Econometrica, Vol. 63, No. 4 (July, 1995), 841–890

AUTOMOBILE PRICES IN MARKET EQUILIBRIUM

BY STEVEN BERRY, JAMES LEVINSOHN, AND ARIEL PAKES[1]

This paper develops techniques for empirically analyzing demand and supply in differentiated products markets and then applies these techniques to analyze equilibrium in the U.S. automobile industry. Our primary goal is to present a framework which enables one to obtain estimates of demand and cost parameters for a class of oligopolistic differentiated products markets. These estimates can be obtained using only widely available product-level and aggregate consumer-level data, and they are consistent with a structural model of equilibrium in an oligopolistic industry. When we apply the techniques developed here to the U.S. automobile market, we obtain cost and demand parameters for (essentially) all models marketed over a twenty year period.

KEYWORDS: Demand and supply, differentiated products, discrete choice, aggregation, simultaneity, automobiles.

1. INTRODUCTION

THIS PAPER DEVELOPS TECHNIQUES for empirically analyzing demand and supply in differentiated products markets and then applies these techniques to analyze equilibrium in the U.S. automobile industry. Our primary goal is to present a framework that enables one to obtain estimates of demand and cost parameters for a class of oligopolistic differentiated products markets. Estimates from our framework can be obtained using only widely available product-level and aggregate consumer-level data, and they are consistent with a structural model of equilibrium in an oligopolistic industry. When we apply the techniques developed here to the U.S. automobile market, we obtain cost and demand parameters for (essentially) all models marketed over a twenty year period. On the cost side, we estimate cost as a function of product characteristics. On the demand side, we estimate own- and cross-price elasticities as well as elasticities of demand with respect to vehicle attributes (such as weight or fuel efficiency). These elasticities, together with the cost-side parameters, play central roles in the analysis of many policy and descriptive issues (see, e.g., Pakes, Berry, and Levinsohn (1993) and Berry and Pakes (1993)).

Our general approach posits a distribution of consumer preferences over products. These preferences are then explicitly aggregated into a market-level demand system that, in turn, is combined with an assumption on cost functions and on pricing behavior to generate equilibrium prices and quantities. The

[1] We would like to thank Don Andrews, Tim Bresnahan, Gary Chamberlain, Zvi Griliches, Jerry Hausman, G. Mustafa Mohatarem, Whitney Newey, Frank Wolak, and the Economic Analysis group at the General Motors Corporation, as well as referees and a Co-Editor of this journal, for helpful comments. While working on this paper, Berry was an Olin Fellow at the NBER and Levinsohn was a National Fellow at the Hoover Institution, Stanford University. Each thanks his respective host. We gratefully acknowledge funding from National Science Foundation Grants SES-8821722 (to Richard Ericson and Ariel Pakes) and SES-9122672. Readers wishing a more extensive discussion of several issues in this paper are referred to our NBER working paper of the same title.

primitives to be estimated are parameters describing the firms' marginal costs and the distribution of consumer tastes. The distribution of tastes determines elasticities, and these, together with marginal cost and a Nash assumption, determine equilibrium prices.

A familiar alternative is to posit a simple functional form for the market-level demand system. This requires some aggregation over products, since, for example, a constant elasticity demand system for 100 products would require estimating 10,000 elasticities. This problem is frequently alleviated by aggregating and/or nesting products into groups, with justification often given by representative consumer theory. Even apart from the appropriateness of the implied restrictions, aggregation methods that might seem useful for one policy experiment are unlikely to be useful for another. For example, an applied researcher investigating tariffs might be tempted to aggregate all foreign and all domestic cars. However the resulting model is unlikely to prove useful when investigating domestic competition or pollution taxes. Further problems associated with the market-level demand approach include an inability to evaluate the impact of the introduction of new goods on demand and the difficulty of incorporating more micro information on the distribution of consumers into a representative agent framework.

One extensively used alternative to the market-level approach is a system that represents consumer preferences over products as a function of individual characteristics and of the attributes of those products (an approach that dates back at least to Lancaster (1971)). Advances in the discrete choice literature over the last two decades have generated much of the econometric methodology needed to use micro level data to estimate the parameters determining individual demands from this characteristics approach (e.g., McFadden (1973) and the literature he cites in his 1986 review article). Moreover a few studies have, by using convenient (but restrictive) assumptions, been able to aggregate the individual demands generated by this approach into a market-level demand system (e.g., Berkovec (1985), Morrison and Winston (1989)). Finally, there is a literature that integrates very simple discrete choice demand systems with an oligopolistic price setting model in a way that allows use of aggregate data to estimate the parameters of marginal cost and demand (Bresnahan (1987)).

We follow in this tradition, consider two problems that arise quite naturally in this framework, and provide computationally tractable methods for solving them. The first of the two problems concerns the imposed functional form of utility and the resulting pattern of cross-price elasticities. We show how, using only aggregate data, to interact consumer and product characteristics, thereby allowing for plausible substitution patterns. The second problem involves the correlation between prices, which are observed by the econometrician, and product characteristics, some of which are observed by the consumer but not by the econometrician, and the bias in estimated elasticities that this induces. This is just the differentiated products analog of the traditional simultaneous equations problem in homogeneous product markets (the classic reference being Working (1926)). The resulting estimation strategy involves solving an aggregation problem in moving from the individual to aggregate demands (solved via simulation, as suggested by Pakes (1986)), and solving a nonlinear simultaneous equations problem to account for endogenous prices (solved via an inversion routine as suggested by Berry (1994)). Both these techniques have precursors in the literature. McFadden, et al. (1977) use simulation to generate aggregate predictions from micro parameter estimates. Hotz and Miller (1993) use a related inversion technique to estimate a dynamic model, and Bresnahan (1987) allows prices to be correlated with a linear disturbance in an equilibrium pricing equation but does not explicitly model the correlation between prices and unobserved characteristics.

Because we rely on mostly aggregate data, we do not have the degrees of freedom associated with more micro-level studies. This naturally raises concerns about obtaining precise estimates of the parameters of interest. We have two suggestions for ameliorating any precision problems that may arise. First, we show how to use widely available data on the distribution of consumer characteristics to augment market level information. Second, we use recent results to describe and compute an approximation to the efficient instrumental variables estimator for our system (Chamberlain (1986), Newey (1990), and Pakes (1994)).

Our framework is based upon: (i) a joint distribution of consumer characteristics and product attributes that determines preferences over the products marketed; (ii) price taking assumptions on the part of consumers; and (iii) Nash equilibrium assumptions on the part of producers. This a very rich framework which we have not fully exploited. In particular, to generate our instruments we use a strong assumption on the orthogonality of observed and unobserved product characteristics. Though we think this is a natural starting place, it is an assumption that can be relaxed in future work. Relatedly, and perhaps more interesting, the framework is rich enough to incorporate nontrivial dynamics and endogenize the distribution of product attributes. We discuss these extensions in Section 8 below.

The Automobile Industry

Few industries have been studied as intensively as the auto industry and with good reason. With sales topping \$150 billion in 1989, the auto market is one of the largest in the U.S. and has ramifications for entire state economies. Moreover it is often at the heart of policy debates (in fields once as diverse as international trade and environmental regulation) and it is a market that has evolved in important ways.

Early work treated autos as a homogeneous product and estimated aggregate demand (e.g. Suits (1958)). Griliches (1971) and later work by Ohta and Griliches (1976) adopted the hedonic approach. Their work was among the first to consider the automobile market at the level of the individual product, a feature that set the tone for much future research (examples include Berkovec and Rust (1985), Toder et al. (1975), and Levinsohn (1988)). None of these studies gave much consideration to the production side of the model, although many of them used consumer micro data (a point we address below).

Perhaps the first attempt at simultaneously modeling and estimating the demand and oligopoly pricing sides of the market was Bresnahan's (1987) study. In that paper, Bresnahan adopted a vertical differentiation model and assumed a uniform density of consumers over the quality line. Feenstra and Levinshohn (1995) extend Bresnahan's work and allow products to be differentiated in multiple dimensions, but retain his assumption of the uniform density of consumers. Manski (1983) investigates the (perfectly competitive) supply side and demand side of the Israeli automobile market. Our goal is to estimate a model that allows for products that are differentiated in multiple dimensions, richer distributions of taste parameters, and unobserved (to the econometrician) product characteristics. We attempt to integrate and extend the advances in this literature, thereby taking a step towards a more detailed understanding of behavior in the auto market. (For a more detailed comparison to previous studies, see Section 2.3.)

A Road Map

The next two sections describe our theoretical model. Section 2 discusses utility and demand, while Section 3 models firm behavior and derives industry equilibrium. Section 4 introduces our instruments, Section 5 formally defines the estimators and describes their properties, while Section 6 provides the required computational techniques. The data and estimation results are discussed in Section 7. This section also provides a quick review of alternative models and compares our estimates to those of some alternative models. We conclude and discuss extensions in Section 8.

2. THEORY: UTILITY AND DEMAND

Our demand system is obtained by aggregating a discrete choice model of individual consumer behavior.[2] We then combine this demand system with a cost function, and embed these two primitives into a model of price setting behavior in differentiated products markets. The demand and pricing equations that this model generates give us the system of equations that we take to the data.

Most of this paper assumes that we do not have data that matches individual characteristics to the products those individuals purchased. Consequently we proceed (as does much of the prior literature on the empirical analysis of *equilibrium* in markets for differentiated products[3]) by considering the problem

[2] For background on demand systems obtained in this manner see McFadden (1981) and the literature cited there as well as the product differentiation literature cited in Shaked and Sutton (1982), Sattinger (1984), Perloff and Salop (1985), Bresnahan (1987), and Anderson, DePalma, and Thisse (1989), among others.

[3] For examples see Bresnahan (1987) and Feenstra and Levinsohn (1995).

of estimating all the parameters of the demand system from product level data (i.e. from information on prices, quantities, and the measurable characteristics of the products). We then extend the discussion to allow for the possibility of incorporating exogenous (and frequently available) information on the *distribution* of individual characteristics (e.g., the distribution of income and/or family size). Only in the extensions section do we come back to the advantages of having data that matches consumer characteristics to the products those consumers purchased.

Our specification posits that the level of utility that a consumer derives from a given product is a function of both a vector of individual characteristics, say ζ, and a vector of product characteristics, say (x, ξ, p). Here p represents the price of the product, and x and ξ are, respectively, observed and unobserved (by the econometrician) product attributes. That is, the utility derived by consumer i from consuming product j is given by the scalar value

$$U(\zeta_i, p_j, x_j, \xi_j; \theta),$$

where θ is a k-vector of parameters to be estimated.

Consumers with different ζ make different choices, and to derive the aggregate demand system we integrate out the choice function over the distribution of ζ in the population. Throughout we will take ζ to have a *known* distribution. This distribution may either be the *empirical* distribution of a characteristic, or a standardized distribution whose standardization parameters are estimated (unit normals for example, whose standardization parameters are a mean vector and covariance matrix). For notational simplicity, we will let θ include any parameters determining the distribution of consumer characteristics, as well as the parameters that describe the utility surface conditional on these characteristics.

Consumer i chooses good j if and only if

$$U(\zeta_i, p_j, x_j, \xi_j; \theta) \geq U(\zeta_i, p_r, x_r, \xi_r; \theta), \qquad \text{for} \qquad r = 0, 1, \dots, J,$$

where alternatives $r = 1, \dots, J$ represent purchases of the competing differentiated products. Alternative zero, or the outside alternative, represents the option of not purchasing any of those products (and allocating all expenditures to other commodities). It is the presence of this outside alternative that allows us to model aggregate demand for autos as a function of prices and auto characteristics. Let

$$(2.1) \qquad A_j = \{\zeta : U(\zeta, p_j, x_j, \xi_j; \theta) \geq U(\zeta, p_r, x_r, \xi_r; \theta), \text{for } r = 0, 1, \dots, J\}.$$

That is, A_j is the set of values for ζ that induces the choice of good "j". Then, assuming ties occur with zero probability, and that $P_0(d\zeta)$ provides the density of ζ in the population, the market share of good "j" as a function of the characteristics of all the goods competing in the market is given by

$$(2.2) \qquad s_j(p, x, \xi; \theta) = \int_{\zeta \in A_j} P_0(d\zeta).$$

Denote the J-element vector of functions whose "jth" component is given by (2.2) as $s(\cdot)$. Then, if M is the number of consumers in the market, the J-vector of demands is $Ms(p, x, \xi; \theta)$.

2.1. *Functional Forms and Substitution Patterns*

This subsection begins by discussing alternative functional forms for the consumer decision problem and then aggregates over consumers to obtain market demand.

A special case of the model in (2.1) and (2.2) is

$$(2.3) \qquad U(\zeta_i, p_j, x_j, \xi_j; \theta) \equiv x_j \beta - \alpha p_j + \xi_j + \epsilon_{i,j} \equiv \delta_j + \epsilon_{ij},$$

where

$$\delta_j = x_j \beta - \alpha p_j + \xi_j,$$

and the mean of the ϵ vector in the population of consumers is assumed to be zero so that for each j, ξ_j is the mean (across consumers) of the unobserved component of utility, $f(\nu_i, \zeta_j)$, while δ_j is the mean of the utility from good j. In (2.3), the ϵ's are the only elements of the vector of consumer characteristics, ζ.

This specification is particularly tractable if the unobserved characteristic $\xi_j = 0$ and the vector ϵ_{ij} is distributed independently across both consumers and products. Note that this implies that the distribution of ϵ_{ij} is independent of the observed characteristics, x_j. The tractability of combining (2.3) with an i.i.d. assumption on the distribution of the ϵ's follows from the ease of computing market shares from

$$(2.4) \qquad s_j = \int_{\epsilon} \prod_{q \neq j} P(\delta_j - \delta_q + \epsilon) P(d\epsilon).$$

Equation (2.4) shows that this computation requires, at most, evaluating a unidimensional integral. We note that if the ϵ are distributed multivariate extreme value (the logit model) then there is a closed form for (2.4) and there is no need to compute any integral.

Despite this computational simplicity, the assumption that the utility function is additively separable into two terms, one determined entirely by the product characteristics (the δ_j in (2.3)) and one determined by the consumer characteristics (the ϵ_{ij} in (2.3)), is problematic. This is because (2.3) generates aggregate substitution patterns, and hence a set of (cross and own) price derivatives, as well as responses to the introduction of new products, that *cannot* possess many of the features that we expect them to have.

Before considering the implications of (2.3) in detail, it may be worthwhile to note that the additive separability assumption just discussed is stronger than the assumptions used in many models of individual consumer behavior. These models often assume i.i.d. additive utility errors, an assumption that has been critiqued extensively in the literature, at least since Debreu's (1960) discussion of the "independence of irrelevant alternatives" property in the logit model.

However, consumer level studies *do* often interact observed consumer characteristics with product characteristics. These interactions mean that market shares do not take the simple form of (2.4) and hence do not have the unnatural implications on demand patterns which we now discuss.

An implication of (2.3) is that all substitution effects depend only on the δ_j's. Since there is a unique vector of market shares associated with each δ-vector, the additively separable specification implies that the cross-price elasticities between any two products, or, for that matter, the similarity in their price and demand responses to the introduction of a new third product, depend only on their market shares. That is, conditional on market shares, substitution patterns do not depend on the observable characteristics of the product.

Thus, if we were using the specification in (2.3) to analyze an automobile market in which an inexpensive Yugo and an expensive Mercedes had the same market shares, then the parameter estimates would have to imply that the two cars have the same cross-price derivative with respect to *any* third car. In particular, the model would necessarily predict that an increase in the price of a BMW would generate equal increases in the demand for Yugos and for Mercedes. This contradicts the intuition which suggests that couples of goods whose characteristics are more "similar" should have higher cross-price elasticities. We expect this to happen because the consumers who would have chosen a BMW at the old prices, but now do not, have a preference for large cars and are therefore likely to move to another large car. Similarly, when a new car enters the market, we expect it to have a large effect on the demand for cars with similar characteristics. Additive separability plus i.i.d. ϵ's, on the other hand, imply that a consumer who substitutes away from any given choice will tend to substitute toward other popular products, not to other similar products. Note that this does not depend on any specific distribution for the ϵ's (e.g. logit).

