The Pure Characteristics Demand Model.[1]

Steven Berry
Dept. of Economics, Yale
and NBER

Ariel Pakes
Dept. of Economics, Harvard
and NBER

July 2, 2007

[1]We are grateful for the opportunity to publish this paper in a volume in honor of Daniel MacFadden. The paper illustrates just one way in which Dan's path breaking work on the economics and econometrics of choice theory has impacted Industrial Organization. We would also like to thank two referees and Charles Manski for helpful comments.

Abstract

In this paper we consider a class of discrete choice models in which consumers care about a finite set of product characteristics. These models have been used extensively in the theoretical literature on product differentiation and the goal of this paper is to translate them into a form that is useful for empirical work. Most recent econometric applications of discrete choice models implicitly let the dimension of the characteristic space increase with the number of products (they have "tastes for products"). The two models have different theoretical properties, and these, in turn, can have quite pronounced implications for both substitution patterns and for the welfare impacts of changes in the number and characteristics of the goods marketed. After developing those properties, we provide alternative algorithms for estimating the parameters of the pure characteristic model and compare their properties to those of the algorithm for estimating the model with tastes for products. We conclude with a series of Monte Carlo results. These are designed to illustrate: (i) the computational properties of the alternative algorithms for computing the pure characteristic model, and (ii) the differences in the implications of the pure characteristic model from the models with tastes for products.

1 Introduction.

The theory and econometrics of demand models that treat products as bundles of characteristics dates back at least to Lancaster (1971) and McFadden (1974). Early applications of these models used restrictive assumptions and a primary concern of the literature that followed (including many contributions by McFadden himself) was that the structure of the discrete choice model used would in some way restrict the range of possible outcomes, thereby providing misleading empirical results. This paper is a continuation of that tradition. In contrast to the empirical models currently available, we consider estimation of a class of discrete choice models in which consumers care about a finite set of product characteristics. These models have been used extensively in the theoretical literature on product differentiation, but have been used much less in empirical work.

Typical discrete choice empirical models (including those in our own work) implicitly assume that the dimension of the product space increases with the number of products. This assumption is often embedded in an otherwise unexplained i.i.d. additive random term in the utility function. One interpretation is that these terms represent variance in the "taste for the product", as opposed to a taste for the characteristics of the products. Though these models can do quite a good job in approximating some aspects of demand, they also have some counter-intuitive implications as the number of products increases. As a result, we worry about their ability to analyze substitution patterns when there are a large number of goods and also about their ability to answer questions about changes in the number of goods.

Recent advances in econometric technique, computing power and data availability have significantly increased the use of characteristic based models with "tastes for products" in analyzing demand in differentiated products markets. If one can condition on the major characteristics that consumers value, the characteristic based demand models have a number of advantages over the more traditional demand models in "product space". One of those advantages concerns the analysis of demand for products not yet marketed.[1] Demand systems estimated purely in product space do not allow us to analyze the demand for potential products. Provided that we are willing to specify the characteristics of the new product and the form of the equilibrium in

[1]The other major advantage of estimating demand systems in characteristic space is that they typically constrain own and cross price (and characteristic) elasticities to be functions of a small number of parameters describing the distribution of tastes for characteristics). In particular, the number of parameters needed is independent of the number of products. In contrast, if we worked in product space a (log) linear demand system for J products would require the estimation of a number of parameters which grows like J^2.

the product market, the characteristic based models do (see, for e.g. Berry Levinsohn and Pakes, 2004, henceforth MicroBLP.)

An important issue closely related to the demand for new products is the problem of estimating the consumer surplus generated by previously introduced products. This type of retrospective analysis dates back at least to Trajtenberg (1989), and its usefulness is illustrated clearly in Petrin's (2002) investigation of the consumer surplus gains from the (privately funded) research expenditures that lead to the development of a major innovation (the Minivan), and Hausman's (1997) investigation of the consumer surplus losses caused by regulatory delay in the introduction of the cell phone. In addition, measuring the consumer surplus gains from new products is an integral part of constructing ideal price indices (see, for e.g. Pakes et al. (1993) or Nevo (2000)). Moreover, at least according to the Boskin Commission Report (1996) , failure to adequately measure the gains from new products is the major source of bias in the Consumer Price Index (see also Pakes (2003)).

The consumer surplus generated by products already introduced can, at least in principal, be analyzed using either product based or characteristic based demand systems. However in either case the results of the analysis are likely to be particularly sensitive to *a priori* modeling assumptions. This is because we typically do not have data on the demand for new products at prices that are high enough to enable us to non-parametrically estimate the reservation prices of a large fraction of consumers. When using product based demand systems the utility gains for the infra-marginal consumers who purchased the good at all *observed* prices are obtained by extrapolating the demand curve estimated at lower prices to a higher price range, and these extrapolations can be very sensitive to the assumptions build into the model[2]. The characteristic based demand model uses slightly more information in its estimation of consumer surplus gains, since it uses the price variance for products with similar characteristics, but analogous problems can and do arise[3]. As a result, the measurement of gains from product introductions is likely to be particularly sensitive to assumptions like those we discuss extensively below.

The next section introduces our pure characteristics model and provides a more extended discussion of the reason we think it might be useful. There are at least two of these. The model with tastes for products implicitly places a limit on substitution patterns between products and hence on markups,

[2]Hausman (1997) for e.g. reports infinite consumer surplus gains from some of his specifications.

[3]See the discussion below or Petrin (2002) who reports large differences in consumer surplus gains from differences in specifications and data sources.

while the pure characteristics model does not. As a result, at least in markets with a large number of products, the substitution patterns implied by the estimates from the two models might be expected to differ. Second, the model with tastes for products has the implication that every consumer's utility grows without bound as the number of products are increased, no matter the characteristics of those products. In the pure characteristic model with standard regularity conditions, the increment in each consumer's welfare as the number of products grows must eventually decline to zero (often at an exponential rate). So one might also expect the two models to give different results for consumer welfare gains.

We then develop some of the properties of our model. These properties enable us to build an algorithm for estimating the pure characteristics model from market level data on prices, quantities, and characteristics (section 3). This section provides an analog to the algorithm developed for the model with "tastes for products" in Berry, Levinsohn and Pakes (2004) (henceforth BLP). The paper concludes with some Monte Carlo evidence. The Monte Carlo studies are designed to give the reader: (i) some indication of the performance of our estimators, (ii) an indication of the computational burden of the pure characteristics model relative to the model with a taste for products, and (iii) an indication of the performance of the pure characteristic model relative to the performance of a model with tastes for products.

The Monte Carlo studies suggest a number of important points for future research. First, we find that the most precise of the computational techniques works well, but can result in time-consuming estimation routines. Simpler and faster techniques that are closer to the original BLP methods sometimes work quite well in providing useful estimates, but not always. In some cases we find that a practical compromise is to use a method very close to the original BLP, but then to compute the predictions of the model using our most precise computational techniques for the pure characteristics model.

2 Discrete Choice Models and Their Implications.

We consider models in which each consumer chooses to buy at most one product from some set of differentiated products. Consumer i's (indirect) utility from the purchase of product j is

$$U_{ij} = U(X_j, V_i, \theta), \tag{1}$$

where X_j is a vector of product characteristics (including the price of the good), V_i is a vector of consumer tastes and θ is a vector of parameters to be estimated.

Probably the earliest model of this sort in the economic literature is the Hotelling (1929) model of product differentiation on the line. In that model X is the location of the product and V is the location of the consumer. Other well known industrial organization models in this class include the vertical model of Mussa and Rosen (1978) (see also Gabszewicz and Thisse (1979), and the model of competition on a circle by Salop (1979).) For early empirical use of these models see Bresnahan (1987), Feenstra and Levinsohn (1995), Greenstein (1996), and more recently Song (2006).

To obtain the market share of good j (our s_j) implied by the model we simply add up the number of consumers who prefer good j over all other goods. That is

$$s_j = Pr\{V_i : U(X_j, V_i, \theta) \geq U(X_k, V_i, \theta), \forall k \neq j)\} \tag{2}$$

To ease the transition to empirical work we follow the notation in Berry, Levinsohn and Pakes (2004) and

- partition the vector of consumer attributes, V_i, into z_i, which an econometrician with a micro data set might observe, and ν_i, which the econometrician does not observe
- and partition product characteristics, X_j, into x_j, which is observed by the econometrician, and ξ_j, which is unobserved to the econometrician (though observed to the agents[4]).

The models taken to data impose more structure. They typically assume that the utility function is additively separable in a deterministic function of the product attributes and the observed consumer data, and a disturbance term, i.e.

$$U_{ij} = f(X_j, z_i; \theta) + \mu_{ij}, \tag{3}$$

where θ is a parameter to be estimated. The model is then completed by making alternative assumptions on the joint distribution of the $\{\mu_{ij}, X_j, z_i\}$ tuples.

[4]Often, especially in the study of consumer (in contrast to producer) goods, the ξ refer to the aggregate impact of a large number of relatively unimportant characteristics, some or all of which may in fact be observed. Even if they are potentially observed, the researcher may not include them in the specification taken to data because of the worry that the data cannot support an investigation of preferences on such a detailed space.

In particular we require: (i) an assumption on the distribution of the μ_{ij} conditional on (X_j, z_i), and (ii) either an assumption on the distribution of the unobserved product characteristic (on ξ) given (x, z) or a procedure which estimates the ξ pointwise[5]. In this section we focus on the assumption on μ_{ij} conditional on (X_j, z_i) as it is this assumption which differentiates the pure characteristic model from the model with tastes for products.

Conditional Distributions for μ_{ij} and Their Implications. Both the pure characteristics model and the model with tastes for products allow the μ_{ij} to contain a set of interactions between the unobserved determinants of consumer tastes (say $\nu_{i,k}$) and the product characteristics $X_{j,k}$ (both observed and unobserved). As is typical, we will assume these interactions are linear: *i.e.* given K characteristics then one component of μ_{ij} is $\sum_k \nu_{i,k} X_{j,k}$. This component allows each consumer to have a different marginal utility for each characteristic, which ameliorates many of the problems caused by the independence of irrelevant alternatives property of earlier models [6]. The model with tastes for products and the pure characteristic model differ in whether μ_{ij} has an additional i.i.d. component.

