A Simple Globally Consistent Continuous Demand Model for Market Level Data*†

Peter Davis
The Brattle Group

Laura Grigolon
University of Mannheim and CEPR

Ricardo Ribeiro
Universidade Católica Portuguesa, Católica Porto Business School and CEGE

February 2023

[PLEASE DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION]

Abstract

This paper considers a new method of uncovering demand information from market level data on differentiated products. In particular, we propose a globally consistent continuous-choice demand model with distinct advantages over the models currently in use and describe the econometric techniques for its estimation. The proposed model combines key properties of both the discrete- and continuous-choice traditions: (i) it is flexible in the sense of Diewert (1973, 1974), (ii) it is globally consistent in the sense it can deal with the entry and exit of products over time, and (iii) incorporates a structural error term. The estimation procedure follows an analogue to the algorithm derived in Berry (1994), Berry et al. (1995). The contraction mapping for matching observed and predicted budget shares is analytic, making it relatively simple and fast to estimate, which can prove a key advantage in some applications such those in competition policy, where time and transparency are important. We provide also a series of Monte Carlo experiments to illustrate the estimation properties of the model. Finally, we discuss how it can be extended to account for consumer heterogeneity and to cope with consumer dynamic behaviour.

JEL Classification: D12, L40

Keywords: Demand, Differentiated Products, Continuous-Choice, Substitution Patterns

---

*An earlier version of this paper circulated under the title "Demand Models for Market Level Data". Thanks are due to Supachoke Thawornkaiwong for helpful comments and suggestions. Ricardo Ribeiro gratefully acknowledges financial support from the Fundação para a Ciência e a Tecnologia.

†Email addresses: Peter.Davis@brattle.com (P. Davis), laura.grigolon@uni-mannheim.de (L. Grigolon), rribeiro@ucp.pt (R. Ribeiro).

1 Introduction

This paper considers a new method of uncovering demand information from market level data on differentiated products. In particular, we propose a continuous-choice demand model with distinct advantages over the models currently in use and describe the econometric techniques for its estimation.

When products are differentiated, the number of parameters required to describe a demand system (without *a priori* restrictions on the substitution patterns) tends to be excessively large to estimate, given the number of observations in a typical dataset. As an illustration of the problem, note that even in the simplest and extremely restrictive of the demand specifications - the linear expenditure model - J products would yield at least J^2 parameters to be estimated, just to capture the substitution patterns. Although implied economic theory's restrictions (like the symmetry of the Slutsky matrix) could be imposed to increase the degrees of freedom available, they do not solve the dimensionality issue. The option for a more flexible functional form would only naturally worsen the problem. Some structure must therefore be placed on the estimation procedure.

The recent industrial economics literature has evolved using primarily discrete-choice models of consumer behaviour. That is, the demand models assume that consumers, while heterogeneous, can purchase at most one unit of one of the available products. Moreover, consumer preferences over products are typically mapped onto a space of characteristics (Lancaster, 1971), reducing the number of parameters to be estimated (since the parameter space is defined by the number of characteristics rather than by the number of products). Within this set of assumptions, we can find the multinomial logit (McFadden, 1974), the nested multinomial logit (McFadden, 1978b), the multinomial probit (Hausman and Wise, 1978), the mixed- or random-coefficients multinomial logit (McFadden, 1981) and the discrete choice analytically flexible (DCAF) models (Davis and Schiraldi, 2014). The most serious drawback of this branch of the literature relates to the typical trade-off between flexibility and computation requirements. On one hand, we have the standard and the nested multinomial logit models which are fully analytic (and thereby relatively simple to estimate), but imply substitution patterns that tend to be model- instead of data-driven. On the other hand, we have the probit and the mixed-coefficients logit multinomial models which provide increased flexibility by introducing unobserved consumer heterogeneity, but require the use of simulation techniques (which in turn increases substantially the computation requirements). The recent discrete choice analytically flexible model seems to present itself as an exception since it appears to combine the good properties from the two groups.

The assumption that consumers purchase at most one unit of one of the available products may be unrealistic in some settings. Cereals, yogurts, soft drinks and wine are all examples where many consumers typically buy more than one product each time they go shopping. While the discreteness assumption can sometimes be rationalised, it is not always natural to do so given the data available. For example, multiple choices are sometimes modelled as discrete by narrowly defining the choice period and, as a consequence, rationalize multiple observed choices by assuming the consumer has made two separate purchasing decisions. Alternatively, a discrete-choice approach can be extended to model multiple choices by defining the choice set to include product bundles. However, often the discrete choice assumption can not easily be motivated given the products being selected and so it may be more natural to model consumers as making continuous quantity choices.

The continuous-choice literature typically assumes a representative agent that might consume all products and uses functional forms that allow flexible substitution patterns. Examples include the translog model of Christensen et al. (1975), the almost ideal demand system (AIDS) due to Deaton and Muelbauer (1980) and the distance metric model from Pinkse and Slade (2004). Consumer preferences can be defined directly over products (as in the translog or in the AIDS cases) or mapped onto a space of characteristics in a manner akin to the discrete-choice literature (as in the distance metric model). However, this set of models presents a serious limitation. In contrast to the discrete-choice framework, they can not be used to uncover demand information from markets with significant entry and exit of products. This limitation has been addressed in the literature before, but never in a way that, to the best of our knowledge, we could categorize as adequate. The typical solutions are largely limited to either consider substitution patterns between broad aggregates of products as, for example, in Christensen et al. (1975), Deaton and Muelbauer (1980), and Hausman et al. (1994), or to estimate the demand system using data only from time periods when all products are present in the market as, for example, in Hausman (1994), Ellison et al. (1997), and Pinkse and Slade (2004).

More recently, the continuous-choice literature has evolved to model consumer heterogeneity explicitly. Chan (2006) constitutes an example of such approach. He develops a continuous hedonic-choice model to investigate the demand for soft drinks that, in contrast to the more traditional line of the literature, is able to cope with entry and exit of products. However, his approach, in addition to requiring simulation techniques, does not model unobserved product characteristics.

In this paper, we follow the traditional continuous-choice literature and develop a representative consumer flexible demand model. We begin by specifying an indirect utility function from which, via Roy's identity, a continuous-choice demand system is derived. The demand function implied by the model is fully analytic and therefore avoids the burden of simulation. The model is flexible in the sense of Diewert (1973, 1974) as the implied own- and cross-price elasticities are capable of capturing the true substitution patterns in the data. In addition, the model can accommodate the use of data on the entry and exit of products. The importance of this last property is twofold as (i) not being able to cope with entry and exit patterns limits the application of the above models and (ii) the ability to deal with variation in the set of choices available to consumers provides pseudo-price variation which can be very helpful in the identification of substitution patterns between products.

For estimation, we propose an analogue to the algorithm derived in Berry (1994), Berry et al. (1995). Following this line of the literature, the error term is structurally embedded in the model and thereby circumvents the critique provided by Brown and Walker (1989) related to the addition of add-hoc errors and their induced correlations. The contraction mapping for matching observed and predicted expenditure shares is analytic.

To sum up, the main contribution of the paper is to propose a new continuous-choice model that combines the centrally desirable properties of both the discrete- and continuous-choice traditions: (i) it is flexible in the sense of Diewert (1973, 1974), (ii) can deal with the entry and exit of products over time, and (iii) incorporates a structural error term. Furthermore, it is relatively simple and fast to estimate which can prove a key advantage in competition policy issues where time and transparency are always crucial factors.

The remainder of the paper is organized as follows. Section 2 describes the proposed demand model and examines its properties. Section 3 discusses identification and presents the proposed estimation procedure. Section 4 provides a series of Monte Carlo experiments to illustrate the estimation properties of the model. Section 5 provides extensions to the proposed demand model so to account for consumer heterogeneity and cope with consumer dynamic behaviour. Finally, Section 6 concludes.

2 The Demand Model

2.1 The General Setup

Let $\Im$ denote the set of $J+1$ choices available to consumers, with options $j = 1, \ldots, J$ referring to the inside choices and $j = 0$ referring to an outside option that aggregates all remaining ones, including that of no purchase. We follow the continuous choice literature and define the demand system by specifying an indirect utility function $V(p, y; \theta, \Im)$ for the representative consumer. p denotes the $(J+1)$ vector of $p_i \in \Re^+$ prices, $y \in \Re^+$ denotes the income of the representative consumer and, finally, θ refers to a set parameters which we assume to have support $\Theta \subseteq R^\theta$. For reasons that will become clear later, we also index the utility function by the set of available goods $\Im$.

$V(p, y; \theta, \Im)$ is assumed to satisfy the properties of an indirect utility function, namely to be a continuous function in p and y, strictly increasing in y and nonincreasing in p_j for any p_j, quasiconvex, and homogeneous of degree zero. If $V(p, y; \theta, \Im)$ has the properties of an indirect utility function, then standard duality results imply that the demand system is easily obtained via Roy's identity. In this paper, we restrict further the class of functions considered by requiring the indirect utility function to satisfy an additional global regularity property, namely the sub-class of indirect utility functions that we define to be globally consistent.

Definition 1. *An indirect utility function $V(p, y; \theta, \Im)$ is globally consistent if and only if for any set of products $\Im' \subset \Im$:*

$$V(p, y; \theta, \Im') = \lim_{p_k \to \infty} V(p, y; \theta, \Im),$$

for all $k \in \Im$ and (simultaneously) $k \notin \Im'$.

This corresponds directly to McFadden (1981)'s social surplus condition. We name this property global consistency since it requires that an indirect utility function, defined on the set of all possible products $\Im$, can be specialized down in an entirely consistent fashion to generate demand systems over arbitrary subsets of the goods $\Im'$. It encapsulates the very natural restriction that removing a good from the choice set is entirely equivalent to increasing its price to infinity.

Proposition 1. *Any indirect utility function $V(p, y; \theta, \Im)$ which satisfies global consistency and non-satiation (that is, non-zero marginal utility of income) generates a demand system via Roy's identity which enjoys the property that:*

$$q_i(p, y; \theta, \Im') = \lim_{p_k \to \infty} q_i(p, y; \theta, \Im),$$

for all $i, k \in \Im$, $i \in \Im'$ and $k \notin \Im'$, where $q_i(p, y; \theta, \Im)$ denotes the Marshallian demand function of good i. Furthermore:

$$q_k(p, y; \theta, \Im) = 0,$$

for all $k \notin \Im'$.