For analogous reasons, the specification in (2.3) implies that two products with the same market share will have the same own-price demand derivatives. For example, if a Jaguar and a Yugo have the same market share, the specification in (2.3) implies that they must have the same own-price derivative. In an oligopoly context, this is troubling for it implies (assuming single-product firms) that the two products must have the same markup over marginal cost. Intuitively, however, we expect markups to be determined by more than market shares. They ought also to be determined by the number of competing products that are "close" in product space, and, because consumers who buy more expensive goods are likely to have lower marginal utilities of income, by the price of the product.[4]

We now consider ways of allowing for interaction between individual and product characteristics. A familiar starting point is to allow each individual to

[4] For earlier discussions of the implications of related specifications on aggregate demand patterns see the Appendix to Hausman (1987), McFadden (1981), and Schmalensee (1985). Note also that (2.4) assumes more than additive separability. It also assumes that δ_j is a linear function of product characteristics and that the distributions of the ϵ_{ij} are identical across "j"; but these assumptions are primarily for expositional simplicity. They can be relaxed with only minor modifications to the discussion that follows (see below).

have a different preference for each different observable characteristic. This generates the traditional random coefficients model

$$(2.5) \qquad U(\zeta_i, p_j, x_j, \xi_j; \theta) = x_j \bar{\beta} - \alpha p_j + \xi_j + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij},$$

where $(\zeta_i, \epsilon_i) = (\nu_{i1}, \nu_{i2}, \dots, \nu_{iK}, \epsilon_{i0}, \epsilon_{i1}, \dots, \epsilon_{iJ})$ is a mean zero vector of random variables with (a known) distribution function. Now the contribution of x_k units of the kth product characteristic to the utility of individual i is $(\bar{\beta}_k + \sigma_k \nu_{ik}) x_k$, which varies over consumers. We scale ν_{ik} such that $E(\nu_{ik}^2) = 1$, so that the mean and variance of the marginal utilities associated with characteristic k are $\bar{\beta}_k$ and σ_k^2 respectively.[5] This specification is particularly tractable if ϵ_i consists of i.i.d. extreme value deviates.

The utility obtained from consuming good j can still be decomposed into a mean

$$\delta_j = x_j \bar{\beta} - \alpha p_j + \xi_j$$

and a deviation from that mean

$$\mu_{ij} = \sum_k \sigma_k x_{jk} \nu_{ik} + \varepsilon_{ij},$$

but now μ_{ij} depends on the interaction between consumer preferences and product characteristics. As a result, consumers who have a preference for size will tend to attach high utility to all large cars, and this will induce large substitution effects between large cars.

Note, however, that though this specification allows for more realistic cross-price elasticities, it re-introduces the problem of computing the integral (in 2.2) that defines market shares as a function of the parameters of the model. We solve this computational problem via aggregation by simulation, a technique introduced by Pakes (1986).

Though familiar, the random coefficients specification in (2.5) is not really suitable for our purposes. We prefer a specification that makes it easy for us to incorporate prior information on both the distribution of the relevant consumer characteristics, and on the functional form of the interaction between those characteristics and product attributes. This is because we have additional information on the distribution of income across households, and a theoretical rationale for the form of the interaction between income and price.

To this end, we now nest the random coefficients specification into a Cobb-Douglas utility function in expenditures on other goods and services and characteristics of the good purchased:

$$(2.6) \qquad U(\zeta_i, p_j, x_j, \xi_j; \theta) = (y_i - p_j)^{\alpha} G(x_j, \xi_j, \nu_i) e^{\epsilon(i,j)},$$

[5] We will assume that the distribution of the $[\nu(i,1), \dots, \nu(i,K)]$ factors into a product of independent densities. This is for expositional convenience; with the addition of some notation we could easily allow for patterns of correlation among them.

where y is income, and ϵ provides the effect of the interactions of unobserved product and individual characteristics.

In our empirical example we assume that $G(\cdot)$ is linear in logs and has the random coefficient specification discussed above, so that if $u_{ij} = \log[U_{ij}]$, then

$$(2.7a) \qquad u_{ij} = \alpha \log(y_i - p_j) + x_j \bar{\beta} + \xi_j + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij},$$

for $j = 1, \ldots, J$, while

$$(2.7b) \qquad u_{i0} = \alpha \log(y_i) + \xi_0 + \sigma_0 \nu_{i0} + \epsilon_{i0}.$$

Note, first, that our current data set does not have information on differences in the value of the outside alternative (differences that would be generated by, among other factors, differences in access to public transportation and differences in used car holdings). Thus, to account for the possibility that there is more unobserved variance in the idiosyncratic component of the outside than of the inside alternatives, we allow for an extra unobserved term in the determination of u_{i0} (the ν_{i0}).[6]

Second, note that the consumer terms that interact with product characteristics are now

$$\nu_i = (y_i, \nu_{i1}, \ldots, \nu_{iK}).$$

We have used special notation for income here both because it enters the utility function in a special way, and because it is a variable whose distribution can be estimated from the March Current Population Survey. As a result, if one assumes a parametric form for the distribution of ν_i conditional on y_i, we can use the CPS to determine the distribution of y_i in our population and reduce the number of parameters that are estimated from our auto data.

Two characteristics of (2.7) are central to the rest of this paper: it allows for interactions between consumer and product characteristics and it allows us to make use of exogenous data on the distribution of income in a natural and parsimonious way. The first characteristic enables us to model reasonable substitution patterns, while the second allows us to get more precise parameter estimates.[7]

2.2. *Endogenous Prices*

If producers know the values of the unobserved characteristics, ξ, even though we do not, then prices are likely to be correlated with them. This

[6] Note that since market shares depend only on differences in utilities, the actual estimation algorithm ends up subtracting the $u(i,0)$ in (2.7b) from the $u(i,j)$, and estimating a model where the outside alternative is "normalized" to zero. Given (2.7b), this implies there is a random coefficient on the constant term in the utility function for the inside goods.

[7] There are also a number of restrictive assumptions in (2.7), including both the decomposition of the interaction of unobserved individual and product characteristics into $\xi_j + \epsilon_{ij}$ with the ϵ_{ij} i.i.d over both i and j, and the separability implied by log-linearity. We are exploring some of these restrictions in related work using more disaggregated data.

generates a differentiated products analog to the classic simultaneity problem in the analysis of demand and supply in homogeneous product markets.[8] The simultaneity problem is complicated by both the discrete choice set for each individual and the interaction of individual and product characteristics. These together make aggregate demand a complicated nonlinear function of product characteristics. Berry (1994) suggests one approach to obtaining estimates of the demand parameters, and proves its viability under certain restrictions. This subsection begins by discussing the importance of unobserved demand characteristics and the resulting endogeneity of prices, then reviews Berry's approach, and finally extends it to allow for the random coefficients in (2.7). The computation section will provide a contraction mapping that allows us to compute the unobserved components and hence use them in estimation.

Much empirical work on discrete choice models of demand (both aggregate and consumer level studies) has specified that the unobserved component in the utility function for each alternative is mean zero and independent across agents.[9] This specification assumes away the simultaneity problem. It also leads to an embarrassing "over fitting" problem on aggregate data (see, for example, Toder et al. (1978)). That is, if there is no "structural" disturbance in the market share equation, then only sampling error can explain differences between the data and the predictions of the model. For sample sizes as large as those typically found in aggregate studies, this variance is just too small to account for any noticeable discrepancy between the data and the model (so that a χ^2 test of the model's restrictions on the multinomial proportions is rejected with probability close to one).[10]

In contrast, aggregate demand in homogeneous product markets is typically specified to have a nonzero disturbance that is generally associated with unobserved determinants of demand that are correlated across consumers in a market. If these disturbances are known to the producers and the consumers (and if demand depends upon them, one expects this to be so), and if there is any equilibrating mechanism in the market, then equilibrium quantities and prices will depend upon the disturbances. It is this relationship between the disturbance and price that generates the simultaneity problem and the need for alternatives to ordinary least squares estimation techniques.

All the utility specifications in the last subsection had disturbances with a product specific mean, ξ, which is the analog of the disturbance in the aggregate demand system in homogeneous product markets. In the automobile example, ξ reflects the difficult to quantify aspects of style, prestige, reputation,

[8] For a history of the econometrics of demand and supply analysis in homogeneous product markets, see Morgan (1990, Ch. 2).

[9] One exception is Berry's (1991) study of airline hubbing, which includes an aggregate market-specific demand error that is correlated with prices. However, that paper uses a very restrictive functional form for utility.

[10] Similar overfitting phenomena have been a source of concern in the biometrics literature for some time; see, for example, Hasemand and Kupper (1979) or Williams (1982). Though they do not worry about simultaneity, their conceptual solution to the overfitting problem is similar to the one we shall use (allowing for unobserved determinants of the cell probabilities).

and past experience that affect the demand for different products, as well as the effects of quantifiable characteristics of the car that we simply do not have in our data. As one might expect, the introduction of ξ will alleviate the overfitting problem. However, our primary concern is that if unobserved characteristics are important, and our data indicate that they are, prices will be correlated with them, and the estimates of price effects will be biased. This is precisely the same logic that leads to biased O.L.S. estimates of price effects in traditional demand systems.

As in traditional homogeneous goods models, we will assume that ξ is mean independent of some set of exogenous instruments and then derive estimators from the orthogonality conditions those assumptions imply. This procedure requires only the same assumptions needed for instrumental variable estimators of demand parameters in homogeneous product markets. In particular we do not require an explicit assumption on the distribution of the ξ, just that they be mean independent of the instruments. Furthermore, the procedure does not depend on the exact form of the pricing rule. On the other hand, since the pricing rule depends in equilibrium on the true values of the demand parameters, joint estimation of the pricing and demand equations should increase efficiency as long as the model is correctly specified.

The difference between our case and the homogeneous product case is that the demand of a given individual, and hence market demand, becomes a nonlinear function of the ξ; i.e. $q_j = Ms_j(x, \xi, p; \theta)$. Consequently the orthogonality between ξ and the x-vector cannot be used for estimation without first transforming the observed quantity, price, and characteristic data into a linear function of ξ. It is this transformation that is the focus of Berry's (1994) paper and we return to it in the computational section. There remains the important issue of the choice of instruments, an issue we come back to after describing the pricing equation.

2.3. *Previous Approaches to Demand Estimation*

Variations on the logit model, discussed at length above, have often formed the basis for micro-data studies of the automobile industry (that is, studies that match consumers to the cars they purchased). The authors of those studies frequently have been aware of the problems that we discuss here: the endogeneity of prices and the need to generate reasonable substitution patterns. With micro-data, there were alternatives to our proposed solutions. In particular, it is possible to interact product characteristics with *observed* consumer characteristics and many studies have done so (for example, Berkovec (1985)). Also, there is a possibility of using nested logit, which in our framework can be shown to be a restricted version of a model with random coefficients on a set of dummy variables that define groups (or "nests") of products (Ben-Akiva (1973), McFadden (1978), and Berkovec and Rust (1985), Goldberg (1993)). Note that this requires a priori information on the order and the contents of the nests. Finally, given recent advances in simulation methodology, one could use a random coefficients specification similar to ours and simulate choice probabilities. Note that the ν_i in the micro model (equation 2.7) underlying our aggregate specification could potentially reflect any combination of observed and unobserved consumer characteristics.

Micro-data does not by itself solve the problem of unobserved product characteristics that are correlated with prices. It does, however, allow one to introduce product-specific dummies to control for unobserved attributes. These dummies correspond to our δ_j's. Note first that this approach runs into an efficiency problem due to the relatively large number of automobile models. For example, there are on average more than 100 products in a given year of our sample, a number that might be compared to the approximately 500 new car purchases observed annually in the Consumer Expenditure Survey (which is on the order of the largest publicly available survey that includes detailed information on automobile purchases).[11] Thus, it is not surprising that we do not know of a study of automobile demand that estimates choice-specific constants, except when choices are artificially aggregated into a small number of alternatives, such as small, medium, and large cars (for more detail, see Train (1986) and his review of the literature on estimating auto demand). In addition, even if product specific dummies could be estimated, these dummies will contain the linear utility components of product characteristics and prices (as in our equation (2.3)). Therefore, to calculate price and characteristic elasticities we would need to separate out the effects of price, x and ξ on the product specific constants. This separation requires additional assumptions—the sort of assumptions that we make here to justify our instrumental variable approach.

With only aggregate data, previous authors have adopted other specifications for utility. In his study, Bresnahan (1987) adopts a pure vertical differentiation model (Shaked and Sutton (1982)). In this model, there is only one characteristic, the marginal valuation of either price or "quality," that varies across consumers. This greatly restricts substitution patterns. In particular, the pattern of cross-price elasticities is determined exclusively by market shares and the rank-order of prices, not by the value of other product characteristics such as size, power, etc. Products have nonzero cross-price elasticities only with the two other products that are adjacent to it in the ranking of prices. Consider, for example, the possibility that the price ranking contains, in order, a \$24,998 family station wagon, a \$24,999 sports car, and another family station wagon priced at \$25,000. In this case the vertical model guarantees that the wagons are *not* substitutes for one another, but that the sports car is. A solution to this is to allow products to be differentiated in multiple dimensions. Feenstra and Levinsohn (1995) adopt this approach while maintaining the rest of Bresnahan's framework.

[11] This suggests the possible advantages of combining consumer and market-level data, an approach that we are currently pursuing.

3. COST FUNCTIONS AND THE PRICING PROBLEM OF THE MULTIPRODUCT FIRM

We take as given that there are F firms, each of which produce some subset, say $\mathcal{F}_f$, of the J products. For simplicity we begin by assuming that the marginal cost of producing the goods marketed is both independent of output levels and log linear in a vector of cost characteristics. These assumptions are made only for expositional convenience and we relax them in our investigation of the robustness of our empirical results.

The cost characteristics are decomposed into a subset which are observed by the econometrician, the vector w_j for model j, and an unobserved component, ω_j. Note that we might expect the observed product characteristics, the x_j, to be part of the w_j, and ω_j to be correlated with ξ_j. This is because larger cars, or cars with a larger unobserved quality index, might be more costly to produce, a possibility we will account for in our estimation algorithm.

Given these assumptions the marginal cost of good j, say mc_j, is written as

$$(3.1) \qquad \ln(mc_j) = w_j\gamma + \omega_j,$$

where γ is a vector of parameters to be estimated.

Given the demand system in (2.1) and (2.2), the profits of firm f, say Π_f, are

$$(3.2) \qquad \Pi_f = \sum_{j \in \mathcal{F}_f} (p_j - mc_j) M s_j(p, x, \xi; \theta),$$

with mc_j given by (3.1). Each firm is assumed to choose prices that maximize its profit given the attributes of its products and the prices and attributes of competing products.[12]

Given our assumptions, any product produced by firm f, or any $j \in \mathcal{F}_f$, must have a price, p_j, that satisfies the first order conditions

$$(3.3) \qquad s_j(p, x, \xi; \theta) + \sum_{r \in \mathcal{F}_f} (p_r - mc_r) \frac{\partial s_r(p, x, \xi; \theta)}{\partial p_j} = 0.$$

The J first-order conditions in (3.3) imply price-cost markups $(p_j - mc_j)$ for each good. To obtain these, define a new J by J matrix, Δ, whose (j, r) element is given by:

$$(3.4) \qquad \Delta_{jr} = \begin{cases} \dfrac{-\partial s_r}{\partial p_j}, & \text{if } r \text{ and } j \text{ are produced by the same firm;} \\ 0, & \text{otherwise.} \end{cases}$$

[12] We assume that a Nash equilibrium to this pricing game exists, and that the equilibrium prices are in the interior of the firms' strategy sets (the positive orthant). While Caplin and Nalebuff (1991) provide a set of conditions for the existence of equilibrium for related models of single product firms, their theorems do not easily generalize to the multiproduct case. However, we are able to check numerically whether our final estimates are consistent with the existence of an equilibrium. Note that none of the properties of the estimates require uniqueness of equilibrium, although without uniqueness it is not clear how to use our estimates to examine the effects of policy and environmental changes.