The empirical specifications to date typically assume that each μ_{ij} has an additional component whose conditional distribution, conditional on (X_j, z_i) *and* all other $\mu_{ij'}$ $(\forall j' \neq j)$, that has support on the entire real line. Letting that component be $\{\epsilon_{ij}\}_{j=1}^{J}$ and indexing the μ_{ij} from these models with a tp superscript (for "tastes for products"), the model is usually written as

$$\mu_{ij}^{tp} \equiv \sum_{k=1}^{K} \nu_{ik} X_{jk} + \epsilon_{ij}. \tag{4}$$

Special cases of this assumption include the random coefficient logit model used in BLP and McFadden and Train (2000), and the random coefficient probit discussed in Hausman and Wise (1978), and McFadden (1981).

One way to obtain a model with an additive component with full support (i.e. to obtain equation 4) is to make the dimension of the characteristic space, K, be a function of the number of products. Indeed the suggestion of Caplin and Nalebuff (1991) is to think of the additive component as being formed from the interaction of a set of product-specific dummy variables and a set of tastes for each product that are distributed i.i.d. both across

[5]The choice typically depends on whether micro data, data which matches individuals to the choice they made, is available; for a discussion see MicroBLP, 2004.

[6]See McFadden (1981) for a statement of the problem, and BLP and Berry, Levinsohn and Pakes (2004) for its implications on estimates of market demand systems used in Industrial Organization.

products, and across individuals for a given product. Thus our labeling of these models as models with "tastes for products".

The pure characteristic model only differs from the model with tastes for products in that it does not contain the $\{\epsilon_{i,j}\}_{j=1}^{J}$. That is if we index the $\mu_{i,j}$ from the pure characteristic model with a superscript of *pc*, the pure characteristic model is written as

$$\mu_{ij}^{pc} \equiv \sum_{k=1}^{K} \nu_{ik} X_{jk}. \tag{5}$$

One can show that equation (5) is related to the "ideal point" models which date back to Hotelling and have been used extensively in the theory literature.[7] In the ideal point models consumers care about the distance between their location (v_i) and the products' location (X_j) in $\mathcal{R}^k$, that is

$$u_{ij} = \|X_j - \nu_i\| - \alpha_i p_j, \tag{6}$$

where $\|\cdot\|$ is some distance metric. If one assumes that $\|\cdot\|$ refers to Euclidean distance and expands the squares, we get a utility functions where the mean utility depends on characteristics and the error term depends on interactions between product and individual characteristics as in equation (6).

Differences In the Implications of the Two Specifications. The two specifications for μ_{ij} place different *a priori* restrictions on the implications of the model estimated. If the model with tastes for products is estimated then the estimated model will imply that

1. there is a limit on substitution possibilities between products, and
2. that as we increase the number of products each individual's utility increases to infinity as the number of new products grows, *regardless* of the observed characteristics of either the products which enter or of the individual.

As is explained below, neither of these implications hold true when the model without tastes for products is estimated.

An implication of the first point is that if we add products whose X characteristic are very similar (indeed they can be identical) to those of product A the markup of product A will remain bounded away from zero (a similar point is made by Anderson, DePalma and Thisse, 1992, in the context of the

[7] See the discussion in Caplin and Nalebuff (1991) and Anderson et al. (1992).

logit model). This follows from the presence of the i.i.d. component, which ensures that there will always be consumers who prefer the new product to the old even if they have to pay a positive price difference, and in virtually all equilibrium pricing models this will imply prices greater than costs.

The fact that markups will not go to zero no matter how much we fill up the "product space" has at least two implications that might be problematic. First, even with a large number of products there will always be a further incentive for product development. Second, market outcomes will not approach a competitive equilibrium as the number of products grow large. One might be particularly worried about using a model that imposes these restrictions when studying markets where there are a large number of products.

The second itemized point, that each consumer's utility must grow without bound as we increase the number of products marketed no matter the characteristics of the products, is a particular concern for the analysis of prospective and retrospective gains from product introductions. Particularly when we use the model to extrapolate outside of the range of the data, as we must do to measure welfare gains from new products, we may obtain results that are not meaningful.

The finite-dimensional pure characteristics model is very different in both the itemized respects. In that model the agent's utility gain from new product introduction is limited by a smooth (usually linear) function of the distance in characteristic space between the new and previously existing products. As the number products increases, the products become increasingly good substitutes for one another and oligopolistic competition will approach the competitive case, with prices driven toward marginal cost and no additional incentive for product development. As the product space fills up, the incremental consumer gain from new product introductions will decline to zero.[8]. That is, the gains to "variety" will be bounded in the pure characteristics model whereas they grow without bound in models with tastes for products.

On the other hand, models that include a taste for products do have a number of important practical advantages. These models

- define all probabilities by integrals with simple limits of integration (see McFadden (1981)),
- insure that all the purchase probabilities are nonzero (at every value

[8]In cases we have worked out this decline will be at an exponential rate. We note that we are implicitly assuming the "environment" does not change as the number of products grows. That is we are ruling out both technological changes and changes in competing and/or complimentary products which alter the relative benefits of producing in different parts of the characteristic space.

of the parameter vector), and (provided certain other regularity conditions are satisfied) have smooth derivatives (McFadden (1981)), and

- aggregate into market shares which can be easily inverted to solve for the unobservable characteristics (the $\{\xi_j\}$) as a linear function of the parameters and the data, enabling the use of instrumental variable techniques to solve simultaneity problem induced by correlation between the unobserved characteristics and price (see BLP).

2.1 Nesting the Two Models.

Looking at equations (4) and (5) one might think it easy to nest the two models and let the data decide which of them is appropriate for the problem at hand. In particular we could introduce the additional parameter σ_ϵ and then assume

$$\mu_{ij} \equiv \sum_{k=1}^{K} \nu_{ik} X_{jk} + \sigma_\epsilon \epsilon_{ij}. \tag{7}$$

The model with tastes for products is the special case of equation (7) with $\sigma_\epsilon = 1$ while the pure characteristic model is the special case with $\sigma_\epsilon = 0$.

To see the problems we run into with specifications like (7) take the familiar case where the $\{\epsilon_{ij}\}$ have extreme value distributions and let the interactions with the observable individual characteristics in the utility function in equation (3) be $f_{ij}(\theta)$, so

$$u_{ij} = f_{ij}(\theta) + \sum_{k=1}^{K} \nu_{ik} X_{jk} + \sigma_\epsilon \epsilon_{ij}. \tag{8}$$

Assume $\sigma_\epsilon > 0$ (though perhaps very small) so that we can define $\mu_\epsilon \equiv \sigma_\epsilon^{-1}$ and multiply all utilities by it (this does not change the ordering of utilities and hence does not change the implications of the model; we come back to a fuller discussion of normalizations below). Then if $F(\cdot)$ provides the distribution of $\nu = [\nu_1, \ldots, \nu_K]$, the familiar "logit" formula gives us

$$Pr(i \text{ chose } j) = \int_\nu \frac{exp\Big(\Big[f_{ij}(\theta) + \sum_{k=1}^{K} \nu_k X_{jk}\Big]\mu_\epsilon\Big)}{\sum_q exp\Big(\Big[f_{i,q}(\theta) + \sum_{k=1}^{K} \nu_k X_{qk}\Big]\mu_\epsilon\Big)}\, dF(\nu). \tag{9}$$

Since we needed to assume $\sigma_\epsilon > 0$ to obtain the probabilities in (9) there is no "special case" of this formula that gives us the probabilities from the pure characteristic model. However it is straightforward to show that

if we consider any sequence of probabilities obtained by letting $\mu_\epsilon \to \infty$ then, under standard regularity conditions, that sequence converges to the probabilities from the pure characteristic model (indeed we can show that the convergence is uniform in θ).[9] That is, the probabilities from the pure characteristics model are a limiting case of the probabilities in (9).

The question then is: what would happen if the data generating process corresponded to the pure characteristics model and yet we estimated the "modified BLP" model represented by (9)? The hope would be that the resulting estimate of μ_ϵ would be large enough for the probabilities in (9) to approximate the probabilities from the pure characteristics model quite well. Indeed, if this were so, then there would be no need for a separate estimation algorithm for the pure characteristics model.

There are, however, reasons to doubt that the modified BLP model would produce estimates that "mimic" the pure characteristic model. First, when we consider consistency we generally assume the true value of the parameter is in a compact set. So were we to apply standard consistency proofs we would have to prove consistency of $1/\mu_\epsilon = \sigma_\epsilon$ (and not μ_ϵ per se). Once we do that, the model's prediction for the sample shares will not converge to the true shares uniformly in θ (the sample shares are undefined at $\sigma_\epsilon = 0$). So standard consistency proofs do not apply.

Second, one might think that to get estimated probabilities from the modified BLP model that well approximate those of the pure characteristic model we might need to set σ_ϵ close to zero. However, as a practical matter, we then have to compute the exponent of $\mu_\epsilon = 1/\sigma_\epsilon$, and when μ_ϵ is large enough the computer can not compute $exp[\mu_\epsilon]$. Third, there is the usual matter of efficiency, i.e. if we know that σ_ϵ is zero, and we had an estimation algorithm which could utilize this fact, we would expect that algorithm to produce more precise estimators of the remaining parameters. Since demand models are frequently estimated on market level data, this "efficiency" issue can be important. Finally, to investigate whether these issues are important in a given setting we need an estimation routine that does not rely on the modified BLP approximation in equation (9).

