New Parametric Demand Systems for Market Level Demand Studies. Preliminary and Incomplete!*

By
Peter Davis †
MIT Sloan

First version: October 2000.

The continuous choice and discrete choice demand literatures have largely evolved separately. However, the models resulting from each literature have very distinct advantages for taking to market level data. I build on the results provided by McFadden (1981) to establish a common framework for developing market level demand systems from either discrete or continuous choices and use the framework to propose new continuous and discrete choice demand models. The resulting 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. Moreover, variation in the observed set of products can be used to help identify substitution patterns between goods in a way recently made popular in the discrete choice demand literature. The proposed discrete choice models are consistent with an underlying discrete choice random utility model. Moreover, the models are flexible functional forms in the sense of being capable of matching any observed market shares, price elasticities, and income elasticities of demand. Approaching aggregate models of discrete choice data in this way is computationally much less demanding than existing approaches because it does not require the use of simulation estimators.

*Thanks are due to Arthur Lewbel and Tom Stoker 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 and have lead to two large empirical literatures, emphasizing continuous- and discrete-choice models respectively. While each class of models is well refined, there remain stark differences between the properties of the existing discrete and continuous choice models. Differences that arise primarily because of their historical areas of application and which suggest productive avenues for further development of the models.

To illustrate these differences, consider the fact that representative agent continous choice demand models are rich enough in parameters that they are flexible functional forms. In contrast, the discrete choice demand literature has resorted 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. On the other hand, discrete choice demand models like the logit can be estimated using datasets where significant product entry and exit occurs, unlike the continuous-choice models like the Translog or Almost ideal demand system. As a result, existing applications of continuous-choice models are largely limited to considering substitution between broad aggregates of goods (eg., food and transportation,) a level of data aggregation which eliminates product entry and exit. This feature of existing continuous choice models presently limits application of these techniques in many areas of marketing and industrial organization where it is not generally possible to aggregate data to a level which removes entry and exit of products. Moreover, this feature of existing continuous-choice models is particularly undesirable since variation in the set of choices available to consumers can provide important information about the substitutability of products; a source of data variation which has been used very effectively in the discrete choice literature (see for example Berry, Levinsohn, and Pakes (1995).)

The aim of this paper is to develop market level parametric discrete- and continuous-choice demand models which have the advantages of both sets of existing models in those literatures.

Continuous choice parametric demand systems are usually generated by specifying a flexible functional form for the indirect utility function and deriving demand equations via Roy's Identity (See Varian (1984) for example.) In contrast, discrete choice demand systems are universally specified using a parametric model for the direct utility function. These starkly different approaches persist in part because an exact equivalence between specifying an indirect and specifying a direct utility function, provided by the duality results, for continuous choice models is not 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 analogous to the continuous choice indirect utility function specification for discrete choice models. In this paper I explore this alternative approach to specifying functional forms for the discrete choice analogue of an indirect utility function. In addition, a natural new functional form for parametric continuous choice demand models emerges with some nice global properties that make it particularly appropriate for market level demand studies of continuous choice data.

Throughout, I attempt to draw out the common features of these apparently disparate literatures. In doing so, I follow and expand on McFadden (1981) andAnderson, de Palma, and Thisse (1992) who emphasize that a continuum of consumers making discrete choices will, under some circumstances, generate the same demand system as a single "representative" consumer making continuous choices. In those cases, these apparently different classes of underlying behavioural choice models can be observationally equivalent.

More specifically, I develop a class of demand generating functions and identify the conditions under which these functions are (i) indirect utility functions and therefore generate continuous choice demand systems and (ii) consistent with an underlying additive random utility discrete choice model. In addition, I 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.

The proposed demand systems share many of the theoretical advantages of each class of demand models. First, unlike most discrete choice models, the proposed discrete-choice demand system has flexible functional form properties in the sence of Diewert (1974), without resorting to the introduction of multiple types of consumers.[1] Throughout, the aim is to develop a concrete empirical strategy based on the observation, implicit in Mcfadden (1978) and explicit in Pudney (1989), 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.[2]

Second, the demand system can be estimated on disaggregated data in which new products are introduced and old products exit. This is in contrast to existing continuous choice models which are typically estimated using broadly aggregated product categories

[1]The advantage of the introduction of this heterogeneity, perhaps using random coefficient models, is the added flexibility in substitution patterns. 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 asymptototic properties of the resulting simulation estimators requires more advanced mathematical arguments than typically taught in the most advanced graduate level econometrics courses (see Pakes and Pollard (1989).)

[2]The closest work in spirit to this paper that I'm aware of is the model proposed by Bresnahan, Stern, and Trajtenbeg (1997) who empirically implement a specification of Mcfadden's Generalised Extreme Value model which is more flexible than the multi-nomial logit model typically used. Thier formulation however, requires the researcher to specify restrictions on the possible substitution patterns between goods, although the restrictions can be tested.

to avoid such introduction and exit of products.[3] The few exceptions to this general rule have involved market level data with some very special characteristics.[4] Moreover, this multiple entry and exit of products has successfully been used as pseudo-price variation in the discrete choice literature[5] but is unavailable in the present generation of continuous choice models.

Finally, in comparison to random coefficient discrete choice models, are simple to compute because they do not *require* estimation via simulation (see Pakes and Pollard (1989) and McFadden (1989).) That said, most standard discrete choice random coefficient models, such as the mixed multinomial logit model (MMNL), are also nested within the framework. The analoguous continous choice models illustrate the way random coefficient models can be introduced to the continous choice literature.[6]

The rest of the paper is as follows. In section 1 I outline the proposed demand system generating function. In section 2, I briefly summarize the existing approaches in the demand literature and introduce the notation I use throughout. In section 3 I briefly introduce the demand system generating function. In section 4, I find conditions under which the new demand system is consistent with a representative consumer making continuous choices. In section 5 I demonstrate the conditions under which the proposed system is consistent with an underlying additive random utility model. In section 6, I demonstrate a practical and fast ways to estimate the respective discrete and continuous

[3] Below, I show that this is because the presently available flexible functional forms do not satisfy a desirable global regularity property.

[4] For example, Hausman (1994) and Ellison, Cockburn, Griliches, and Hausman (1997) each estimate variants of the Almost Ideal Demand System proposed by Deaton and Muellbauer (1980). However, in both cases, the demand system must be estimated using the data only from the period when *all* goods are present in the market.. Naturally, in dynamic markets with large numbers of products this is not always an option since products frequently enter and exit simultaneously.