Proof. If $V(p, y; \theta, \Im)$ has the properties of an indirect utility function then standard duality results imply that, under non-satiation, the demand system is easily obtained via Roy's identity:

$$q_i(p, y; \theta, \Im) = -\frac{V_i(p, y; \theta, \Im)}{V_y(p, y; \theta, \Im)},$$

where $V_i(p, y; \theta, \Im)$ denotes the marginal utility with respect to p_i and $V_y(p, y; \theta, \Im)$ denotes the marginal utility of income. It is easy to verify that if $V(p, y; \theta, \Im)$ satisfies in addition the global consistency property, then:

$$\begin{aligned} V_i(p, y; \theta, \Im') &= \lim_{p_k \to \infty} V_i(p, y; \theta, \Im) \\ V_y(p, y; \theta, \Im') &= \lim_{p_k \to \infty} V_y(p, y; \theta, \Im), \end{aligned}$$

for all $i, k \in \Im$, $i \in \Im'$ and $k \notin \Im'$. This yields the first part of the proposition:

$$q_i(p, y; \theta, \Im') = -\frac{V_i(p, y; \theta, \Im')}{V_y(p, y; \theta, \Im')} = -\frac{\lim_{p_k \to \infty} V_i(p, y; \theta, \Im)}{\lim_{p_k \to \infty} V_y(p, y; \theta, \Im)} = \lim_{p_k \to \infty} q_i(p, y; \theta, \Im).$$

For all $k \notin \Im'$, we have, under global consistency, that $V_k(p, y; \theta, \Im) = 0$, which together with non-satiation yields that $q_k(p, y; \theta, \Im) = 0$. In other words, removing a good from the choice set explicitly forces the level of demand for that good to zero. □

Proposition 1 establishes that a demand system derived from an indirect utility function that satisfies global consistency and non-satiation *can* be estimated using datasets where significant product entry and exit occurs. The possibility of using datasets where significant product entry and exit occurs provides a potentially useful source of pseudo-price variation to help identify substitution patterns.

Surprisingly, the extremely mild and intuitive global consistency condition is *not* satisfied by the vast majority of existing continuous-choice models like the translog model of Christensen et al. (1975), the almost ideal demand system (AIDS) due to Deaton and Muelbauer (1980) and the distance metric model from Pinkse and Slade (2004).[1] In all of these examples, demand depends linearly on prices (or on its logarithms) and if a product is not present in a given market such a demand system can not be estimated.

The typical solutions include either consider substitution patterns between broad aggregates of products (a level of aggregation which eliminates product entry and exit) as, for example, in Christensen et al. (1975), Deaton and Muelbauer (1980), and Hausman et al. (1994), or to estimate the demand system using data only from markets (e.g. time periods) when all goods are present as, for example, in Hausman (1994), Ellison et al. (1997), and

[1] 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, the Translog Reciprocal Indirect Utility Function and Diewert's Reciprocal Indirect Utility Function. See for example Varian (1984) for a discussion of these models and further references. More recently, Chan (2006) develops a continuous hedonic-choice model that is able to cope with entry and exit of products. However, his approach, has the undesirable feature of not modelling unobserved product characteristics.

Pinkse and Slade (2004). The former type of solution involves resorting to an analysis of aggregate data which clearly limits our ability to describe the substitution patterns between the goods actually being purchased by consumers. The latter type of solution, on the other hand, involves resorting to data only from markets when all goods are present, which while effective (if potentially inefficient) in some markets, such as the pharmaceutical markets studied by Ellison et al. (1997) where generic entry is driven by loss of patent protection so all entry occurs within a very constrained period in the data, would be largely impractical in other arena where product entry and exit occur simultaneously.

In contrast with existing literature, Proposition 1 establishes that a demand system derived from an indirect utility function that satisfies global consistency and non-satiation *can* be estimated using datasets where significant product entry and exit occurs. Removing a good from the choice set is entirely equivalent to increasing its price to infinity.

We now move on to specify a computationally convenient change of variable. Let, without loss of generality, $V(p, y; \theta, \Im) = H(r, y; \theta, \Im)$ where r denotes the element-by-element inverse $(J+1)$ vector of p_i: $r_i = 1/p_i \in \Re$. As it is easy to verify, under the new indirect utility function, removing a good k from the choice set is entirely equivalent to decreasing the corresponding r_k to zero. Definition 2 below provides a restatement of Definition 1 in terms of the set of (relatively easy to verify) conditions on the function $H(r, y; \theta, \Im)$ 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.

Definition 2. *A globally consistent indirect utility function* $V(p, y; \theta, \Im) = H(r, y; \theta, \Im)$ *possesses the following properties:*

1. *Continuous in* p *and* y*.*

2. *Homogeneous of degree zero in* p *and* y*.*

3. *Strictly increasing in* y *and nonincreasing in* p_i *for any* $i \in \Im$*.*

4. *Quasiconvex, that is, the set* $\{(p, y) : V(p, y; \theta, \Im) \leq \bar{v}\}$ *is convex for any* $\bar{v}$*.*

5. $\lim_{p_k \to \infty} V(p, y; \theta, \Im) = \lim_{r_k \to 0} H(r, y; \theta, \Im) = V(p, y; \theta, \Im') = H(r, y; \theta, \Im')$ *for any set of products* $\Im' \subset \Im$ *and for all* $k \notin \Im'$*.*

Following the vast majority of the continuous-choice literature, we will describe the demand system derived from the globally consistent indirect utility function $H(r, y; \theta, \Im)$ in terms of the Marshallian budget share functions:

$$w_i(r, y; \theta, \Im) = \frac{H_i(r, y; \theta, \Im)\, r_i}{H_y(r, y; \theta, \Im)\, y}, \quad (1)$$

for $i \in \Im$ and where $H_i(r, y; \theta, \Im)$ denotes the marginal utility with respect to r_i, $H_y(r, y; \theta, \Im)$ denotes the marginal utility of income, and finally $w_i(r, y; \theta, \Im) = p_i q_i(r, y; \theta, \Im)/y$ denotes the budget share of good i.

The actual algebraic functional form for the indirect utility function $H(r, y; \theta, \Im)$ is unknown to the econometrician. Following the continuous-choice literature, we approximate it with a flexible functional form so not to restrict the derived price substitution patterns. In particular, we approximate the indirect utility function with a normalized quadratic function, following Berndt et al. (1977), McFadden (1978a), and Pinkse and Slade (2004), in which both the representative consumer income y and the inside goods r_i have been normalized by the outside option r_0. The normalized approximation is given by:

$$H(r, y; \theta, \Im) = a_0 + \sum\nolimits_{m=1}^{J} a_m r_m + 1/2 \sum\nolimits_{m=1}^{J} \sum\nolimits_{n=1}^{J} b_{mn} r_m r_n + c_0 y + \sum\nolimits_{m=1}^{J} c_m r_m y, \quad (2)$$

where, with a slight abuse of notation, $H(r, y; \theta, \Im)$ denotes the normalized indirect utility function, and both r and y denote the vectors of the corresponding normalized variables. a_0 is a scalar parameter, $a = [a_i]$ is a J vector of parameters that, as we will discuss below, will be used to capture the vector of observed budget shares, $B = [b_{ij}]$ is a $J \times J$ matrix of parameters that will be used to capture the price substitution patterns, and finally $c = [c_i]$ is a J vector of parameters that will be used to capture the income effects.

It is easy to verify that the function $H(r, y; \theta, \Im)$ satisfies properties 1, 2 and 5 for a globally consistent indirect utility function. It is an homogeneous of degree zero functional form and continuous in both p and y. Moreover, it can be specialized down in an entirely consistent fashion to generate demand systems over arbitrary subsets of the goods (by setting $r_k = 0$ for those goods not in the current choice set). However, if we allow the parameters to be unrestricted in sign and magnitude, it does not necessarily satisfy properties 3 and 4. We can either impose a set of restrictions on the parameters that, given the vector of prices and income, ensure $H(r, y; \theta, \Im)$ satisfies those properties (the Mathematical Appendix lists the set of those implied restrictions) or, alternatively, we may choose, even in the absence of an underlying model of utility, not to impose such restrictions *a priori*, but rather test whether the data is consistent with them.

If $H(r, y; \theta, \Im)$ has the properties of an indirect utility function then standard duality results imply that the demand system is easily obtained via Roy's identity for each inside good $i = 1, \ldots, J$:

$$w_i(r, y; \theta, \Im) = \frac{a_i r_i + 1/2 \sum_{m=1}^{J} (b_{im} + b_{mi}) r_i r_m + c_i r_i y}{c_0 y + \sum_{m=1}^{J} c_m r_m y}, \tag{3}$$

where $w_i(r, y; \theta, \Im)$ denotes the budget share of good i. The derived budget share function has three characteristics that are important to discuss. First, it satisfies Proposition 1 and hence removing a good from the choice set (equivalent to set $r_i = 0$) explicitly forces the level of demand to zero. As a consequence, the model is able to match the shares of those goods with zero observed demand and hence can be estimated using datasets where significant product entry and exit occurs. Second, although it is only defined explicitly for the J inside choices (given the normalization with respect to the outside option), the budget constraint implies that a complete model of demand for the $J + 1$ budget shares is derived implicitly. Finally, because it is homogeneous of degree zero in the parameters, identification requires a normalization. Without loss of generality, we normalize c_0.

In many policy applications, including merger simulations, the key object of interest is the matrix of own- and cross-price demand elasticities. The analytical expressions for the budget share, price and income elasticities implied by the model with respect to any given inside goods i and j, are the following:

$$\begin{aligned} \varepsilon_{ij}^{s}(r, y; \theta, \Im) &= -1(j = i) - \frac{1/2 (b_{ij} + b_{ji}) r_j r_i (w_i y)^{-1} - c_j r_j}{c_0 + \sum_{m=1}^{J} c_m r_m} \\ \eta_i^{s}(r, y; \theta, \Im) &= -1 + \frac{c_i r_i}{c_0 w_i^A + \sum_{m=1}^{J} c_m r_m w_i}, \end{aligned} \tag{4}$$

where ε_{ij}^{s} and η_i^{s} denote the price and income elasticities, respectively. $[b_{ii}] > 0$ constitutes a sufficient condition (although not necessary) for a downward sloping own demand curve. We may expect $[b_{ij}] < 0$ if goods i and j are substitutes, but this constitutes neither a necessary nor a sufficient condition, since the total price effect depends on the size of the income effect. From the *budget share* elasticities, we can straightforwardly obtain the implied *demand* elasticities for the corresponding inside goods, as follows: $\varepsilon_{ii}^{d} = \varepsilon_{ii}^{s} - 1$, $\varepsilon_{ij}^{d} = \varepsilon_{ij}^{s}$ and $\eta_i^{d} = \eta_i^{s} + 1$, where ε_{ij}^{d} and η_i^{d} denote the demand price and income elasticities, respectively. While the elasticities involving the outside good can not be estimated directly (as the model is defined only over the J inside goods), the budget constraint implies that those elasticities can, nevertheless, still be recovered from equilibrium behavior, given the elasticities for the inside goods.