In vector notation the first order conditions can then be written as

$$s(p,x,\xi:\theta)-\Delta(p,x,\xi;\theta)[p-mc]=0.$$

Solving for the price-cost markup gives

$$p=mc+\Delta(p,x,\xi;\theta)^{-1}s(p,x,\xi;\theta).$$

Note that prices are additively separable in marginal cost and the markup defined as

$$b(p,x,\xi;\theta)\equiv\Delta(p,x,\xi;\theta)^{-1}s(p,x,\xi;\theta). \tag{3.5}$$

The vector of markups in (3.5) depends only on the parameters of the demand system and the equilibrium price vector. However, since p is a function of ω, $b(p,x,\xi;\theta)$ is a function of ω, and cannot be assumed to be uncorrelated with it (the correlation of ξ with ω also generates a dependence between the markups and ω). Substituting in the expression for marginal cost, we obtain the pricing equation we take to the data:

$$\ln(p-b(p,x,\xi;\theta))=w\gamma+\omega. \tag{3.6}$$

Just as in estimating demand, estimates of the parameters of (3.6) can be obtained if one assumes orthogonality conditions between ω and appropriate instruments. We now move on to a discussion of appropriate instruments.

4. INSTRUMENTS

We need to specify instruments for both the demand and pricing equations. Any factors that are correlated with specific functions of the observed data, but are not correlated with the demand or supply disturbances, ξ and ω, will be appropriate instruments. Our procedure is to specify a list of variables that are mean independent of ξ and ω and use the logic of the estimation procedure to derive appropriate instruments.

Our mean independence assumption is that the supply and demand unobservables are mean independent of both observed product characteristics and cost shifters. Formally, if $z_j=[x_j,w_j]$ and $z=[z_1,\dots,z_J]$, then

$$E\left[\xi_j|z\right]=E\left[\omega_j|z\right]=0. \tag{4.1}$$

Note first that we do not include price or quantity in the conditioning vector, z. This is because our model implies that price and quantity are determined in part by ξ and ω. In contrast, we do not model the determination of product characteristics and cost shifters.

On the other hand, one might think that there is a "true" underlying model which jointly determines both observed and unobserved product characteristics. Assumption (4.1) will only be correct if that model has very specific properties. It is relatively easy to formulate other assumptions that would also formally identify the model. For example, with a panel data set such as ours, one could assume that the ξ's and ω's of a given auto model evolve as a first-order Markov process, with the innovation in that process independent of the auto's initial characteristics. It is possible to modify our procedures to use this assumption as the basis of our estimation algorithm. However, this modified procedure would be much more demanding of the data. As a result, we started with the simpler assumption in (4.1).

Given (4.1) and some additional regularity conditions, we show in Section 5 that the model generates an optimal set of instruments. While those instruments are hard to compute, we suggest an approximation to them. It is important to realize that the instruments associated with product j include functions of the characteristics and cost shifters of all other products. The intuition here follows from a natural feature of oligopoly pricing: products that face good substitutes will tend to have low markups, whereas other products will have high markups and thus high prices relative to cost. Similarly, because Nash markups will respond differently to own and rival products, the optimal instruments will distinguish between the characteristics of products produced by the same multi-product firm versus the characteristics of products produced by rival firms. Similar intuition has been used to motivate identification assumptions in several previous models, e.g. Bresnahan (1987).

Given the fact that demand for any product is, via the functional form of the demand system, a function of the characteristics of all products, our instruments cannot rely on "exclusion" restrictions. However, in our model the utility of consuming product j depends only on the characteristics of that product. Given this restriction, it is natural that the number of utility parameters grows with the dimension of the product characteristics space and not with the number of products. For example, if we approximated utility via a polynomial in characteristics, the number of utility parameters would be determined solely by the order of the polynomial and the number of characteristics. This restriction, combined with specific functional form and distributional assumptions, is what allows us to identify the demand system even in the absence of cost shifters that are excluded from the x vector.[13]

We turn now to a formal description of the estimation algorithm.

5. THE ESTIMATION ALGORITHM

To keep the exposition simple, we begin by maintaining some simplifying assumptions that we later remove. In particular, although we will actually use panel data, we start by assuming that our data consist of a single cross section of the autos marketed in a given year. If J is the number of autos marketed, the data set then contains J vectors (x_j, w_j, p_j, q_j), and a number of households sampled, n, which, when combined with the information on purchases, can be used to compute the share of the outside alternative. Thus, the observed vector of sampled market shares, denoted s^n, belongs to the $J+1$ dimensional unit simplex. (This includes the share of the outside alternative).

[13] Note, however, that we use identification in an informal sense; a formal identification argument requires further regularity conditions.

The assumptions on the data generating process are as follows. Market shares are calculated from the purchases of a random sample of n consumers from a population with a distribution of characteristics, ν, given by $P_0(\cdot)$. This population abides by the model's decision rules at $\theta = \theta_0$. Letting s^0 denote the vector of shares in the underlying population, the multinomial sampling process implies that s^n converges to s^0 at rate $\sqrt{n}$, or $(s^n - s^0) = O_p(1/\sqrt{n})$. The $(\xi_j, \omega_j, x_j, w_j)$ vectors that characterize the primitive product characteristics are independent draws from some larger population of possible characteristic vectors.[14] The distribution of these vectors in this population has the mean independence property of (4.1), namely that $E[\xi_j|z] = E[\omega_j|z] = 0$.[15] We also assume that

$$E\left[(\xi_j, \omega_j)'(\xi_j, w_j)|z\right] = \Omega(z_j), \tag{5.1}$$

with $\Omega(z_j)$ finite for almost every z_j.

The logic behind the estimation procedure is simple enough. Appendix I shows that given the data on the prices and the observed characteristics of the products, *any* choice of a triple consisting of an observed vector of positive market shares, say s, a distribution of consumer characteristics, say P, and the parameters of the model, say θ, implies a unique sequence of estimates for the two unobserved characteristics of our products, say $\{(\xi_j(\theta, s, P), \omega_j(\theta, s, P)\}_{j=1}^J$. Assume, temporarily, that we can actually calculate $\{(\xi_j(\theta, s^0, P_0), \omega_j(\theta, s^0, P_0)\}_{j=1}^J$ for alternative values of θ. In fact, we do not actually observe s^0 (though we do observe s^n), and for most of the models we consider we cannot actually compute the disturbances generated by P_0, but rather only from a (simulation) estimator of it. So our actual estimation procedure will be based on substituting estimates of s^0 and of P_0 into the algorithm we now develop.

Assuming we can compute $\{\xi_j(\theta, s^0, P_0), \omega_j(\theta, s^0, P_0)\}$, then at $\theta = \theta_0$ our computation will reproduce the true values of the unobserved car characteristics. Consequently, the conditional moment restrictions in (4.1) imply that any function of z must be uncorrelated with the vector $\{\xi(\theta, s^0, P_0), \omega(\theta, s^0, P_0)]\}$

[14] In fact all we require is that the draws on $y_j = (\xi_j, \omega_j, z_j)$ be exchangeable draws from some population. That is, if the joint distribution of $\{y_j\}$ is $f[J](\cdot)$, then we require that

$$f[J][y_1, \ldots, y_J] = f[J][y_{\pi(1)}, \ldots, y_{\pi(J)}]$$

for any permutation $[\pi(1), \ldots, \pi(J)]$ of $[1, \ldots, J]$. A reason for using this assumption (rather than the more restrictive assumption of independence) is to allow the (at least in part, chosen) characteristics of a product to be related to the characteristics of other products, and to allow for the outcomes of environmental processes that are likely to affect many products. The assumption that we can permute the y vector without changing our model [i.e., $f(\cdot)$] amounts to assuming that the y-vectors include all characteristics that are determinants of the choices made (a strong, but not unfamiliar, assumption in applied work, especially given our allowance for the unobservables, ξ and ω).

[15] In reference to the representation in the last footnote, we note that exchangeability implies the existence of a random variable, say $q(J)$, and distribution functions, say $g[J](\cdot)$ such that the $\{y_j\}$ are independent conditional on (the "aggregate") random variable $q(J)$ or $f[J](y_1, \ldots, y_j) = \Pi_j g[J](y_j|q)$ (see Kingman (1978)). One can place (different sets of) restrictions on this representation that imply (4.1) and (5.1), though this returns us to the discussion in Section 4.

when that vector is evaluated at $\theta = \theta_0$. As in Hansen (1982), we can use this fact to generate a method of moments estimator of θ_0. That is, we can form the sample analog to some set of covariance restrictions and find that value of θ that sets this sample analog "as close as possible" to zero (see below).

To be more precise let $T(z_j)$ be a 2 by 2 matrix of functions of z_j, and $H_j(z)$ be an L by 2 matrix of functions of z (the j index here indicates that the function may differ with the observation). The matrix $T(\cdot)$ is introduced to standardize $[\xi(\theta_0), \omega(\theta_0)]$; so we will assume that

$$(5.2) \qquad T(z)'T(z) = \Omega(z)^{-1}.$$

$H_j(\cdot)$ is a matrix of instruments for two standardized disturbances. Now define

$$(5.3) \qquad G^J(\theta) = E\left[H_j(z)T(z_j) \begin{pmatrix} \xi_j(\theta, s^0, P_0) \\ w_j(\theta, s^0, P_0) \end{pmatrix} \right]$$

and note that (4.1) guarantees $G^J(\theta_0) = 0$. So form

$$(5.4) \qquad G_J(\theta; s^0, P_0) = \frac{1}{J} \sum_{j=1}^{J} H_j(z)T(z_j) \begin{pmatrix} \xi_j(\theta, s^0, P_0) \\ \omega_j(\theta, s^0, P_0) \end{pmatrix},$$

and choose, as an estimate of θ, the value that minimizes, up to a term of $o_p(1/\sqrt{J})$,

$$\left\| G_J(\theta; s^0, P_0) \right\|,$$

where for any vector y, $\|y\| = y'y$.

We need to account for the fact that we cannot actually compute the moment conditions, $G_J(\theta; s^0, P_0)$, needed to minimize the objective function. There are two separate problems here. The first is that we do not observe s^0 but just s^n, so for any P we actually calculate $G_J(\theta, s^n, P)$. Second, for most of our models we will not be able to calculate $G_J(\theta; s, P_0)$ explicitly but will have to suffice with a simulation estimator of it. We show in Section 6 that this is equivalent to using $G_J(\theta; s, P_{ns})$ where P_{ns} provides the empirical distribution of ns simulation draws from P_0.[16] Consequently, the objective function that our estimator θ minimizes is

$$(5.5) \qquad \left\| G_J(\theta, s^n, P_{ns}) \right\|.$$

In a separate paper, Berry and Pakes (in process) provide conditions that insure that our estimate is consistent and asymptotically normal. Three problems arise in deriving the limiting properties of this estimator. First, the interdependence implicit in the demand system generates dependence in the quantities that we average over to form moment conditions. Indeed these quantities are not mean independent at values of θ different from θ_0, so consistency requires us to bound the moment conditions away from zero

[16] Actually for increased efficiency we use an importance sampling simulator; see equation (5.3).

uniformly for θ different from θ_0. Given consistency, asymptotic normality follows from mean independence and smoothness of the objective function at θ_0. Second, the quantities entering the moment conditions are nonlinear functions of the disturbances generated by the consumer sampling and simulation processes. As a result, consistency requires both the number of simulation draws, *ns*, and the size of the consumer sample, n, to grow large. In addition, both the consumer sampling and simulation processes generate disturbances whose effects on the variance of our parameter estimates we want to quantity. Third, as J goes to infinity all but a finite number of the choice probabilities must go to zero, which makes it particularly difficult to evaluate the impact of the simulation and sampling errors on the inverse market share function that defines ξ. To accommodate these last two points more detailed assumptions must be made on the rate at which n and ns grow with respect to J.

The covariance matrix, provided in Berry and Pakes (in process), for our estimator is[17]

$$(5.6)\qquad (\Gamma'\Gamma)^{-1}\Gamma'\left(\sum_{i=1}^{3} V_i\right)\Gamma(\Gamma'\Gamma)^{-1}.$$

Here

$$\Gamma = \lim_{J\to\infty} \frac{\partial E[G_J(\theta, s^0, P_0)]}{\partial\theta'}\bigg|_{\theta=\theta_0},$$

while if

$$V_1^J = E_z\left[H_j(z)T(z_j)\begin{pmatrix}\xi_j(\theta_0, s^0, P_0)\\ \omega_j(\theta_0, s^0, P_0)\end{pmatrix}\begin{pmatrix}\xi_j(\theta_0, s^0, P_0)\\ \omega_j(\theta_0, s^0, P_0)\end{pmatrix}' T(z_j)'H_j(z)'\right],$$

$$V_2^J = \frac{J}{n}E\Big(\sqrt{n}\,[G_J(\theta_0, s^0, P_0) - G_J(\theta_0, s^n, P_0)] \times\sqrt{n}\,[G_J(\theta_0, s^0, P_0) - G_J(\theta_0, s^n, P_0)]'\Big),$$

and

$$V_3^J = \frac{J}{ns}E\Big(\sqrt{ns}\,[G_J(\theta_0, s^n, P_{ns}) - G_J(\theta_0, s^n, P_0)] \times\sqrt{ns}\,[G_J(\theta_0, s^n, P_{ns}) - G_J(\theta_0, s^n, P_0)]'|s^n\Big),$$

then

$$V_1 = \lim_{J\to\infty} V_1^J,\qquad V_2 = \lim_{J\to\infty} V_2^J,\qquad \text{and}\qquad V_3 = \lim_{J\to\infty} V_3^J.$$

[17] Berry and Pakes (in process) provide expressions for V_2 and V_3 in terms of the model's primitives. The conditions in their paper include an identification condition, conditions which insure the existence of limits, a condition on the form of the covariance matrix for a single draw of the simulation process as J goes to ∞, conditions on the rate at which the derivatives of the market share vector with respect to ξ go to zero as J goes to ∞, conditions on the rate at which *ns* and n grow as J goes to ∞ and smoothness conditions on the map from $\Theta \times R^J$ to $s(\theta, \xi, P_0)$.

The matrices V_1, V_2, and V_3 arise from the three independent sampling processes. V_1 arises from the process generating the product characteristics (the (x, w, ξ, ω)), V_2 from the consumer sampling process (which generates the difference between s^n and s^0) and V_3 from the simulation process (which generates the difference between P_{ns} and P_0).

From the utility specification used here, the results in Berry and Pakes (in process) require n and ns to grow quite rapidly, on the order of J^3. Despite this, the fact that n in our sample is so large (the number of households in the U.S. economy is on the order of 100 million) implies that V_2 is negligible in our problem. On the other hand, we are concerned about the variance due to simulation error. Section 6 develops variance reduction techniques that enable us to use relatively efficient simulation techniques for our problem. Even so, we found that with a reasonable number of simulation draws the contribution of the simulation error to the variance in our estimates (V_3) is not negligible.

To calculate standard errors, we estimate V_1 by substituting θ_J for θ_0 and taking the sample analog of the expression above. To estimate V_3, we substitute θ_J for θ_0 and employ a Monte Carlo procedure. Specifically, we draw P_{ns} independently times. For each of these samples, we calculate the vector of moment conditions (5.4) and use the empirical variance of these moment conditions as our estimate. Correcting for the variance due to simulation increases our reported standard errors in Table IV of Section 7 by about 5–20% (with the exception of one parameter, whose reported standard error doubles).