[5] Altering the choice sets facing consumers tells the researcher a considerable amount about the substitutability of products.

[6] Although it is important to bear in mind that one important advantage of a model that allows for unobservably heterogeneous consumers is no longer true since a representative agent discrete choice model can account for flexible substitution patterns.

choice models. In section 7 I establish some flexibility properties of the model. Finally I conclude.

2 Previous Literatures and some Notation

Early demand studies endowed consumers with preferences over products directly and assumed that consumers solved a utility maximization problem subject to a budget constraint

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

where p denotes the vector of prices, y denotes the consumers' income. The solution to this problem is a vector of demand equations for each product, $x(\mathrm{p}, y; \theta)$. Standard duality results establish conditions on the function $V(\cdot)$ sufficient to ensure that one could equivalently specify a parametric form for the indirect utility function, $V(p, y; \theta)$ and solve for the demand system using Roy's identity.[7] By taking this dual approach, the resulting parametric demand systems were assured to be consistent with utility maximization, at least for some parameter values. In addition, the resulting demand systems were easy to estimate, capable of generating flexible substitution patterns, but successfully avoided the explicit solution to the non-linear direct utility maximization problem.

When the consumer must make a discrete choice from a set of alternatives, this may be represented by the additional constraints on the problem, that $x_j x_k = 0$ for all $j \neq k$ (see Small and Rosen (1981).) Imposing this discreteness and maximizing over the set

[7] 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.

of continuous choices, which typically includes the option of buying *only* the outside option (since $y - p_j I(j > 1)$ is spent on the outside good if option j is chosen), yields a set of conditional indirect utility functions which I denote $(v_1, .., v_m)$. Using the budget constraint to determine the quantity of the outside good purchased provides a discrete choice model

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

where, naturally $\mathcal{J}$ includes the option to purchase none of the inside goods, here denoted by choice 1.

Standard empirical approaches to specifying empirical models for this discrete choice problem, such as the multinomial logit or probit models, specify a parametric functional form for the conditional indirect utility functions of each type of consumer c, perhaps, $v_j(y - p_j I(j > 1), p_1, c, \theta) = v_j(y - p_j I(j > 1), p_1, w_j, \theta) + \epsilon_j$ for each consumer type $c = (\epsilon_1, .., \epsilon_J)$ and posit a parametric distribution of consumer types in the population. The consumer is assumed to know his own type 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 $r_j = \exp\{\psi_j(y, p_j, p_1, \delta_j)\}$. Although many functional forms are possible, for concreteness, I will consider specifications that include a generalized linear form for $H(\cdot)$ (see Diewert (1971)) and a quadratic form for $H(\cdot)$

$$H(r_1,\ldots,r_J) \equiv \sum_{j=1}^{J} a_j r_j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b_{jk} r_j^{\frac{1}{2}} r_k^{\frac{1}{2}} = a'r + \frac{1}{2} r^{\frac{1}{2}\prime} B r^{\frac{1}{2}}, or \quad (2)$$

$$H(r_1,\ldots,r_J) \equiv \sum_{j=1}^{J} a_j r_j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} b_{jk} r_j r_k \quad (3)$$

where $a = (a_1, \ldots, a_J)$, B is a JxJ matrix with jk^{th} element b_{jk}, $r = (r_1, .., r_J)$, $r^{\frac{1}{2}} = (r_1^{\frac{1}{2}}, .., r_J^{\frac{1}{2}})$, and $\psi_j()$ is a known function. Again, for concreteness, specifications of ψ that will be of particular interest include

$$\psi_j(y, p_j, p_1, \delta_j) = \frac{y - p_j I(j > 1)}{p_1} - \delta_j, \text{ or}$$
$$\psi_j(y, p_j, p_1, \delta_j) = lny - lnp_j - \delta_j$$

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, 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 A Continuous Choice Demand System

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.

If $V(\mathrm{p}, y, w, s)$ has the properties of an indirect utility function then standard duality results imply that the demand system is easily obtained via Roy's identity.

Proposition 1 *Let $V(r) = \frac{1}{m} ln(H(r))$ with $r_j = \exp\{\psi_j(y, p_j, \delta_j)\}$ and $H(r)$ is an homogeneous of degree m function, continuous, and non-decreasing in r. 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 and non-decreasing in y*
3. *$\frac{\partial^2 H}{\partial r_j \partial r_k} \leq 0 \; \forall \; j \neq k$, and*
4. *$r_j(p_j, y)$ is a convex function of p_j, with y normalized to 1.*

then, $V(\mathrm{p}, y) = \frac{1}{m} ln(H(r(\mathrm{p}, y, \cdot), \cdot))$ is an indirect utility function.

Proof We must establish each of the properties of the indirect utility function in turn.

1. V is clearly a continuous function at all postive prices and incomes provided H is. In turn, $H()$ is a has the appropriate continuity properties provided each r_j does.
2. Since $H(r)$ is non-decreasing in r, r is non-increasing in p and increasing in y, $H(\cdot)$ and hence $V(\cdot)$ is non-increasing in p and non-decreasing in y.

3. By construction, $\psi_j(\cdot)$, is homogeneous of degree zero in (y, p_j). Trivially therefore, each ψ_j is homogenous of degree zero in (y, p). Hence $V()$, is homogeneous of degree zero in (y, p).

4. Convexity, and thus Quasi-convexity, follows from the result in **Lemma 2** under the conditions above.[6]

4.1 Advantages of the "new" demand system

Many existing demand systems are closely related to the broad class of indirect utility functions considered here. For example, the "indirect addilog" model considered by Houthakker (1960) sets all elements in the matrix B to zero in the generalized linear form and sets $r_j(y, p_j) = \frac{(p_j/y)^{\beta_j+1}}{\beta_j+1}$ where β_j are parameters. Similarly, the Log Translog model would be obtained by setting $r_j(y, p_j) = log(\frac{p_j}{y})$ in 3 and imposing restrictions on the parameters (a, B) to ensure that $V()$ is convex in the ratio, $\frac{p_j}{y}$.