This paper first develops that algorithm and then compares Monte Carlo results from it to that from the modified BLP model. The comparison is done in two ways. First we use the probabilities from (9) and a μ_ϵ exogenously set to be as large as our computer can handle. Second we use the probabilities

[9] The convergence result follows from the fact that for every ν, except for a measure zero set of ν's that generate a tie, the integrand in (9) converges to one or zero according as the probability of the pure characteristic model is zero or one, and then applying dominated convergence. A covering argument shows that the convergence is uniform in θ.

from (9) to jointly estimate the parameters of the model and μ_ϵ. We will show that sometimes one obtains acceptable estimators for the pure characteristics model using the modified BLP model in (9), and sometimes we do not. Moreover the difference typically depends on characteristics of the data generating process.

We now proceed to provide an algorithm for estimating a pure characteristics model that does not rely on the approximation in equation (9). We then generate data from the pure characteristic model and ask whether we can characterize: (i) when is it easy to estimate the pure characteristic model, and (ii) when would the estimates from the model with tastes for products, possibly adjusted to allow for a μ_ϵ as in the modified BLP model in (9), generate implications which are similar to the implications for the pure-characteristics model. The implications we consider in this context are own and cross-price elasticities, and the welfare gains from new goods.

3 Estimating The Pure-Characteristic Model.

Utility in our model is given by

$$u_{ij} = x_j\beta_i - \alpha_i p_j + \lambda_i \xi_j, \tag{10}$$

for $j = 0, \ldots, J$, where 0 designates the outside good and $j = 1, \ldots, J$ are the goods competing in the market. Here $\beta_{i,k} = \beta_k + \nu_{i,k}$, $\alpha_i = \alpha + \nu_{i,p}$ and we assume that $\lambda_i > 0$ for all i.[10] That is β_k is the mean of the utility for an increment in x_k and $\nu_{i,k}$ is the individual specific deviation from that mean.

Two assumptions implicit in equation (10) are worth emphasizing:

- A1. there is only one unobserved product characteristic; i.e. $X = (x, \xi) \in R^k \times R^1$, and
- A2. that ξ is a "vertical" characteristic in the sense that every individual would prefer more of it.

3.1 Further Constraints and Normalizations.

Since the utility functions of consumers can only be identified up to a monotone (in our case affine) transformation, theory implies that we can take each consumer's utility from every choice and

[10] We note that this would be identical to the model used in Das et al. (1995) were we to *omit* their i.i.d. disturbances with full support. Also if there were consumer level data we would let β_i be a function of those variables as in MicroBLP.

- multiply them by a consumer specific positive constant, and
- add to them a consumer specific constant.

We add $-u_{i,0}$ to the utility of each choice so that the utility of the outside option is zero and the utility of the inside options should be interpreted as their utility relative to that of the outside option. Our second normalization is to divide each u_{ij} by λ_i (so that the coefficient of ξ is one for all consumers). Imposing these normalizations and (for notational simplicity) reinterpreting the characteristic values for option j to be the value of the j^{th} option for that characteristic *minus* the value of the outside option for that characteristic, we have

$$u_{ij} = x_j\beta_i - \alpha_i p_j + \xi_j, \tag{11}$$

and

$$u_{i,0} = 0.$$

This is identical to the model in BLP *without* their i.i.d. additive component with full support. Our change of notation implies that the ξ variables represent the difference in the unobserved quality of the inside options from that of the outside option[11].

We still need to chose units[12]. One useful normalization is to set the mean of $\alpha_i = 1$, so that the units of utility and of ξ are in terms of the mean price coefficient.

Our previous work emphasized the reasons for (and the empirical importance of) allowing for unobserved product characteristics, ξ, in estimating discrete choice demand (see Berry (1994), BLP and Berry, Levinsohn and Pakes (2004). The ξ are the analog of the disturbances in the standard demand model, i.e. they account for the factors that are unobserved to the econometrician but affect demand. Without them it will typically be impossible to find parameter values that make the implications of the model consistent with the data. Moreover if we incorporate them and consider any realistic model of market equilibrium we are lead to worry about a simultaneous equations bias resulting from price being a function of the ξ.

[11] This becomes important when we try to measure welfare gains over time, as we are implicitly doing when we construct a price index. This because the difference in the average estimated level of ξ across periods can be due either to a change in the average unobserved quality of the products in the market being analyzed, or to a change in the average value of the outside alternative. For more details and an attempt to decompose movements in ξ into its components see Pakes, Berry and Levinsohn (1993) and Song (forthcoming).

[12] Multiply all $\xi_j, \alpha_i,$ and β_i, by the same positive constant and the implications of the model are not changed.

Of course in reality there may be more than one unobserved characteristic that is omitted from the empirical specification and, provided consumers varied in their relative preferences for the different unobserved characteristics, the model in (11) is, at least strictly speaking, misspecified. Though allowing for one unobserved factor seems to be particularly important in obtaining realistic implications for own and cross price elasticities, our ability to estimate "multi-unobserved factor" discrete choice models with such data seems extremely limited[13].

As noted above the fact that the model with a single unobserved factor might provide good fits "in sample" does not imply that it delivers meaningful welfare measures. On the other hand, if we allow for as many unobserved factors as there are products, then the pure characteristics model with multiple unobserved characteristics has the traditional models with tastes for products as a special case. In this sense the pure characteristics model with *one* unobserved characteristic is an opposite extreme to the model with tastes for products. We might hope that the two models would bound the impacts of unobserved product heterogeneity.

4 Estimating the Model.

The issues that arise when estimating the pure characteristics model are similar to those found in estimating more traditional discrete choice models. As a result we use the techniques in BLP (1995) and Berry, Levinsohn and Pakes (2005), and the vast literature cited therein, as starting points. There are, however, four modifications of those techniques we will need in order develop an estimator for the pure characteristics model. The modifications provide

1. a method of calculating the aggregate market share function conditional on the vectors of characteristics and parameters (θ),

2. an argument that proves existence of a unique ξ vector conditional on any vector of model parameters and observed market shares,

3. an algorithm for computing that ξ vector, and

[13]We note here that this is likely to change when richer data is available. For e.g. Goettler and Shachar (2001) and a related literature in the field of marketing, successfully estimate multiple unobserved product characteristics from data that observes the same consumers making a repeated set of choices, and Heckman and Snyder have used multi-factor discrete choice models to analyze political choice.

4. a limiting distribution of the estimated parameter vector.

The modifications we use to accomplish these tasks imply different computational tradeoffs as opposed to the model with tastes for products – differences that play out differently when using different types of data.

4.1 Computing Market Shares.

In the model with product-specific tastes, market shares can be calculated by a two step method. The first step conditions on preferences for the product characteristics (the β_i) and integrates out the product-specific tastes. This provides market shares conditional on the β_i. When the additive product-specific tastes has a "logit" form the market shares conditional on the β_i have an analytic form, so that there is no approximation error in calculating them, and they are a smooth function of the β_i. The second step follows Pakes (1986) and uses simulation to provide an approximation to the integral defining the expectation (over β_i) of the conditional market shares (i.e. to the aggregate market share).

When there are no additive product-specific tastes we must compute market shares in a different way. A simple two-step replacement is to use the structure of the vertical model to integrate out of one of the dimensions of heterogeneity in the pure characteristics model, thereby producing market shares conditional on the rest of the β_i, and then use the suggestion in Pakes (1986) again to compute the aggregate share. Given an appropriate distribution for the dimension of heterogeneity integrated out in the first step (see below), this produces a smooth objective function.

To see how to do this we first consider the simple vertical model

$$u_{ij} = \delta_j - \alpha_i p_j, \tag{12}$$

for $j =, \ldots, J$, where δ_j is product "quality"

$$\delta_j = x_j\beta + \xi_j, \tag{13}$$

and $u_{i,0} = 0$.

Order the goods in terms of increasing price. Then good j is purchased *iff* $u_{ij} > u_{ik}, \forall k \neq j$, or equivalently

$$\delta_j - \alpha_i p_j > \delta_k - \alpha_i p_k, \;\; \Rightarrow \;\; \alpha_i(p_j - p_k) < \delta_j - \delta_k, \;\; \forall k \neq j. \tag{14}$$

Recall that $(p_j - p_k)$ is positive if $j > k$ and negative otherwise. So a consumer endowed with α_i will buy product j *iff*

$$\alpha_i \;\; < \;\; \min_{k<j} \frac{\delta_j - \delta_k}{(p_j - p_k)} \equiv \overline{\Delta}_j(\delta, p), \;\; \text{and}$$

$$\alpha_i \quad > \quad \max_{k>j} \frac{\delta_k - \delta_j}{(p_k - p_j)} \equiv \underline{\Delta}_j(\delta, p). \tag{15}$$

These formula assume that $0 < j < J$. However if we set

$$\overline{\Delta}_0 = \infty, \text{ and } \underline{\Delta}_J = 0 \tag{16}$$

they extend to the $j = 0$ (the outside good) and $j = J$ cases.

If the cdf of α is $F(\cdot)$, then the market share of product j is

$$s_j(x, p, \xi; \theta, F) \equiv \left(F(\overline{\Delta}_j(x, p, \xi)) - F(\underline{\Delta}_j(x, p, \xi)\right) 1\left[\overline{\Delta}_j > \underline{\Delta}_j\right], \tag{17}$$

where here and below $1[\cdot]$ is the indicator function for the condition in the brackets and θ is a vector containing all unknown parameters.

If $\overline{\Delta}_j \leq \underline{\Delta}_j$, then $s_j(\cdot) = 0$. Since the data has positive market shares the model should predict positive market shares at the true value of the parameters. Note that the vertical model behaves differently then does the model with tastes for products; the latter *never* predicts zero market shares for any parameter value. In the vertical model any parameter vector which generates an ordering which leaves one product with a higher price but lower quality than some other product predicts a zero market share for that product.

4.1.1 The Extension to K Dimensions.

Recall that the difference between the vertical model and the pure characteristics model is that in the pure characteristics model characteristics other than price can have coefficients which vary over consumers (the β_i). However if $u_{ij} = x_j\beta_i - \alpha_i p_j + \xi_j$ then, conditional on β_i, the model is once again a vertical model with cut-off points in the space of α_i (but now the quality levels in those cut-offs depend on the β_i). So to obtain market shares in this case we do the calculation in (17) conditional on the β_i, and then integrate over the β_i distribution.