5.1. *Optimal Instruments*

In Section 4, we propose using as instruments functions of z, the cost and demand characteristics of all products in a given year. In this section we consider the form of those functions. Because we use only market level data and are therefore concerned with efficiency, we are guided in our choice by the optimal instrument literature.

Using an i.i.d. sampling scheme and other mild regularity conditions Chamberlain (1986) shows that the efficient set of instruments when we have only conditional moment restrictions is equal to the conditional expectation of the derivative of the conditional moment condition with respect to the parameter vector (conditioning on the same set of variables that condition the moment restriction, and evaluated at θ_0). The analogous instruments for our case are

$$(5.7) \qquad H_j(z) = E\left[\frac{\partial \xi_j(\theta_0, s^0, P_0)}{\partial \theta}, \frac{\partial \omega_j(\theta_0, s^0, P_0)}{\partial \theta} \middle| z \right] T(z_j) \equiv D_j(z) T(z_j),$$

in which case the variance covariance matrix of the estimated parameter vector is

$$\left\{E_z\left[D_j(z)\Omega(z_j)^{-1} D_j(z)'\right]\right\}^{-1}.$$

The formula in (5.7) is very intuitive: larger weights should be given to the observations that generate disturbances whose computed values are very sensitive to the choice of θ (at $\theta = \theta_0$). Unfortunately $D_j(z)$ is typically very difficult, if not impossible, to compute. To calculate $D_j(z)$ we would have to calculate the pricing equilibrium for different $\{\xi_j, \omega_j\}$ sequences, take derivatives at the equilibrium prices, and then integrate out over the distribution of such sequences. In addition, this would require an assumption that chooses among multiple equilibria when they exist, and either additional assumptions on the joint distribution of (ξ, ω), or a method for estimating that distribution.[18]

Newey (1990) considers the special case where $T(z) = T$ (for all z), and shows that, again under mild regularity conditions, one can circumvent the problem of computing $D_j(z)$ by using a semiparametric estimator of it, and still generate an estimator whose limiting variance-covariance matrix is $\{E_z[D_j(z)\Omega^{-1}D_j(z)']\}^{-1}$ (see also the related work on feasible GLS by Robinson (1987); and the literature cited in both of these articles). The first stage of this procedure uses an initial consistent estimate of θ_0 to compute a nonparametric estimate of $H_j(z)$. Newey (1990) provides results from a Monte Carlo experiment that shows that this procedure tends to work well when a polynomial series approximation to the efficient instrument vector is used.

Though polynomial approximations are easy to compute, there is a dimensionality problem in using them to approximate functions whose arguments include the characteristics of all competing products. An unrestricted polynomial series approximation of a given order will have a number of basis functions that grows polynomially in the number of products in the market, J. In our case J is also the limiting dimension of the problem. This implies that the dimension of the basis needed for the approximation grows polynomially in sample size. This in turn both creates a practical problem in forming the estimator and violates the regularity conditions required for the consistency of the first stage estimator of the efficient instruments.

As shown in Pakes (1994) this dimensionality problem can be circumvented if ξ and ω are symmetric, or more precisely exchangeable, in some of their arguments. By exchangeable we mean that we can permute the order in which those variables enter a function without changing the value of that function. Recall that ξ and ω are determined by the demand function, the cost function, and the pricing assumption. By construction, both the demand and the cost functions for product j are exchangeable in vectors of characteristics of all other products. This is true trivially of cost functions that only depend on own-product characteristics, and is true for any differentiated products demand system in which the demand for a product does not depend on the ordering of rival products but just on their characteristics.

The pricing function for a given firm's product will change, however, if we permute the order of a product produced by the given firm and a product

[18] In an early version of this paper, we proposed alternative ways of approximating $H_j(z)$ and we have found some of these useful in subsequent work; see Berry, Levinsohn, and Pakes (1994).

produced by a rival firm. So this function is not exchangeable in the characteristics of all other products. On the other hand, any unique Nash equilibrium is still partially exchangeable: that is, exchangeable in the characteristics of the firm's other products and exchangeable in the state vectors of its competitors products. In fact, a unique Nash equilibrium would imply the following three forms of exchangeability for the ξ and ω functions:

(i) exchangeable in the order of the competing firms (e.g., the prices of GM's products would not change if instead of listing the characteristics of Ford's products before Chrysler's, we listed the characteristics of Chrysler's products before Ford's).

(ii) for a given competitor, exchangeable in the order of that competitors products, and

(iii) for a given product, exchangeable in the order of the other products marketed by the same firm.

Theorem 32 in Pakes (1994) shows that the dimension of the basis for polynomials of a given order that are partially exchangeable is independent of the number of exchangeable arguments. For example, given the properties above, the first order basis functions associated with characteristic z_{jk}, the kth characteristic of product j produced by firm f, are

$$z_{jk}, \sum_{r \neq j, r \in \mathscr{F}_f} z_{rk}, \sum_{r \neq j, r \notin \mathscr{F}_f} z_{rk}. \tag{5.8}$$

(Remember that $\mathscr{F}_f$ is the set of products produced by firm f). Note that the dimension of the first order terms in this basis is $3K$, where K is the dimension of z_j. In contrast, the dimension of the first order terms in the unrestricted basis is JK.

For each of the separate cost and demand characteristics in our model, we compute the three terms in (5.8) and include these three terms as potential instruments. For example, if one of our characteristics is the size of a car, then the instrument vector for product j includes the size of car j, the sum of size across own-firm products, and the sum of size across rival firm products. Note the two sums vary across products in our sample because (i) they exclude different own-products j, (ii) different firms produce different sets of products, and (iii) there is variation across time in the products in our panel data set. Note also that one of our characteristics is a constant term, so that the number of own-firm products and rival-firm products become instruments.

We could also include second and higher order basis functions, but in practice we found these extra terms to be nearly collinear with the terms in (5.8). In fact, the entire matrix of these linear terms is also nearly not of full rank. We faced a somewhat arbitrary choice of what terms to leave out, but given the near multicollinearity the choice should not greatly affect our estimates.

In constructing a set of instruments to interact with the demand error, ξ_j, we began with the three terms in (5.8) for each of the five demand variables described in the data section below, as it seemed reasonable to insure that x_j

entered the demand side moment conditions. The two variables that in our specification enter cost but not demand (miles per gallon and a trend) could be added to this list, but we found them to be so nearly collinear as to cause numerical problems in inversion and therefore we left them out, giving 15 demand-side instruments. To construct a list of variables to interact with ω, we began with the three terms of (5.8) for each of the six elements of w_j, giving at least eighteen cost side instruments. We were able to add the excluded demand variable, miles per dollar, to this list without causing a problem with near collinearity. Therefore, there are nineteen cost side instruments, giving a total of $34 = 15 + 19$ sample moment restrictions.

5.2. *Other Details*

The method outlined by Newey (1990) would suggest projecting the derivatives in (5.7) onto the basis functions in (5.8). Instead, we enter the basis functions directly into the instrument vector. To see why, let $f_j(z) \in \boldsymbol{R}^R$ provide the values of the basis functions in (5.8) for the jth observation, let $\otimes$ be the Kronecker product operator, and let I_2 be an identity matrix of order two. It is helpful to consider the special case in which $T(z) = T$ and the conditional expectation of the derivative matrix, $D_j(z)$, is a linear function of a finite dimensional basis. That is, $D_j(z)$ exactly equals $(f_j'(z) \otimes I_2)B$ for some matrix B. In this case algebraic manipulation shows that the estimator found by first projecting the derivatives in (5.7) onto $(f_j'(z) \otimes I_2)$, and then using the fitted values from this projection as the estimate of $D_j(z)$, has the same limiting distribution as the generalized method of moments (or GMM) estimator (Hansen (1982)) that uses $\{[\xi(\theta), \omega(\theta)]' \otimes f_j(z)\}$ as moments and a consistent estimate of $E(\{[\xi(\theta), \omega(\theta)]' \otimes f_j(z)\}\{[\xi(\theta), \omega(\theta)]' \otimes f_j(z)\}')$ as its weighting matrix. Since the method of moments estimator is easier to compute, we use it in the actual estimation subroutine.[19]

Finally, note that the data we actually use are not a single cross section, but a panel data set that follows car models over all years they are marketed. It is likely that the demand and cost disturbances of a given model are more similar across years than are the disturbances of different models (so model-year combinations are not exchangeable). Though correlation in the disturbances of a given model marketed in different years does not alter the consistency or asymptotic normality of the parameter estimates from our algorithm, it does affect their variance-covariance matrix. As a result, we use estimators that treat the sum of the moment restrictions of a given model over time as a single observation from an exchangeable population of models. That is, replacing product index j by indices for model m and year t, we define the sample

[19] Additionally, if $T(z) \neq T$ we do not know of a proof which insures that the two step estimator is more efficient than this GMM estimator.

moment condition associated with a single model as

$$g_m(\theta) \equiv \sum_t \left[f_{mt}(z)' \otimes I_2 \right] \begin{bmatrix} \xi_{mt}(\theta) \\ w_{mt}(\theta) \end{bmatrix},$$

and then obtain our GMM estimator by minimizing our quadratic form in the average of these moment conditions across models. Although this is probably not the most efficient method for dealing with correlation across years for a given model, it does produce standard errors that allow for arbitrary correlation across years for a given model and arbitrary heteroskedasticity across models.

6. COMPUTATION

The method of moments estimation algorithm outlined in the last section requires computation of the moments, $G_J(\theta, s^n, P_{ns})$, for different values of θ. Most of this section is devoted to providing an algorithm that computes the $G_J(\theta, s^n, P_{ns})$. The reader who is not interested in computational details can go directly to the empirical results in Section 7.

We focus throughout on two special cases. The first is the pure logit model, while the second adds interactions between consumer and product characteristics as in (2.7). The advantages of carrying along the logit model, despite its unreasonable substitution patterns, stem from its computational simplicity. This makes it easy to use the logit model to illustrate both the logic of the overall estimation procedure and the likely importance of unobserved product characteristics.

There are four steps to each evaluation of $G_J(\theta, s^n, P_{ns})$ in both models. For each θ:

(i) estimate (via simulation) the market shares implied by the model;

(ii) solve for the vector of demand unobservables [i.e. $\xi(\theta, s^n, P_{ns})$] implied by the simulated and observed market shares;

(iii) calculate the cost side unobservable, $\omega(\theta, s^n, P_{ns})$, from the difference between price and the markups computed from the shares; and finally

(iv) calculate the optimal instruments and interact them with the computed cost and demand side unobservables (as in (5.3)) to produce $G_J(\theta, s^n, P_{ns})$.

Both models are nested to the utility specification,

$$u_{ij} = \delta(x_j, p_j, \xi_j, \theta_1) + \mu(x_j, p_j, \nu_i, \theta_2) + \epsilon_{ij}, \tag{6.1}$$

where the ϵ_{ij} are draws from independent extreme value distributions (independent over both i and j). Here $\delta_j = \delta(x_j, p_j, \xi_j; \theta_1)$ is a product-specific component that does not vary with consumer characteristics, while $\mu_{ij} = \mu(x_j, p_j, \nu_i; \theta_2)$ contains the interactions between product specific and consumer characteristics. We begin with the logit model.

6.1. *The Logit Model*

Our first model will assume no interaction effects: i.e. $\mu_{ij} = 0$. Given that we are assuming that ϵ_{ij} has the Weibull (or type I extreme value) distribution function, $\exp[-\exp(-\epsilon)]$, the assumption that $\mu_{ij} \equiv 0$ gives us the traditional logit model for market shares. In addition we assume that the mean utility level is linear in product characteristics, or

$$(6.2) \qquad \delta_j = x_j \beta - \alpha p_j + \xi_j,$$

so that $u_{ij} = x_j \beta - \alpha p_j + \xi_j + \epsilon_{ij}$. Since $\mu_{io} = \epsilon_{io}$ (that is, we normalize δ_0 to zero), the market-share functions are given by

$$(6.3) \qquad s_j(p, x, \xi, \theta, P_0) = \frac{e^{\delta_j}}{\left(1 + \sum_{j=1}^{J} e^{\delta_j}\right)}$$

for $j = 0, 1, \dots, J$ (McFadden (1973)).

Also, since (6.3) implies that

$$(6.4) \qquad \delta_j = \ln(s_j) - \ln(s_0),$$

our estimate of δ_j for the logit model is $\ln(s_j^n) - \ln(s_0^n)$, and, consequently, our estimate of the demand-side unobservable is

$$(6.5) \qquad \xi(s^n, p, x, \theta, P_0) = \ln\left(s_j^n\right) - \ln\left(s_0^n\right) - x_j \beta + \alpha p_j.$$

That is, there are analytic formulae for both the market share and the inverse functions for the logit model (see (i) and (ii) above).

The demand-side parameters can be estimated by interacting the demand-side unobservables from (6.5) with instruments and applying a method of moments procedure to the resulting moment conditions. For joint estimation of the demand and pricing equations we also need to compute the markups (see (iii) above) from the logit demand system and then use them to compute the cost-side unobservables (as in (3.6)).

6.2. *A Model with Interactions*

We now reintroduce a nontrivial interaction term $\mu = \mu(x_j, p_j, \nu_i, \theta_2)$. For the reasons noted, we focus on the "Cobb-Douglas" specification in (2.7).[20]

For this model, it is useful to obtain the market share function in two stages. First, condition on the ν and integrate out over the extreme value deviates to obtain the conditional (on ν) market shares as

$$(6.6) \qquad f_j(\nu_i, \delta, p, x, \theta) = \frac{e^{\delta_j + \mu(x_j, p_j, \nu_i, \theta_2)}}{1 + \sum_{j=1}^{J} e^{\delta_j + \mu(x_j, p_j, \nu_i, \theta_2)}}.$$

Second, integrate out over the distribution of ν to obtain the market shares

[20] The computational techniques provided here generalize to handle a variety of other cases. For example, at an additional computational cost we can allow for an interaction between unobserved product (ξ_j) and consumer (ν_i) characteristics, and/or do away with the extreme value, or idiosyncratic, error (the ϵ_{ij}). Also it is straightforward to generalize to less restrictive functional forms for utility (at least subject to mild regularity conditions).

conditional only on product characteristics as

$$(6.7) \qquad s_j(p,x,\xi,\theta,P_0)=\int f_j(\nu_i,\delta(x,p,\xi),p,x,\theta)P_0(d\nu).$$

Note that (6.6) has a closed form, while (6.7) does not. Since we cannot compute (6.7) exactly we will substitute a simulation estimator of its value into the estimation algorithm. Integrating out the ϵ analytically in the first stage allows us to limit the variance in the estimator of $s_j(p, x, \xi, \theta, P_0)$ to the variance induced by the ν. It also produces simulated market shares that are: positive, sum to one, and are smooth functions of their argument. We come back to the problem of efficiently simulating (6.7) in the next subsection; for now we simply assume we have a good simulation estimator and label the vector of simulated shares $s(p, x, \delta, P_{ns}; \theta)$.

Next we have to combine our estimates of the market share function with the observed market shares to solve for δ as a function of θ (see (ii) above). Once we add the interaction term we cannot solve for δ analytically, so we will have to solve for it numerically each time we evaluate the objective function at a different θ. Recall that δ solves the nonlinear system $s^n = s(p, x, \delta, P_{ns}, \theta)$, or equivalently

$$\delta=\delta+\ln(s^n)-\ln\left[s(p,x,\delta,P_{ns};\theta)\right].$$

In Appendix I, we show that for any triple (s, θ, P), such that s is in the interior of the $J+1$ dimensional unit simplex, $\theta \in \Theta \subset R^k$, and P is a proper distribution for ν, the operator $T(s, \theta, P): R^J \to R^J$ defined pointwise by

$$(6.8) \qquad T(s,\theta,P)\left[\delta_j\right]=\delta_j+\ln(s_j)-\ln\left[s_j(p,x,\delta,P;\theta)\right],$$

is a contraction mapping with modulus less than one. This implies that we can solve for δ recursively. That is, we begin by evaluating the right-hand side of (6.8) at some initial guess for δ, obtain a new δ' as the output of this calculation, substitute δ' back into the right hand side of (6.8), and repeat this process until convergence.