One global regularity property of the model, that arises with an appropriate choice of functions for r_j that is worth emphasizing, is that $\lim_{p_j \to \infty} V(p_j, \mathrm{p}_{-j}, \cdot) = V(p_{-j}, \cdot)$ provided $\lim_{p_j \to \infty} r_j = 0$. If this condition holds, removing goods from the model is equivalent to increasing their prices to infinity. This means that the model may be estimated using the psuedo-variation in prices that results from alterations over time in the set of goods available to consumers. In the discrete choice literature, this source of variation has proven very useful for identifying paramters in demand systems. Because the demand models previously available from the continuous choice literature do not

[6] An alternative, direct, proof of quasi-convexity is also available for the generalized linear form. Specifically, B symmetric with non-positive off diagonal elements is necessary and sufficient for the global convexity of $H(r)$ in r (see Diewert (1974) for example.) Normalize one element in (y, p). Since $H(r)$ is a non-decreasing convex function of r, provided r is a convex function of the normalized (p, y), $H(\mathrm{p}, y)$ is convex in the normalized (p, y). Finally, since V is a monotonic increasing function of H, V is quasi-convex in the normalized (p, y).

have this property, applying them to anything but wide aggregates of products has proven difficult.[9]

5 Additive Random Utility 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.[10] That is, each consumer with individual characteristics c solves the maximization problem

$$max_{j \subset J} v_j(y - p_j I(j > 1), p_1, w_j, c, \theta) - \delta_j$$

where as before w_j denotes the product characteristics observed by the consumer (only some of which may be observed by the econometrician), y denotes income, p_1 denotes the price of the outside alternative, $I(\cdot)$ is an indicator function, c denotes a vector of this consumer's characteristics, and δ_j denotes a (possibly composite) product characteristic.

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

[9]For example, the translog has an indirect utility function which has many terms like $\alpha_j lnp_j$ and $\beta_{jk} lnp_j lnp_k$. Evidently, such a specification does not satisfy this regularity property. As a result, the demand systems also tend to have poor properties as the amount of price, or pseudo-price, variation is large. Following Hausman (1994), a typical response to this problem has been to use data only on the period when all goods are observed in the market. While this approach is 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, in other arena's product entry and exit occur simultaneously making this approach impractical.

[10]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.$\delta_j = w'_{1j}\beta + \xi_j$.

(1994), Berry, Levinsohn, and Pakes (1995), and Berry, Levinsohn, and Pakes (1997). In that case, $v_j(y - p_j I(j > 1), p_1, w_j, c, \theta) = v_j(y - p_j I(j > 1), p_1, w_j, c_1, \theta) + \epsilon_j$ and the vector of individual characteristics $\epsilon_{\mathcal{J}} \equiv (e_1, \dots, e_J)$ 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.

Define a class of functions, $\mathcal{V}$, whose members satisfy the following expected maximum random utility ($EMRU$) properties:[11]

1. For each choice set, $\mathcal{J} = \{1, \dots, J\}$, $V(\cdot)$ 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 $j \in \mathcal{J}$.

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

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 $V(\delta, \cdot) = E_{\nu|}[max_j\{v_j(z, y - p_j, p_1, w_j, s, c, \theta) - \delta_j\}]$

[11] 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.

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. In particular, I explore the class of expected maximum utility fuctions described by Equations 2 and 3. I show later that particular specifications of the expected utility function can generate parametric demand systems that are cabable of generating *arbitrary* substitution patterns between goods and arbitrary income elasticities of demand. Of course, since 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. Moreover, 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 ERUM class of functions.

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

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

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*

$$V(\delta) = E \max_{j \subset \mathcal{J}} v_j - \delta_j \tag{4}$$

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,

$$Pr\{j|w, y, \mathrm{p}, c, \mathcal{J}\} = -\frac{\partial V(\cdot)}{\partial \delta_j}. \tag{5}$$

Converse: *Suppose that $V(\delta; w, \mathrm{p}, y, c)$ is any function with the EMRU properties. Then equation (5) defines a probability choice system. Further, there exists an ARUM form such that $V(\delta)$ could be generated by equation (4).*

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 2 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.

Proposition 2 *A Set of Sufficient Conditions for ERUM*

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

1. *$H(r)$ is a non-negative, homogeneous of degree m function of $(r_1, \ldots, r_J) \geq 0$.*

2. *Suppose for any distinct $(j_1, \ldots, j_k)$ from $\{1, \ldots, J\}$, $\frac{\partial^k H}{\partial r_{j_1}, \ldots, \partial r_{j_k}}$ 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* $H_i(\cdot)$*,* $i=1,..,I$ *each satisfy all three of these conditions and each* H_i *is of the same degree of homogeneity* $m_i=m$*, then* $H(\cdot)=\Pi_{i=1}^{I}H_i(\cdot)$ *is the in the class of ERUM functions.*

Using this proposition, it is easy to establish the conditions required for the specifications provided in Equations 2 and 3 to correspond to an additive random utility model.

Proposition 3 *If* $V(\Psi)=lnH(r)$*, where* $H(r)$ *is a generalized linear form generalized linear form in Equation 2 and* $r_j=e^{\phi_j(y-p_jz_{mj})-\delta_j}$. *Then* $V(\delta)$ *is in the class of EMRU functions provided (i)* $a'r+\frac{1}{2}r^{\frac{1}{2}'}Br^{\frac{1}{2}}\geq 0$ *so* $H(\cdot)$ *is non-negative, (ii)* $\frac{\partial H}{\partial r_j}=a_j+\frac{1}{2}r_j^{-\frac{1}{2}}\sum_{l=1}^{J}b_{lj}r_l^{\frac{1}{2}}\geq 0$ *for all* j *(iii)* $b_{jk}\leq 0$ *for all* $j\neq k$.

Proof 1 *See Appendix*

These conditions together with the requirement that $lim_{r_j\to\infty}H(r_1,..,r_J)=+\infty$ are provide a slight relaxation of the conditions provided by Mcfadden (1978) as sufficient for the Generalized Extreme Value Model.[12]

Proposition 4 *GEV model (Very slight generalization of Mcfadden (1978))*

[12] 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 homogenous 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 $m\leq 1$.