An important advantage of this specification is that, as we show below, it is relatively simple and fast to estimate which can prove a key advantage in competition policy issues, where time and transparency are typically crucial factors.

2.2 The More General Budget Share Function

Sometimes, it may be interesting to estimate a model which is more general than the model described thus far. The reason is twofold. First, as discussed above, we may choose, even in the absence of an underlying model of utility, to estimate a model without imposing *a priori* restrictions to the parameters (in sign and/or magnitude) and test whether the data is consistent with the properties of a globally consistent indirect utility function.

Second, we may choose to estimate a specification that is asymmetric with reference to the matrix B of parameters. As it is easy to note, the specification described above is observationally equivalent to a symmetric model with $b_{ij} = b_{ji} = 1/2\,(b_{ij} + b_{ji})$. Both the budget share and elasticities functions do not depend specifically on the individual $[b_{ij}]$ parameters, but only on their sum. Although this property of the model provides a great advantage in terms of the estimation procedure (as the number of parameters to be estimated decreases substantially), it implicitly restricts the flexibility properties of the model. The price substitution patterns of a demand system with $J+1$ goods can assume up to $(J+1)^2$ arbitrary values (or J^2 if we focus only on the inside goods). Under a symmetric specification, the model requires estimates of $J\,(J+2)\,/2$ parameters $[b_{ij}]$, which are clearly insufficient to assume the required arbitrary values. Following the continuous-choice tradition, we may often wish to estimate a model which does not impose such symmetry restrictions.

Specifically, consider the following *more general budget share function* for the case in which properties 3 and 4, as well as the symmetry assumptions, are not imposed *a priori*:

$$s_i\,(r, y; \theta, \Im) = \frac{a_i r_i + \sum_{m=1}^{J} b_{im} r_m r_i + c_i r_i y}{c_0 y + \sum_{m=1}^{J} c_m r_m y}. \qquad (5)$$

In these circumstances, the model is obviously not consistent with consumer utility maximization, but rather it nests a model which is and hence we can test the validity of such assumptions. If these restrictions are consistent with the patterns in the data, we can subsequently impose them on the model.

The analytic expressions for the price and income elasticities implied by the more general budget share are:

$$\begin{aligned}
\varepsilon_{ij}^{s}\left(r,y;\theta,\Im\right) &= -1\left(j=i\right)-\frac{b_{ij}r_jr_i\left(s_iy\right)^{-1}-c_jr_j}{c_0+\sum_{m=1}^{J}c_mr_m} \\
\eta_{i}^{s}\left(r,y;\theta,\Im\right) &= -1+\frac{c_ir_i}{c_0s_i+\sum_{m=1}^{J}c_mr_ms_i}.
\end{aligned} \tag{6}$$

2.3 Product Characteristics

Lancaster (1971) suggests that consumers are interested in goods because of the characteristics they provide. Classical choice models can be generalized to incorporate such proposal and introduce preferences directly over product characteristics. The model described in this paper is no exception. This introduction places no additional restrictions or structure on the form of the indirect utility function $H\left(r,y;\theta,\Im\right)$ as, in principle, such characteristics may enter through any of the parameters of the model in an arbitrary fashion. Moreover, it carries three important advantages.

First, it allows the introduction of a structural error term addressing the important critique offered by Brown and Walker (1989) which warns about the potential risks of simply 'tagging' on linear error terms to the end of budget share equations.

Let the set of parameters θ be decomposed, for notational purposes, into $\theta=(\theta_1,\theta_2)'$, where θ_1 refers to the $[a_i]$ parameters for $i=1,\ldots,J$ and θ_2 refers to the remaining ones ($[b_{ij}]$ and $[c_i]$). We introduce the random utility hypothesis by assuming that all product characteristics are observed by the consumer, but not necessarily by the econometrician. In particular, we follow Pinkse et al. (2002) and map the θ_1 parameters onto the characteristics space:

$$\theta_1\equiv a_i\left(x_i,\xi_i;\beta\right), \tag{7}$$

where x_i denotes the K-dimensional vector of characteristics associated with good i, observed by both the consumer and the econometrician, ξ_i denotes a one-dimensional vector of characteristics that are observed by the consumer, but not by the econometrician, and finally β refers to the K-dimensional vector of taste parameters associated with the observed characteristics. The precise functional form for $a_i\left(x_i,\xi_i;\beta\right)$ is an issue that can be examined using conventional testing procedures. We follow the industrial economics literature and assume a linear specification:

$$a_i\left(x_i,\xi_i;\beta\right)=\sum_{k=1}^{K}\beta_kx_{ki}+\xi_i, \tag{8}$$

which has the desirable property of being monotonic in the value of a given product's characteristics. The presence of unobserved product characteristics allows for a product-level source of sampling error, giving an explicit structural interpretation to the error term.

The second advantage of introducing product characteristics is related to the fact that it can substantially reduce the number of parameters to be estimated. If the number of goods $J+1$ is large, a dimensionality problem may arise as the model may yield too many parameters to be estimated with the available data. In that case, the number of parameters can be reduced by also mapping the θ_2 parameters onto the characteristics space (whenever the number of characteristics is smaller than the number of goods):

$$\begin{aligned} b_{ii} &= g_1\left(x_{1i};\alpha_1\right) \\ b_{ij} &= g_2\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right) \\ c_i &= g_3\left(x_{3i};\alpha_3\right), \end{aligned} \tag{9}$$

where $g_1\left(x_{1i};\alpha_1\right)$ is a function of a set $L1$ of good i's characteristics, $g_2\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right)$ is a function of a distance metric between goods i and j in the set of characteristics space $L2$, and finally $g_3\left(x_{3i};\alpha_3\right)$ is a function of a set $L3$ of good i's characteristics. In theory, all observed characteristics could be mapped onto both θ_1 and θ_2 sets of parameters. In this case, the sets $L1$, $L2$ and $L3$ will all coincide with the set of K-observed characteristics. In real-world applications, however, we hypothesize that if we do so, some high correlation may introduce biases in the estimation procedure. For this reason, we suggest that if all characteristics are allocated to all sets, some transformation should be used (see Pinkse and Slade, 2004).

The precise mapping is, again, a functional form issue that can be examined using conventional testing procedures. The following specifications are among the possible alternatives:

$$\begin{aligned} b_{ii} &= x_{1i}'\alpha_1 \\ b_{ij} &= d_{ij}\left(x_{2i},x_{2j};\alpha_2\right) \\ c_i &= x_{3i}'\alpha_3, \end{aligned}$$

or:

$$\begin{aligned} b_{ii} &= \exp\left(x_{1i}'\alpha_1\right) \\ b_{ij} &= \exp\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right) \\ b_{ij} &= \exp\left(x_{3i}'\alpha_3\right), \end{aligned}$$

where the distance metric could be defined as $d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)=\sqrt{\sum_{l=1}^{L2}\alpha_{2l}\left(x_{2li}-x_{2lj}\right)^2}$ or

$d_{ij}(x_{2i}, x_{2j}; \alpha_2) = \sum_{l=1}^{L2} \alpha_{2l} |x_{2li} - x_{2lj}|$.[2] Another obvious alternative is to estimate the functions nonparametrically as in Pinkse et al. (2002). Independently of the specification chosen though, the important fact is that the θ_2 parameters are mapped onto the characteristics space and hence the number of parameters to be estimated is reduced while allowing the estimates to still be data-driven.

The third and final advantage of introducing product characteristics is related to the new goods problem (Ackerberg et al., 2007). If a demand system is defined over the product space, we can not investigate demand behavior for goods not yet introduced. The introduction of product characteristics solves this problem and makes the analysis of issues related to incentives for entry, possible.

2.4 Flexibility

An algebraic functional form for a complete system of consumer budget share functions $s_i(r, y; \theta, \Im)$ is said to be Diewert flexible (see Diewert, 1973, 1974; and Lau, 1986) if, at any given set of non-negative prices and income, the parameters can be chosen so that the complete system of consumer budget share functions, their own- and cross-price demand and income elasticities are capable of assuming arbitrary values at the given set of prices and income (subject only to the requirements of theoretical consistency). Barnett (1983) proved that flexibility in the sense defined by Diewert is necessary and sufficient for a function to satisfy the mathematical definition of a local second order approximation. In this section, we follow Diewert (1973, 1974) and show that both our specification is flexible in that sense.[3]

Matching Predicted to Observed Shares

The first step in establishing flexibility is to show that, for every set of θ_2 parameters, there is a unique value of θ_1 that equates the shares predicted by the model $s_i(r, y; \theta, \Im)$ with the observed shares s_i^{true}, where $s_i(r, y; \theta, \Im)$ denotes the *more general budget share function*. Recall that θ_1 refers to the $[a_i]$ parameters for $i = 1, \ldots, J$ and θ_2 refers to the remaining ones ($[b_{ij}]$ and $[c_i]$). This step ensures the model can always match the vector of observed shares, one requirement for a model to be a Diewert flexible functional form. We then proceed by showing that the set of θ_2 parameters is such that the predicted elasticities are capable of assuming arbitrary values.

[2] As a technical note, if *a priori* we want to impose $b_{ij} \leq 0$, the following alternatives are possible: $b_{ij} = -\exp(d_{ij}(x_{2i}, x_{2j}; \alpha_2))$ or $b_{ij} = -d_{ij}(x_{2i}, x_{2j}; \alpha_2)$.

[3] Our problem differs from Diewert (1971, 1973, 1974) in the sense that we consider flexible functional forms as approximations to indirect utility and demand functions rather than to cost and production functions (Diewert, 1971) or profit and transformation functions (Diewert, 1973) or revenue and factor requirements functions (Diewert, 1974).

Let $\Im_0^+$ denote the subset of $\Im$ that includes the goods that are effectively present in the market and exhibit, as a consequence, strictly positive observed shares: $\Im_0^+ \equiv \left\{i | s_i^{obs} > 0, i \in \Im\right\}$. Further, let J_0^+ denote the number of goods in set $\Im_0^+$. Finally, let $\Im^+$ define the subset of $\Im_0^+$ that includes only the inside goods (in other words, the subset of $\Im_0^+$ that excludes the outside good).

Proposition 2. *The* ***more general budget share function*** $s_i\left(r, y; \theta, \Im\right)$ *can match any vector of observed budget shares. Furthermore, there exists a unique set of* θ_1 *parameters that matches the subset* $\Im_0^+$ *of goods with strictly positive observed demand.*

Proof. Proposition 1 establishes that the model explicitly matches predicted and observed demand for all goods that exhibit zero observed shares and hence we can proceed by restricting our analysis to the subset of remaining goods in $\Im$.