Given $\delta_j(\theta, s, P)$, it is easy to solve for the demand-side unobservable as $\xi_j(\theta, s, P) = \delta_j(\theta, s, P) - x_j\beta$. Next we calculate the cost-side unobservable. To do so, we need to solve for the markup, which in turn requires the derivatives of the market share function with respect to price. Equation (6.7) implies that those derivatives are

$$(6.9a) \qquad \partial s_j(p,x,\xi,\theta,P_0)/\partial p_j$$
$$=\int f_j(\nu,\delta,x,p,\theta)\left(1-f_j(\nu,\delta,x,p,\theta)\right)\left[\partial\mu_{ij}/\partial p_j\right]P_0(d\nu),$$

$$(6.9b) \qquad \partial s_j(p,x,\xi,\theta,P_0)/\partial p_q$$
$$=\int -f_j(\nu,\xi,x,p,\theta)f_q(\nu,\delta,x,p,\theta)\left[\partial\mu_{ij}/\partial p_q\right]P_0(d\nu).$$

6.3. *Simulators for Market Shares*

The integral in (6.7) becomes difficult to calculate as the dimension of the consumer characteristic grows much beyond two or three. As a result we form a simulation estimator of that integral and use it in the estimation algorithm. One simple simulation estimator would replace the population density, $P_0(d\nu)$ in (6.7), with the empirical distribution obtained from a set of ns pseudo-random draws from P_0, say, $(\nu_1, \ldots, \nu_{ns})$ and calculate

$$(6.10) \qquad s_j(p, x, \xi, \theta, P_{ns}) \equiv \frac{1}{ns} \sum_{i=1}^{ns} f_j(\nu_i, \delta, p, x, \theta).$$

The derivatives of market shares have similar, simple analytic forms. Although this simulation estimator does have a smaller variance than the standard frequency simulator, we looked for a simulator with yet smaller variance.

The importance sampling literature notes that we can often reduce the sampling variance of a simulation estimator of an integral by transforming both the integrand and the density we are drawing from in a way that reduces the variance of a simulation draw but leaves its expectation unchanged (see Rubinstein (1981), and the literature cited there). To see this take any function $h(\cdot, \theta)$ that is strictly positive on the support of P_0, and note that the integral in (6.7) can be rewritten as

$$(6.11) \qquad s_j(\theta, P_0) = \int \left[\frac{f_j(\nu, \theta)}{h(\nu, \theta)} p_0(\nu) h(\nu, \theta) \right] d\nu$$

$$\equiv \int f_{hj}(\nu, \theta) P_{hj}(d\nu, \theta) \equiv s_j(\theta, P_{hj}),$$

where

$$P_{hj}(d\nu, \theta) \equiv h(\nu, \theta)\, d\nu \quad \text{and}$$

$$f_{hj}(\nu, \theta) \equiv \left[f_j(\nu, \theta) p_0(\nu) \right] / h(\nu, \theta),$$

and we have assumed, for simplicity, that P_0 has a density with respect to Lebesgue measure (denoted by p_0).

Let $s_j(\theta, P_{hj,ns})$ be the $h(\cdot)$-based unbiased estimator of $s_j(\theta, P_0)$ formed from a simulated analogue to (6.11). Since there are many feasible $h(\cdot)$ the literature has focused on finding an $s_j(\theta, P_{hj,ns})$ with minimum variance. The solution is to set

$$(6.12) \qquad P_{hj}^*(d\nu, \theta) = \left[f_j(\nu, \theta) p_0(\nu)\, d\nu \right] / s_j(\theta, P_0),$$

as, in this case, $s_j(\theta, P_{hj,ns})$ equals $s_j(\theta, P_0)$ exactly (no matter ns). Intuitively, $P_{hj}^*(d\nu, \theta)$ places proportionately higher weight (relative to P_0) on draws of ν that result in larger values of the integrand. That is, we over sample consumers whose characteristics would lead them to buy product j.

Unfortunately, the optimal importance sampling simulator cannot be used directly. The most obvious problem with it is that to use it we need to know the integral itself, i.e., $s_j(\theta, P_0)$. Also, it depends on θ, while the limit properties of simulation estimators (and indeed the performance of the search algorithms used to find them) require the use of simulation draws that do not change as the minimization algorithm varies θ (see Pakes and Pollard (1989)). Finally, the contraction property that allows us to solve for the unobservables as a function of θ requires the vector of simulated shares to sum to one. However, the optimal importance sampling estimator changes with the share we are trying to simulate. If we use draws that change across shares in this way, it is difficult to guarantee that the shares sum to one.

Though these problems make direct use of the simulator in (6.12) impossible, that formula does suggest how to build an importance sampling simulator with low variance. First, note that though we do not know $P^*_{hj}(d\nu, \theta)$, we can obtain a consistent estimator of it, at least about $\theta = \theta_0$, by taking an initial consistent estimate of θ_0, say θ', calculating a good estimate of the share at θ', say $s_j(\theta', P_{nsi})$, and then drawing from $[f_j(\nu, \theta')p_0(\nu)\,d\nu]/s_j(\theta', P_{nsi})$. Note that the estimate $s_j(\theta', P_{nsi})$ is calculated only once, so *nsi* (the number of simulation draws for the initial step) can be quite large without imposing too much of a computational burden.

To implement this suggestion we need a way of drawing from $[f_j(\nu, \theta')p_0(\nu)\,d\nu]/s_j(\theta', P_{nsi})$. A simple *acceptance/rejection* procedure which accomplishes this is to draw ν from P_0 and "accept" it with probability $f_j(\nu, \theta')$. It is easy to use Bayes Rule to show that the accepted draws have the required density.

Lastly, to insure that the vector of simulated market shares sums to one, we used the same simulation draws to calculate each market share. Thus, we had to base the importance sampling estimators for the shares of all choices on the market share for a particular choice. We focus on the share of households who purchase automobiles, that is, on $\bar{s}(\theta) = [1 - s_o(\theta)] = \Sigma_{j=1}^J s_j$.

Thus we proceed as follows. We obtain an initial estimator of θ_o, say θ', using the simple smooth simulator in (6.10). Next we draw ν from P_0 and accept it with probability $\bar{f}(\nu, \theta') = \Sigma_{j=1}^J f_j(\nu, \theta')$. The vector of simulated market shares are then calculated as

$$(6.13) \qquad s_j[\theta, P_h^*(\theta')_{ns}] = \sum_{i=1}^{ns} \frac{\bar{s}(\theta', P_0)}{\bar{f}(\nu_i, \theta')} f_j(\nu_i, \theta)$$

where the sum is over accepted ν draws. This oversamples (relative to P_0) the ν's that are more likely to lead to (some) auto being purchased and then weights the purchase probabilities, f_j, by $s(\theta', P_0)/\bar{f}(\nu_i, \theta')$, the inverse of the sampling weights.

6.4. *The Empirical Distribution of Income and the Final Form of the Simulator*

Recall that consumer preferences in our interactive "Cobb-Douglas" model of (2.7) are determined by the marginal utility of characteristics [the vectors $\nu_i' = (\nu_{i0}, \ldots, \nu_{ik})$] and income (y_i). We assume that the ν_i are random draws from a normal distribution with mean vector zero and an identity covariance matrix independent of the level of consumer's income (y_i). The income distribution is assumed to be lognormal and we estimate its parameters from the March Current Population Survey (CPS) for each year of our panel (we denote the estimated mean by m_t and the estimated standard deviation by $\hat{\sigma}_y$). This allows us to use the exogeneously available information on the income distribution to increase the efficiency of our estimation procedure.[21]

Using this procedure our utility model is written as

$$(6.14a) \quad u_{itj} = \alpha \ln\left(e^{m_t + \hat{\sigma}_y \nu_{iy}} - p_{jt}\right) + x_{jt}\bar{\beta} + \xi_{jt} + \sum_k \sigma_k x_{jkt} \nu_{ik} + \epsilon_{ijt},$$

$$(6.14b) \quad u_{i0t} = \alpha \ln\left(e^{m_t + \hat{\sigma}_y \nu_{iy}}\right) + \xi_{0t} + \sigma_0 \nu_{i0} + \epsilon_{i0t},$$

where the vectors $(\nu_{iy}, \nu_{i0}, \ldots, \nu_{ik})$ are random draws from a multivariate normal distribution with mean 0 and an identity covariance matrix. Note that we held the vector of characteristics $(\nu_{iy}, \nu_{i0}, \ldots, \nu_{ik})$ fixed over the time period of the panel.

6.5. *Minimization*

Finally, we need a minimization routine that searches to find the value of θ that minimizes the objective function in (5.5). The minimization routine can be simplified by noting that the first order conditions for a minimum to (5.5) for our specifications are linear in β and γ for any given (α, σ). As a result β and γ can be "concentrated out" of those conditions, allowing us to confine the nonlinear search to a search over (α, σ) couples. This search was performed using the Nelder-Mead (1965) nonderivative "simplex" search routine.

7. DATA AND RESULTS

7.1. *The Data*

We use data on product characteristics obtained from annual issues of the *Automotive News Market Data Book*.[22] Product characteristics for which we have data include the number of cylinders, number of doors, weight, engine displacement, horsepower, length, width, wheelbase, EPA miles per gallon rating (MPG), and dummy variables for whether the car has front wheel drive, automatic transmission, power steering, and air conditioning as standard equipment.

[21] We could have taken *ns* draws from the CPS for each year and used these draws directly to simulate the market shares. This places fewer restrictions on the empirical distribution of income, but is inefficient if the true income distribution is in fact lognormal. We found the less restrictive procedure led to quite imprecise simulators (it did a particularly bad job of estimating changes in the upper tail of the income distribution), and, as a result, we kept the lognormal assumption. Also, we did not attempt to estimate a different standard deviation of income in each year because such estimates were imprecise.

[22] The data set combines data collected by us with a similar data set graciously made available to us by Ernie Berndt of MIT.

The price variable is the list retail price (in \$1000's) for the base model. This is clearly not ideal; we would prefer transaction prices, but these are not easy to find. All prices are in 1983 dollars. (We used the Consumer Price Index to deflate.) The sales variable corresponds to U.S. sales (in 1000's) by name plate.[23] The product characteristics correspond to the characteristics of the base model for the given name plate.

The data set includes this information on (essentially) all models marketed during the 20 year period beginning in 1971 and ending in 1990 (the only models excluded are "exotic" models with extremely small market shares, such as the Ferrari and the Rolls Royce). Since models both appear and exit over this period, this gives us an unbalanced panel. Treating a model/year as an observation, the total sample size is 2217. Throughout we shall assume that two observations in adjacent years represent the same model if (a) they have the same name; and (b) their horsepower, width, length, or wheelbase do not change by more than ten percent. With these definitions the 2217 model/years represent 997 distinct models (as noted in Section 5, different models are assumed to have unobservables whose conditional distributions are independent of one another, but the unobservables for different years of the same model are allowed to be freely correlated).

Aside from these product characteristics, we obtain additional data from a variety of sources. Because the cost of driving may matter to consumers (as opposed to just the MPG rating), we gathered data on the price of gasoline (the real price of unleaded gasoline as reported by the U.S. Department of Commerce in *Business Statistics*, 1961–1988). One of our product characteristics is then miles per dollar (MP\$), calculated as MPG divided by price per gallon. Also, our measure of market size (M) was the number of households in the U.S. and this was taken for each year from the *Statistical Abstract of the U.S.*, while, as noted in the computation section, the parameters of the distribution of household income were estimated from the annual March Current Population Surveys. We also obtained *Consumer Reports* reliability ratings. This variable is a relative index that ranges from 1 (poor reliability) to 5 (highest reliability).[24]

The multi-product pricing problem requires us to distinguish which firms produce which models. We assume that different branches of the same parent company comprise a single firm. For example, Buick, Oldsmobile, Cadillac, Chevrolet, and Pontiac are all part of one firm, General Motors. This follows Bresnahan (1981) and Feenstra and Levinsohn (1995). For some results, we also assign a country of origin to each model, which is simply the country associated with the producing firm.[25]

[23] We do not observe fleet sales, which include sales to rental car companies. In ignoring fleet sales, we effectively assume that fleet purchasers are acting as agents for households.

[24] Unfortunately, this variable is not available for every product in our sample and, more importantly, the rating was rescaled in every year of our sample. For example, the absolute level of reliability of a "3" rating changes every year in an unreported way, as does the absolute increment in reliability represented by a one point increase in the index.

[25] For example, we treat Hondas as Japanese and VW's as German, although, by the end of our sample, some of each were produced in the U.S.

TABLE 1
DESCRIPTIVE STATISTICS

Year	No. of Models	Quantity	Price	Domestic	Japan	European	HP/Wt	Size	Air	MPG	MP$
1971	92	86.892	7.868	0.866	0.057	0.077	0.490	1.496	0.000	1.662	1.850
1972	89	91.763	7.979	0.892	0.042	0.066	0.391	1.510	0.014	1.619	1.875
1973	86	92.785	7.535	0.932	0.040	0.028	0.364	1.529	0.022	1.589	1.819
1974	72	105.119	7.506	0.887	0.050	0.064	0.347	1.510	0.026	1.568	1.453
1975	93	84.775	7.821	0.853	0.083	0.064	0.337	1.479	0.054	1.584	1.503
1976	99	93.382	7.787	0.876	0.081	0.043	0.338	1.508	0.059	1.759	1.696
1977	95	97.727	7.651	0.837	0.112	0.051	0.340	1.467	0.032	1.947	1.835
1978	95	99.444	7.645	0.855	0.107	0.039	0.346	1.405	0.034	1.982	1.929
1979	102	82.742	7.599	0.803	0.158	0.038	0.348	1.343	0.047	2.061	1.657
1980	103	71.567	7.718	0.773	0.191	0.036	0.350	1.296	0.078	2.215	1.466
1981	116	62.030	8.349	0.741	0.213	0.046	0.349	1.286	0.094	2.363	1.559
1982	110	61.893	8.831	0.714	0.235	0.051	0.347	1.277	0.134	2.440	1.817
1983	115	67.878	8.821	0.734	0.215	0.051	0.351	1.276	0.126	2.601	2.087
1984	113	85.933	8.870	0.783	0.179	0.038	0.361	1.293	0.129	2.469	2.117
1985	136	78.143	8.938	0.761	0.191	0.048	0.372	1.265	0.140	2.261	2.024
1986	130	83.756	9.382	0.733	0.216	0.050	0.379	1.249	0.176	2.416	2.856
1987	143	67.667	9.965	0.702	0.245	0.052	0.395	1.246	0.229	2.327	2.789
1988	150	67.078	10.069	0.717	0.237	0.045	0.396	1.251	0.237	2.334	2.919
1989	147	62.914	10.321	0.690	0.261	0.049	0.406	1.259	0.289	2.310	2.806
1990	131	66.377	10.337	0.682	0.276	0.043	0.419	1.270	0.308	2.270	2.852
All	2217	78.804	8.604	0.790	0.161	0.049	0.372	1.357	0.116	2.099	2.086

Note: The entry in each cell of the last nine columns is the sales weighted mean.

Tables I and II provide some summary descriptive statistics of variables that are used in the specifications we discuss below. These variables include quantity (in units of 1000), price (in $1000 units), dummies for where the firm that produced the car is headquartered, the ratio of horsepower to weight (in *HP* per 10 lbs.), a dummy for whether air conditioning is standard (1 if standard, 0 otherwise), the number of ten mile increments one could drive for $1 worth of gasoline (*MP$*), tens of miles per gallon (*MPG*), and size (measured as length times width).

