delta_j}e^{-\delta_l} = \frac{\partial s_j^*}{\partial y} - 2s_j^* + s_j^* = \frac{\partial s_j^*}{\partial y} - s_j^*$, $\sum_{j=1}^{J}\sum_{l=1}^{J} b_{jl}e^{-\delta_j}e^{-\delta_l} = \sum_{j=1}^{J}\left(\frac{\partial s_j^*}{\partial y} - s_j^*\right) = 0$, and $\sum_{j=1}^{J} e^{-2\delta_j} = 1$. Finally notice that at this solution B is a symmetric matrix provided slutsky symmetry holds.□

Proposition Flexibility of the discrete choice model. Consider the model $H(r) = \frac{1}{2}r(I+B)r$, where $r_j = \exp\{\ln y - \ln p_j - \delta_j\}$, B is a matrix of parameters with jk^{th} element b_{jk}. This model can match any observed vector of market shares, matrix of cross price elasticities, and vector of income elasticities provided the data generating process satisfies additivity and homogeneity of degree zero in income and prices.

Proof

I take the only requirements of theoretical consistency to be additivity and homogeneity. Given these constraints, we want to show that at an arbitrary point (p^*, y^*), if we observe some s^*, $\frac{\partial lns_j^*}{\partial lnp_k}$, and $\frac{\partial lns_j^*}{\partial lny}$ that satisfy the additivity and homogeneity conditions, then we can always choose the parameters of the model $\theta = (\delta, B)$ that satisfy the following equations:

$$s_j^* = s_j(p^*, y^*, \theta) \quad j = 1, ..., J \tag{12}$$

$$\frac{\partial lns_j^*}{\partial lnp_k} = \frac{\partial lns_j(p^*, y^*, \theta)}{\partial lnp_k} \quad j,k = 1, ..., J \tag{13}$$

$$\frac{\partial lns_j^*}{\partial lny} = \frac{\partial lns_j(p^*, y^*, \theta)}{\partial lny} \quad j = 1, .., J \tag{14}$$

or, in terms of the model $s_j^* = -\frac{r_j}{H}\frac{\partial H}{\partial r_j}$, $\frac{\partial s_j^*}{\partial p_k} = \left(\frac{r_j r_k H_{jk}}{H} - \frac{r_j r_k H_j H_k}{H^2}\right)\frac{1}{p_k} + \frac{r_j H_j}{H}\frac{1\!\!1(j=k)}{p_k}$ and $\frac{\partial s_j^*}{\partial y} = \frac{r_j}{H}\frac{\partial H_j}{\partial y} - \frac{r_j H_j}{H}\frac{1}{H}\frac{\partial H}{\partial y} + \frac{H_j r_j}{H}\frac{1}{y}$.

Additivity of the market shares implies that $\sum_{j=1}^{J} s_j^* = 1$, $\sum_{j=1}^{J}\frac{\partial s_j^*}{\partial p_k} = 0$ and $\sum_{j=1}^{J}\frac{\partial s_j^*}{\partial y} = 0$, while the fact that the true demands are homogeneous of degree zero in (y, p) implies that $\sum_{k=1}^{J} p_k \frac{\partial s_j^*}{\partial p_k} = -y\frac{\partial s_j^*}{\partial y}$

Without loss of generality we can choose $(p^*, y^*) = (1, .., 1)$ since the physical units of each demand equation can be chosen. At that point, the equations we must satisfy becomes

$$\begin{aligned} s_j^* &= e^{-2\delta_j} + \sum_{l=1}^{J} b_{jl} e^{-\delta_j} e^{-\delta_l} \\ \frac{\partial s_j^*}{\partial p_k} &= b_{jk} e^{-\delta_j} e^{-\delta_k} - s_j^* s_k^* + s_j^* \mathbb{I}(j=k) \\ \frac{\partial s_j^*}{\partial y} &= \sum_{l=1}^{J} e^{-\delta_j} e^{-\delta_l} b_{jl} \end{aligned}$$

Choosing $e^{-2\delta_j} = s_j^* - \frac{\partial s_j^*}{\partial y}$ and $b_{jk} e^{-\delta_j} e^{-\delta_k} = \frac{\partial s_j^*}{\partial p_k} + s_j^* s_k^* - s_j^* \mathbb{I}(j=k)$ it is easy to see that all three sets of equations are satisfied. Thus, the model provides a flexible functional form.□