We begin by showing that the system of inside goods $J_0^+ - 1$ equations $s_i^{true} = s_i\left(r, y; \theta, \Im\right)$ for $i \in \Im^+$ has exactly one solution $a\left(r, y, s^{true}; \theta_2, \Im\right)$ that equates the shares predicted by the model to the observed shares. In order to see why this is the case, note that the system of equations:

$$s_i^{true} = \frac{a_i r_i + \sum_{m=1}^{J} b_{im} r_m r_i + c_i r_i y}{c_0 y + \sum_{m=1}^{J} c_m r_m y}$$

for $i \in \Im^+$ is linear in the vector $a = \left(a_1, \ldots, a_{J^+}\right)'$ and can be rewritten as $Da = G$, where D denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ diagonal matrix with diagonal elements $d_{ii} = r_i$ and G denotes a $J_0^+ - 1$ vector with elements $g_i = s_i^{true}\left(c_0 y + \sum_{m=1}^{J} c_m r_m y\right) - \sum_{m=1}^{J} b_{im} r_m r_n - c_i r_i y$.

It is well known that a sufficient and necessary condition for uniqueness of a system of $J_0^+ - 1$ linear equations with $J_0^+ - 1$ unknowns is that the matrix D is nonsingular. And a square matrix is nonsingular if and only if its determinant is nonzero.

The determinant of the diagonal matrix D is the product of its diagonal elements. In the case above, $\det\left(D\right) = \prod_{i \in \Im^+} r_i$. Because the subset $\Im^+$ only includes those goods that exhibit strictly positive observed shares, we know that $r_i \neq 0$ for all $i \in \Im^+$. As a consequence, matrix D is nonsingular. The nonsingularity of D establishes that there is a unique vector of a's associated with each good that solves the system of equations $s_i^{true} = s_i\left(r, y; \theta, \Im\right)$ for $i \in \Im^+$. Once the shares of the $J_0^+ - 1$ inside goods are matched, the share of the outside good will automatically be matched as a consequence. The model is therefore able to equate predicted and actual budget shares for all goods (both inside and outside) with strictly positive observed shares.

We have already noted that Proposition 1 establishes that the model is able to match the shares of those goods with zero observed demand. Thus, as the proposition claims, the model can match any vector of observed budget shares. □

Arbitrary Elasticities

Given that the model is capable of matching observed with predicted budget shares, we proceed by investigating if the set of θ_2 parameters are such that the model is able to also assume arbitrary values for the predicted elasticities - which concludes the flexibility result.

Proposition 3. *There exists a set of θ parameters such that the* ***more general budget share function*** $s_i(r, y; \theta, \Im)$ *can match* any vector of budget shares*,* any vector of income elasticities for the inside goods*, and finally* any matrix of own- and cross-price elasticities for the inside goods*.*

Proof. In order to establish Proposition 3, we want to show that we can choose the set of parameters θ so that the model satisfies simultaneously the following equations:

$$\begin{aligned} s_i^{true} &= s_i(r, y; \theta, \Im) \\ \eta_i^{true} &= \eta_i^s(r, y; \theta, \Im) \\ \varepsilon_{ij}^{true} &= \varepsilon_{ij}^s(r, y; \theta, \Im), \end{aligned}$$

where s_i^{true}, η_i^{true} and ε_{ij}^{true} denote any true vector of budget shares, any true vector of income elasticities for the inside goods, and finally any true matrix of own- and cross-price elasticities for the inside goods.

Proposition 2 establishes that the more general budget share function $s_i(r, y; \theta, \Im)$ can match any vector of observed budget shares. It establishes also that, for any set of $[b_{ij}]$ and $[c_i]$ parameters (θ_2), there exists a unique set of $[a_i]$ parameters (θ_1) that matches the subset $\Im_0^+$ of goods with strictly positive observed demand. Given such solution for θ_1, we want to show that the θ_2 parameters can be chosen simultaneously to equate the inside goods' (i) own- and cross-price elasticities as well as (ii) income elasticities predicted by the model to the observed ones.

We begin by investigating the ability of the model to match predicted with true income elasticities for inside goods. Since the model can match any vector of budget shares, the income elasticities predicted by the model are given by:

$$\eta_i^s(r, y; \theta, \Im) = -1 + \frac{c_i r_i}{c_0 s_i^{true} + \sum_{m=1}^{J} c_m r_m s_i^{true}},$$

for $i \in \Im^+$. It is easy to show that the system of $J_0^+ - 1$ inside goods equations $\eta_i^{true} = \eta_i^s(r, y; \theta, \Im)$ is linear in the $[c_i]$ parameters and can be rewritten as follows: $Dc = G$, where D denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ matrix with diagonal elements $d_{ii} = r_i s_i^{true}\left(\eta_i^{true} + 1\right) - r_i$

and cross-diagonal elements $d_{ij} = r_j s_i^{true}(\eta_i^{true} + 1)$, and finally G denotes a $J_0^+ - 1$ vector with elements $g_i = c_0 s_i^{true}(\eta_i^{true} + 1)$.

It is well know that a sufficient and necessary condition for uniqueness of a system of linear $J_0^+ - 1$ equations with $J_0^+ - 1$ unknowns is that the matrix D is nonsingular. A square matrix is nonsingular if and only if its determinant is nonzero. We begin by notting that the matrix D can rewritten as the product of two square matrices: $D = D_1 D_2$, where D_1 denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ matrix with diagonal elements $d_{ii}^1 = s_i^{true}(\eta_i^{true} + 1) - 1$ and cross-diagonal elements $d_{ij} = s_i^{true}(\eta_i^{true} + 1)$, while D_2 denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ diagonal matrix with diagonal entries $d_{ii}^2 = r_i$. The determinant of the product of two square matrices is the product of the individual determinants.

In order to compute the determinant of D_1, we note that a property of the determinant is that adding a multiple of one row to another row, or a multiple of one column to another column, does not affect its value. In particular, consider the following sequence of linear operations: (i) subtract the first column to all remaining columns, and (ii) add to the first row all other rows. This sequence yields a triangular matrix with the first diagonal element given by $d_{11} = -1 + \sum_{i \in \Im^+} s_i^{true}(\eta_i^{true} + 1)$ and the remaining diagonal elements given by $d_{ii} = -1$. The determinant of a triangular matrix equals the product of its diagonal elements: $\det(D_1) = \prod_{i \in \Im^+} d_{ii}^1 = (-1)^{J_0^+ - 1} s_0^{true}(\eta_0^{true} + 1)$, which is nonzero for an outside good with a strictly positive share and a *demand* income elasticity different from zero.[4] If this was not the case, the argument still follows if we renormalize the parameters in terms of an inside good with a strictly positive market share and a *demand* income elasticity different from zero. The determinant of the diagonal matrix D_2 is the product of its diagonal elements: $\det(D_2) = \prod_{i \in \Im^+} d_{ii}^2 = \prod_{i \in \Im^+} r_i$. Because the subset $\Im^+$ only includes those goods that exhibit strictly positive observed shares, we know that $r_i \neq 0$ for all $i \in \Im^+$. As a consequence, matrix D_2 is nonsingular. The determinant of matrix D is therefore the product of the individual determinants: $\det(D) = (-1)^{J_0^+ - 1} s_0^{true}(\eta_0^{true} + 1) \prod_{i \in \Im^+} r_i$, which is nonzero for an outside good with a strictly positive share and a *demand* income elasticity different from zero. Again, if this was not the case, the argument still follows if we renormalize the parameters in terms of an inside good with a strictly positive market share and a *demand* income elasticity different from zero.

The nonsingularity of D establishes that there is a unique vector of the $[c_i]$ parameters, independent of the set of $[b_{ij}]$ parameters, that solves the system of equations $\eta_i^{true} = \eta_i^s(r, y; \theta, \Im)$ and matches the $J_0^+ - 1$ vector of predicted to true income elasticities associated with the inside goods.

We proceed by considering the price elasticities. Under the asymmetry assumption of

[4] Recall the relationship between the budget share and the demand income elasticities. $\eta_i^d = \eta_i^s + 1$. A budget share income elasticity $\eta_i^{s,true} = -1$ is equivalent to a demand income elasticity $\eta_i^{d,true} = 0$.

the more general budget share function, we show that the model predicts elasticities that are able to assume arbitrary $\left(J_0^+ - 1\right)^2$ values for the inside goods. We start by noting that since the model can match any vector of budget shares, the own- and cross-price elasticities predicted by the model are given by:

$$\varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right) = -1\left(j = i\right) - \frac{b_{ij} r_j r_i \left(s_i^{true} y\right)^{-1} - c_j r_j}{c_0 + \sum_{m=1}^{J} c_m r_m}.$$

It remains to show that there is a vector of $[b_{ij}]$ parameters that solve all the inside goods $\varepsilon_{ij}^{true} = \varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right)$ equations simultaneously for any $i, j \in \Im^+$. The problem is that the system of price elasticities equations depends on the $[c_i]$ parameters. However, we have shown that the vector of $[c_i]$ parameters that matches the $J_0^+ - 1$ vector of predicted to true income elasticities associated with the inside goods is independent of the set of $[b_{ij}]$ parameters and depends only on the vectors r, s_i^{true} and η_i^{true}. As a consequence, we can solve the joint system of equations by recursion. In particular, we have that setting:

$$\begin{aligned} b_{ii} &= \left(s_i^{true}\eta_i^{true} + s_i^{true} - 1 - \varepsilon_{ij}^{true}\right)\frac{s_i^{true} y}{r_i r_i} f\left(r, s^{true}, \eta^{true}\right) \\ b_{ij} &= \left(s_j^{true}\eta_j^{true} + s_j^{true} - \varepsilon_{ij}^{true}\right)\frac{s_i^{true} y}{r_j r_i} f\left(r, s^{true}, \eta^{true}\right), \end{aligned}$$

for any $i, j \in \Im^+$ matches the $\left(J_0^+ - 1\right)^2$ vector of predicted to true price elasticities for the inside goods, where $f\left(r, s^{true}, \eta^{true}\right)$ denotes the expression $\left(c_0 + \sum_{m=1}^{J} c_m r_m\right)$ after substitution the $[c_i]$ parameters for the corresponding implicit solution that matches the vector of true income elasticities for the inside goods. □

3 Identification and Estimation

The identification of the different parameters requires a set of normalizations we now describe. We have already noted that the budget share functions are homogeneous of degree zero in the parameters and hence are identified up to a scalar. Without loss of generality, we normalize c_0. Given this normalization, the identification of the remaining parameters is standard given a large enough sample. The $[a_i]$ parameters in θ_1 are identified from variation in the budget shares across the different goods. Having identified the $[a_i]$ parameters, the taste parameters in vector β are identified from variations in the observed product characteristics. The $[b_{ii}]$ and $[b_{ij}]$ parameters in θ_2 are identified from variation in prices, with the identification of the former relying in variation from own-prices, and the latter in variation from competitors prices. Finally, the $[c_i]$ parameters are identified from variations in income.