Suppose $V(\delta, \cdot) = \frac{1}{m} \ln H(r(\delta))$ *and* $H()$ *satisfies the conditions in Proposition 2. Suppose in addition,* $\lim_{r_j \to \infty} H(r_1, .., r_J) = +\infty$ *for* $j = 1, .., J$*. Then,*

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

defines a probabilistic choice model from alternatives $j = 1, .., J$ *which is consistent with utility maximization. Moreover, this probability choice model is generated by an ARUM with the utility provided by good j* $u_j = \delta_j + \epsilon_j$*, where* $(\epsilon_1, .., \epsilon_J)$ *is distributed as* $F(\epsilon_1, .., \epsilon_J) = e^{-H(e^{-\epsilon_1}, .., e^{-\epsilon_J})}$.

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 5 *If* $V(\mathrm{p}, y, \delta) = \ln H(r)$ *with* $r_j = \exp\{\psi_j(y, p_j, p_1, \delta_j)\}$ *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* $\psi_j(y, p_j, p_1, \delta) = \frac{y - p_j I(j>1)}{p_1} - \delta_j$ *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 Roys identity to* $V()$.

Proof

In general, the demand systems corresponding to discrete choice behavior and continuous choice behaviour will not be the same.[13] 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 $j > 1$ and the shares of the outside goods match. Under the conditions in the proposition for all $j > 1$,

[13] One could probably imagine developing tests of the "continuity" of choice behaviour or clearly for the existence of a representative agent in a discrete choice model.

$$
\begin{aligned}
\frac{\partial V}{\partial p_j} &= \frac{\partial V}{\partial \psi_j}\frac{\partial \psi_j}{\partial p_j} = \frac{-1}{p_1}\frac{\partial V}{\partial \psi_j}, \text{ while} \\
\frac{\partial V}{\partial \delta_j} &= \frac{\partial V}{\partial \psi_j}\frac{\partial \psi_j}{\partial \delta_j} = -\frac{\partial V}{\partial \psi_j} \\
\frac{\partial V}{\partial y} &= \frac{1}{p_1}\sum_{k=1}^{J}\frac{\partial \ln H}{\partial r_k}\frac{\partial r_k}{\partial \psi_k} = \frac{1}{p_1}\sum_{k=1}^{J}\frac{\partial \ln H}{\partial r_k} r_k = \frac{1}{p_1}
\end{aligned}
$$

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 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, $\frac{y - p_j I(j>1)}{p_1}$ is spent on the outside good. Thus, in a discrete choice model, total demand for the outside good is $s_{outside}^{discrete} = \sum_{j=1}^{J} \frac{y - p_j I(j>1)}{p_1} s_j = \frac{y}{p_1} - \sum_{j>1} \frac{p_j}{p_1} s_j^{discrete}$. Since we have already established that $s_j^{discrete} = s_j^{continuous}$ for all $j > 1$, 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. □

6 Estimation

6.1 Discrete Choice Models

Consider the program $\max_\delta V(\delta) + s'\delta$. If $V(\delta)$ is convex in δ so is the objective function.[14] Thus, subject to minor regularity conditions, the program has a unique solution, δ^* which satisfies the first order conditions

[14] Sufficient conditions for $V(\delta)$ to be convex are provided in Lemma .

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

Thus, provided $V(\delta)$ is 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

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

6.2 Continuous Choice Models

Errors are typically added onto the demand system in continous choice models. This is clearly one option here.

$$s_j = s_j(\delta) + \xi_j.$$

On the other hand, 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 $\Psi_j = \phi_j(y, p_j, p_0) - \delta_j$ then the terms in the diagonal matrix are independent of δ.

Thus the solution to this program satisfyies the first order conditions

$$s_j = \frac{\partial \Psi_j}{\partial p_j} \left(\frac{\partial \Psi_j}{\partial \delta_j} \right)^{-1} \frac{\partial E(p, \delta, u)}{\partial \delta_j}. \tag{6}$$

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, $\frac{\partial V}{\partial \delta_j} = \frac{\partial V}{\partial p_j} \frac{\partial \Psi_j}{\partial \delta_j} \left(\frac{\partial \Psi_j}{\partial p_j} \right)^{-1}$. 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 6 yeilds 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 $V(\cdot) = lnH(\cdot)$ 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

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

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, $D_y V(\cdot) = \frac{1}{y}$, so $D_\delta E(\mathrm{p}, u, \delta) = -y D_\delta V(p, y, \delta)$ 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[15] are 'produced' from products themselves, $w = f(x)$. Thus, in a lancastrian world the consumer is assumed to solve the choice problem[16]

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

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

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

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.

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

[16] 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.

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 Choice Models

Consider the model $V(p_j, p_{-j}, w_j, w_{-j}) = lnH(p_j, p_{-j}, w_j, w_{-j})$ where $H(r) = \frac{1}{2}r'(I + B)r = \frac{1}{2}\sum_{l=1}^{J}\sum_{m=1}^{J} b_{lm} r_l r_m$ 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 $\frac{\partial^2 H}{\partial p_j \partial p_k} = (I(j = k) + b_{jk}) r_j r_k \frac{1}{p_j p_k}$.

Hence a polynomial specification with $b_{jk} = \sum_{l=1}^{q} \gamma_l d(w_j, w_k; \alpha)^l$, 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 $\sum_{l=1}^{m} \alpha_l = 1$, will be able to flexibly capture the relationship between the distance between two products in characteristics space and the resulting substitution pattern between those 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.

7.2 Product Characteristics in Discrete Choice Models

Then, the δ_j parameters in the model can only depend on the the characteristics of good j while the γ_{jk} parameters can only depend on the characteristics of goods j and k.[17]

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 discrete choice model considered here,

$$\begin{aligned}\frac{\partial ln s_j}{\partial ln p_k} &= \gamma_{jk} + \sum_{m=1}^{J} \frac{r_m(\delta,\Gamma)}{\sum_{l=1}^{J} r_m(\delta,\Gamma)}\gamma_{mk} \\ &= \gamma_{jk} + \sum_{m=1}^{J} s_m(\delta,\Gamma)\gamma_{mk}\end{aligned}$$

Hence a polynomial specification with $\gamma_{jk} = \sum_{l=1}^{p} \tau_l d(w_j, w_k; \alpha)^l$, 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 $\sum_{l=1}^{m} \alpha_l = 1$, will be able to flexibly capture the relationship between the distance between two products in characteristics space and the resulting substitution pattern. One important caveat to this specification is that imposes symmetry on the Γ matrix. Thus, there may be better alternatives to this mapping between observed product characteristics and the parameters

[17]Otherwise, $lim_{p_j \to \infty} V(p_j, p_{-j}, w_j, w_{-j}) = \sum_{l \neq j} r_l$ would depend on the characteristics of good j, w_j.

of the model.