More precisely we begin as before by ordering the goods by their price. Then for a fixed β we can compute the cut-off points $\overline{\Delta}(x, p, \xi, \beta)_j$ and $\underline{\Delta}(x, p, \xi, \beta)_j$, so the market share function is

$$s_j(x, p, \xi; \theta, F, G) \equiv \tag{18}$$

$$\int \left(F(\overline{\Delta}_j(\delta, p, X, \beta)|\beta) - F(\underline{\Delta}_j(\delta, p, X, \beta)|\beta)\right) 1\left[\overline{\Delta}_j(\cdot, \beta) > \underline{\Delta}_j(\cdot, \beta)\right] dG(\beta),$$

where $F(\cdot|\beta)$ is the cdf of α given β and $G(\cdot)$ is the cdf of β.

That is the pure characteristic model's market share function can be expressed as a mixture of the market share functions of the pure vertical

models. The conditioning argument used here avoids the difficult problem of solving for the exact region of the β_i space on which a consumer prefers product j[14]. It does, however, produce an integral which is typically not analytic. So we use a simulation estimator to approximate it. That is we obtain an unbiased estimator of the integral by taking ns draws from the distribution G of the random coefficients β and then calculating

$$s_j(x, p, \xi; \theta, F, G_{ns}) \equiv \tag{19}$$

$$\frac{1}{ns} \sum_i \left(F(\overline{\Delta}_j(x, p, \xi, \beta_i)|\beta_i) - F(\underline{\Delta}_j(x, p, \xi, \beta_i)|\beta_i) \right) 1 \left[\overline{\Delta}_j(\cdot, \beta_i) > \underline{\Delta}_j(\cdot, \beta_i) \right], \tag{20}$$

where $G_{ns}(\cdot)$ is notation for the empirical distribution of the simulated β_i.[15]

Note that if $F(\cdot)$ has a density (with respect to Lebesgue measure) which is a differentiable function of the parameter vector, then the market share function is a continuously differentiable function of the parameter vector. Of course, this introduces simulation error into the calculated shares, just as in the original BLP.

4.2 Existence and Uniqueness of the $\xi(\cdot)$ Function.

Recall that BLP proceed in three steps. First they show that their model associates a unique $\xi(\cdot)$ with any triple consisting of a vector of parameters, a vector of observed shares, and a distribution over individual characteristics. They then provide a contraction mapping which computes the $\xi(\cdot)$. Finally they make an identifying assumption on the distribution of ξ, and estimate by finding that value of θ that makes the theoretical restrictions implied by the identifying assumption as "close as possible" to being satisfied.

We will mimic those steps. The first task is to show that for every θ and distribution of consumer characteristics there is a unique value of ξ that equates the model's predicted shares to the observed shares, s^o. As in BLP, we assume that s^o, the (J + 1-dimensional) vector of observed market shares, is in the interior of the J-dimensional unit simplex (all mar-

[14]Feenstra and Levinsohn directly calculate the region of integration, $A_j \subset \mathcal{R}^K$ such that if $(\beta, \alpha) \in A_j$ then good j is purchased directly, but this becomes quite complicated.

[15]The calculation of market shares is further simplified by noting that a necessary and sufficient condition for the indicator function's condition for product j to be one conditional on a particular value for β is that $max_{q<j} \Delta_q(\cdot, \beta_i) < \Delta_j(\cdot, \beta_i)$ (recall that the $j-$ ordering is the price ordering). As a result for a given value of β our program first computes the $\{\Delta_j(\cdot, \beta)\}$, then drops those goods for whom $\Delta(\cdot, \beta)$ are "out of order", and then computes the shares.

ket shares are strictly between zero and one.) Simplify notation and let $s(\theta, \xi) \equiv s(x, p, \xi; \theta, F, G)$ for any fixed (F, G, x, p).

Consider the system of $J + 1$ equations

$$s(\theta, \xi) = s^o. \tag{21}$$

Our goal is to provide conditions under which, given the normalization $\xi_0 = 0$, this system has exactly one solution, $\xi(\theta, s^o)$.

Let the discrete choice market share, as a function of all unobserved product characteristics (including ξ_0) be

$$s_j(\xi_j, \xi_{-j}, 0), \tag{22}$$

where ξ_j is the own-product characteristic, ξ_{-j} is the vector of rival-product characteristics and 0 is the (normalized) value of ξ_0.

Now define the "element-by-element" inverse for product j, $r_j(\mathrm{s}_j, \xi_{-j})$, as

$$s_j(r_j, \xi_{-j}, 0) = \mathrm{s}_j \tag{23}$$

The vector of element-by-element inverses, say $r(\mathrm{s}, \xi)$, when viewed as a function of ξ, takes $R^J \to R^J$. It is more convenient to work with a fixed point defined by the element-by-element inverse than to work directly with the equations defined by (21). In particular, the inverse of the market share function (i.e. $\xi(\cdot)$) exists and is unique if there is a unique solution to the fixed point

$$\xi = r(\mathrm{s}, \xi). \tag{24}$$

Theorem 1 *Suppose the discrete choice market share function has the following properties:*

1. Monotonicity. *s_j is weakly increasing and continuous in ξ_j and weakly decreasing in ξ_{-j}. Also, for all ξ_{-j} there must be values of ξ_j that set s_j arbitrarily close to zero and values of ξ_j that set s_j arbitrarily close to one.*

2. Linearity of utility *in ξ. If the ξ for every good (including the outside good) is increased by an equal amount, then no market share changes.*

3. Substitutes with Some Other Good. *Whenever s is strictly between 0 and 1, every product must be a strict substitute with some other good.*

In particular, if $\xi' \leq \xi$, with strict inequality holding for at least one component, then there is a product (j) such that

$$s_j(\xi_j, \xi_{-j}, 0) < s_j(\xi_j, \xi'_{-j}, 0). \tag{25}$$

Similarly, if $\xi' \geq \xi$, with strict inequality holding for at least one component, then there is a product (j) such that

$$s_j(\xi_j, \xi_{-j}, 0) > s_j(\xi_j, \xi'_{-j}, 0). \tag{26}$$

Then, for any market share vector s that is strictly interior to the unit simplex: (i) an inverse exists, and (ii) this inverse is unique. ♠

Comments.
1. The theorem is true independent of the values of (θ, F, G, x, p) that go into the calculation of $s(\cdot)$ provided those values imply an $s(\cdot)$ function that satisfies conditions (1) and (3). It is easy to verify that those conditions will be satisfied for any finite θ as long as F has a density which is positive on the real line (a.e. β). In particular $G(\cdot)$ need not have a density (w.r.t. Lebesgue measure), and indeed the simulated $G_{ns}(\cdot)$ we typically use in computation will not have a density.
2. The *Linearity* assumption is redundant if we stick with the model in (11). That is, a non-linear model that was linear only in ξ would have exactly the same properties as those developed here provided it satisfied the usual identification conditions.

Proof. Existence follows from the argument in Berry (1994). Our first step in proving uniqueness is to show that the map $r(\xi, s)$ is a weak contraction (a contraction with modulus ≤ 1), a fact which we use later in computation.

Take any ξ and $\xi' \in R^J$ and let $\|\xi - \xi'\|_{sup} = d > 0$. From (23) and Linearity

$$s_j(r_j + d, \xi_{-j} + d, d) = \mathrm{s}_j. \tag{27}$$

By Monotonicity

$$s_j(r_j + d, \xi', 0) \geq \mathrm{s}_j, \tag{28}$$

and by (3) there is at least one good, say good q, which for which this inequality is strict (any good that substitutes with the outside good). By Monotonicity, this implies that for all j,

$$r'_j \leq r_j + d \tag{29}$$

with strict inequality for good q. A symmetric argument shows that the condition

$$s_j(r_j - d, \xi_{-j} - d, -d) = \mathrm{s}_j \tag{30}$$

implies that for all j,

$$r'_j \geq r_j - d \tag{31}$$

with strict inequality for at least one good. Clearly then $\|r(\xi', s) - r(\xi, s)\| \leq d$, which proves that the inverse function is a weak contraction.

Now assume that both ξ and ξ' satisfy (24), and that $\|\xi - \xi'\|_{sup} = \kappa > 0$, i.e. that there are two distinct solutions to the fixed point. In particular let $\xi_q - \xi'_q = \kappa$. Without loss of generality assume that q substitutes to the outside good (if this were not the case then renormalize in terms of the good that substitutes with q and repeat the argument that follows). From above, $s_q(r_q + \kappa, \xi'_{-q}, 0) > s_q$. But this last expression equals $s_q(\xi'_q, \xi'_{-q}, 0)$, which, since ξ' is a fixed point, equals s_q, a contradiction. ♠

4.3 Computation of $\xi(\cdot)$.

We provide three different methods for computing $\xi(\cdot)$. They are

- use of BLP's contraction mapping for the modified BLP model given in equation (9) with μ_ϵ set as high as is computationally practical,
- use of the element by element inverse introduced in the proof of uniqueness, and
- a homotopy method similar to those used in the literature on the computation of general equilibrium models.

We introduce all three of these methods because no single method works well in all situations. Each of the methods given above has a limitation which has proven to be problematic in certain Monte Carlo examples. The first method is fast but will be inaccurate if μ_ϵ can not be set sufficiently high, and there are data designs where good approximations require a μ_ϵ larger than this algorithm appears able to handle. The second method is not guaranteed to converge at any particular speed and in fact appears to converge very slowly in some of our Monte Carlo runs. The third method is, at least in principal, guaranteed to be converge, but, as we shall show, can also be fairly slow.

As a result our suggestion is to use the first method when practical, and a combination of all three when necessary. As is explained below, the combination begins with the modified BLP model with a large fixed μ_ϵ. It

then switches to a homotopy when the rate of change in the sup norm for the contraction used in modified BLP is too small. The homotopy is constructed from the element by element inverse introduced in the proof of uniqueness. Finally when the homotopy is close to the truth and is not progressing quickly enough, we switch to a Newton method.