Table I gives sales-weighted means. Several interesting trends are evident. The number of products available generally rises from a low of 72 in 1974 to its high of 150 in 1988. Sales per model, on the other hand trend downward (though here there is some movement about the trend). In real terms, the sales-weighted average list price of autos has risen almost 50 percent during the 1980's after having remained about constant during the 1970's. On the other hand, the characteristics of the cars marketed are also changing (so the cost of a car with a given vector of characteristics need not be increasing). The ratio of horsepower to weight fell in the early 1970's and has since trended upward. Most of the changes in this ratio are attributable to changes in weight as horsepower has remained remarkably constant. It appears that prior to the first oil price shock, cars were becoming heavier, while after the mid-1970's cars became lighter. Along with the change in the ratio of horsepower to weight, cars have also become more fuel cost-efficient. In 1971, the average new car drove

TABLE II
THE RANGE OF CONTINUOUS DEMAND CHARACTERISTICS
(AND ASSOCIATED MODELS)

Variable	Percentile 0	25	50	75	100
Price	90 Yugo 3.393	79 Mercury Capri 6.711	87 Buick Skylark 8.728	71 Ford T-Bird 13.074	89 Porsche 911 Cabriolet 68.597
Sales	73 Toyota 1600CR .049	72 Porsche Rdstr 15.479	77 Plym. Arrow 47.345	82 Buick LeSabre 109.002	71 Chevy Impala 577.313
HP/Wt.	85 Plym. Gran Fury 0.170	85 Suburu DH 0.337	86 Plym. Caravelle 0.375	89 Toyota Camry 0.428	89 Porsche 911 Turbo 0.948
Size	73 Honda Civic 0.756	77 Renault GTL 1.131	89 Hyundai Sonata 1.270	81 Pontiac F-Bird 1.453	73 Imperial 1.888
MP$	74 Cad. Eldorado 8.46	78 Buick Skyhawk 15.57	82 Mazda 626 20.10	84 Pontiac 2000 24.86	89 Geo Metro 64.37
MPG	74 Cad. Eldorado 9	79 BMW 528i 17	81 Dodge Challenger 20	75 Suburu DL 25	89 Geo Metro 53

Notes: The top entry for each cell gives the model name and the number directly below it gives the value of the variable for this model.

18.50 miles on a (1983) dollar of gasoline, while by 1990 that figure was 28.52 miles. Also, while no cars had air conditioning as standard equipment at the start of the sample, 30.8 percent had it by the end. This is indicative of a general trend toward more extensive standard equipment. The market share of domestic cars has fallen from a 1973 high of 93.2 percent to a 1990 low of 68.2 percent. European market share has been fairly constant since the demise of the popular VW Beetle in the mid-1970's hovering around 4 to 5 percent. The Japanese market share has risen from a low of 4.0 percent in 1973 to a high of 27.6 percent in 1990. An automobile's size, given by its length times width trends generally downward with this measure falling about 17 percent over the sample.

Table II associates some names with the numbers. This table provides an indication of the range of the continuous product attributes by presenting the quartiles of their distribution. The least expensive car in the sample is the 1990 Yugo at $3393 (1983 dollars) while the top-of-the-line Porsche 911 Turbo Cabriolet costs $68,597. The 1989 Geo Metro has the highest MPG and MP$ while the 1974 Cadillac Eldorado has the lowest. The ratio of horsepower to weight varies tremendously from 0.170 for the (questionably named) 1985 Plymouth Gran Fury to .948 for the Porsche 911 Turbo. The smallest car in the sample was the 1973 Honda Civic.

7.2. *Some Results*

We will report three basic sets of results together with some auxiliary calculations. These are a simple logit specification, an instrumental variables logit specification, and the Cobb-Douglas specification in (6.14) above. For simplicity, we will refer to the first as logit, the second as IV logit, and the third as BLP.

The logit model, discussed first, provides an easy to compute reference point. One advantage of presenting logit results is that we can explore the effects of controlling for the endogeneity of prices in a very simple framework. The IV logit maintains the restrictive functional form of the logit (and hence must generate the restrictive substitution patterns that this form implies), but allows for unobserved product attributes that are correlated with price, and therefore corrects for the simultaneity problem that this correlation induces. The BLP results allow both for unobserved product characteristics and a more flexible set of substitution patterns. Results from each specification will be discussed in turn.

7.3. *The Logit and the IV Logit*

The first set of results are based on the simplest logit specification for the utility function. They are obtained from an ordinary least squares regression of $\ln(s_j) - \ln(s_0)$ on product characteristics and price (see (5.5)).

The choice of which attributes to include in the utility function is, of course, ad hoc. For the BLP specification, computational constraints dictate a parsimonious list. Since we wish to compare results across different specifications, we adopt a short list of included attributes in the logit specifications also. Included characteristics are the ratio of horsepower to weight (*HPWT*), a dummy for whether air conditioning is standard, miles per dollar (*MP*\$), size, and a constant. Horsepower over weight and *MP*\$ are obvious measures of power and fuel efficiency, while air conditioning proxies for a measure of luxury. Size is intended as a measure of both itself and safety. Other measures of size such as interior room are not available for much of the sample period while government crash test results are only available for a small subsample of the data. Though there are surely solid arguments for including excluded attributes, their force is somewhat diminished by our explicit treatment of product attributes unobserved by the econometrician but known to the market participants. Still, we investigate how robust results are to the choice of included attributes in sensitivity analyses that are presented below.

In the first column of Table III, we report the results of OLS applied to the logit utility specification. Most coefficients are of the expected sign, although the (imprecisely estimated) negative coefficients on air conditioning and size are anomalies, as one would expect these attributes to yield positive marginal utility. On the other hand these estimates have a distinctly implausible set of implications on own price elasticities. The estimated coefficient on price in Table III implies that 1494 of the 2217 models have inelastic demands. This is inconsistent with profit maximizing price choices. Moreover this is not simply a problem generated by an imprecise estimate of the price coefficient. Adding and subtracting two times the estimate of the standard deviation of the price coefficient to its value and recalculating the price elasticities still leaves 1429 and 1617 inelastic demands respectively.

TABLE III
RESULTS WITH LOGIT DEMAND AND MARGINAL COST PRICING
(2217 OBSERVATIONS)

Variable	OLS Logit Demand	IV Logit Demand	OLS ln (*price*) on *w*
Constant	−10.068	−9.273	1.882
	(0.253)	(0.493)	(0.119)
*HP/Weight**	−0.121	1.965	0.520
	(0.277)	(0.909)	(0.035)
Air	−0.035	1.289	0.680
	(0.073)	(0.248)	(0.019)
MP$	0.263	0.052	—
	(0.043)	(0.086)	
*MPG**	—	—	−0.471
			(0.049)
*Size**	2.341	2.355	0.125
	(0.125)	(0.247)	(0.063)
Trend	—	—	0.013
			(0.002)
Price	−0.089	−0.216	—
	(0.004)	(0.123)	
No. Inelastic Demands	1494	22	*n.a.*
(+/−2 *s.e.'s*)	(1429–1617)	(7–101)	
R^2	0.387	*n.a.*	.656

Notes: The standard errors are reported in parentheses.
* The continuous product characteristics—hp/weight, size, and fuel efficiency (*MP$* or *MPG*)—enter the demand equations in levels, but enter the column 3 price regression in natural logs.

In the second column of Table III, we re-estimate the logit utility specification, this time allowing for unobservable product attributes that are known to the market participants (and hence can be used to set prices), but not to the econometrician. To account for the possible correlation between the price variable and the unobserved characteristics, we use an instrumental variable estimation technique, using the instruments discussed at the end of Section 5.1.

The use of instruments generates substantial changes in several of the parameter estimates. All characteristics now enter utility positively and all but *MP$* are statistically significant. Moreover, just as the simultaneity story predicts, the coefficient on price increases in absolute value (indeed it more than doubles). Our interpretation of this finding is familiar: products with higher unmeasured quality components sell at higher prices. Note that now only 22 products have inelastic demands—a significant improvement from the OLS results. Seven to 101 demands are estimated to be inelastic when we evaluate elasticities at plus and minus two standard deviations of the parameter estimate.

These results seem to indicate that correcting for the endogeneity of prices matters. One can also see the importance of unobservable characteristics by examining the fit of the logit demand equation. The simple logit specification gives an R^2 of 0.387. This implies that 61 percent of the variance in mean utility levels is due to the unobserved characteristics.

As noted in Section 2, the separability, in product and consumer characteristics, of the logit functional form applied to aggregate data implies that neither the IV nor the simple logit estimates can possibly generate plausible cross price elasticities, or for that matter differences in markups across products. Thus, the IV logit estimates reported in Table III imply that all models have about the same mark-up (ranging from \$4630 for the BMW to \$4805 for the Chevy Cavalier). Markups are related to the model's market share (which, as noted, are about equal in absolute terms for all products) and how many products are made by the same parent firm. GM produces the most models and therefore its estimated markups are highest, while BMW produces the fewest models and its estimated markups are, quite counter-intuitively, the lowest.[26]

Table III also presents results from a very simple model of "supply." For the purposes of Table III we assume marginal cost pricing, with the specification for marginal cost found in (3.1). The marginal cost pricing equation is obtained by setting the markup term in our pricing equation (3.6) to zero, and regressing log price on w (the characteristics that shift the cost surface).[27]

The third column of Table III presents the results of this simple regression. In Table III (and in subsequent cost-side results), included cost shifters (w_j) are the same attributes that appear in utility with three modifications. First, miles per gallon replaces miles per dollar, as the production cost of fuel efficient vehicles presumably does not change with the retail price of gasoline (at least in the short-run). Second, we include a trend term to capture technical change and other trending influences (e.g. government regulation) on real marginal cost. Third, we use the log of continuous attributes, not their level, in the cost function. Thus the cost function parameters have the interpretation of elasticities of marginal cost with respect to associated product characteristics.

Note that the cost function adopted here is both simple and restrictive. In particular, it implies a constant elasticity of marginal cost with respect to all attributes and does not permit marginal cost to vary with output. Though our robustness tests provide some results with more flexible cost functions (see Table IX), we hesitate to use a more detailed specification of the cost surface without having more direct information on costs.

As is typical in similarly estimated hedonic pricing regressions, each of the coefficients on characteristics (except *MPG*) is estimated to be positive and all are significantly different from zero. (We comment on the *MPG* coefficient below). For example, a 10% increase in the ratio of horsepower to weight is

[26] A referee has noted that we could generate variation in markups by putting $\ln(p)$ instead of p into the logit utility function, which might also more closely match the $\ln(y-p)$ specification in the full model. We implemented this suggestion and found that markups are indeed more reasonable, but that substitution patterns are still quite unreasonable.

[27] The hedonic pricing literature, e.g. Griliches (1971), frequently presents similar regressions of log price on product characteristics. Of course, these regressions are motivated much differently from the marginal cost pricing argument we give here.

associated with a 5.2% increase in prices (and, in this context, in marginal costs). Also familiar from hedonic results is the fact that the R^2 from this regression is fairly high (at 0.66); simple functions of observable characteristics seem to be much better able to explain differences in the log of prices, than they are able to explain differences in the mean utility levels that rationalize the logit demand structure.

We turn now to results from our full model.

7.4. *Results from the Full Model*

The demand system for the full model is derived from the utility function in (5.14). The attributes that enter the utility function (the x-vector) for our base case scenario are the same as in Table III. Now, the marginal utility of each attribute varies across consumers so that we estimate a mean and a variance for each of them.[28] The pricing equation is given in (3.6) and the cost-side variables (the w-vector) are the same as in the third column of Table III.

The results from jointly estimating the demand and pricing equations from our specification are provided in Table IV.[29] As noted, the reported standard errors have been corrected for simulation error and for serial correlation of unobserved characteristics within models across years (but not for any correlation across models). The first and second panels of the table provide the estimates of the means and standard deviations of the taste distribution of each attribute, respectively. The third panel provides the estimate of the coefficient of $\ln(y-p)$, and the last panel provides the estimates of the parameters of the cost functions.

We begin with a discussion of the cost-side parameters. The coefficients on $\ln(HP/\text{Weight})$, Air, and the constant are positive and significantly different from zero. The term on trend is also positive and significant. The coefficient on ln(size) is not significantly different from zero. The coefficient on *MPG* is negative and significant, just as it is in the regression of log price on product characteristics reported in Table III.

Indeed, recall that our pricing equation is essentially an instrumental variable regression of $\ln[p-b(p,x,\xi;\theta)]$ on the cost side characteristics, where $b(p,x,\xi;\theta)$ is the markup (see (3.5)). Since $\ln[p-b(p,x,\xi;\theta)] \cong \ln(p) - b(p,x,\xi;\theta)/p$, if our model is correct, the marginal cost pricing, or "hedonic," regression should, by the traditional omitted variable formula, produce coeffi-

[28] In this context, we remind the reader that a positive variance of the random coefficient on the constant term implies that the distribution of the outside good has more idiosyncratic variance than that of the extreme value deviates generating idiosyncratic variance for the inside alternatives.

[29] We should note here that we have also estimated the demand side of our specification separately, and that we have run specifications that allowed for firm specific dummy variables on both the demand and cost side. Since there are 22 firms in our data set this latter specification generates 66 additional parameters (a mean and variance for each firm on the demand side, and one cost elasticity on the supply side). Neither of these changes generated point estimates that were much different from the point estimates in Table IV, but both generated much larger estimated standard errors.

TABLE IV

ESTIMATED PARAMETERS OF THE DEMAND AND PRICING EQUATIONS: BLP SPECIFICATION, 2217 OBSERVATIONS

Demand Side Parameters	Variable	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Means ($\bar{\beta}$'s)	*Constant*	−7.061	0.941	−7.304	0.746
	HP/Weight	2.883	2.019	2.185	0.896
	Air	1.521	0.891	0.579	0.632
	MP$	−0.122	0.320	−0.049	0.164
	Size	3.460	0.610	2.604	0.285
Std. Deviations (σ_β's)	*Constant*	3.612	1.485	2.009	1.017
	HP/Weight	4.628	1.885	1.586	1.186
	Air	1.818	1.695	1.215	1.149
	MP$	1.050	0.272	0.670	0.168
	Size	2.056	0.585	1.510	0.297
Term on Price (α)	$\ln(y-p)$	43.501	6.427	23.710	4.079
Cost Side Parameters					
	Constant	0.952	0.194	0.726	0.285
	$\ln(HP/Weight)$	0.477	0.056	0.313	0.071
	Air	0.619	0.038	0.290	0.052
	$\ln(MPG)$	−0.415	0.055	0.293	0.091
	$\ln(Size)$	−0.046	0.081	1.499	0.139
	Trend	0.019	0.002	0.026	0.004
	$\ln(q)$			−0.387	0.029

cients that are approximately the sum of the effect of the characteristic on marginal cost and the coefficient obtained from the auxiliary regression of the percentage markup on the characteristics. Comparing the cost side parameters in Table IV with the hedonic regression in Table III we find that the only two coefficients that seem to differ a great deal between tables are the constant term and the coefficient on size. The fall in these two coefficients tells us that there is a positive average percentage markup, and that this markup tends to increase in size.

The coefficients on *MPG* and size may be a result of our constant returns to scale assumption. Note that, due to data limitations, neither sales nor production enter the cost function. Almost all domestic production is sold in the U.S., hence domestic sales is an excellent proxy for production. The same is not true for foreign production, and we do not have data on model-level production for foreign automobiles. The negative coefficient on *MPG* may result because the best selling cars are also those that have high *MPG*. By imposing constant returns to scale, we may force these cars to have a smaller marginal cost than they actually do. Due, to the positive correlation between both *MPG* and size and sales, conditional on other attributes, the coefficients on *MPG* and size are driven down. We can attempt to investigate the accuracy of this story by including ln(sales) in the cost function, keeping in mind that for foreign cars this is not necessarily well measured. (Note, though, in Table I that about 80%

of the cars in our sample are domestic). When we include $\ln(q)$, so that the cost function is given by

$$\ln(c_j) = w_j\gamma_w + \gamma_q \ln(q_j) + \omega_j,$$

and re-estimate with the same instruments, all cost shifters are positive and significantly different from zero. These estimates are presented in the last two columns of Table IV. The coefficient on $\ln(q)$ is very significantly negative, giving implied returns to scale that seem implausibly high. Adding higher order terms in $\ln(q)$ reduces this problem, but we hesitate to take this approach too far since the data are inaccurate for about a fifth of our sample.