8 Consumer Characteristics

If $V(\delta, c)$ is an EMRU function and is convex in δ for all consumer types, then the problem $\max_\delta \int V(\delta, c) f(c) dc + s'\delta$ 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 $E(p, \delta, c, u)$. Then provided $E(p, \delta, c, u)$ is concave in δ for each c, the problem $\max_\delta \int E(p, \delta, u, c) f(c) dc - s' diag\{\left(\frac{\partial \Phi_t}{\partial p_t}\right)^{-1} \frac{\partial \Phi_t}{\partial \delta_t}\}\delta$ 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.[16]

---

[16] 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, $V^*(\delta) = \int V^c(c) f(c) dc$ will also be a member of the EMRU set of functions with aggregate demand system, $s_j = \int s_j f(c) dc$.

9 Flexibility Properties of the Models

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 elasticitities 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).) In this section of the paper I propose particular functional forms for the discrete- and continuous-choice demand models and show that these specifications are flexible functional forms.

Proposition 6 *Flexibility of the Discrete Choice Model Consider the model* $H(r) = r'\mathbf{1}_J$, *where* $r_j = \exp\{-\delta_j\}\Pi_{l=1}^{J}\exp\{\gamma_{jl}(\ln y - \ln p_l)\}$, Γ *is a matrix of parameters with* jk^{th} *element* γ_{jk}, *and* $\mathbf{1}_J$ *denotes a* Jx1 *vector of ones. 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 (See Appendix.)

Proposition 7 *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 homogeneity, additivity, and slutsky symmetry.*

Proof (See Appendix.)

10 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 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 two 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 extremly disaggregated in nature and product introduction and exit are both extremely frequent and very informative about substitution patterns.

Second, the discrete-choice model proposed is a member of the class of GEV models and is therefore 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.

11 Appendix A

Lemma 1 *If $\Phi(r_1,..,r_J)$ is a real valued convex function that is either*

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

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

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

Proof Choose any pair of vectors $\mathbf{x}$ and $\mathbf{x}'$ and let $0 \leq \lambda \leq 1$.

1. If $\underline{r}(\mathbf{x})$ is convex in $\mathbf{x}$

$$\underline{r}(\lambda x + (1-\lambda)x') \leq \lambda \underline{r}(x) + (1-\lambda)\underline{r}(x').$$

However, since Φ is non-decreasing in $\underline{r}$

$$\begin{aligned}\Phi(\underline{r}(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}') &\leq \Phi(\lambda \underline{r}(\mathbf{x}) + (1-\lambda)\underline{r}(\mathbf{x}')) \\ &\leq \lambda\Phi(\underline{r}(\mathbf{x})) + (1-\lambda)\Phi(\underline{r}(\mathbf{x}'))\end{aligned}$$

where the latter inequality follows since Φ is convex in r.

2. If $\underline{r}(\mathbf{x})$ is concave in $\mathbf{x}$

$$\underline{r}(\lambda x + (1-\lambda)x') \geq \lambda \underline{r}(x) + (1-\lambda)\underline{r}(x').$$

However, since Φ is non-increasing in $\underline{r}$

$$\begin{aligned}\Phi(\underline{r}(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}') &\leq \Phi(\lambda \underline{r}(\mathbf{x}) + (1-\lambda)\underline{r}(\mathbf{x}')) \\ &\leq \lambda\Phi(\underline{r}(\mathbf{x})) + (1-\lambda)\Phi(\underline{r}(\mathbf{x}'))\end{aligned}$$

where the latter inequality follows since Φ is convex in r.

□

Lemma 2 *Let $V(r) = c + d\ln H(r)$ 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 1 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

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

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 yeilds

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

If $\frac{\partial^2 \ln H}{\partial \psi_k \partial \psi_j} \leq 0$ for all $j \neq k$, then $\frac{\partial^2 \ln H}{\partial \psi_k^2} \geq 0$ and the matrix of second deriatives of $\ln H(\psi)$ 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 $\ln H(\psi)$ is convex.

Thus, it suffices to establish that the conditions in the lemma ensure that $\frac{\partial^2 \ln H}{\partial \psi_k \partial \psi_j} \leq 0$ 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 2

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 $j = 1, \ldots, J$, 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 $\frac{\partial V}{\partial\delta_1,\ldots,\delta_k} = (-1)^k r_1 \ldots r_k \frac{\partial V(r)}{\partial r_1,\ldots,\partial r_k}$. Thus, to ensure that the mixed cross partials of $\frac{\partial V(\delta)}{\partial\delta_1,\ldots,\delta_k}$ 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 $V_{1,\ldots,k}$ denotes $\frac{\partial V}{\partial r_1\ldots\partial r_k}$ and $H_{1,\ldots,k} \equiv \frac{\partial H}{\partial r_1\ldots\partial r_k}$, define, recursively, $Q_1 = H_1$ and $Q_k = \frac{\partial Q_{k-1}}{\partial r_k} - \frac{1}{H}Q_{k-1}H_k$.[19] Suppose $V_{1,\ldots,k-1} = \frac{Q_{k-1}}{H}$. Then differentiating with respect to r_k yeilds $V_{1,\ldots,k} = \left(\frac{\partial Q_{k-1}}{\partial r_k} - \frac{1}{H}Q_{k-1}H_k\right)\frac{1}{H}$. Since $V_1 = \frac{Q_1}{H}$, $V_{1,\ldots,k} = \frac{Q_k}{H}$ for all k by induction.

[19]Thus, for example,

$$\begin{aligned} V_1 &= \frac{H_1}{H} \\ V_{12} &= \frac{H_{12}}{H} - \frac{1}{H^2}H_1H_2. \end{aligned}$$

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.