BLP's Contraction and the Modified BLP Model. For a fixed μ_ϵ this method is just the BLP method applied using the normalization of this paper. The question that remains is how to chose μ_ϵ. As is explained above if one chooses a value of μ_ϵ which is too small the approximation will not be adequate whereas if one chooses a value which is too high the computer will not be able to calculate the needed exponents. Our Monte Carlo results experiment with two ways of setting μ_ϵ. In one we set μ_ϵ as high a value as seems initially practical, and then reduce it if necessary. In the second, we estimate μ_ϵ together with the rest of the parameters of the model.

Element-by-element Inverse. This procedure uses the element by element inverse shown to lead to a weak contraction in the proof of the theorem in section 4.2 (i.e the $r(s, \xi)$ in the proof of Theorem 1). If the weak contraction had modulus that was strictly less than one, this contraction would be guaranteed to converge to the fixed point at a geometric rate. Unfortunately we have been unable to prove that the modulus is strictly less than one, and in Monte Carlo studies we find that it sometimes contracts so slowly as to become useless.

To implement this procedure we use the simulated pure characteristic share function as given in (19). That is we begin with a candidate ξ vector, hold ξ_{-1} fixed, then find the value of ξ_1 that makes the simulated share for the first product match the actual share for that produce, and proceed similarly for $j = 2, \ldots J$ (always holding the ξ_{-j} vector at its initial value). This provides the new ξ vector, which is then passed through the same algorithm. We continue in this fashion until the predicted shares at the new vector match the observed shares.

In practice, sometimes this method provides large improvements for the first few iterations and then slows nearly to a stop. If the predicted shares at that point are positive, we can try a simple Newton method to look for the solution. If that fails, we then require our third method.

Homotopy. Homotopy methods are frequently used with great success to find fixed points in the computable general equilibrium literature.[16] The basic idea is to consider a sequence of objective functions formed as a weighted average of the true fixed point problem and a similar problem with a known solution. We begin with the objective function which places all weight on the fixed point with a known solution, and slowly change the weights to place more weight on the true problem.

Starting with the standard fixed-point homotopy we have

$$h(\xi, t, \xi_0) = (1 - t) * (\xi - \xi_0) + t * (\xi - r(s, \xi)) \tag{32}$$

where

- t is a parameter that takes values between zero and one,
- the function $r(\cdot)$ returns the element-by-element inverse of the market share function (see equation (23)), and
- ξ_0 is a good initial guess at the true ξ, in particular take it as the output one of the other methods when those have failed to converge.

For each value of t, consider the value of ξ that sets $h(\xi, t, \xi_0)$ to zero. Call this $\xi(t, \xi_0)$. For $t = 0$ the solution is trivially the starting guess of ξ_0. For $t = 1$ the solution is the fixed point that we are looking for. The homotopy methods suggest starting at $t = 0$, where the solution is trivial, and slowly moving t toward one. The series of solutions $\xi(t)$ should then move toward the fixed-point solution $\xi(1)$. If t is moved slowly enough, then by continuity the new solution should be close to the old solution and therefore "easy" to find (say by a Newton method starting at the prior solution).

In our problem a version of the fixed-point homotopy is a strong contraction when $t < 1$ making it easy to compute the ξ from successive t. I.e. the homotopy implies that

$$\xi(t, \xi_0) = (1 - t) * \xi_0 + t * r\left(s, \xi(t, \xi_0)\right), \tag{33}$$

which, when viewed as an operator which takes $\mathcal{R}^J$ into itself, is a contraction mapping with modulus less than t. This suggests a recursive solution method, taking an initial guess, ξ, for the solution $\xi(t, \xi_0)$ and then placing this guess on the RHS of (33) to create a new guess, ξ':

$$\xi' = (1 - t) * \xi_0 + t * r(s, \xi). \tag{34}$$

[16] See, for example, Whitehead (1993) and Eaves and Schmedders (1999).

Equation (34) is a recursive method suggested by the logic of the homotopy. We can set t close to one and get a reasonably good approximation to ξ.

Our "homotopy" method involves using fixing a value for t and then using equation (34) to recursively compute a value of ξ that solves the fixed-point implied by that equation. For t sufficiently less than one, this is very fast. As t approaches one, the modulus of contraction for fixed point in (34) also approaches one.

In practice, we can start with a fairly large value for t, say t = 0.99, because that still gives us a quick answer. However, as t gets closer and closer to one, it is sometimes necessary to move t very slowly while solving for the next $\xi^*(t_k)$ via Newton's method. However, whereas the element-by-element inverse can slow down fairly far away from the truth, the homotopy method in practice gets us much closer to the correct answer before slowing down.

When we refer to the "homotopy" method in remainder of the paper, we actually mean an integrated routine that incorporates all of the methods of this subsection. First, we fix a "high" value of μ and use the BLP contraction (switching to a lower value of μ if numeric errors arise.) This is quite fast. Then, if necessary we switch to the element-by-element inverse, in conjunction with periodic Newton steps when possible, to see if this gives us a quick answer. If not, we only then move to the homotopy method defined by (34). For small problems (like our initial Monte Carlo examples), we look for a very precise answer with t very close to or equal to one and often succeed in solving the original fixed point nearly exactly. For large and difficult problems (like our last Monte Carlo example below), we apply (34) only for a limited set of values of t (say, t = 0.99, t = 0.995 and t = 0.998) without trying to push t closer to one. These few steps move the computed δ's quite far and the resulting Monte Carlo estimates seem good, whereas the computational burden is still manageable.

The computational burden in the pure characteristics model also turns out to be related to asymptotic behavior as the number of products increases. The next subsection provides an overview of related results on limit theorems.

4.4 Limit Theorems and Simulation Error

Berry, Linton and Pakes (2004) provides limit theorems for the parameter estimates from differentiated product models both with and without tastes for products. The limit theorems for the different models differ, which leads to a better understanding of the differences in the computational properties

of the estimators for these models. Indeed, together with the discussion above, they imply computational tradeoffs which will make different models easier to compute in different situations.

In particular, the argument in Berry, Linton and Pakes (2004) shows that in BLP-style random coefficient logit models, the calculated unobservables, ξ, are very sensitive to simulation error. Therefore, the number of simulation draws has to increase rapidly with the number of products. On the other hand, in pure characteristic models, the calculation of ξ is much less sensitive to simulation error. A practical implication is that we may be able to use many fewer simulation draws to estimate pure characteristics models, as opposed to random coefficient logit models. This advantage can partly offset the otherwise more burdensome computational problem presented by the pure characteristics model.

The difference in the limit properties of the estimators from the two model stems from differences in the mapping from market shares to $\xi(\cdot)$ as the number of products, J, grows large (and therefore at least some of the market shares become small.) Take the logit model as a special case of the original BLP. In the pure logit model, no simulation is necessary, but errors in observed market shares, s_j, might similarly be introduced by sampling error from a finite sample of consumers. Berry, Linton and Pakes (2004) show that simulation error introduces the same sort of problem in the random coefficients logit model that sampling error introduces in the logit model.

In the pure logit model, it is well-known that the solution for ξ is analytic:

$$\xi_j = ln(s_j) - ln(s_0) - x_j\beta, \tag{35}$$

so that

$$\frac{\partial \xi_j}{\partial s_j} = \frac{1}{s_j}. \tag{36}$$

As J becomes large, by necessity many market shares must become small and so the calculated ξ_j's become very sensitive to any error in s_j, whether from a small sample of consumers or from simulation error.

This effect is natural in logit-style models, where substitution patterns between products are very diffuse. When there are many products in such models, small changes in ξ produce only small changes in observed market shares – i.e. $\partial s_j/\partial \xi$ is small. It may be intuitive, then, that the derivative of the inverse mapping $\partial \xi_j/\partial s_j$, can be correspondingly large under the same circumstances. But when ξ is sensitive to errors in computed market shares, a large number of simulation draws will be necessary to compute accurate ξ's.

Pure characteristics models have different properties. In these models, competition becomes very localized and "fierce" as the number of products increases. That is, $\partial s_j/\partial \xi_j$ becomes very large. It may be intuitive, then, that the derivative of the inverse mapping $\partial \xi_j/\partial s_j$, can be correspondingly small under the same circumstances. So, a relatively small number of simulation draws may be adequate to compute accurate accurate ξ's.

For example, Berry, Linton and Pakes (2004) formally show that to obtain asymptotically normal estimates of the parameter vector in the vertical model the number of simulation draws has to grow only at rate J, whereas in logit-style models the number of draws has to grow at the much faster rate J^2.

Similarly, in Monte Carlo exercises they show that as a practical matter pure characteristics models can be estimated with many fewer simulation draws than in BLP-style models. The differences between the two models increase in the number of products marketed, but it is clear that they can be large for numbers of products that are relevant for empirical work. For example, for the limiting distribution to provide an accurate description of the Monte Carlo distribution of the vertical model with up to two hundred products, fifty simulation draws seem to suffice. In contrast, the logit model with one hundred products requires over two thousand simulation draws.

The advantage held by pure-characteristics models in controlling simulation error provides a partial offset to their otherwise greater computational complexity. Readers desiring a (much) more formal treatment of simulation error, sample size and limit theorems in these models should consult Berry, Linton and Pakes (2004).

4.5 Computational Comparisons.

Gathering the results of prior sections, we have two theoretical reasons for expecting the computational burden of the pure characteristics model to differ from the computational burden of the model with tastes for products, but they have opposite implications.

- First, the number of simulation draws needed to get accurate estimates of the moment conditions must grow at rate J^2 in the model with a taste for products, while it need only grow at rate J in the pure characteristics model.

- Second the contraction mapping used to compute the inverse is expected to converge at a geometric rate for the model with tastes for products, but we do not have such a rate for the pure characteristics model.