Lemma 4 *(see McFadden (1981), p 266.)*

If $\mathbf{v}_{\mathcal{J}} = (v_1, ..., v_J)$ has a conditional c.d.f., $F(\mathbf{v}_{\mathcal{J}}; w, y-p, p_o, c)$, with first moments, then

$$E \max_{j \in \mathcal{J}} (v_j - \delta_j) = E \max_{j \in \mathcal{J}} v_j + \int_{-\infty}^{\infty} [F(\mathbf{0}_{\mathcal{J}} + t) - F(\delta_{\mathcal{J}} + t)] dt$$

Proof

$$\begin{aligned} \int_{-\infty}^{\infty} [F(\mathbf{0}_{\mathcal{J}} + t) - F(\delta_{\mathcal{J}} + t)] dt &= \lim_{M \to \infty} \int_{-M}^{M} [F(\mathbf{0}_{\mathcal{J}} + t) - F(\delta_{\mathcal{J}} + t)] dt \\ &= -\lim_{M \to \infty} [t(F(\delta_{\mathcal{J}} + t) - F(0_{\mathcal{J}} + t))]_{-M}^{M} \\ &\quad + \lim_{M \to \infty} \int_{-M}^{M} t \frac{d}{dt}(F(\mathbf{0}_{\mathcal{J}} + t) - F(\delta_{\mathcal{J}} + t)) dt \\ &= \int_{-\infty}^{+\infty} t \frac{dF(\delta_{\mathcal{J}} + t)}{dt} dt - \int_{-\infty}^{+\infty} t \frac{dF(0_{\mathcal{J}} + t)}{dt} dt \\ &= E \max_{j \in \mathcal{J}} (v_j - \delta_j) - E \max_{j \in \mathcal{J}} v_j \end{aligned}$$

where the last equality follows from the definition of the expectation of the maximum component of a random vector.[27]

Proof of Theorem 1 (McFadden (1981), p 260- p268)

(i) Consider the additive random utility model,

$$\max_{j \in \mathcal{J}} v_j - \delta_j$$

[27]Let $(X_1, .., X_n)$ be a random vector with joint c.d.f. $F_X(x)$ and define $Y = \max_{i=1,..,n} X_i - \delta_i$ with cdf $F_Y(t)$. Then $F_Y(t) = Pr\{Y \leq t\} = Pr\{X_1 \leq \delta_i + t, .., X_n \leq \delta_n + t\} = F_X(\delta_1 + t, .., \delta_n + t)$ and hence the density of Y is $\frac{dF_Y(t)}{dt} = \frac{dF_X(\delta_1+t,..,\delta_n+t)}{dt}$ and hence $E[Y] = \int t \frac{dF_X(\delta_1+t,..,\delta_n+t)}{dt} dt$.□

where $\mathbf{v}_{\mathcal{J}} = (v_1, .., v_J)$ is distributed in the population with conditional cumulative distribution function, $F(\mathbf{v}_{\mathcal{J}}; w, y, \text{p}, p_0, c, \mathcal{J})$ and conditional density function $f(\mathbf{v}_{\mathcal{J}}; w, y, \text{p}, p_0, c, \mathcal{J})$. For notational simplicity, for the remainder of this section, I leave implicit the dependence of $\mathbf{v}_{\mathcal{J}}$, $F(\mathbf{v}, \cdot)$ and $f(\mathbf{v}, \cdot)$ on the vector $(w_j, y, p_j, p_0, c, \mathcal{J})$ $j = 1, .., J$.

Then,

$$\begin{aligned} Pr\{j|w, y, \text{p}, p_0, c, \mathcal{J}\} &= Pr\{\mathbf{v}_j - \delta_j \geq \mathbf{v}_k - \delta_k \text{ for all } k \in \mathcal{J}\} \\ &= \int_{v_i=-\infty}^{+\infty} \int_{v_2=-\infty}^{v_i-\delta_i+\delta_2} \cdots \int_{v_J=-\infty}^{v_i-\delta_i+\delta_J} f(\mathbf{v}_{\mathcal{J}}) d\mathbf{v}_{\mathcal{J}} \end{aligned}$$

Hence the RUM generates a system of choice probabilities which are non-negative, sum to one, and depend only on the variables $(w, y, \text{p}, \text{p}_0, c, \mathcal{J})$ through $\mathbf{v}$.