Our estimate of the variance of the cost-side unobservable, ω, implies that it accounts for about 22% of the estimated variance in log marginal cost. Thus, though our estimates do imply that there are some differences in "productivity" across firms, most of the differences in (the log of) marginal costs can be accounted for by a simple linear function of observed characteristics. As one might expect, the correlation between the demand-side error, ξ, and ω is positive implying that products with more unmeasured quality were more costly to produce. On the other hand, that correlation was only .17, implying that most of the (substantial) variance in ξ could not be accounted for by a linear function of differences in *marginal* costs of production.

Before discussing the demand-side coefficients in the first three panels of Table IV, we briefly review the structure of purchases in a discrete-choice model. Recall that these are driven by the maximum, and not by the mean, of the utilities associated with the given products. Thus there are, in general, two ways to explain why, say, products with high levels of horsepower to weight (*HPWT*), are popular. One can explain this by either positing a high mean for the distribution of tastes for *HPWT*, or by positing a large variance of that same distribution, for both an increase in the mean and an increase in the variance of tastes will increase the share of consumers who purchase cars with high *HPWT*. However, the two explanations have different implications for substitution patterns, and thus different implications for how market share will change with product attributes and prices. If there were, for example, a zero standard deviation for the distribution of marginal utilities of *HPWT*, we would find that when a high *HPWT* car increases its price, consumers who substitute away from that car have the same marginal utilities for *HPWT* as any other consumer and hence will not tend to substitute disproportionately toward other high *HPWT* cars. If, on the other hand, the standard deviation of tastes for *HPWT* was relatively large, the consumers who substitute away from the high *HPWT* cars will tend to be consumers who placed a relatively high marginal utility on *HPWT* originally, and hence should tend to substitute disproportionately towards other high *HPWT* cars.[30]

[30] This same reasoning leads to an interesting set of questions regarding the nonparametric identification of the parameters of the taste distribution, which we have not yet begun to investigate. We should note, however, that we had much more difficulty estimating separate mean and variance terms from a single cross section than we did from the panel; indeed, this was one motivation for using a panel data set.

We now move on to the estimates of the means, $\bar{\beta}_k$, and the standard deviations, σ_k, of the marginal utility distributions. For expositional simplicity, we will focus on the estimates in the first two columns. The demand-side estimates in the nonconstant returns to scale case imply elasticities and substitution patterns similar to the constant returns case. We find that the means ($\bar{\beta}$'s) on *Air* and *Size* are positive and are estimated precisely enough to be significant at traditional significance levels. The estimate of the constant is precise and negative, while the mean utility levels associated with *HPWT* and *MP*\$ are insignificantly different from zero. On the other hand, the estimate of the standard deviations of the distribution of marginal utilities for *HPWT* and *MP*\$ are substantial and estimated precisely enough to be considered significant at reasonable significance levels. Thus, each of the included attributes is estimated to have either a significantly positive effect on the mean of the distribution of utilities, or a significant positive effect on the standard deviation of that distribution (and in the case of *Size* on both). We turn next to providing some figures on the economic magnitude of these effects.

Table V presents estimates of elasticities of demand with respect to the continuous attributes, including prices. Each row in this table corresponds to a model. The top number in each cell is the actual value of that attribute for that model, while the bottom number is the elasticity of demand with respect to the attribute. For example, the Mazda 323 has a *HP/weight* ratio of 0.366 and its elasticity of market share with respect to *HP/weight* is 0.458.

The elasticities with respect to *MP*\$ illustrate the importance of considering both the mean and standard deviation of the distribution of tastes for a characteristic. The results here are quite intuitive. The elasticity of demand with respect to *MP*\$ declines almost monotonically with the car's *MP*\$ rating. While a 10 percent increase in *MP*\$ increases sales of the Mazda 323, Sentra, and Escort by about 10 percent, the demand for the cars with low *MP*\$ are actually falling with an increase in *MP*\$. The decreases, though, are quite close to zero. Hence, we conclude that consumers who purchase the high mileage cars care a great deal about fuel economy while those who purchase cars like the BMW 735i or Lexus LS400 are not concerned with fuel economy. Similarly, the demand elasticities with respect to size are generally declining as cars get larger.

The elasticity of demand with respect to *HP/weight*, our proxy for acceleration, is also small (about 0.1) for the largest cars in the sample, the Lincoln, Cadillac, Lexus, and BMW. On the other hand, it appears that consumers who purchase the smallest cars place a greater value on increased acceleration. For the Mazda 323, Sentra, and Escort, a 10 percent increase in *HP/weight* increases demand by about 4.5 percent. The relationship between the elasticities and the value of *HP/weight* is not monotonic though. For midsize cars, the elasticities are varied. The Maxima (a fairly sporty midsize car) has a relatively high elasticity (0.322) while the similarly sized but more sedate Taurus has an elasticity of 0.180.

The term on $\ln(y-p)$, α, is of the expected sign and is measured precisely enough to be highly significant. Its magnitude is most easily interpreted by

TABLE V
A SAMPLE FROM 1990 OF ESTIMATED DEMAND ELASTICITIES WITH RESPECT TO ATTRIBUTES AND PRICE (BASED ON TABLE IV (CRTS) ESTIMATES)

	Value of Attribute/Price Elasticity of demand with respect to:				
Model	*HP/Weight*	*Air*	*MP$*	*Size*	*Price*
Mazda323	0.366	0.000	3.645	1.075	5.049
	0.458	0.000	1.010	1.338	6.358
Sentra	0.391	0.000	3.645	1.092	5.661
	0.440	0.000	0.905	1.194	6.528
Escort	0.401	0.000	4.022	1.116	5.663
	0.449	0.000	1.132	1.176	6.031
Cavalier	0.385	0.000	3.142	1.179	5.797
	0.423	0.000	0.524	1.360	6.433
Accord	0.457	0.000	3.016	1.255	9.292
	0.282	0.000	0.126	0.873	4.798
Taurus	0.304	0.000	2.262	1.334	9.671
	0.180	0.000	−0.139	1.304	4.220
Century	0.387	1.000	2.890	1.312	10.138
	0.326	0.701	0.077	1.123	6.755
Maxima	0.518	1.000	2.513	1.300	13.695
	0.322	0.396	−0.136	0.932	4.845
Legend	0.510	1.000	2.388	1.292	18.944
	0.167	0.237	−0.070	0.596	4.134
TownCar	0.373	1.000	2.136	1.720	21.412
	0.089	0.211	−0.122	0.883	4.320
Seville	0.517	1.000	2.011	1.374	24.353
	0.092	0.116	−0.053	0.416	3.973
LS400	0.665	1.000	2.262	1.410	27.544
	0.073	0.037	−0.007	0.149	3.085
BMW 735i	0.542	1.000	1.885	1.403	37.490
	0.061	0.011	−0.016	0.174	3.515

Notes: The value of the attribute or, in the case of the last column, price, is the top number and the number below it is the elasticity of demand with respect to the attribute (or, in the last column, price.)

examining the elasticities and markups it, together with the other estimated coefficients, imply. We note first that the estimates imply that demands for *all* 2217 models in our sample are elastic. The last column of Table V lists prices and price elasticities of demand for our subsample of 1990 models. We find that the most elastically demanded products are those that are in the most "crowded" market segments—the compact and subcompact models. (The Buick Century is an exception to this pattern.) The Sentra and Mazda 323 face demand elasticities of 6.4 and 6.5 respectively, while the $37,490 BMW and $27,544 (in 1983 dollars) Lexus face demand elasticities of 3.5 and 3.0 respectively.

Table VI presents a sample of own and cross price semi-elasticities. Each semi-elasticity gives the percentage change in market share of the row car associated with a $1000 increase in the price of the column car. Looking down the first column, for example, we note that a thousand dollar increase in the price of a Mazda 323 increases the market share of a Nissan Sentra by .705

TABLE VI

A SAMPLE FROM 1990 OF ESTIMATED OWN- AND CROSS-PRICE SEMI-ELASTICITIES: BASED ON TABLE IV (CRTS) ESTIMATES

	Mazda 323	Nissan Sentra	Ford Escort	Chevy Cavalier	Honda Accord	Ford Taurus	Buick Century	Nissan Maxima	Acura Legend	Lincoln Town Car	Cadillac Seville	Lexus LS400	BMW 735i
323	−125.933	1.518	8.954	9.680	2.185	0.852	0.485	0.056	0.009	0.012	0.002	0.002	0.000
Sentra	0.705	−115.319	8.024	8.435	2.473	0.909	0.516	0.093	0.015	0.019	0.003	0.003	0.000
Escort	0.713	1.375	−106.497	7.570	2.298	0.708	0.445	0.082	0.015	0.015	0.003	0.003	0.000
Cavalier	0.754	1.414	7.406	−110.972	2.291	1.083	0.646	0.087	0.015	0.023	0.004	0.003	0.000
Accord	0.120	0.293	1.590	1.621	−51.637	1.532	0.463	0.310	0.095	0.169	0.034	0.030	0.005
Taurus	0.063	0.144	0.653	1.020	2.041	−43.634	0.335	0.245	0.091	0.291	0.045	0.024	0.006
Century	0.099	0.228	1.146	1.700	1.722	0.937	−66.635	0.773	0.152	0.278	0.039	0.029	0.005
Maxima	0.013	0.046	0.236	0.256	1.293	0.768	0.866	−35.378	0.271	0.579	0.116	0.115	0.020
Legend	0.004	0.014	0.083	0.084	0.736	0.532	0.318	0.506	−21.820	0.775	0.183	0.210	0.043
TownCar	0.002	0.006	0.029	0.046	0.475	0.614	0.210	0.389	0.280	−20.175	0.226	0.168	0.048
Seville	0.001	0.005	0.026	0.035	0.425	0.420	0.131	0.351	0.296	1.011	−16.313	0.263	0.068
LS400	0.001	0.003	0.018	0.019	0.302	0.185	0.079	0.280	0.274	0.606	0.212	−11.199	0.086
735*i*	0.000	0.002	0.009	0.012	0.203	0.176	0.050	0.190	0.223	0.685	0.215	0.336	−9.376

Note: Cell entries i, j, where i indexes row and j column, give the percentage change in market share of i with a \$1000 change in the price of j.

percent but has almost no effect on the market share of a Lincoln Town Car, Cadillac Seville, Lexus LS400, or a BMW 735i.

In general, Table VI shows cross-price elasticities that are large for cars with similar characteristics. Perhaps not surprisingly, the magnitudes of the effects of a $1000 price increase of the higher priced cars are much smaller than they are for the lower priced cars. The general pattern of cross-price semi-elasticities accords well with intuition. For example, the Lexus is the closest substitute (measured by magnitude of cross price semi-elasticities) to the BMW 735, the Cadillac is the closest substitute to the Lincoln, and the Accord is the closest substitute to the Taurus. Since the demand elasticities will play a crucial role in policy analysis, the sensible elasticities in Table VI are encouraging.

Next we consider the substitutability of our auto models with the "outside good," that is ds_0/dp_j. To give some idea of the magnitude of this derivative, we express it as a percentage of the absolute value of the own-price derivative:

$$\frac{100*(ds_0/dp_j)}{|ds_j/dp_j|}.$$

For a small increase in the price of product j, this gives the number of consumers who substitute from j to the outside good, as a percentage of the total number of consumers who substitute away from j. The results of this exercise are given in Table VII. There we report results concerning substitution to the outside good for our subsample of 1990 models under both the logit and the BLP specifications. The first column in Table VII indicates that for every model, about 90 percent of the consumers who substitute away from a model opt instead for the outside good. This figure is just $s_0/(1-s_j)$. The results under the BLP specification are not nearly as uniform across models. Here, the numbers still seem a bit large to us, which may point to the need for improve-

TABLE VII

SUBSTITUTION TO THE OUTSIDE GOOD

Model	Given a price increase, the percentage who substitute to the outside good (as a percentage of all who substitute away.) Logit	BLP
Mazda 323	90.870	27.123
Nissan Sentra	90.843	26.133
Ford Escort	90.592	27.996
Chevy Cavalier	90.585	26.389
Honda Accord	90.458	21.839
Ford Taurus	90.566	25.214
Buick Century	90.777	25.402
Nissan Maxima	90.790	21.738
Acura Legend	90.838	20.786
Lincoln Town Car	90.739	20.309
Cadillac Seville	90.860	16.734
Lexus LS400	90.851	10.090
BMW 735*i*	90.883	10.101

TABLE VIII

A SAMPLE FROM 1990 OF ESTIMATED PRICE-MARGINAL COST MARKUPS AND VARIABLE PROFITS: BASED ON TABLE 6 (CRTS) ESTIMATES

	Price	Markup Over MC $(p - MC)$	Variable Profits (in \$'000's) $q*(p - MC)$
Mazda 323	\$5,049	\$ 801	\$18,407
Nissan Sentra	\$5,661	\$ 880	\$43,554
Ford Escort	\$5,663	\$1,077	\$311,068
Chevy Cavalier	\$5,797	\$1,302	\$384,263
Honda Accord	\$9,292	\$1,992	\$830,842
Ford Taurus	\$9,671	\$2,577	\$807,212
Buick Century	\$10,138	\$2,420	\$271,446
Nissan Maxima	\$13,695	\$2,881	\$288,291
Acura Legend	\$18,944	\$4,671	\$250,695
Lincoln Town Car	\$21,412	\$5,596	\$832,082
Cadillac Seville	\$24,353	\$7,500	\$249,195
Lexus LS400	\$27,544	\$9,030	\$371,123
BMW 735*i*	\$37,490	\$10,975	\$114,802

ments in our treatment of the outside good (see the extensions section below). However, our estimates are much smaller than the corresponding figures for the logit model. Our results also show the expected pattern that consumers of lower priced cars are more likely to stay with the outside good when the price of their most preferred model increases.

Table VIII presents the estimated price-marginal cost markups implied by the estimates of the constant returns to scale case reported in Table IV. In 1990, the average markup is \$3,753 and the average ratio of markup to retail price is 0.239.[31] The pattern and magnitudes of the markups reported in Table VIII are quite plausible. The models with the lowest markups are the Mazda (\$801), Sentra (\$880), and Escort (\$1077). At the other extreme, the Lexus and BMW have estimated markups of \$9,030 and \$10,975 respectively. In general, markups rise almost monotonically with price.

In the third column of Table VIII, we list variable profits for each model (since marginal costs are assumed to be constant in output, variable profits are just sales multiplied by price minus marginal cost). Given our estimates, large markups do not necessarily mean large profits, as the sales of some of the high markup cars are quite small. The models that, according to our estimates, are the most profitable (by a factor of two, relative to the other models reported in the table) are the Honda Accord and the Ford Taurus. Both are widely regarded as essential to each firm's financial well-being.

It seems to us that Tables IV through VIII demonstrate that allowing more flexible utility specifications generates a more realistic picture of equilibrium in the U.S. automobile industry. Conditional on allowing for a more flexible utility specification, there are, however, a number of different variables one might

[31] Interestingly, while the pattern of markups differs considerably between the logit case and the BLP specification, the average level of markups is similar across the two sets of results.