The first argument implies that computation should be easier in the pure characteristics model, the second that computation should be easier in the model with tastes for products. Of course which of the two effects turns out to dominate may well depend on features of the data being analyzed: the number of products, the number of times the inverse must be evaluated in the estimation algorithm (which typically is related to the number of parameters), and so on.

There is a great deal of evidence on the speed and accuracy of BLP's algorithm for estimating the model with tastes for products (and we will provide a bit more below). As a result the next section focuses on the properties of the algorithms available for estimating the pure characteristic model.

5 Evidence from Simulated Data

Our goal in this section is to investigate the properties of alternative algorithms for estimating the pure characteristic model. As noted the computational difficulties that arise in the pure characteristic model are a result of the need to compute the δ which solve the fixed point in equation (24). We compare algorithms based on the following three different methods for computing this fixed point.

- The "homotopy" method outlined in the text. This begins with the weak contraction obtained from the element by element inverse in equation (24), moves to the homotopy in equation (33) when the element by element inverse fails to improve the objective function, and then moves to a Newton method when the homotopy method gets close enough to the true solution.

- The second method sets the μ_ϵ in equation (9) to some fixed number, and proceeds using BLP's contraction.

- The third method only differs from the second in that it estimates μ_ϵ along with the other parameters of the model.

The comparison will be done in terms of both compute times and the precision of various estimates. We consider precision in three steps: first of the estimates of the δ themselves, then of the estimates of the parameters of the underlying model, and finally of the estimates of the implications of interest to applied studies (own and cross price elasticities, and welfare).

5.1 A Model for Simulation

For most of the results we report, data is drawn from a model with utility function

$$u_{ij} = \delta_j + \sigma_x \nu_{ix} x_j - (\alpha_i * p_j) \tag{37}$$

where

$$ln(\alpha_i) = \sigma_p \nu_{ip} \tag{38}$$

and

$$\delta_j = \beta_0 + \beta_x x_j + \xi_j. \tag{39}$$

The consumer-specific random terms (ν_{ix}, ν_{ip}) are distributed standard normal (so that α_i is log normal, with a normalized mean.)

The x variable is drawn as twice a uniform (0,1) draw that is firm-specific plus 1/2 a uniform (0,1) draw that is common to all firms in a market. This is to allow for within market correlation across observables. Initially ξ is drawn as a uniform on (-0.5,0.5). Note that the variance of x is then greater than that of ξ, which is likely to help the Monte Carlo find good parameter estimates with limited data. Later we will increase the variance of ξ and see how this affects the results. Price, p, is set equal to a convex function of δ, $p_j = e^{\delta_j}/20$, and this insures positive shares for all goods (though some of the shares get very small when we consider markets with a large number of products). Note also that p is a function of δ and δ is a function of ξ, so that p and ξ are correlated in the simulated datasets. Finally the actual consumer choice data is generated from 5000 independent draws on ν_i who chose optimally given the true values of the parameter.

5.2 Calculating δ.

We begin with the calculation of delta. We can illustrate our results here by looking at a simple example with only five products. The example uses randomly drawn data from the base model of the last subsection. The first column of Table 1 shows the "true" values of the randomly drawn δ.

With 5 products and our data-creation model our homotopy method can almost always find a δ vector that exactly reproduces the "true" market shares. The first homotopy column uses the same 5000 draws on ν used to create the data, so it has no simulation error and it recovers the exact values of the true δ's. The second homotopy column uses only 500 simulation draws and so some error is introduced. While the homotopy method with simulation error fits the shares to computer precision, it does so with δ's that vary from the originals. On the other hand even with simulation error the order of

the δ's is preserved. The last row of the table gives the computational time relative to using the full homotopy method with 5000 draws. Decreasing the number of draws on ν decreases compute time almost in the proportion of the decrease in the number of draws.

The columns labeled "Modified BLP Contraction: Fixed μ_ϵ" use the contraction in BLP with the shares modified as in equation (9). Recall that this multiplies the variance of the Type II extreme value errors by μ_ϵ^{-1}. In Table 1 we first look at columns that use the same 5000 draws on ν with; μ_ϵ set at 1, which gives us back BLP's contraction, and then μ_ϵ set to 10 and 50 respectively. In the last column, simulation error is again introduced by using only 500 draws with $\mu_\epsilon = 50$.

With $\mu_\epsilon = 1$ there is little correlation between the true δ and those obtained from the contraction. On the other hand the compute time is only $(1/50)^{th}$ of the compute time for the full homotopy. With $\mu_\epsilon = 10$ only one of the δ obtained from the contraction is "out of order". When we get to $\mu_\epsilon = 50$ the order of the δ obtained from the contraction is correct, though they still have a one or two percent error in their values. Now however the compute time is 60% of the compute time for the full homotopy. There is also some indication that even at $\mu_\epsilon = 50$ the modified BLP contraction is more sensitive to simulation error.

We note that this is illustrative of the results we got on computing δ with other sample designs. So we conclude that from the point of view of estimating the δ it might be efficient to go to the modified BLP contraction, but only if μ_ϵ is kept very high.

Table 1: An Example of Calculating δ Using Different Methods

	True	Homotopy.		Modified BLP Contraction; Fixed μ_ϵ .			
nsim	5000	5000	500	5000	5000	5000	500
				$\mu_\epsilon = 1$	$\mu_\epsilon = 10$	$\mu_\epsilon = 50$	$\mu_\epsilon = 50$
δ_1	2.99	2.99	3.14	3.08	3.02	3.04	3.25
δ_2	3.21	3.21	3.36	2.09	3.19	3.26	3.51
δ_3	3.56	3.56	3.71	3.13	3.60	3.62	3.94
δ_4	4.04	4.04	4.27	4.13	4.10	4.10	4.61
δ_5	4.10	4.10	4.32	2.68	4.06	4.16	4.65
Rel. Time:		1	0.15	0.02	0.12	0.58	0.03

5.3 Sample Designs

The results for the remainder of the issues we investigated depended somewhat on the sample designs. The major feature of the design that seemed relevant was the number of products marketed. So we focused on two sample designs: one where there are a small number of products marketed but a reasonably large number of markets, and one with a large number of products marketed but a small number of markets.

The sample with a small number of products marketed consists of 20 markets, and for each market the number of products was chosen randomly from a distribution which put equal weight on $[2, ..., 10]$. With this design estimation is very fast and we have experimented with a number of alternative assumptions some of which will be illustrated below. The sample with a large number of products has one hundred products per market, but only three markets. The homotopy's compute time is quite large under this sample design, and as a result we have done less experimentation with it. All samples used five thousand simulation draws to construct the model's predictions for market shares (i.e. this is the size of the consumer sample).

5.3.1 Parameter Estimates: Small Number of Products.

First we look at estimates of parameter values in Tables 2 and 3. The instruments used are a constant, x, x^2, the mean x in the market and the minimum distance to the nearest x[17].

Starting with Table 2 we see that with a small number of products per market, and a small variance on the unobservable, the modified BLP contraction mapping with a μ_ϵ set exogenously to thirty does quite well; not noticeably worse than our full homotopy. However to get the performance of the two algorithms to be comparable when the distribution of ξ had more variance we needed to push μ_ϵ up to fifty. Computational problems made it difficult to push μ_ϵ much higher than this.

The numbers reported underneath the coefficient estimates are the standard errors of the estimates across different Monte Carlo samples. Since the estimates are a mean across these one hundred samples, the standard error of this mean should be about one tenth of the reported standard error. So the asymptotic approximation of the distribution of the estimator is underestimating the estimator's variance by quite a bit with these sample sizes. Also all algorithms do worse when we increase the variance of ξ; i.e. intuitively

[17] We also tried an approximation to "optimal instruments" as in Berry et al. (1999) but this had little effect on the results.

we need more data to obtain a given level of precision as the unobservable variance increases.

One problem with using the modified BLP contraction with a fixed μ_ϵ is that we would not know how high we needed to set μ_ϵ to get reasonable approximations if we were working with a real data set. Moreover we experimented some to find the best values for these runs. We rejected any value for the fixed scale that resulted in numeric errors, and in the first experiment $\mu = 30$ worked a bit better than $\mu = 50$ even though there were no obvious numeric errors[18].

An alternative which avoids the problem of choosing the scale is to let the data try to estimate μ_ϵ. Table 3 presents the results from this exercise. We note that in several cases the scale parameter, which is now set by the search algorithm, increased to the point where numeric errors occurred. In those cases, we fixed μ_ϵ at 50. In one of those cases, even $\mu = 50$ caused problems and so we fixed μ_ϵ at 25[19].

Table 3 presents both the estimates of the parameter of interest, and of the "auxiliary" parameter μ_ϵ. In particular it provides both the mean and median of the estimates of μ_ϵ across runs. The first column of results uses the base specification for the unobservable, while the second column increases the variance of the unobservable. In both cases this gives us results which are worse than those obtained in Table 2, though probably still acceptable, especially for the sample design with less variance in ξ.

We conclude that with a small number of products, the modified BLP contraction with μ_ϵ fixed at a large value we may do well, especially if the variance of of the unobservable is small. However at high values of μ_ϵ numeric errors are quite common, and when we estimate μ_ϵ instead of fix it exogenously, we do seem to do noticeably worse.

5.3.2 Substitution Effects: Small Number of Products.

Table 2 shows that, for the example data-generating process, if we could fix μ_ϵ at a large enough value the modified BLP estimation does about as well as the homotopy. On the other hand if we did not know what value was large enough and consequently decided to estimate μ_ϵ, then Table 3 indicates that how well we do depends on the variance in ξ relative to the variance in x. On the other hand there is an advantage of the modified BLP algorithm that

[18]This could be because of approximation errors in the computer routines that calculate the exponents of large numbers.

[19]We also tried experiments where we imposed the traditional logit scale normalization of one, while dropping our current normalization on the price coefficient. Those runs where much less likely to converge without numeric errors.