Of course, in many instances it will be appropriate to use instruments rather than (say) the variation in the actual prices to empirically identify the model's parameters, an approach we discus further below.

The estimation algorithm that we propose is based in Berry (1994) and Berry et al. (1995), and includes four steps that we now describe.[5] Although the sample dimension will not impact the general setup of the algorithm, it will have an affect on the set of θ_2 parameters. If the number of goods is relatively small so that the dimensionality problem does not constitute an issue, θ_2 will include the set of $[b_{ij}]$ and $[c_i]$ parameters, whereas if, on the other hand, the number of goods yields a too great number of parameters to be estimated, a mapping needs to be defined and θ_2 will include instead the vector α.

Step 1 *Set initial values for the parameters in θ_2.*

The choice of the initial values for θ_2 is always arbitrary and ideally would not have an impact on the final parameter estimates. Unfortunately, this may fail to happen on the class of highly nonlinear demand models for differentiated products that involve simulation - of which Berry et al. (1995) constitutes the most extensively used example in empirical exercises. Despite the work on identification by Berry et al. (2004), Berry and Haile (2008), and Fox and Gandhi (2016), numerical problems difficulties can arise. Knittel and Metaxoglou (2014) explore the issue of potential multiple local minima using Berry et al. (1995)'s estimator and find that convergence may occur at a number of local extrema, at saddles and in regions of the objective function where the first-order conditions are not satisfied.

Our estimation procedure, although motivated by and closely related to Berry (1994) and Berry et al. (1995), will prove to be globally convex and hence avoid some of the numerical issues related to local extrema - as our monte carlo simulations below demonstrate. Furthermore, we propose an alternative algorithm in the spirit of Davis and Schiraldi (2014) and Dubé et al. (2012) that improves the rate of convergence substantially.

Step 2 *Computation of the budget shares conditional on the θ_2 parameters*

Propositions 2 and 3 establish that, conditional on the θ_2 parameters, the more general budget share function can match any vector of observed budget shares. In Step 2, we solve for the unique θ_1 parameters that, given each guess for the set of parameters in θ_2, match the observed s_i^{true} with the predicted $s_i\left(r, y; \theta, \Im\right)$ budget shares in *each* market or time period.

[5] The marginal utility of income is given by the price index $c_0 + \sum_{m=1}^{J} c_m r_m$. If, following Pinkse and Slade (2004), the researcher is willing to assume this price index to be constant, then the specification can be estimated by a simple linear instrumental regression approach.

The solution to this problem is analytical:

$$a_i = \left(s_i^{true}/r_i\right)\left(c_0 y + \sum\nolimits_{m=1}^{J} c_m r_m y\right) - \sum\nolimits_{m=1}^{J} b_{im} r_m - c_i y. \tag{10}$$

Let $a\left(r, y, s^{true}; \theta_2, \Im\right)$ denote the solution vector of the a's that ensure that the observed s_i^{true} and the predicted $s_i\left(r, y; \theta, \Im\right)$ budget shares are equated.

Step 3 *Computation of the structural error*

Having solved for the unique θ_1 parameters that match, in *each* market or time period, the observed budget shares s_i^{true} with the predicted ones $s_i\left(r, y; \theta, \Im\right)$, we proceed by running a Berry (1994) style regression on the following relationship:

$$a_{it}\left(r, y, w^{true}; \theta_2, \Im\right) = \sum\nolimits_{i=1}^{K} \beta_k x_{kit} + \xi_{it}, \tag{11}$$

and obtain estimates for the vector β of parameters and for the $[\xi_{it}]$ unobserved characteristics. The latter estimates will be a function of both β and θ_2 vectors of parameters and will be used to compute the objective function of a Generalized Method of Moments (GMM) procedure.

Step 4 *Estimation of the θ_2 parameters*

Estimate the θ_2 parameters by GMM. The approach relies on an identifying restriction on the distribution of the true unobserved characteristics and is based on the sample analogue to the population condition.

The standard identifying restriction states that, at the true values of the parameters, $\theta^{true} = \left(\theta_1^{true}, \theta_2^{true}\right)'$, the true unobserved characteristics are mean independent of a set of M instruments $Z_{it} = \left(z_{1it}, \ldots, z_{Mit}\right)$:

$$E\left(\xi_{it}\left(\theta^{true}\right)|Z_{it}\right) = 0. \tag{12}$$

Please note that other identifying restrictions would also enable the estimation of the model. In particular, given the typical panel structure of the data, an alternative assumption could incorporate the likelihood of the econometric error and the set of instruments to be more similar for a given product across time, than for those of different products. Please see Berry et al. (1995) and Davis and Schiraldi (2014) for a more detailed analysis on this subject.

The above population moment condition can be used, akin to Hansen (1982), to render a method of moments estimator of θ by interacting the estimated unobserved characteristics with the set of instruments, and then search for the value of the θ parameters that set the sample analogues of the moment conditions as closed as possible to zero. Let $G_n(\theta)$ denote the sample analogues of the moment conditions:

$$G_n(\theta) = \frac{1}{n} \sum\nolimits_{i=1}^{J} \sum\nolimits_{t=1}^{T} \tilde{\xi}_{it}(\theta)\, \tilde{Z}'_{it}, \tag{13}$$

where for notational purposes $\tilde{\xi}_{it}(\theta) = \xi_{it}(\theta)\,\chi_{it}$, $\tilde{Z}_{it} = (z_{1it}\chi_{it}, \ldots, z_{Mit}\chi_{it})$, and $\chi_{it} = 1$ if good i is sold in market t and zero otherwise. χ_{it} provides, thereby, a missing value indicator used to compute $n = \sum_{i=1}^{J} \sum_{t=1}^{T} \chi_{it}$.

Formally, the method of moments estimator for $\hat{\theta}$ is the argument that minimizes the weighted norm criterion of $G_n(\theta)$, for some weighting matrix A_n with rank at least equal to the dimension of θ:

$$\hat{\theta} = \arg\min_{\theta} \quad \|G_n(\theta)\|_{A_n} = G_n(\theta)' A_n G_n(\theta). \tag{14}$$

The strong non-linearity of the objective function requires a minimization routine. The non-linear search over θ can be simplified by making use of the fact that the first order conditions for a minimum of $\|G_n(\theta)\|_{A_n}$ are linear for the subset β of the θ_1 parameters of estimation in $\theta = (\theta_1, \theta_2)$. In particular, it is possible, given the standard instrumental variables results, to express the vector β as a function of θ_2, and limit the non-linear search over θ_2:

$$\hat{\beta} = \left(\tilde{X}'\tilde{Z}A_n^{-1}\tilde{Z}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{Z}A_n^{-1}\tilde{Z}'a\left(r, y, s^{obs}; \theta_2, \Im\right), \tag{15}$$

where $\tilde{X}$ denotes the $n \times K$ matrix of $[x_{kjt}\chi_{it}]$ observed characteristics, and $\tilde{Z}$ denotes the $n \times M$ matrix of $[z_{mit}\chi_{it}]$ instruments.

Standard Errors

In contrast to a model based on simulation, the GMM estimator for this model does not need to be corrected for simulation error and hence the standard formulae for a GMM estimator apply. Hansen (1982) establishes the formal conditions under which $\hat{\theta}$, the method of moments estimator, is consistent and asymptotically normal with bounded variance, consistently estimated as follows:

$$\sqrt{n}\left(\hat{\theta} - \theta^*\right) \sim N\left(0, \left(\widehat{\Gamma}' A_n \widehat{\Gamma}\right)^{-1} \widehat{\Gamma}' A_n \hat{\Phi} A_n \widehat{\Gamma} \left(\widehat{\Gamma}' A_n \widehat{\Gamma}\right)^{-1}\right), \tag{16}$$

where $\widehat{\Gamma}$ denotes a consistent estimator of the gradient of the objective function:

$$\widehat{\Gamma}' = \widehat{\Gamma_n\left(\hat{\theta}\right)} = \frac{1}{n}\sum\nolimits_{i=1}^{J}\sum\nolimits_{t=1}^{T}\left(\frac{\partial\tilde{\xi}_{it}\left(\hat{\theta}\right)\tilde{Z}_{it}'}{\partial\theta'}\right),$$

and $\hat{\Phi}$ denotes a consistent estimator of the variance-covariance matrix of the moment conditions $\hat{\Phi} = \widehat{Var\left(G_n\left(\hat{\theta}\right)\right)}$. The optimal weighting matrix is proportional to Φ^{-1}, giving less weigh to those moments with a higher variance.

We may follow the traditional literature on the continuous demand for differentiated products, wherein asymptotic arguments are assumed to work in the number of markets (or time periods) or, alternatively, we may follow the Berry et al. (1995) assumption that asymptotic arguments work in the number of products. Furthermore, the structure of the matrix $Var\left(G_n\left(\hat{\theta}\right)\right)$ will depend on our assumptions about the covariance structure of the unobserved product characteristics. We may, for example, assume that they are independent across markets, but allow for correlation across products within a given market or, alternatively, we may assume that they are correlated across markets and independent across products. Naturally, it is also possible - and probably desirable - to assume richer error variance structures. See, for example, Davis (2002).

4 Monte Carlo Experiment

4.1 Data-Generating Process

In this section, we describe the data-generating process for a Monte Carlo experiment designed to analyze the convergence properties of the demand model. We consider a setting with J goods in T markets. We allow the econometrician to observe the income in each market as well as the price and $K = 6$ characteristics of each good in each market. The income in market t, y_t, is drawn from a $(1000, 5000)$ uniform distribution. The observed characteristic k for good i in market t, x_{kit}, is drawn from a $(0, 1)$ uniform distribution.

Not all characteristics of a good are observed by the econometrician. We allow also for unobserved characteristics, ξ_{it}, which we draw from a $(-l_\xi, l_\xi)$ uniform distribution. The set of $[a_{it}]$ parameters are defined as $a_{it} = \beta_0 + \sum_{k=1}^{6}\beta_k x_{kit} + \xi_{it}$, with $\beta_0 = 50$ and $\beta_k = 1$ for all $k \geq 1$. With the variance of the observed characteristics fixed, l_ξ controls the 'noise-to-signal' ratio in the model. If l_ξ is *small*, we expect to require relatively smaller samples to consistently estimate the parameters than when compared with the *large* l_ξ case. The baseline case focus on $l_\xi = 0.50$, but we investigate the sensitivity of the estimates to other assumptions.