4. As $\delta_k \rightarrow -\infty$, 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

$$
\begin{aligned}
lim_{\delta_j \rightarrow -\infty} \frac{\partial V}{\partial \delta_j} &= -lim_{\delta_j \rightarrow -\infty} \frac{r_j H_j(r, \cdot)}{mH(r, \cdot)} \\
&= -\lim_{\delta_j \rightarrow -\infty} \frac{e^{-(m-1)\delta_j} r_j H_j(e^{\delta_1} r_1, .., e^{\delta_J} r_J)}{m e^{-m\delta_j} H(e^{\delta_1} r_1, .., e^{\delta_J} r_J)} \\
&= -\lim_{\delta_j \rightarrow -\infty} e^{\delta_j} r_j \frac{H_j(e^{\delta_1} r_1, .., e^{\delta_J} r_J)}{mH(e^{\delta_1} r_1, .., e^{\delta_J} r_J)} \\
&= 1 \frac{H_j((0, \ldots, 1, 0..0), \cdot)}{mH((0, \ldots, 0, 1, 0...0), \cdot)} \\
&= -1
\end{aligned}
$$

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

$$
\begin{aligned}
1 = lim_{\delta_j \rightarrow -\infty} \sum_{k=1}^{J} \frac{r_k H_k(r_1, .., r_J)}{mH(r_1, .., r_J)} &= lim_{\delta_j \rightarrow -\infty} \sum_{k=1}^{J} e^{-\delta_k} \frac{e^{\delta_j} H_k(e^{\delta_1} r_1, .., e^{\delta_J} r_J)}{mH(e^{\delta_1} r_1, \ldots, e^{\delta_J} r_J)} \\
&= lim_{\delta_j \rightarrow -\infty} \sum_{k=1}^{J} e^{-\delta_k + \delta_j} \frac{H_j(e^{\delta_1} r_1, .., e^{\delta_J} r_J)}{mH(e^{\delta_1} r_1, \ldots, e^{\delta_J} r_J)}
\end{aligned}
$$

$$= \lim_{\delta_j \to \infty} \frac{H_j(e^{\delta_j} r_1, .., e^{\delta_j} r_J)}{m H(e^{\delta_j} r_1, ..., e^{\delta_j} r_J)}$$

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

Proof to proposition 3 I show that $H(r) = a'r + r^{\frac{1}{2}'} B r^{\frac{1}{2}}$ has the properties in Proposition 2 and is therefore in the class of ERUM functions.

Notice that $H(r, \cdot)$ is non-negative by assumption, linear homogeneous by construction, and defined for $r \geq 0$. The second condition follows from $H() \geq 0$ and the fact that terms of order r_j^1 will dominate the asymptotic properties of the function as r_j becomes sufficiently large. Thus $\lim_{r_j \to +\infty} H(r) = +\infty$ provided $a_j + b_{jj} > 0$. The mixed partial derivatives of $H(r)$ clearly exist provided all the elements in a and B are finite. 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. All higher mixed partial deriatives are clearly zero and therefore satisfy the partial derivative conditions. The fourth condition is trivially satisfied for the generalized linear form while the fifth condition is satisfied since terms of order r_j dominate in H and $\frac{\partial H}{\partial r_j}$. Consequently, $\lim_{r_j \to \infty} \frac{\partial H}{\partial r_j} = a_j + b_{jj}$ while $\lim_{r_j \to \infty} \frac{H}{r_j} = a_j + b_{jj}$. The result follows.

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,

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

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 homogenous degree one function $\bar{H}(\cdot) = H(\cdot)^{1/m}$, the sign properties of the derivatives of $H(\cdot)$ are not generally inherited by $\bar{H}(\cdot)$.)

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 $\epsilon_j \rightarrow -\infty$, then $H \rightarrow +\infty$, implying $F \rightarrow 1$. 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 hypothesised 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, $\frac{\partial F}{\partial \epsilon_1} = e^{-\epsilon_1} Q_1 F$. Suppose $\frac{\partial^{k-1} F}{\partial \epsilon_1, .., \partial \epsilon_{k-1}} = e^{\epsilon_1} Q^{k-1} F$. Then $\frac{\partial^k F}{\partial \epsilon_1, .., \partial \epsilon_k} = e^{-\epsilon_1} ... e^{-\epsilon_k} \{Q_{k-1} H_k F - F \frac{\partial Q_{k-1}}{\partial r_k}\} = e^{-\epsilon_1} ... e^{-\epsilon_k} Q_k F$. By induction, $\frac{\partial^J F}{\partial e^{-\epsilon_1}, .., \partial e^{-\epsilon_J}} Q_J F \geq 0$. 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

$$
\begin{aligned}
\pi_1 &= \int_{\epsilon=-\infty}^{+\infty} F_1(\epsilon, \delta_1-\delta_2+\epsilon, \ldots, \delta_1-\delta_J+\epsilon)d\epsilon \\
&= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon} H_1(e^{-\epsilon}, e^{-\delta_1+\delta_2-\epsilon}, \ldots, e^{-\delta_1+\delta_J-\epsilon}) exp\{-H(e^{-\epsilon}, e^{-\delta_1+\delta_2-\epsilon}, \ldots, e^{-\delta_1+\delta_J-\epsilon})\}d\epsilon \\
&= \int_{\epsilon=-\infty}^{+\infty} e^{-\epsilon}(e^{\epsilon+\delta_1})^{-(m-1)} H_1(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J}) exp\{-(e^{\epsilon+\delta_1})^{-m} H(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})\}d\epsilon \\
&= e^{\delta_1}\int_{u=-\infty}^{+\infty} e^{-m(u)} H_1(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J}) exp\{-(e^{-m(u)} H(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})\}du \\
&= e^{\delta_1}\frac{H_1(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})}{H(e^{\delta_1}, .., e^{\delta_J})}\int_{u=-\infty}^{+\infty} e^{-m(u)+lnH(e^{\delta_1}, .., e^{\delta_J})} exp\{-(e^{-m(u)+lnH}\}du \\
&= \frac{1}{m}e^{\delta_1}\frac{H_1(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})}{H(e^{\delta_1}, .., e^{\delta_J})}\int_{u=-\infty}^{+\infty} m e^{-m(u-\frac{1}{m}lnH(e^{\delta_1}, .., e^{\delta_J}))} exp\{-(e^{-m(u-\frac{1}{m}lnH(e^{\delta_1}, e^{\delta_2}, .., e^{\delta_J}))}\}du \\
&= \frac{e^{\delta_1}H_1(e^{\delta_1}, \ldots, e^{\delta_J})}{mH(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})}
\end{aligned}
$$

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 $\xi = \frac{1}{m}lnH(e^{\delta_1}, e^{\delta_2}, \ldots, e^{\delta_J})$ 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. □.