By lemma (4), $V(\delta) = E[\max_{i \in \mathcal{J}} v_i] + G(\delta)$, where $G(\delta) \equiv \int_{-\infty}^{\infty} [F(\mathbf{0}_{\mathcal{J}} + t) - F(\delta_{\mathcal{J}} + t)] dt$. Moreover, since $F()$ has a first moment, $E \max_{j \in \mathcal{J}} v_j$ exists, is a real valued function and is constant with respect to δ, thus $V(\delta)$ is a member of the EMRU class of functions if and only if G is a member of the EMRU class of functions.[28]

Defining $G(\delta)$ as above, McFadden (1981) first shows that G exists.[29]

---

[28]Notice, also that if $G(\delta)$ is differentiable (shown below), then $V_i(\delta) = G_i(\delta)$.

[29]Let F^i denote the marginal cdf of v_i. If $\lambda = \max_{i \in \mathcal{J}} |\delta_i - \delta_i'|$, then

$$F(t + \delta) - F(t + \delta + \lambda) \quad \leq \quad F(t + \delta) - F(t + \delta') \leq F(t + \delta) - F(t + \delta - \lambda)$$

implying that $|F(t + \delta) - F(t + \delta')| \leq F(t + \delta + \lambda) - F(t + \delta - \lambda)$. Since F is a cdf,

$$F(t + \delta + \lambda) - F(t + \delta - \lambda) \leq \sum_{i=1}^{M} [F^i(t + \delta_i + \lambda) - F^i(t + \delta_i - \lambda)].$$

For any scalar, $M \geq 0$, and positive integer K,

$$\begin{aligned} &\int_M^{M+K\lambda} [F^i(t + \delta_i + \lambda) - F^i(t + \delta_i - \lambda)] dt \\ &= \textstyle\sum_{k=1}^{K} \int_{M+(k-1)\lambda}^{M+k\lambda} [F^i(t + \delta_i + \lambda) - F^i(t + \delta_i - \lambda)] dt \\ &= \textstyle\int_{M+(K-1)\lambda}^{M+(K+1)\lambda} F^i(t + \delta_i) dt - \int_{M+\lambda}^{M-\lambda} F^i(t + \delta_i) dt \\ &\leq 2\lambda \{F^i(M + (K+1)\lambda + \delta_i) - F^i(M - \lambda + \delta_i)\} \end{aligned}$$

Letting $K \to +\infty$, these imply that

$$\int_M^{\infty} |F(t + \delta) - F(t + \delta')| dt \leq 2\lambda \sum_{i=1}^{m} (1 - F^i(M - \lambda + \delta_i).$$

A similar argument yields,

$$\int_{-\infty}^{-M} |F(t + \delta) - F(t + \delta')| dt \leq 2\lambda \sum_{i=1}^{m} F^i(-M + \lambda + \delta_i).$$

Taking $\delta' = O_{\mathcal{J}}$ and $M = 0$ implies

$$\int_{-infty}^{+\infty} |F(t + \delta) - F(t)| dt \leq 4m \max |\delta_i|$$

Next show that $G(\delta)$ (and consequently V) satisfies the additivity property of an EMRU function. For $\theta > 0$ (An analogous argument establishes the result for $\theta < 0$.),

$$\begin{aligned} G(\delta) - G(\delta+\theta) &= \lim_{K\to\infty} \sum_{i=-K}^{K-2} \int_{i\theta}^{(i+1)\theta} [F(t+\delta+\theta) - F(t+\delta)]dt \\ &= \lim_{K\to\infty} \left\{ \int_{(K-1)\theta}^{K\theta} [F(t+\delta)dt - \int_{(1-K)\theta}^{-K\theta} [F(t+\delta)dt \right\} \\ &= \theta. \end{aligned}$$