In order to simulate the endogeneity that arises from profit-maximizing price setting, we define prices to follow $p_{it} = \left|2.5 + \sum_{k=1}^{6} \beta_k x_{kit} + \xi_{it} + e_{it}\right|$, where e_{it} denotes a $(0,1)$ uniform innovation. In order to deal with this endogeneity problem, we construct a number M instruments correlated with price, but not with the unobserved characteristics. The instruments are derived following $z_{mit} = \sum_{k=1}^{6} \beta_k x_{kit} + e_{it} + 0.5 g_{mit}$, with g_{mit} being drawn from a $(0,1)$ uniform distribution. As we discuss below, we consider several sample designs in our Monte Carlo experiment in order to investigate the estimation properties of the model. However, for comparison purposes, we maintain the number of instruments constant across the sample designs. In particular, we consider a conservative number of instruments following a combination of Hausman and Bresnahan approaches, where observed product characteristics of a good in other markets become instruments for its price in a given market: $M = K\left[\min(T) - 1\right]$, where $\min(T)$ denotes the minimum of markets across the different sample designs.

The variation in the set of choices available to consumers can provide important information about substitution patterns. We allowed goods to be missing at random from a given market in order to mimic the entry and exit behavior characteristic of typical real-world datasets. In particular, a good was assumed missing in market t if the realization of a standard continuous-uniform random variable was less than 0.2 subject to the constraint of having at least one inside good.

4.2 θ_2 Parameters

We consider two broad sample designs: one where the number of different goods in a market is small, but a reasonably large number of markets exist, and another with a large number of goods marketed on a small number of markets. In both designs, we use a limited-memory BFGS optimization algorithm with analytical derivatives and a strict tolerance level of 1e-12 to find the estimates of θ_2 that make the sample analogue of the moment conditions as close to zero as possible.

We begin by addressing the small number of goods case and hence assume that the asymptotic arguments work in the number of independent markets. The true values of the parameters of the indirect utility function are assumed to be $b_{ii} = 40$, $b_{ij} = -10$ and $c_i = 6$ for all i, j. For expositional convenience, the symmetry assumption was imposed on the estimation of the $[b_{ij}]$ parameters.

Table 1 presents the mean and standard deviation (in parenthesis) of the GMM estimates of the θ_2 parameters across 50 sample experiments. All the results are conditional on $J = 3$ goods. Panel A explores the sensitivity of the estimates to the number of markets, with columns (i)-(iii) considering the following three cases: $T = 100$, $T = 200$ and $T = 400$. It

Table 1: *Monte-Carlo Results: θ_2 Estimates, Small J Case*

Panel A: Sensitivity to Number of Markets T				
Parameter	True	(i): 100	(ii): 200	(iii): 400
b_{11}	40.00	34.30 (2.08)	36.89 (1.86)	37.92 (1.37)
b_{21}	-10.00	-9.94 (0.60)	-10.03 (0.81)	-10.09 (0.67)
b_{22}	40.00	33.89 (2.18)	36.92 (1.73)	38.05 (1.36)
b_{31}	-10.00	-10.10 (0.86)	-9.93 (0.78)	-10.17 (0.78)
b_{32}	-10.00	-9.85 (0.69)	-10.19 (0.62)	-10.16 (0.82)
b_{33}	40.00	34.25 (2.30)	36.80 (1.60)	38.02 (1.38)
c_1	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_2	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_3	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
Panel B: Sensitivity to Noise-to-Signal Ratio l_ξ				
Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
b_{11}	40.00	29.47 (2.77)	37.92 (1.37)	40.00 (0.00)
b_{21}	-10.00	-10.35 (1.54)	-10.09 (0.67)	-10.00 (0.00)
b_{22}	40.00	29.65 (3.26)	38.05 (1.36)	40.00 (0.00)
b_{31}	-10.00	-10.37 (2.05)	-10.17 (0.78)	-10.00 (0.00)
b_{32}	-10.00	-10.42 (1.91)	-10.16 (0.82)	-10.00 (0.00)
b_{33}	40.00	29.46 (2.58)	38.02 (1.38)	40.00 (0.00)
c_1	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_2	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_3	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
Panel C: Sensitivity to Starting Values				
Parameter	True	(vi): $(1/2)\,\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
b_{11}	40.00	37.92 (1.37)	37.92 (1.37)	37.92 (1.37)
b_{21}	-10.00	-10.09 (0.67)	-10.09 (0.67)	-10.09 (0.67)
b_{22}	40.00	38.05 (1.36)	38.05 (1.36)	38.05 (1.36)
b_{31}	-10.00	-10.17 (0.78)	-10.17 (0.78)	-10.17 (0.78)
b_{32}	-10.00	-10.16 (0.82)	-10.16 (0.82)	-10.16 (0.82)
b_{33}	40.00	38.02 (1.38)	38.02 (1.38)	38.02 (1.38)
c_1	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_2	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)
c_3	6.00	6.00 (0.00)	6.00 (0.00)	6.00 (0.00)

* The table depicts the mean and standard deviation (in parenthesis) of the estimates of the θ_2 parameters across 50 Monte Carlo experiments. Panel A considers the case of $J = 3$, $l_\xi = 0.50$ and starting values at θ_2^{true}. Panel B considers the case of $J = 3$, $T = 400$ and starting values at θ_2^{true}. Panel C considers the case of $J = 3$, $T = 400$ and $l_\xi = 0.50$.

seems that the algorithm performs reasonably well, with both price and income parameters converging to the true values even at small sample sizes, although in what the price parameters are concerned, it seems to be easier to identify cross-price than own-prices parameters.

Panel B explores the sensitivity of the algorithm to the variance of the unobserved characteristics ξ_{it} that controls the noise-to-signal ratio. We consider the following three cases: $l_\xi = 1.50$, $l_\xi = 0.50$ and $l_\xi = 0.00$, with column (iii) just reproducing the baseline case from panel A for comparison easiness. As expected, the results suggest that an increase in the variance of ξ_{it} (and consequent increase in the noise-to-signal ratio) deteriorates the performance of the algorithm - which implies that we require additional data in order to obtain a given level of statistical precision. We should note also that, for the case where no unobserved characteristics exist ($l_\xi = 0$), the model identifies the exact true values of the parameters at every single replication of the Monte Carlo experiment. The reason is that the $[a_{it}]$ parameters are retrieved analytically and, as a result, no numerical error is introduced in the contraction inner loop (please see Dubé et al., 2012 for more details on numerical contraction induced errors).

The results presented on panels A and B of Table 1 are computed using the true parameters as starting points of the algorithm. Panel C explores the potential multiple local minima property of the GMM objective function in the lines of Dubé et al. (2012) and Knittel and Metaxoglou (2014) by considering multiple starting points. In particular, we consider two alternative starting points: $(1/2)\, \theta_2^{true}$ and $2\theta_2^{true}$, with column (iii) again just reproducing the baseline case from panel A. The results point to the robustness of the algorithm to the different starting points since it converges to the same parameter values each time.

To sum up, the results suggest three key features of the estimation procedure: (i) the estimators seem to be consistent, (ii) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally (iii) the GMM objective function seems to be have an apparent global minimum.

We now move to sample experiments with a number of goods that yield a too great number of parameters to be estimated. As a consequence, the $[b_{ij}]$ and $[c_i]$ parameters need to be mapped onto the characteristics space. In theory, all observed characteristics can be mapped onto both θ_1 and θ_2 sets of parameters. We choose to consider the case where the set of characteristics that affect the parameters in θ_1 is to some extent coincident to the set of characteristics that affect the parameters in θ_2. In particular, we consider that the $[b_{ii}]$ parameters are mapped onto the first two observed characteristics, the $[b_{ij}]$ parameters are mapped onto the second two observed characteristics, and the $[c_i]$ parameters are mapped onto the last two observed characteristics. The setup is identical to the one for the small

Table 2: *Monte-Carlo Results: θ_2 Estimates, Large J Case*

Panel A: Sensitivity to Number of Goods J				
Parameter	True	(i): 30	(ii): 60	(iii): 120
α_{10}	20.00	16.41 (3.39)	17.49 (3.26)	19.62 (3.39)
α_{11}	20.00	16.99 (5.50)	19.30 (6.26)	18.17 (6.22)
α_{12}	20.00	17.09 (5.65)	18.67 (6.01)	19.16 (5.25)
α_{20}	-5.00	-5.08 (0.09)	-4.99 (0.08)	-5.02 (0.09)
α_{21}	-5.00	-4.99 (0.11)	-5.01 (0.06)	-5.00 (0.04)
α_{22}	-5.00	-4.78 (0.11)	-5.00 (0.06)	-5.00 (0.03)
α_{30}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
α_{31}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
α_{32}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
Panel B: Sensitivity to Noise-to-Signal Ratio l_ξ				
Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
α_{10}	20.00	15.85 (5.30)	19.62 (3.39)	20.00 (0.00)
α_{11}	20.00	12.93 (9.58)	18.17 (6.22)	20.00 (0.00)
α_{12}	20.00	14.60 (9.52)	19.16 (5.25)	20.00 (0.00)
α_{20}	-5.00	-5.02 (0.19)	-5.02 (0.09)	-5.00 (0.00)
α_{21}	-5.00	-4.99 (0.09)	-5.00 (0.04)	-5.00 (0.00)
α_{22}	-5.00	-4.99 (0.08)	-5.00 (0.03)	-5.00 (0.00)
α_{30}	3.00	3.00 (0.01)	3.00 (0.00)	3.00 (0.00)
α_{31}	3.00	3.00 (0.01)	3.00 (0.00)	3.00 (0.00)
α_{32}	3.00	3.00 (0.01)	3.00 (0.00)	3.00 (0.00)
Panel C: Sensitivity to Starting Values				
Parameter	True	(vi): $(1/2)\,\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
α_{10}	20.00	19.62 (3.39)	19.62 (3.39)	19.62 (3.39)
α_{11}	20.00	18.17 (6.22)	18.17 (6.22)	18.17 (6.22)
α_{12}	20.00	19.16 (5.25)	19.16 (5.25)	19.16 (5.25)
α_{20}	-5.00	-5.02 (0.09)	-5.02 (0.09)	-5.02 (0.09)
α_{21}	-5.00	-5.00 (0.04)	-5.00 (0.04)	-5.00 (0.04)
α_{22}	-5.00	-5.00 (0.03)	-5.00 (0.03)	-5.00 (0.03)
α_{30}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
α_{31}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)
α_{32}	3.00	3.00 (0.00)	3.00 (0.00)	3.00 (0.00)