Table 2: Monte Carlo Results, Small Number of Products
(Modifed BLP Contraction with Fixed μ_ϵ vs Homotopy)

	(1)	(2)	(3)	(4)
Method	Mod. BLP	Homotopy	Mod. BLP	Homotopy
Scale (μ_ϵ):	30	n.r.	50	n.r.
$\xi_j =$	$U(-\frac{1}{2}, \frac{1}{2})$	$U(-\frac{1}{2}, \frac{1}{2})$	U(-1.5,1.5)	U(-1.5,1.5)
σ_x (= 1)	1.04	1.03	1.24	1.26
	(0.04)	(0.03)	(0.06)	(0.06)
σ_p (= 1)	1.00	0.98	1.02	1.02
	(0.01)	(0.01)	(0.03)	(0.03)
β_0 (= 2)	2.06	2.00	2.34	2.33
	(0.05)	(0.05)	(0.10)	(0.09)
β_x (= 1)	0.99	1.00	1.04	1.05
	(0.01)	(0.01)	(0.02)	(0.03)

estimates μ_ϵ over the fixed point homotopy; it is much easier to compute. So just on the basis of parameter estimates which procedure seems the best one to apply depends on the nature of the problem. Of course parameter estimates are not usually the objects of interest; rather it is their implications we are usually concerned with.

Own and cross-price elasticities are one frequent focus of empirical applications in I.O. If we use a modified BLP contraction to estimate parameters we still have more than one option for computing own and cross-price elasticities. The standard procedure would be to use the parameter estimates obtained from the modified contraction together with its functional forms to compute the elasticities. Alternatively, we could subsitute the parameter estimates from the modified BLP contraction that estimates $\mu_{epsilon}$ into the share equations generated from the pure characteristic models and calculate the own and cross price elasticities from this "hybrid" model.

There is some intuition for this hybrid model. We know that we could chose a scale parameter large enough to make the predictions of the modified BLP algorithm as close as we like to those of the pure characteristic model that generates our data. The problem is that the BLP estimation algorithm may not be able to pick out a value of μ_ϵ which is large enough for the approximation in equation (9) to be accurate. However if the only error in the BLP estimates were in the estimate of μ_ϵ, so that the problem in estimating the scale did not "spill over" to other coefficients, we would expect the hybrid model to produce exact estimates of the appropriate elasticities.

Table 3: Small Number of Products, Estimated μ_ϵ*

	True	(1)	(2)
ξ distribution		$U(-\frac{1}{2}, \frac{1}{2})$	$U(-1.5, 1.5)$
nsim		500	500
σ_x	1	1.14	1.64
		(0.04)	(0.08)
σ_p	1	1.03	1.09
		(0.01)	(0.03)
β_0	2	2.19	2.79
		(0.06)	(0.12)
β_x	1	1.00	1.03
		(0.01)	(0.03)
Scale, μ_ϵ	∞	34.08	15.50
		(3.31)	(1.98)
μ_ϵ (Median)	∞	17.81	4.67

*All estimates are means across 100 simulated datasets. Estimated standard deviations of the mean estimates are given in parentheses. The homotopy estimates took on the order of 10 times as long to compute.

Table 4 provides an example of results on how well the methods reproduce the true substitution patterns. The example considers one market from the data-generating process that has more spread in the unobservables and therefore less precise estimates (i.e. the one in the last column of Tables 2 and 3). The column labeled "true" gives the true derivatives of the market share of the first (lowest-priced) product in that market with respect to the row product. The first entry is therefore an own-price derivative, the second entry is the cross-price derivative of product one with respect to product two, and so forth.

The last five columns of Table 4 recompute those derivatives using five different methods which differ from one another in either the parameters used or in the functional forms used to compute the derivatives conditional on the parameter estimates. They are: (i) estimates from the homotopy and functional forms from the pure characteristic model, (ii) estimates from the modified BLP algorithm with μ_ϵ set to fifty and those functional forms, (iii) estimates from the modified BLP algorithm that estimates μ_ϵ and those functional forms, (iv) estimates from the modified BLP algorithm that estimates μ_ϵ *but* functional forms from the pure charactersitic model, and (v) estimates from the modified BLP algorithm that estimates μ_ϵ but the functional forms from the modified BLP algorithm with μ_ϵ set to fifty. We do the computation for each of the parameters generated by a single Monte Carlo

run in the earlier tables, and then average across the Monte Carlo sample of parameters. This approximates the answer we would get by averaging over the asymptotic distribution of the parameters as estimated by the appropriate method. We also provide the standard deviation of the calculated sample mean.

Table 4 indicates that use of the homotopy estimates and the functional form from from the pure characteristics model does a much better job of reproducing the true substitution pattern than using either of the modified BLP estimates and their functional forms. This is especially true relative to the model with an estimated μ_ϵ, but even the model which fixed μ_ϵ at a large number, and consequently does not have great computational advantages, misses rather dramatically on the fourth elasticity. The surprising result in this table is how well we do using the parameters of the modified BLP algorithm that estimates μ_ϵ and the functional forms from the pure characteristic model. Indeed even using the functional forms from the modified BLP algorithm which sets μ_ϵ to fifty, does fairly well, again except for the fourth elasticity.

Of course this is just one example, and it need not be indicative of what would happen under different data designs. Still this example does have the following implications

- "good" parameter estimates from a particular model need not imply that use of those parameter estimates and the model's functional forms lead to good predictions for the implications of interest; i.e. in our case the functional form approximation used to compute derivatives matters, and
- even if we use the modified BLP contraction for estimation, we may not want to use the functional form in equation (9) to compute our estimates of the implications of the parameter estimates.

5.4 A Large Number of Products

In this section we consider a sample with a larger number of products (an average of 100 per market) and a smaller number of markets (3) structurred as a time series on a market with improving characteristics. In particular we use the data on the evolution of megahertz in the computer market from 1995 to 1997, taken from Pakes (2003), to pattern the x's in our sample.

Year:	1	2	3
Min Mhz:	25	25	33
Max Mhz:	133	200	240

Table 4: Examples of Substitution Patterns.

Product	True	Homotopy	Modified BLP		Blp Parms, but $\partial s/\partial p$ via	
			Fix $\mu_\epsilon = 50$	Estimate μ_ϵ	Homotopy	Mod. BLP
1	-0.6027	-0.6198	-0.7313	-0.4130	-0.5241	-0.6353
		(0.0337)	(0.0220)	(0.0232)	(0.0267)	(0.0226)
2	0.1981	0.2093	0.2531	0.1656	0.1629	0.2034
		(0.0233)	(0.0140)	(0.0139)	(0.0133)	(0.0144)
3	0.1621	0.1624	0.1183	0.0331	0.1461	0.1096
		(0.0106)	(0.0022)	(0.0059)	(0.0129)	(0.0024)
4	0.0000	0.0002	0.0278	0.0348	0.0001	0.0172
		(0.0001)	(0.0038)	(0.0042)	(0.0001)	(0.0031)
5	0.0204	0.0252	0.0292	0.0235	0.0212	0.0296
		(0.0026)	(0.0024)	(0.0023)	(0.0023)	(0.0031)

The Monte Carlo sample lets the number of products increase across the three years, from 75 to 100 and then to 125, and has x's drawn from uniform on min-max range of Mhz (divided by 10). So that the ξ's scale with x, the ξ's are set equal to the range of the megahertz for the year times $(2u_i - 1)$ wheer u_i is a draw from uniform on $[0, 1]$.

The δ's are determined as above ($\delta_j = \beta_0 + \beta_1 x_j + \xi_j$) as are the parameters and the distribution of α_i. To mimic a high tech market we let prices fall over time, t, according to

$$\ln(p_j) = (t-1) * \ln(0.7) + 1.1 * \ln(\delta_j). \tag{40}$$

Note that again price is convex in δ thus assuring positive shares. However the shares we generate are sometimes very small, smaller than we typically observe in actual data sets[20] (see below). Price for the same set of characteristics declines at roughly 30% per year. This together with the increase in the number of products and the improvement in the product characteristics over time are consistent with a market which generates "large" increases in welfare over time.

The instruments used here are a constant and x, both interacted with a set of dummies for each market. That these seem to suffice is probably a result of the fact that we are requiring the "same utility model" to fit in each time period and letting the products in the market change rather dramatically over time.

[20]This may well be more of a result of the data not reflecting reality than our simulation not reflecting reality. That is goods which truly have very small market shares might exist and simply not be included in traditional data sets.

The fact that market shares are so small makes computation of the δ more difficult in this example. Partly as a result, we estimate on only one example dataset, and use asymptotic standard errors. We obtain estimates from all three algorithms discussed above. In the homotopy method we begin with the δ outputted by the modified BLP contraction with fixed μ_ϵ, say δ_0, and then iterate only 25 times on the "homotopy" equation:

$$\delta' = 0.05\delta + 0.95r(\delta), \tag{41}$$

where $r(\delta)$ is the element-by-element inversion.

We note that even this limited search greatly improves the calculated δ's. In particular the mean of the calculated δ's are much too small if we stop at the solution to the modified BLP contraction with fixed μ_ϵ. This contrasts with the case with a small number of products: the modified BLP contraction with fixed μ_ϵ did rather well with that sample design. The difference is that the current sample design generates products with small market shares, and though the pure-characteristic model will do that if the characteristics of products are close enough to one another, the model with tastes for products can only generate small shares if the $\delta's$ are very small.

Table 5: Parameter Estimates
(Dataset with a Large Number of Products)

Parameter	True	Homotopy	Modified BLP	
			Fix μ_ϵ	Estimate μ_ϵ
σ_x	1	0.833	0.862	0.832
		(0.194)	(0.380)	(6.956)
σ_p	1	1.192	1.188	1.207
		(0.556)	(0.621)	(1.783)
β_0	2	1.956	1.354	-6.455
		(2.013)	(2.066)	(66.864)
β_x	1	0.984	0.986	0.879
		(0.209)	(0.198)	(1.102)
scale, μ		∞	10*	0.934
				(6.680)

* The scale was initially set to 10, but some combinations of parameter values and markets this caused numeric problems and the scale in those cases was halved until the numeric problems went away. In a few cases, a scale as low as 2.5 was necessary. Asymptotic standard errors are in parenthesis.