Mcfadden shows that G is differentiable with[30]

$$\begin{aligned} G_i(\delta) &= -\int_{+\infty}^{\infty} F(\delta+t)dt \\ &= -\int_{+\infty}^{\infty} F(\delta - \delta_i + t)dt \\ &= -\int_{v_i=-\infty}^{+\infty} \int_{v_2=-\infty}^{v_i-\delta_i+\delta_2} \cdots \int_{v_J=-\infty}^{v_i-\delta_i+\delta_J} f(v_{\mathcal{J}};\cdot)d\mathbf{v}_{\mathcal{J}} \\ &= -Prob[v_i - \delta_i \geq v_j - \delta_j \text{ for } j \in \mathcal{J}] \\ &= -Prob[i|\cdot] \end{aligned}$$

with the second equality following a change in the variable of integration from t to $t-\delta_i$. Since $G_i() = -\int_{-\infty}^{+\infty}\int_{v_2=-\infty}^{v_i-\delta_i+\delta_2}\cdots\int_{v_2=-\infty}^{v_i-\delta_i+\delta_2} f(\mathbf{v}_{\mathcal{J}})d\mathbf{v}_{\mathcal{J}}$, the mixed partial derivatives of G exist and are non-positive and independent of the order of differentiation. Moreover, since $G_i = -Pr[i|\cdot]$, it follows immediately that $\sum_{i\in\mathcal{J}} G_i = -1$.

(Aside: *Additional sufficient conditions for convexity in* δ*:* If $G_{ij} = \frac{\partial G_i}{\partial \delta_j} \leq 0$, $\sum_{i\in\mathcal{J}} G_{ij} = 0$, and hence $G_{jj} = -\sum_{i\neq j} G_{ij}$. Hence the Hessian of G has a weakly dominant positive diagonal and G is therefore convex in δ.)

Hence G defined above exists.

[30]Differentiability: For $\delta = \delta' + \theta\delta''$ and $\lambda = \max_i |\delta_i''|$,

$$\left| \frac{G(\delta'+\theta\delta'') - G(\delta')}{\theta} + \int_{-M}^{M} \frac{F(\delta'+\theta\delta''+t) - F(\delta'+t)}{\theta} dt \right|$$
$$\leq 2\lambda \sum_{i=1}^{m} [1 - F^i(M - \lambda + \delta_i') + F(-M + \lambda + \delta_i')].$$

The right-hand side of this inequality converges to zero as $M \to +\infty$, uniformly in θ. For each M the left hand side converges to

$$\left| \frac{G(\delta'+\theta\delta'') - G(\delta')}{\theta} + \int_{-M}^{M} \sum_{i=1}^{m} F_i(\delta'+t)\delta_i'' dt \right|$$

since F has a density and is therefore differentiable. This establishes that G is differentiable.

Proof of Converse Suppose $V(\delta) \in \mathcal{V}$, the class of functions with EMRU properties. Define

$$\pi_j(\delta) \equiv -V_j, \text{ and} \tag{15}$$
$$F(\mathbf{v}_{\mathcal{J}}) \equiv \int_{-\infty}^{v_1} \pi_1(0, v_2 - t, ..., v_J - t)\phi(t)dt \tag{16}$$

where $\phi(t)$ is an arbitrary univariate probability density. First note that $\sum_{j=1}^{J} V_j = -1$ follows from the linearity property since

$$\frac{dV(\delta+\theta)}{d\theta} = \sum_{j=1}^{J} \frac{\partial V}{\partial \delta_j} = \frac{dV(\delta)}{d\theta} - 1 = -1$$

where the first equality follows from the definition of the total differential with respect to θ, the second from the linearity property, and the final equality since the first term on the left hand side is independent of θ. Thus $(\pi_j)_{j=1}^{J}$ defines a probability choice system.

Next, we show that $F()$ is a cumulative distribution function. Since $\lim_{\delta_j \to -\infty} V_j(\delta) = -1$ for all $j \in \mathcal{J}$ we know $\lim_{v_1 \to -\infty} \pi_1(\mathbf{v}_{\mathcal{J}}) = +1$. Consequently $\lim_{v_1 \to -\infty} F(\mathbf{v}_{\mathcal{J}}) = \lim_{v_1 \to -\infty} \int_{-\infty}^{v_1} \pi_1(0, v_2 - t, ..., v_J - t)\phi(t)dt = 0$. In addition,

$$\begin{aligned}
\lim_{\mathbf{v}_{\mathcal{J}} \to +\infty} F(\mathbf{v}_{\mathcal{J}}) &= \lim_{\mathbf{v}_{\mathcal{J}} \to +\infty} \int_{-\infty}^{v_1} \pi_1(0, v_2 - t, ..., v_J - t)\phi(t)dt \\
&= \lim_{v_1 \to +\infty} \int_{-\infty}^{v_1} \pi_1(0, +\infty, ..., +\infty)\phi(t)dt \\
&= \lim_{v_1 \to +\infty} \int_{-\infty}^{v_1} \pi_1(0, +\infty, ..., +\infty)\phi(t)dt \\
&= \pi_1(0, +\infty, ..., +\infty) \lim_{v_1 \to +\infty} \int_{-\infty}^{v_1} \phi(t)dt \\
&= \pi_1(0, +\infty, ..., +\infty) \\
&= 1
\end{aligned}$$