* The table depicts the mean and standard deviation (in parenthesis) of the estimates of the θ_2 parameters across 50 Monte Carlo experiments. Panel A considers the case of $T = 10$, $l_\xi = 0.50$ and starting values at θ_2^{true}. Panel B considers the case of $J = 120$, $T = 10$ and starting values at θ_2^{true}. Panel C considers the case of $J = 120$, $T = 10$ and $l_\xi = 0.50$.

number of goods case above with the exception that:

$$\begin{aligned} b_{ii,t} &= \alpha_{10} + \alpha_{11} x_{1it} + \alpha_{12} x_{2it} \\ b_{ij,t} &= \alpha_{20} + \alpha_{21} \left| x_{3it} - x_{3jt} \right| + \alpha_{22} \left| x_{4it} - x_{4jt} \right| \\ c_{i,t} &= \alpha_{30} + \alpha_{31} x_{5it} + \alpha_{32} x_{6it}, \end{aligned} \tag{17}$$

with $\alpha_{10} = \alpha_{11} = \alpha_{12} = 20$, $\alpha_{20} = \alpha_{21} = \alpha_{22} = -5$ and $\alpha_{30} = \alpha_{31} = \alpha_{32} = 3$. We choose linear specifications for the different mappings as they have the potential to introduce eventual biases (see Pinkse et al., 2002) as the set of $[a_{it}]$ parameters are also assumed linear. This option is, thus, less likely to sugarcoat the monte-carlo results. In practice, we may find it useful to consider alternative functional forms or to consider that the observed characteristics that affect the parameters in θ_2 are some transformation of the ones that affect the parameters in θ_1. Alternatively, we could also consider the case where the set of characteristics that affect the parameters in θ_1 is disjoint from the set of characteristics that affect the parameters in θ_2. We chose not to, again, to avoid sugarcoating the monte-carlo results.

Table 2 presents the results for the large number of goods case and hence assumes that the asymptotic arguments work in the number of goods. All the results are conditional on $T = 10$ markets. The first panel explores the sensitivity of the estimates to the number of goods. We consider three cases: $J = 30$, $J = 60$ and $J = 120$. Similarly to the previous case, the results point to the convergence of the algorithm, with both price and income parameters converging to the true values even at small sample sizes. Similarly, in what the price parameters are concerned, it seems to be easier to identify cross-price than own-prices parameters. The second panel again explores the sensitivity of the algorithm to noise-to-signal ratio. We consider three cases: $l_\xi = 1.50$, $l_\xi = 0.50$ (the level under which the results in panel A are derived) and $l_\xi = 0.00$. The results suggest that, as before, an increase in the noise-to-signal ratio deteriorates the performance of the algorithm. Again, when there are no unobserved characteristics ($l_\xi = 0.00$), the algorithm identifies the exact true values of the parameters at every single replication of the Monte Carlo experiment, since the $[a_{it}]$ parameters are retrieved analytically and no numerical error is introduced in the contraction inner loop. The third panel explores the potential multiple local minima property of the GMM objective function by considering two multiple starting points: $(1/2)\, \theta_2^{true}$ and $2\theta_2^{true}$, and the algorithm, again, seems robust to those alternative starting points.

The results for the large number of goods case point to same three key features of the estimation procedure outlined for the previous case: (i) the estimators seem to be consistent, (ii) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally (iii) the GMM objective function seems to be have an apparent global minimum.

4.3 Substitution Patterns

Real-world applications are typically concerned with the estimation of demand own- and cross-price elasticities (and less often with the estimation of income elasticities). In order to examine the quality of those estimates, we compute the mean and standard deviation of the bias implied for the different elasticities across the 50 sample experiments. Table 3 presents the median of those means and standard deviations (in parenthesis) across own-price elasticities, cross-price elasticities and income elasticities. The labels on the columns correspond to those described in Tables 1 and 2. The results suggest in all cases, the estimates seem to capture extremely well the price and income effects as well as the general pattern and level of substitution across goods.

5 Extensions

In this section, we briefly discuss two possible extensions of our proposed demand model.

5.1 Consumer Heterogeneity and Welfare Analysis

An important advantage of a structural model is that it can be used for welfare analysis. In this section, we extend the model to account for consumer heterogeneity in order for the indirect utility function to constitute a *social* indirect utility function. The model can easily account for such heterogeneity by allowing the set of parameters θ_2 to be consumer-specific. In theory, consumer heterogeneity may extend any specification of the model. In practice, however, since this extension increases the vector of parameters of interest, it may make sense to introduce consumer heterogeneity into a model where the θ_2 parameters are initially mapped onto the characteristics space.

Consider a population of I households, where each household h is described by income y_h and a set of observed and unobserved consumer characteristics, d_h and u_h. Let the preferences of household h be given by $H_h\left(r, y_h; \theta_h, \Im\right)$, where θ_h denotes the set of individual-specific parameters. Following the random-coefficients discrete-choice literature, we may assume that θ_h can be decomposed into a set θ_1 that will capture the mean aggregate budget shares and is common to all households, and a set θ_{2h} that will capture the price and income substitution patterns and is individual-specific. In particular, let the θ_{2h} parameters be defined as functions of the observed and unobserved consumer characteristics, d_h and u_h:

$$\theta_{2h} = \theta_2 + \theta_2^d d_h + \theta_2^u u_h, \tag{18}$$

Table 3: *Monte-Carlo Results: Predicted Elasticity Bias*

Panel A: Small J Case							
Median Elasticity	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Own-Price	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Cross-Price	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Income	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Panel B: Large J Case*							
Median Elasticity	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Own-Price	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Cross-Price	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Income	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

* The table depicts the median across own-price elasticities, cross-price elasticities and income elasticities of the means and standard deviations (in parenthesis) of the bias implied for the different elasticities across the 50 sample experiments. The labels on the columns correspond to those described in Tables 1 and 2.

where θ_2^d and θ_2^u represent coefficients on the interactions with observed and unobserved individual attributes, respectively.

The social indirect utility function $H(r, y_1, \ldots, y_I; \theta, \Im) = (1/I)\sum_{h=1}^{I} H_h(r, y_h; \theta, \Im)$ will be, in general, a function of the vector of prices and the complete distribution of household income. If $H_h(r, y_h; \theta_h, \Im)$ has the properties of an indirect utility function then standard duality results imply that, under non-satiation, the demand system is easily obtained via Roy's identity. Let $w_{ih}(r, y_h; \theta_h, \Im)$ denote the budget share of good i for household h.

The introduction of consumer heterogeneity is not obtained costlessly, since there is no closed form expression for the market-level budget share, $w_i(r, y_h; \theta_h, \Im)$, that aggregates the I individual budget shares:

$$w_i(r, y_h; \theta_h, \Im) = \int w_{ih}(r, y_h; \theta, \Im)\, dP^*(y, d, u) = \int w_{ih}(r, y_h; \theta, \Im)\, dP^*_{yd}(y, d)\, dP^*_u(u), \tag{19}$$

where $P^*(y, d, u)$ denotes the population distribution of the consumer types (y_h, d_h, u_h) and the last equality is a direct consequence of the independence assumptions that can be made on (y, d) and u. Following Pakes (1986), Pakes and Pollard (1989), and McFadden (1989), we can approximate the above integral by simulation using, for example, a smooth estimator. Such computation requires drawing ns pseudo-random vectors of consumer characteristics from $P^*_{yd}(y, d)$ and $P^*_u(u)$ and simulate the aggregate budget shares by:

$$w_i(r, y_h; \theta_h, \Im, P^{ns}) = \frac{1}{ns}\sum_{h=1}^{ns} w_{ih}(r, y_h; \theta_h, \Im), \tag{20}$$

where P^{ns} denotes the empirical distribution of the simulation draws. A drawback of this type of solution is that it introduces simulation error, which, as Berry et al. (2004) point out, influences the asymptotic distribution of the GMM estimator and, therefore, needs to be explicitly taken it account. However, apart from these modifications, the introduction of consumer heterogeneity would not impact significantly the general setup of the estimation algorithm.

If the aggregate demand functions are generated, via Roy's identity, from a social indirect utility function, they have welfare significance and can be used to make welfare judgments via the standard welfare measurement techniques. By estimating the parameters of the demand functions, we have the required parameters of the social indirect utility function, which we can then easily invert - either algebraically or numerically - to derive the expenditure function - and compute *compensating* and *equivalent variations*. Of course this approach only makes sense if the estimated parameters satisfy the various restrictions that ensure an underlying model of utility.

Although the social indirect utility function will be, in general, a function of the vector of prices and the complete distribution of household income, Gorman (1953)'s seminal article on exact aggregation establishes, however, that when consumers have indirect utility of the Gorman form, aggregate demand can always be thought of as being generated by a normative representative consumer with indirect utility function $H(r, y; \theta, \Im) = (1/I)\sum_{h=1}^{I} H_h(r, y_h; \theta, \Im)$ defined in terms of y, the average income, regardless of the form of the social welfare function.

An indirect utility function for consumer h is said to be of the Gorman form if it can be written in terms of functions $d_h(r)$, which may depend on the specific consumer, and $k(r)$, which is common to all consumers:

$$H_h(r, y_h; \theta, \Im) = d_h(r) + k(r)\, y_h.$$

We now show that the specification for $H(r, y; \theta, \Im)$ satisfies the Gorman polar form and, as a consequence, the preferences of our *fictional* representative consumer constitutes a measure of aggregate social welfare. As a consequence, for this particular specification, welfare judgments can be made without incorporating consumer heterogeneity.

Proposition 4. *The specification for the indirect utility function $H(r, y; \theta, \Im)$ constitutes an admissible social indirect utility function for the normative representative consumer.*

Proof. Let the indirect utility function of each household h, $H_h(r, y_h; \theta, \Im)$, be given by a generalization (to the household-level) of the specification chosen for the indirect utility function:

$$H_h(r, y_h; \theta, \Im) = a_0 + \sum\nolimits_{m=1}^{J} a_m r_m + 1/2 \sum\nolimits_{m=1}^{J} \sum\nolimits_{n=1}^{J} b_{mn} r_m r_n + c_0 y_h + \sum\nolimits_{m=1}^{J} c_m r_m y_h.$$

It is easy to verify that $H_h(r, y_h; \theta, \Im)$ satisfies the Gorman form with $d_h(r) = a_0 + \sum_{m=1}^{J} a_m r_m + 1/2 \sum_{m=1}^{J} \sum_{n=1}^{J} b_{mn} r_m r_n$ and $k(r) = c_0 + \sum_{m=1}^{J} c_m r_m$. $\square$

5.2 Dynamic Demand

In some markets, consumers may engage into forward-looking behavior. In this section we briefly discuss how could we extend the model to account for dynamic behaviour. A more extensive study of the properties of this extension seems more appropriately considered as a separate one and hence is left for future research.