Corrollary

Under the hypotheses of Theorem 4, expected maximum utility, defined by

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

(with f the density of F), satisfies

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

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

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

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 11 can be partitioned into regions where each alternative has maximum utility, yeilding

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

Let $\xi = \frac{1}{m} ln H(e^{\delta_1}, e^{\delta_2}, ..., e^{\delta_J})$ 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 11 becomes

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

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 elasticitities 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^*, $\frac{\partial ln s_j^*}{\partial ln p_k}$, and $\frac{\partial ln s_j^*}{\partial ln y}$ that satisfy additivity and slutsky symmetry, then we can always choose the parameters of the model, $\theta = (\delta, B)$ that satisfy the following equations:

$$
\begin{aligned}
s_j^* &= s_j(p^*, y^*, \theta) \quad j = 1, \ldots, J && (7) \\
\frac{\partial ln s_j^*}{\partial ln p_k} &= \frac{\partial ln s_j(p^*, y^*, \theta)}{\partial ln p_k} \quad j, k = 1, \ldots, J && (8) \\
\frac{\partial ln s_j^*}{\partial ln y} &= \frac{\partial ln s_j(p^*, y^*, \theta)}{\partial ln y} \quad j = 1, .., J && (9)
\end{aligned}
$$

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

$$
\begin{aligned}
\sum_{j=1}^{J} s_j^* &= 1 && (10) \\
\sum_{j=1}^{J} \frac{\partial s_j^*}{\partial p_k} &= -s_k^* && (11) \\
\sum_{j=1}^{J} \frac{\partial s_j^*}{\partial y} &= 1 && (12)
\end{aligned}
$$

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 (7),(8), and (9), 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

$$\sum_{j=1}^{J} s_j(p^*, y^*, \theta) = 1 \tag{13}$$

$$\sum_{j=1}^{J} \frac{\partial s_j(p^*, y^*, \theta)}{\partial p_k} = -s_k(p^*, y^*, \theta) \tag{14}$$

$$\sum_{j=1}^{J} \frac{\partial s_j(p^*, y^*, \theta)}{\partial y} = 1 \tag{15}$$

or,

$$\sum_{j=1}^{J} -\frac{\partial H(p^*, y^*, \theta)}{\partial p_j} = \frac{\partial H(p^*, y^*, \theta)}{\partial y} \tag{16}$$

$$-\sum_{j=1}^{J} \frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial p_k} - \frac{\partial H(p^*, y^*, \theta)}{\partial p_k} = \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y \partial p_k} \tag{17}$$

$$-\sum_{j=1}^{J} \frac{\partial^2 H(p^*, y^*, \theta)}{\partial p_j \partial y} = \frac{\partial H(p^*, y^*, \theta)}{\partial y} + \frac{\partial^2 H(p^*, y^*, \theta)}{\partial y^2} \tag{18}$$

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

$$
\begin{aligned}
\frac{\partial H(p^*,y^*,\theta)}{\partial y} &= 1 \\
-\frac{\partial H(p^*,y^*,\theta)}{\partial p_j} &= s_j^* \\
\frac{\partial^2 H}{\partial p_j \partial p_k} &= -\frac{\partial s_j^*}{\partial p_k} - 2 s_j^* s_k^* + s_j \frac{\partial s_k^*}{\partial y}
\end{aligned}
$$

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,

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

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

By choosing $b_{jk} e^{-\delta_j} e^{-\delta_k} = -\frac{\partial s_j^*}{\partial p_k} - 2 s_j^* s_k^* + s_j \frac{\partial s_k^*}{\partial y}$ and $e^{-2\delta_j} = 2 s_j^* - \frac{\partial s_j^*}{\partial y}$ it is easy to verify that each of these constraints are satisfied since $\sum_{l=1}^J b_{jl} e^{-\delta_j} e^{-\delta_l} = \frac{\partial s_j^*}{\partial y} - 2 s_j^* + s_j^* = \frac{\partial s_j^*}{\partial y} - s_j^*$, $\sum_{j=1}^J \sum_{l=1}^J b_{jl} e^{-\delta_j} e^{-\delta_l} = \sum_{j=1}^J \left(\frac{\partial s_j^*}{\partial y} - s_j^* \right) = 0$, 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 $H(r) = r' \mathbf{1}_J$, where $r_j = \exp\{-\delta_j\} \Pi_{l=1}^J \exp\{\gamma_{jl} (\ln y - \ln p_l)\}$, Γ is a matrix of parameters with jk^{th} element γ_{jk}, and $\mathbf{1}_J$ denotes a Jx1 vector of ones. 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^*, $\frac{\partial \ln s_j^*}{\partial \ln p_k}$, and $\frac{\partial \ln s_j^*}{\partial \ln y}$ that satisfy the additivity and homogeneity conditions, then we can always choose the parameters of the model $\theta = (\delta, \Gamma)$ that satisfy the following equations:

$$s_j^* = s_j(p^*, y^*, \theta) \quad j = 1, \ldots, J \tag{19}$$

$$\frac{\partial \ln s_j^*}{\partial \ln p_k} = \frac{\partial \ln s_j(p^*, y^*, \theta)}{\partial \ln p_k} \quad j, k = 1, \ldots, J \tag{20}$$

$$\frac{\partial \ln s_j^*}{\partial \ln y} = \frac{\partial \ln s_j(p^*, y^*, \theta)}{\partial \ln y} \quad j = 1, .., J \tag{21}$$

or, in terms of the model

$$s_j^* = -\frac{1}{H}\frac{\partial H}{\partial \delta_j} \tag{22}$$

$$\frac{\partial s_j^*}{\partial p_k} = -\frac{1}{H}\frac{\partial^2 H}{\partial \delta_j \partial p_k} - \frac{s_j^*}{H}\frac{\partial H}{\partial p_k} \tag{23}$$

$$\frac{\partial s_j^*}{\partial y} = -\frac{1}{H}\frac{\partial^2 H}{\partial \delta_j \partial y} - \frac{s_j^*}{H}\frac{\partial H}{\partial y} \tag{24}$$

Additivity of the market shares implies that

$$\sum_{j=1}^{J} s_j^* = 1, \;\; \sum_{j=1}^{J} \frac{\partial s_j^*}{\partial p_k} = 0, \;\; \sum_{j=1}^{J} \frac{\partial s_j^*}{\partial y} = 0$$