where the latter equality follows since $V(\delta_1, +\infty, ..., +\infty) = V_1(0, +\infty, ..., +\infty) - \delta_1$ by additivity, and hence $\pi_1(0, +\infty, ..., +\infty) = -\frac{\partial V}{\partial \delta_1} = 1$.

Since $V \in \mathcal{V}$ has all mixed partial derivatives positive,

$$F_{1...J}(\mathbf{v}) = \pi_{1,2,...,m}(0, v_2 - v_1, ...v_J - v_1)\phi(v_1) \geq 0$$

and so this is a joint density function for $\mathbf{v}$, and hence Equation (16) defines a cumulative distribution function. (Aside note to self: Note that this looks like a slightly strong set of sufficient conditions for $F()$ to be a c.d.f.)

Now consider the function, $\tilde{V}$, that would be generated by an additive random utility model with this cdf, $F()$.

$$
\begin{aligned}
\tilde{V}(\delta) &\equiv \int_{-\infty}^{+\infty} F(t+0_{\mathcal{J}}) - F(t+\delta)dt \\
&= -\delta_1 - \int_{-\infty}^{+\infty} F(t+\delta-\delta_1) - F(t+0_{\mathcal{J}})dt \\
&= -\delta_1 - \int_{-\infty}^{\infty}\int_{-\infty}^{t} (\pi_1(0, t+\delta_2-\delta_1-\tau, ..., t+\delta_J-\delta_1-\tau) - \pi_1(0, t-\tau, ..., t-\tau))\phi(t)d\tau dt \\
&= -\delta_1 - \int_{\tau=-\infty}^{+\infty}\int_{0}^{s=\infty} (\pi_1(0, \delta_2-\delta_1+s, ..., \delta_J-\delta_1+s) - \pi_1(0, s, ..., s))\phi(s+t)dsdt \\
&= -\delta_1 - \int_{0}^{+\infty} (\pi_1(0, \delta_2-\delta_1+s, ..., \delta_J-\delta_1+s) - \pi_1(0, s, ..., s)) \int_{t=-\infty}^{+\infty} \phi(s+t)dtds \\
&= -\delta_1 + [V(-t, 0..., 0) - V(-t, \delta_2-\delta_1, ..., 1_m-\delta_1)]_0^{+\infty} \\
&= -\delta_1 + V(0, \delta_2-\delta_1, ..., \delta_m-\delta_1) = V(\delta)
\end{aligned}
$$

where the first equality follows since $\tilde{V}$ has the additive property by construction. The second equality follows from the definition of $F()$, evaluated at $t+\delta-\delta_1$. The third equality follows from a change of variable $s = t-\tau$. The fourth since the order of integration is exchangeable and the fifth since $\int_{-\infty}^{+\infty} \phi(t+s)dt = 1$ for all s and $\frac{\partial V(\delta)}{\partial \delta_1} = -\pi_1()$. Thus, $\tilde{V}(\delta)$ defined by this ARUM form equals $V(\delta)$.□

Lemma If a $(J\text{x}J)$ matrix M is a symmetric matrix with the gross substitutes sign pattern (negative diagonal elements and positive off diagonal elements) and we have $M\alpha = M'\alpha = 0$ for some vector of constants $\alpha >> 0$, then $\hat{M}$ is negative definite, where $\hat{M}$ is any $(J-1)\text{x}(J-1)$ matrix obtained from M by deleting any row and the corresponding column. (See for example, Theorem M.D.5 in Mas-Colell, Whinston, and Green (1995).)

Proposition Set $\delta_1 = 0$. If the mixed partial derivatives $\frac{\partial^2 V}{\partial \delta_j \partial \delta_k}$ $j, k = 2, .., J$ are negative for all $j \neq k$ and positive for $j = k$, then $V(\delta_1, \delta_{-1})$ is strictly convex in the $((J-1)\text{x}1)$ vector δ_{-1}.