In Table 5 we see much bigger differences between the parameter estimates from the different estimation algorithms than we did when we had a small number of products. The estimates from the modified BLP algorithm with estimated μ_ϵ are clearly the worst of the three. The estimated value of the scale parameter μ is relatively small which implies that the logit error is being assigned a relatively important role. To counteract the effects of the logit and still match the small shares, the constant in δ is driven down to less than -6. The modified BLP contraction with a fixed μ_ϵ does better, but still suffers from a β_0 which is too low.

Substitution Patterns and Own Price Elasticities In this example all the models do a good job of capturing the general pattern of substitution across products, although the BLP model is a bit more diffuse, as expected. However, only the homotopy method does a good job of capturing the overall level of elasticities. This is shown in Table 6 which gives, for the first five products in the year 1 data (the lowest price products), actual share and price data and then price elasticities (calculated from a discrete 1% change in price.) The products with a very small share have excellent substitutes in the pure characteristics model, but the modified BLP contraction with estimated μ_ϵ does not capture this effect and even the model with a fixed μ_ϵ has a lot of trouble reproducing this result. Note that for the 4th product, the share is truly tiny and that a 1% increase in price would wipe out all of the sales of that product[21].

The last column of the table is particularly interesting. This column uses the BLP estimates of σ as given in the last column of Table 5. However, given those σ estimates we now solve for δ via the "homotopy" contraction described above and re-calculate the linear β parameters from that new δ. Because the BLP σ's are close to the truth, the recalculated β's are also close to the truth. Finally, we use the BLP σ's and the recalculated β's to compute the price elasticities in the last column of Table 6. These elasticities, unlike the pure BLP elasticities, are quite close to the correct values.

The last column of Table 5 suggests the possibility that BLP might sometimes get the σ's correct even when the scale parameter is badly estimated – i.e. there is information in the data on the relative importance (in determining substitution patterns) of various x's, but there is not a lot of information on exactly how "local" are the substitution patterns. In such a case, the method of the last column of Table 6 may provide a good, and relatively

[21]Such a product would likely only survive in the market if produced by a multi-product firm, so that some markup could be sustained and fixed-costs perhaps shared across products.

easy to compute, estimate of elasticities. If nothing else, the calculation in that column provides a robustness check in answering the question about what elasticities would be if we held σ fixed and took μ off to infinity.

Table 6: Predicted Elasticities
(Dataset with a Large Number of Products)

		% Change in Share from a 1% Price Chg.				
% Share	Price	True	Homotopy	Modified BLP		BLP, but $\partial s/\partial p$ via
				Fixed μ_ϵ	Est. μ_ϵ	Homotopy
5.2109	1.26	-11.0	-14.6	-6.8	-1.8	-14.6
4.0180	1.58	-26.8	-24.5	-9.2	-2.3	-24.7
0.1078	2.93	-30.8	-46.3	-11.0	-2.9	-45.3
0.0038	3.85	-100.0	-100.0	-14.8	-3.7	-100.0
0.8855	4.01	-63.4	-58.5	-19.3	-4.9	-57.9

Welfare Effects. We now calculate the average per-person welfare gain (in dollars) of moving from the year one choice set to the year three choice set. Recall that we greatly increase both the number of goods as well as their quality at the same time as lowering prices. As a result there is both a large influx of consumers from the outside good over time, and a "large" true welfare increase (much of it going to the new consumers).

Results on the total welfare gain are given in Table 7. The rows of that table correspond to the last 5 columns of Table 6.

The surprising result here is that all methods do very well. The homotopy is within .5% of the true result, but even the modified BLP algorithm with an estimated μ_ϵ is within three per cent of the truth. The modified BLP methods do not do as well on the parameters, or on the elasticities, but the fact that the contraction fits the shares exactly means that the extra gain from the logit errors is offset by lower δ's and this roughly counteracts the problems generated for welfare measurement by the model with tastes for products.

6 Summary and Conclusion

This paper largely focuses on the practical and computational problems of estimating the pure characteristics model. We provide several possible algo-

Table 7: Welfare Effects

Method	Gain
True	266.1
Homotopy	265.3
Modified BLP $\mu_\epsilon = 10$	270.0
Modified BLP μ_ϵ estimated	259.0
Mod. BLP, μ_ϵ est., but final homotopy contraction	272.1

rithms.

Of these algorithms, the easiest is to simply use the existing BLP method, perhaps using the "units" normalization of this paper as opposed to the traditional normalization on the scale of the i.i.d. term. One can hope that even if the pure characteristics model is correct, then the estimated scale parameter, μ, will be large enough to closely reproduce the implications of the true model. In Monte Carlo exercises, we did find this result when the data was "high quality" in the sense that the number of products was not large and the relative variance of the unobservable ξ was not too high.

If one prefers to impose the restriction of the pure characteristics model, then one could consider the modified BLP algorithm with the scale parameter μ fixed at a large value. This method leads to relatively easy computation using already-existing methods and we find in Monte Carlo exercises that it can work quite well. As compared to tradtional BLP, it also gains efficiency by estimating one fewer parameter. However, it is difficult to know *a priori* what value of μ is sufficiently large to provide a good approximation and in practice there are limits to how large a value of μ can be used before numeric errors start to arise.

Our most complicated computation method uses a homotopy argument to compute an accurate but slow value for the mean utility δ. For large numbers of products, it proved difficult to let that algorithm fully converge, but the method still provided good estimates, although at large computational cost.

We also had some success in estimating the model via either traditional or modified BLP methods, and then using the homotopy computation only to compute the implications of the model.

One conclusion, then, is that it would be strictly best to use the homotopy method when that is feasible, but that the other methods may also work well then the data quality is good. However, another practical suggestion is to simply estimate the model via traditional BLP, but then also compute the

predictions of the pure characteristics model via the homotopy method. At the least this provides a robustness check, and in some cases may provide a good approximation to the fully estimated pure characteristics model. The exact conditions under which this idea will work well is a good topic for further research.

References

Anderson, S., A. DePalma, and F. Thisse, *Discrete Choice Theory of Product Differentiation*, Cambridge MA: MIT Press, 1992.

Berry, Steven, "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, Summer 1994, *23* (2), 242–262.

_ , James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, July 1995, *60* (4), 889–917.

_ , _ , and _ , "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," *American Economic Review*, 1999, *89* (3), 189–211.

_ , _ , and _ , "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, February 2004, *112* (1), 68–105.

_ , Oliver Linton, and Ariel Pakes, "Limit Theorems for Differentiated Product Demand Systems," *Review of Economic Studies*, 2004, *71* (3), 613–614.

BoskinCommission, *Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index*, Senate Finance Committee, 1996.

Bresnahan, Timothy, "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War," *Journal of Industrial Economics*, 1987, *35*, 457–482.

Caplin, Andrews and Barry Nalebuff, "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, January 1991, *59* (1), 1–23.

Das, S., G. Steven Olley, and Ariel Pakes, "The Market for TVs," Technical Report, Yale Univeristy 1995.

Eaves, BC and K. Schmedders, "General equilibrium models and homotopy methods," *Journal of Economic Dynamics and Control*, 1999, *23* (9), 1249–1279.

Feenstra, Robert and James Levinsohn, "Estimating Markups and Market Conduct with Multidimensional Product Attributes," *Review of Economic Studies*, 1995, *62* (1), 19–52.

Gabszewicz, J.J. and J.-F. Thisse, "Price Competition, Quality and Income Disparity," *Journal of Economic Theory*, 1979, *20*, 340–359.

Goettler, Ronald L. and Roni Shachar, "Spatial Competition in the Network Television Industry," *RAND Journal of Economics*, Winter 2001, *32* (4), 624–56.

Greenstein, Shane M., "From Superminis to Supercomputers: Estimating the Surplus in the Computing Market," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods*, Chicago: University of Chicago Press, 1996.

Hausman, J.A. and D. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 1978, *46*, 403–426.

Hausman, Jerry, "Valuing the Effect of Regulation on New Services in Telecommunications," *The Brookings Papers on Economic Activity: Microeconomics*, 1997, pp. 1–38.

Hotelling, H., "Stability in Competition," *Economic Journal*, 1929, *39*, 41–57.

Lancaster, K., *Consumer Demand: A New Approach*, New York: Columbia University Press, 1971.

McFadden, D. and K. Train, "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 2000, *15* (5), 447–470.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka, ed., *Frontiers of Econometrics*, New York: Academic Press, 1974.

_ , "Econometric Models of Probabilistic Choice," in Charles Manski and Daniel McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press, 1981.

Mussa, M. and S. Rosen, "Monopoly and Product Quality," *Journal of Economics Theory*, 1978, *18*, 301–307.

Nevo, Aviv, "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *RAND Journal ofEconomics*, 2000, *31* (3), 395–421.

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 1986, *54*, 755–784.

_ , "A Reconsideration of Hedonic Price Indices with an Application to PC's," *American Economic Review*, 2003, *93* (5), 1578–1596.

_ **, Steven Berry, and James Levinsohn**, "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Price Indexes and the Analysis of Environmental Change," *American Economic Review, Paper and Proceedings*, May 1993, *83*, 240–246.

Petrin, Amil, "Quantifying the Benefits of New Products: The Case of the Minivan," *JPE*, August 2002, *110* (4), 705–729.

Salop, S., "Monopolistic Competition with Outside Goods," *RAND Journal of Economics*, 1979, *10*, 141–156.

Song, Minjae, "Measuring Consumer Welfare in the CPU Market: An Application of the Pure Characteristics Demand Model," Working Paper, Georgia Tech 2006. forthcoming *RAND Journal of Economics*.

Trajtenberg, M., "The Welfare Analysis of Product Innovations: with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 1989, *97*, 444–479.

Whitehead, G.W., *Elements of homotopy theory*, Springer-Verlag, 1993.