Imposing the analogous constraints on the model implies that we may choose a specification that enforces

$$-H = \sum_{j=1}^{J} \frac{\partial H}{\partial \delta_j} \tag{25}$$

$$-\frac{\partial H}{\partial p_k} = \sum_{j=1}^{J} \frac{\partial^2 H}{\partial \delta_j \partial p_k} \tag{26}$$

$$-\frac{\partial H}{\partial y} = \sum_{j=1}^{J} \frac{\partial^2 H}{\partial \delta_j \partial y} \tag{27}$$

while the fact that the true $V = lnH$ is homogeneous of degree zero in (y, p) implies that we may restrict our choice of H to satisfy

$$-y\frac{\partial H}{\partial y} = \sum_{j=1}^{J} \frac{\partial H}{\partial p_j} p_j \tag{28}$$

$$-y\frac{\partial^2 H}{\partial \delta_k \partial y} = \sum_{j=1}^{J} \frac{\partial^2 H}{\partial \delta_k \partial p_j} p_j \tag{29}$$

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, these homogeneity constraints become

$$-\frac{\partial H(p^*, y^*, \theta)}{\partial y} = \sum_{l=1}^{J} \frac{\partial H(p^*, y^*, \theta)}{\partial p_l} \tag{30}$$

$$-\frac{\partial^2 H(p^*, y^*, \theta)}{\partial \delta_k \partial y} = \sum_{l=1}^{J} \frac{\partial^2 H(p^*, y^*, \theta)}{\partial \delta_k \partial p_l} \tag{31}$$

Using these constraints together with the constraints given in Equations (25) (26), and (27), we may re-write Equations (19), (20), and (21) as

$$s_j^* = -\frac{1}{H}\frac{\partial H}{\partial \delta_j} \tag{32}$$

$$\frac{\partial s_j^*}{\partial p_k} = -\frac{1}{H}\frac{\partial^2 H}{\partial \delta_j \partial p_k} + \frac{s_j^*}{H}\sum_{l=1}^{J}\frac{\partial^2 H}{\partial \delta_l \partial p_k} \tag{33}$$

$$\frac{\partial s_j^*}{\partial y} = -\frac{1}{H}\frac{\partial^2 H}{\partial \delta_j \partial y} - \frac{s_j^*}{H}\sum_{l=1}^{J}\frac{\partial^2 H}{\partial \delta_l \partial y} \tag{34}$$

$$= \frac{1}{H}\sum_{l=1}^{J}\frac{\partial^2 H(p^*, y^*, \theta)}{\partial \delta_j \partial p_l} - \frac{s_j^*}{H}\sum_{l=1}^{J}\sum_{m=1}^{J}\frac{\partial^2 H}{\partial \delta_l \partial p_m} \tag{35}$$

Using the proposed specification, $H(r) = \sum_{l=1}^{J} r_j$ first note that $\frac{\partial H(p^*, y^*, \theta)}{\partial \delta_j} = -r_j(p^*, y^*) = -e^{\delta_j}$ and consequently $\frac{\partial^2 H}{\partial \delta_j \partial p_k} = -\frac{\partial r_j}{\partial p_k}$. Moreover,

$$\frac{\partial r_j}{\partial p_k} = \frac{r_j \gamma_{jk}}{-p_k} \text{ so that } \frac{\partial r_j(p^*, y^*)}{\partial p_k} = -e^{\delta_j}\gamma_{jk} \tag{36}$$

Thus, establishing flexibility means showing that there is a value of (Γ, δ) that solve the equations

$$s_j^* = \frac{1}{H} e^{\delta_j} \tag{37}$$

$$\frac{\partial s_j^*}{\partial p_k} = -\frac{1}{H} e^{\delta_j} \gamma_{jk} + \frac{s_j^*}{H} \sum_{l=1}^{J} e^{\delta_l} \gamma_{lk} \tag{38}$$

$$\frac{\partial s_j^*}{\partial y} = \frac{1}{H} \sum_{l=1}^{J} e^{\delta_j} \gamma_{jl} - \frac{s_j^*}{H} \sum_{l=1}^{J} \sum_{m=1}^{J} e^{\delta_l} \gamma_{lm} \tag{39}$$

$$\tag{40}$$

Rearranging and re-writing the latter two equations in matrix form yields

$$H(p^*, y^*, \theta) s_j^* = e^{\delta_j} \quad j = 1, .., J \tag{41}$$

$$H(p^*, y^*, \theta) D_p lns^* = -\Gamma + 1_J s' \Gamma \tag{42}$$

$$= -(I_J - 1_J s') \Gamma \tag{43}$$

$$H(p^*, y^*, \theta) D_y lns^* = \Gamma 1_J - 1_J s' \Gamma 1_J$$

$$= (I_J - 1_J s') \Gamma 1_J \tag{44}$$

First, without loss of generality, we can normalize $H(p^*, y^*, \theta) = \sum_{j=1}^{J} e^{\delta_j} = 1$ there are J parameters in δ and only $J - 1$ market shares to determine, given the additivity constraint.

Next, notice that setting $\gamma_{jk} = -\frac{\partial ln s_j^*}{\partial p_k}$ solves Equation (43) since this solution ensures that, by additivity, $\sum_{l=1}^{J} s_l^* \gamma_{lk} = \sum_{l=1}^{J} s_l^* \frac{\partial ln s_l^*}{\partial p_k} = 0$.

Finally notice that this solution also satisfies Equation 44 since, provided the share equations are homogeneous of degree zero in prices and income, Euler's Theorem ensures that

$$\sum_{k=1}^{J} p_k \frac{\partial s_j}{\partial p_k} + y \frac{\partial s_j}{\partial y} = 0.$$

Evaluating at (p^*, y^*) and dividing both sides by s_j^* establishes that $-\sum_{k=1}^{J} \frac{\partial \ln s_j^*}{\partial p_k} = \frac{\partial \ln s_j^*}{\partial y}$. Ie., $\sum_{k=1}^{J} \gamma_{jk} = \frac{\partial \ln s_j^*}{\partial y}$. This is exactly Equation 44 at the proposed solution since $s^{*\prime} \Gamma 1_J = s^{*\prime} D_y ln s^* = 0$.

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.

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

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. (1971): "An Application of the Shephard Duality Theorem: A Generalized Leontieg Production Function," *The Journal of Political Economy*, 79, 481–507.

——— (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. 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.

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.

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

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

Varian, H. R. (1984): *Microeconomic Analysis*. Norton.

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