Proof Define the functions $f_j(\delta, s_j) = s_j(\delta) - s_j = -\frac{\partial V}{\partial \delta_j} - s_j$ for $j = 1, .., J$. Notice that the matrix of derivatives of this function $\frac{\partial f_j}{\partial \delta_k} = -\frac{\partial^2 V}{\partial \delta_j \delta_k}$. Moreover, the column and row sums of the matrix of second derivatives are zero since $\sum_{j=1}^{J} s_j(\delta) = 1$. Since the matrix of second derivatives of V have negative off diagonal elements, then the diagonal elements must be positive. Applying the lemma above, if we normalize $\delta_1 = 0$ and delete the first row and column of the matrix of second derivatives of V to construct the matrix $\hat{M} = -\frac{\partial^2 \hat{V}}{\partial \delta_j \partial \delta_k}$ $j, k = 2, .., J$. Then $-\hat{M}$ is must be positive definite so that the function $V(\delta_1, \delta_{-1})$ is consequently strictly convex in δ_{-1}. □.

References

Anderson, S. P., A. de Palma, and J. F. Thisse (1992): *Discrete Choice Theory of Product Differentiation.* MIT Press.

Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (1997): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," NBER Working Paper # 6481.

Berry, S., and A. Pakes (1999): "Estimating the Pure Hedonic Discrete Choice Model," Discussion paper, Yale University.

Bresnahan, T., S. Stern, and M. Trajtenberg (1997): "Market Segmentation and the Sources of Rents From Innovation: Personal Computers in the late 1980s," *RAND.*

Brown, B., and M. Walker (1989): "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica*, 59, 815–829.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975): "Transcendental Logarithmic Utility Functions," *American Economic Review*, 65, 367–83.

Daly, A., and S. Zachary (1979): *Identifying and Measuring the Determinants of Mode Choice*chap. Improved Multiple Choice Models. London: Teakfield.

Deaton, A., and J. Muellbauer (1980): "An Almost Ideal Demand System," *The American Economic Review*, pp. 312–326.

Diewert, W. E. (1974): "Applications of Duality Theory," in *Frontiers of Quantitative Economics*, ed. by M. D. Intriligator, and D. A. Kendrick, vol. II, chap. 3. North Holland.

Ellison, S. F., I. Cockburn, Z. Griliches, and J. Hausman (1997): "Characteristics of Demand for Pharmaceutical products: an examination of four cephalosporins," *Rand Journal of Economics*, 28(3), 426–446.

Hausman, J., and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(2).

Hausman, J. A. (1994): "Valuation of New Goods Under Perfect and Imperfect Competition," *NBER Working Paper # 4970.*

Houthakker, H. (1960): "Additive Preferences," *Econometrica*, 28(2), 244–257.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995): *Continuous Univariate Distributions*, vol. 2. John Wiley & Sons, Inc., second edn.

Judd, K. (1999): *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

Lancaster, K. (1966): "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74(2), 132–155.

Lancaster, P., and M. Tismenetsky (1985): *The Theory of Matrices*, Computer Science and Applied Mathematics. Academic Press, 2 edn.

Lau, L. (1986): *Handbook of Econometrics*vol. 3, chap. Functional Forms in Econometric Model Building.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*. Oxford University Press.

Mcfadden, D. (1978): *Spatial Interaction Theory and Planning Models*chap. Modelling The Choice of Residential Location. North Holland, Amsterdam.

McFadden, D. (1981): "Structural Analysis of Discrete Data with Econometric Applications," in *Economic Models of Probabilistic Choice*, ed. by C. Manski, and D. McFadden. MIT Press, Cambridge Mass.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

McFadden, D., and K. Train (1998): "Mixed MNL Models For Discrete Response," Mimeo, University of California at Berkeley.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Pinkse, J., M. E. Slade, and C. Brett (1997): "Spatial Price Competition: A Semiparametric Approach," *University of British Columbia Discussion Paper No; 97-15.*

Pudney, S. (1989): *Modelling Individual Choice*. Blackwell.

Slade, M. (2001): "Assessing Market Power in UK Brewing," *Mimeo, University of British Columbia.*

Small, K. A., and H. S. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1).

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Frontiers Of Economic Research. Princeton University Press.

Varian, H. (1984): *Microeconomic Analysis*. Norton, New York, USA, 2 edn.

Williams, H. (1977): "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," *Environment Planning*, A(9), 285–344.