Let us consider a demand setting in the lines of Gorman (1996)'s multi-stage budgeting approach as an example. Consider that our representative consumer is faced with J different brands of a storable good and has to decide, in each period, how much of each brand to purchase. Following Gorman (1996)'s approach (see Aguirregabiria, 2002; and Hendel and

Nevo, 2006 for similar dynamic applications), we can separate the quantity decision from the brand decision. Assume that the purchased amount, denoted by x_t, is simply a choice of size, with $x_t = 0$ standing for no purchase. Let $d_{xt} = 1$ denote a purchase of size x in period t, and assume $\sum_x d_{xt} = 1$. Because the good is storable, the consumer does not need to consume all the purchased quantity in a given period. As a consequence, the consumer has also to decide how much to consume in each period. Quantity not consumed is stored as inventory.

In each period, the consumer's problem can be represented by:

$$
\begin{aligned}
V(s_1) &= \max_{\{c(s_t),x(s_t)\}} \sum\nolimits_{t=1}^{\infty} \delta^{t-1} E\left(u(c_t, v_t; \theta) - C(I_{t+1}; \theta) + \sum\nolimits_x d_{xt} H_{xt}(r, y; \theta, \Im_x) \,|s_1\right), \\
&\quad \text{s.t.} \\
0 &\leq I_t, \quad 0 \leq c_t, \quad 0 \leq x_t, \quad \sum\nolimits_x d_{xt} = 1, \quad I_{t+1} = I_t + x_t - c_t
\end{aligned}
$$

where s_t denotes the state at time t, $\delta > 0$ denotes the discount factor, $u(c_t + v_t; \theta)$ denotes the utility from consumption, v_t denotes a shock to utility that impacts the marginal utility from consumption, $C(I_{t+1}; \theta)$ denotes the cost of storage, and $H_{xt}(r, y; \theta, \Im_x)$ denotes the utility from a size x purchase ($\Im_x$ denotes the set of brands available with size x).

The state vector s_t consists, at time t, of the current (or beginning-of-period) inventory I_t, current prices, and the shock v_t. We can, following Hendel and Nevo (2006), make the simplifying assumptions that v_t is independently distributed over time, and prices follow an exogenous Markov process.

The estimation algorithm could be based in the procedure outlined in Hendel and Nevo (2006). The first step would consist of estimating the parameters in $H_{xt}(r, y; \theta, \Im_x)$. This could be achieved by estimating a static continuous choice model conditional on the choice set $\Im_x$. This procedure yields consistent, although potentially inefficient, parameters for the indirect utility function. The second step would consist of computing the indirect utility associated with each size and their transition probabilities from period to period. Finally, the third step would consist of solving a simplified version of the full dynamic problem - restricted to the remaining parameters only - by maximizing the likelihood of the observed sequence of sizes purchased.

6 Conclusions

In this paper, we consider a new method of uncovering demand information from market level data on differentiated products. We follow the continuous-choice literature and develop a globally consistent continuous-choice demand model that combines desirable properties of both the discrete- and continuous-choice literatures: (i) it is flexible in the sense of Diewert (1974), (ii) it is globally consistent in the sense it can deal with entry and exit of products over time, and (iii) incorporates a structural error term. The estimation procedure follows an analogue to the algorithm derived in Berry (1994) and Berry et al. (1995). The contraction mapping for matching observed and predicted budget shares is analytical, making it relatively simple and fast to estimate, which can prove a key advantage in competition policy issues, where time and transparency are typically crucial factors.

We provide a series of Monte Carlo experiments to illustrate the estimation properties of the model and discuss how it can be extended to cope with consumer dynamic behaviour. A more extensive study of the properties of this extension seems more appropriately considered as a separate one and hence is left for future research.

References

Ackerberg, D., Benkard, L., Berry, S., and Pakes, A. (2007). *Handbook of Econometrics*, volume 6, chapter Econometric Tools for Analyzing Market Outcomes, pages 4171–4276. Elsevier.

Aguirregabiria, V. (2002). Sales promotions in supermarkets: Estimating their effects on profits and consumer welfare. Unpublished Manuscript, Boston University.

Barnett, W. A. (1983). Definitions of 'second order approximation' and of 'flexible functional form'. *Economics Letters*, 12(1):31–35.

Berndt, E. R., Fuss, M. A., and Waverman, L. (1977). Dynamic models of the industrial demand for energy. Technical report, Electric Power Research Institute.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Berry, S. and Haile, P. A. (2008). Nonparametric identification of multinomial choice models with heterogeneous consumers and endogeneity. Unpublished Manuscript, Yale University.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S., Linton, O. B., and Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *Review of Economic Studies*, 71(3):613–654.

Brown, B. and Walker, M. (1989). The random utility hypothesis and inference in demand systems. *Econometrica*, 57(4):815–829.

Chan, T. Y. (2006). Estimating a continuous hedonic-choice model with an application to demand for soft drinks. *RAND Journal of Economics*, 37(2):466–482.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1975). Transcendantal logarithm utility functions. *American Economic Review*, 65(3):367–383.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics*, 106(1):67–95.

Davis, P. and Schiraldi, P. (2014). The flexible coefficient multinomial logit (fc-mnl) model of demand for differentiated products. *RAND Journal of Economics*, 45(1):32–63.

Deaton, A. and Muelbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3):312–326.

Diewert, W. E. (1971). An application of the shephard duality theorem: a generalized leontief production function. *Journal of Political Economy*, 79(3):481–507.

Diewert, W. E. (1973). Functional forms for profit and transformation functions. *Journal of Economic Theory*, 6(3):284–316.

Diewert, W. E. (1974). Functional forms for revenue and factor requirements functions. *International Economic Review*, 15(1):119–130.

Dubé, J.-P., Fox, J. T., and Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

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

Fox, J. T. and Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *RAND Journal of Economics*, 47(1):118–139.

Gorman, W. M. (1953). Community preference fields. *Econometrica*, 21(1):63–80.

Gorman, W. M. (1996). *Separability and Aggregation: The Collected Works of W. M. Gorman*, volume 1, chapter Two-Stage Budgeting, pages 19–30. Oxford University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hausman, J., Leonard, G., and Zona, D. (1994). Competitive analysis with differentiated products. *Annales d'Economie et de Statistique*, 34:159–180.

Hausman, J. A. (1994). *The Economics of New Goods*, chapter Valuation of New Goods under Perfect and Imperfect Competition, pages 207–248. University of Chicago Press.

Hausman, J. A. and Wise, D. A. (1978). A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46(2):403–426.

Hendel, I. and Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673.

Knittel, C. R. and Metaxoglou, K. (2014). Estimation of random coefficient demand models: two empiricists' perspective. *Review of Economics and Statistics*, 96(1):34–59.

Lancaster, K. (1971). *Consumer demand: a new approach.* Columbia University Press, New York.

Lau, L. (1986). *Handbook of Econometrics*, volume 3, chapter Functional forms in econometric model building, pages 1515–1566. Elsevier.

McFadden, D. (1974). *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior, pages 105–142. Academic Press.

McFadden, D. (1978a). *Production Economics: A Dual Approach to Theory and Applications*, volume 1, chapter The general linear profit function, pages 269–286. North-Holland.

McFadden, D. (1978b). *Transportation forecasting and travel behavior*, chapter Modelling the choice of residential location, pages 72–77. Transportation Research Board.

McFadden, D. (1981). *Structural Analysis of Discrete Data and Econometric Applications*, chapter Structural discrete probability models derived from theories of choice, pages 198–272. MIT Press.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

Pakes, A. (1986). Patents as options: some estimates of the value of holding european patent stocks. *Econometrica*, 54(4):755–784.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057.

Pinkse, J. and Slade, M. (2004). Mergers, brand competition, and the price of a pint. *European Economic Review*, 48(3):617–643.

Pinkse, J., Slade, M., and Brett, C. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, 70(3):1111–1153.

Varian, H. R. (1984). *Microeconomic Analysis*. W. W. Norton & Co., New York.

Mathematical Appendix

In this mathematical appendix, we present the set of parameter restrictions that ensure function $H(r, y; \theta, \Im)$, satisfies properties 3 and 4 for a globally consistent indirect utility function. The set of restrictions resemble conditions 3.15 and 3.16 in Diewert (1971) that ensure the generalized Leontief function can be interpreted as the cost function corresponding to some underlying production possibilities set.

Assume that the vectors r and y are strictly positive. We begin by addressing property 3. In order for $H(r, y; \theta, \Im)$ to be strictly increasing in y and nonincreasing in p_i (or in other words, nondecreasing in r_i) for any $i \in \Im$, the following $J + 1$ inequalities need to be satisfied:

$$\begin{aligned} H_y(r, y; \theta, \Im) &= c_0 + \sum\nolimits_{j=1}^{J} c_j r_j > 0 \\ H_{r_i}(r, y; \theta, \Im) &= a_i + 1/2 \sum\nolimits_{j=1}^{J} (b_{ij} + b_{ji}) r_j + c_i y \geq 0, \end{aligned}$$

for any $i \in \Im$.

We now turn to property 4 and present the set of restrictions that ensure $H(r, y; \theta, \Im)$ is quasiconvex. It is well known that every convex function is quasiconvex. Furthermore, a twice differentiable function is convex over a convex set if and only if its matrix of second partial derivatives is positive semidefinite for every point in the set.

If symmetry is imposed on matrix B (please see the symmetry subsection for more details), the second partial derivatives matrix of $H(r, y; \theta, \Im)$ will be symmetric. A symmetric matrix is positive definite (and automatically positive semidefinite) if and only if all its leading principal minors are strictly positive. If, on the other hand, no symmetry on matrix B is assumed *a priori*, the second partial derivatives matrix of $H(r, y; \theta, \Im)$ will not necessarily be symmetric. An arbitrary matrix is positive definite (and automatically positive semidefinite) if and only if its Hermitian part is positive definite. In other words, an arbitrary matrix is positive definite if and only if all the leading principal minors of its Hermitian part are strictly positive.

Let D_k denote the determinant of the kth order principal submatrix of the second partial derivatives matrix of $H(r, y; \theta, \Im)$. For the algebraic functional form to be quasiconvex, the following $J + 1$ inequalities (the formulation nests both the case where symmetry is imposed on matrix B and the case where it is not) need to be satisfied:

$$\begin{aligned} D_1 &= b_{11} > 0 \\ D_2 &= b_{11} b_{22} - (1/4)(b_{12} + b_{21})^2 > 0 \\ &\vdots \\ D_{J+1} &> 0. \end{aligned}$$