TABLE IX

RESULTS FROM SOME ALTERNATIVE SPECIFICATIONS: PRICE-MARGINAL COST MARKUPS

	Base Case (reported in Table VIII)	Include $\ln(q)$ in cost function	Use *AT* instead of *AIR*	*Weight and HP* instead of *HP/Wt*	Include interaction terms in cost function	Use 3 region dummies and add *Reliability*	Add *weight* and include interactions in cost function
Mazda 323	\$ 801	\$1,616	\$1,012	\$1,073	\$ 828	\$1,125	\$1,389
Nissan Sentra	\$ 880	\$1,769	\$1,153	\$1,271	\$ 912	\$1,308	\$1,487
Ford Escort	\$1,077	\$2,043	\$1,326	\$1,470	\$1,111	\$2,094	\$1,690
Chevy Cavalier	\$1,302	\$2,490	\$1,729	\$1,655	\$1,329	\$2,593	\$2,020
Honda Accord	\$1,992	\$3,059	\$2,629	\$2,703	\$2,059	\$3,839	\$2,327
Ford Taurus	\$2,577	\$3,721	\$2,528	\$3,344	\$2,585	\$4,094	\$2,898
Buick Century	\$2,420	\$4,162	\$3,161	\$2,939	\$2,405	\$4,030	\$3,321
Nissan Maxima	\$2,881	\$4,674	\$4,565	\$2,085	\$2,911	\$6,941	\$3,513
Acura Legend	\$4,671	\$7,105	\$6,563	\$3,059	\$4,661	\$8,305	\$5,081
Lincoln Town Car	\$5,596	\$8,029	\$6,778	\$4,765	\$5,508	\$7,114	\$6,518
Cadillac Seville	\$7,500	\$10,733	\$8,635	\$4,863	\$7,439	\$9,182	\$8,015
Lexus LS400	\$9,030	\$10,510	\$8,411	\$4,791	\$8,585	\$10,925	\$7,398
BMW 735i	\$10,975	\$13,646	\$9,122	\$7,605	\$10,713	\$12,153	\$12,202
No. of demand side variables significant at 95% level[a]	5 of 5	4 of 5	4 of 5	6 of 6	4 of 5	8 of 8	4 of 6

[a] A demand side variable is considered significant if *either* its mean or standard deviation (σ_β) is significant. See text for details.

include in the utility and cost functions. We now ask how sensitive our results are to our admittedly ad hoc choice of included variables. Table IX begins to address this issue.

There are many ways one might summarize the implications of the estimated parameters. We choose to report the estimated price-marginal cost markups that result from alternative specifications, since these markups embody information from both the cost and demand sides of the model, and they are easily interpretable. The first column of Table IX replicates the results in Table VIII and is included to make comparisons more convenient. In the second column, we report the markups that result when we include the natural log of output in the cost function. The vector of other cost-shifters, w, is unchanged from the base case. This is the specification reported in the last 2 columns of Table IV and, as previously noted, the quantity variable is problematic. Nonetheless, the markups follow the same pattern in the base case. The main difference is that the markups are uniformly higher. This results from the decreasing returns to scale. The markups over average variable cost (not reported) are much lower. Indeed, without higher order terms in $\ln(q)$ entering the cost function, the markups over average variable cost are implausibly low. Of all the alternate specifications we investigated, this one yielded the highest price-marginal cost markups, and yet even these markups are not extraordinarily high. For this and all the other alternate specifications, we also report the number of demand side variables whose means or σ's are significantly different from zero at standard levels.

In the Table IV results, the σ associated with air conditioning was not significantly different from zero. We believe the *AIR* variable is proxying for a degree of luxury. It is possible that there really is little disagreement in the population about this attribute, but perhaps it is a poor proxy. In column 3 of Table IX, we report the markups that result from using another proxy—whether automatic transmission is standard equipment. The pattern and magnitudes of the markups are quite similar to the base case results. Markups are slightly higher for the less expensive cars and slightly lower for the high-end cars, but not dramatically so.

In the fourth column of Table IX, we report the results from a specification that replaces the ratio of horsepower to weight with the two variables entered separately and linearly. Of all the alternative specifications investigated, this one gave the largest change in estimated markups. While the patterns of markups is the same, this specification gave implausibly low markups for the more costly cars. This might result if cost were not linear in horsepower and weight, since these cars have large values of each attribute, hence forcing marginal cost to be higher than it perhaps actually is.

In our model, adding additional terms to the cost function is computationally cheap, while adding additional demand side random coefficients is computationally demanding. In column 5 of Table IX, we include interaction terms in the cost function between all the continuous characteristics. This captures the notion that the cost of a characteristic may depend on the level of another characteristic. The results of this exercise give markups very similar to our base case results. For most models, the markups are within a few percent of one another. We found that most interaction terms were statistically significant at the usual levels and the elasticities of marginal cost with respect to the continuous attributes were virtually identical to those that resulted with no interaction terms in the cost function. Further, the parameters associated with one of the five demand side variables was no longer significantly different from zero.

In the sixth column of Table IX, we report the markups that result when we replace the constant in the utility function with a set of dummy variables indicating whether the car was built by a firm from the U.S., Japan, or Europe. We also include the Consumer Reports reliability rating. Problems with this variable are noted above, but we include it in this particular specification because we suspect that the region dummies may be highly correlated with reliability. If we did not include a measure of reliability, it would mean that an instrument would be correlated with the unobservables, contrary to the assumptions we need for the consistency of our estimator. In this specification there is a separate mean and variance for the dummy associated with each region. Once again, the markups exhibit the same pattern as in the other specifications. We do find, though, that the markups for a number of the models in the middle of the price range are substantially higher. Since these models are not from just one region, it is not clear what drives this change.

In the final column of Table IX, we report the results when we add *weight* to the list of regressors (instead of sufficing with the ratios of horsepower to weight), and then allow for interactions in all the cost side variables. Here the linear coefficient of the *weight* variable came in insignificant on the cost side with a significant mean and insignificant standard deviation on the demand side. The pattern and magnitude of markups was quite similar to the base case results.

8. APPLICATIONS, PROBLEMS, AND EXTENSIONS

8.1. *Applications*

Our model is defined in terms of four primitives and a Nash equilibrium assumption in prices. The primitives are the utility surface that assigns values to different possible combinations of product characteristics as a function of consumer characteristics, a cost function which determines the production cost associated with different combinations of product characteristics, a distribution of consumer characteristics, and a distribution of product characteristics. Conditional on these primitives the model can solve for the distribution of prices, quantities, variable profits, and consumer welfare. There are, therefore, at least two ways one might use the estimated parameters. One is to investigate changes in one of the primitives *assuming* that the others are held fixed, while the other is to determine the extent that changes in the various primitives can account for historical movements in the data. The first corresponds to traditional policy analysis, while the second provides an interpretation of the changes that have occurred in the industry.

It is easy to list policy questions that our estimates might be used to help analyze. These include: trade policy (e.g., the effect of import restrictions), merger policy, environmental policy (e.g., carbon and gas guzzler taxes as well as Auto Emission and Corporate Average Fuel Efficiency Standards) and the construction of price indices. For a start on these issues, see Berry, Levinsohn, and Pakes (1994) and Berry and Pakes (1993) for the first two and Pakes, Berry, and Levinsohn (1993) for the last two. Demand elasticities play a crucial role in each of these issues and hence the methods developed in this paper might provide more realistic analyses than some more traditional models.

On the other hand, all of the models, including our own, are limited in that they provide only a "conditional" analysis of each issue. That is, to do policy analysis we will have to perturb a small number of parameters and compute new equilibria conditional on the other primitives of the model remaining unchanged. In fact in many cases these other "primitives" will change in response to a change in policy or in the environment.

For example, Pakes, Berry, and Levinsohn (1993) used our model's estimates to predict the effect of the 1973 gas price hike on the average MPG of new cars sold in subsequent years. We found that our model predicted 1974 and 1975 average MPG almost exactly. This is because the characteristics of cars, treated as fixed in our predictions, did not change much in the first two years after the gas price hike and our model did well in predicting responses conditional on the characteristics of cars sold. However, by 1976 new small fuel efficient models began to be introduced and our predictions, based on fixed characteristics, became markedly worse and deteriorated further over time. We return to the problem of endogenizing characteristics in the next subsection.

8.2. *Extensions*

Our methods have been developed on the premise that consumer and producer level data are not always available. This seems an important concession to the realities facing empirical researchers investigating many, but not all, markets. We do note that information on the distribution of many of the relevant consumer characteristics is generally available and we illustrate how to make use of the empirical distribution of this information in the estimation algorithm. (In addition to income, consumer characteristics that might be expected to interact with product attributes and for which distributional information is available include household size, geographic region in which the household resides, and age of head of household.)

There are, however, several industries in which some consumer and/or producer-level micro data are available, and the auto industry is one of them. Though production costs for autos are not publicly available at the product-level, the Longitudinal Research Data (LRD) maintained by the Bureau of the Census do contain plant-level cost data. Since industry publications link automotive models to specific plants, we are exploring the possibility of using this information to improve our estimates. Note that separate information on costs would allow for a more detailed examination of the relationship of prices to marginal costs, and, therefore, for a more detailed analysis of the nature of the appropriate equilibrium in the spot market for current output. The cost information would also enable a more flexible analysis of functional forms for the cost surface, and, perhaps, an analysis of how that surface has changed over time in response to changes in both R&D investments and in government policies. As noted above, there is also consumer survey information on automobile purchases and we are investigating how to integrate survey data with the aggregate data used here.

The other, perhaps more important and certainly more difficult, direction for future work is incorporating a realistic treatment of dynamics. On the producer side there are two aspects of this problem. The first and possibly easier one is obtaining consistent estimates of the parameters of the static profit function while allowing for a correlation between observed and unobserved characteristics. This correlation may result from the fact that both sets of characteristics are, in part, determined by related decision-making processes. The second, and richer, part of the problem is to endogenize the actual choice of the characteristics of the models marketed. Even the more detailed models of dynamic industry equilibrium (see, for example, the theory in Ericson and Pakes (1995) and the computational algorithm in Pakes and McGuire (1994)) still have to be enriched before we can provide a realistic approximation to the multiproduct, multi-characteristic nature of the auto industry.

On the consumer side, a complete model of dynamic decision making would incorporate both the transaction costs of buying and selling a car and uncertainty about the future. In particular, a dynamic model of consumer decision-making would highlight the important role played by our outside alternative, which for many consumers is simply an older model car. Treating the outside alternative in a realistic way would require building a demand system for durable goods and incorporating a used car market.

Dept. of Economics, Yale University, 37 Hillhouse Ave., New Haven, CT 06520, U.S.A., and National Bureau of Economic Research,

Dept. of Economics, University of Michigan, 319 Lorch Hall, Ann Arbor, MI 48109, U.S.A., and National Bureau of Economic Research,

and

Dept. of Economics, Yale University, 37 Hillhouse Ave., New Haven, CT 06520, U.S.A., and National Bureau of Economics Research.

Manuscript received November, 1992; final revision received September, 1994.

APPENDIX I: THE CONTRACTION MAPPING

In this appendix, we will establish the contraction argument used in the computational algorithm. We will show that the function (6.8) has a unique fixed point. Furthermore, we want to establish that (6.8) is a contraction mapping. In fact, our proof will require us to impose an upper bound on the value taken by the function in (6.8), although in practice we never had to impose this bound. (This upper bound appears in the definition of $\hat{f}$ in the statement of the following theorem.)

THEOREM: *Consider the metric space* $(\boldsymbol{R}^K, d)$ *with* $d(x,y) = \|x-y\|$ (*where* $\|\cdot\|$ *is the sup-norm*). *Let* $f: \boldsymbol{R}^K \to \boldsymbol{R}^K$ *have the properties*:

(1) $\forall x \in \boldsymbol{R}^K$, $f(x)$ *is continuously differentiable, with*, $\forall j$ *and* k,

$$\partial f_j(x)/\partial x_k \geq 0$$

and

$$\sum_{k=1}^{K} \partial f_j(x)/\partial x_k < 1.$$

(2) $\min_j \inf_x f(x) \equiv \underline{x} > -\infty$.

(3) *There is a value*, $\bar{x}$, *with the property that if for any* j, $x_j \geq \bar{x}$, *then for some* k (*not necessarily equal to* j), $f_k(x) < x_k$.

Then, there is a unique fixed point, x_0, *to* f *in* $\boldsymbol{R}^K$. *Further, let the set* $X = [\underline{x}, \bar{x}]^K$, *and define the truncated function*, $\hat{f}: X \to X$, *as* $\hat{f}_j(x) = \min\{f_j(x), \bar{x}\}$. *Then*, $\hat{f}(x)$ *is a contraction of modulus less than one on* X.

PROOF: We will first show the contraction mapping property that $\exists \beta < 1$ such that $\forall x$ and $x' \in X$, $\|\hat{f}(x) - \hat{f}(x')\| \leq \beta \|x - x'\|$. To see this, choose any x and x' in X and define the scalar $\lambda = \|x - x'\|$. Consider the jth element of $\hat{f}$, $\hat{f}_j(x)$ and WLOG assume $\hat{f}_j(x') - \hat{f}_j(x) \geq 0$. Then, $x + \lambda \geq x'$ implies

$$\hat{f}_j(x') - \hat{f}_j(x) \leq \hat{f}_j(x+\lambda) - \hat{f}_j(x) \leq f_j(x+\lambda) - f_j(x) = \int_0^{\lambda} \left[\sum_{k=1}^{K} \partial f_j(x+z)/\partial x_k \right] dz \leq \beta\lambda,$$

where

$$\beta \equiv \max_j \max_{x \in W} \sum_{k=1}^{K} \partial f_j(x)/\partial x_k$$

and the set W is defined as

$$W=\{y \in \boldsymbol{R}^K : y=(x+z), x \in X, z \in [0, \bar{x}-\underline{x}]\}.$$

The second inequality follows from the fact that $\hat{f}_j(x+\lambda) \leq f_j(x+\lambda)$, while $\hat{f}_j(x)=f_j(x)$. The scalar β exists, as it is the maximum of a continuous function over a compact set. β is the maximum value of the integrand over the set of $(x+z)$ values that can possibly be reached when $x \in X$ and the scalar z is less than the possible difference between any two points in the set X. The final inequality and the fact that $\beta<1$ follow from Assumption (1).

We have now established that $\hat{f}$ is a contraction of modulus $\beta<1$ on X. Therefore, there is a unique fixed point, x_o, to $\hat{f}$ on X and for any x in X, the sequence $\hat{f}^n(x)$ converges to x_o. Assumptions (2)–(3) rule out the existence of fixed points to either f or $\hat{f}$ that are outside the interior of X. Thus, x_o cannot be on the boundary of X; x_o is a fixed point of f and there can be no other fixed point to f. *Q.E.D.*

We will now show that the function $f(\delta)=\delta+\ln(s)-\ln(s(\delta))$ satisfies the hypotheses of the theorem. The function f is differentiable by the differentiability of the function $s(\delta)$. To check the monotonicity condition of Assumption 1 note that

$$\partial f_j(\delta)/\partial \delta_j = 1-\frac{1}{s_j}\frac{\partial s_j}{\partial \delta_j},$$

while for $k \neq j$,

$$\partial f_j(\delta)/\partial \delta_k = -\frac{1}{s_j}\frac{\partial s_j}{\partial \delta_k}.$$

By differentiating our specific market share function, it is easy to show that both $\partial f_j/\partial \delta_j$ and $\partial f_j/\partial \delta_k$ are positive and that $\sum_{k=1}^{J} \partial s_j/\partial \delta_k < s_j$. This in turn establishes that the derivatives of f sum to less than one, establishing all the conditions of Assumption 1.

It is easy to find the lower bound for f (Assumption 2). First note that we can rewrite $s_i(\delta)$ as

$$s_j(\delta)=e^{\delta_j}D_j(\delta), \qquad \text{where}$$

$$D_j(\delta) \equiv \int \frac{e^{\mu_i}}{1+\sum_k e^{\delta_k+\mu_i}}\, d\Phi(\mu).$$

Plugging this into the definition of f gives

$$f_j(\delta)=\ln(s_j)-\ln(D_j(\delta)).$$

Note that D_j is declining in all the δ_k. As all of the δ_k approach $-\infty$, $D_j(\delta)$ goes to $\int e^{\mu_i}\,d\Phi(\mu)$. Thus a lower bound for f_j is $\underline{\delta}_j \equiv \ln(s_j)-\ln(\int e^{\mu_i}\,d\Phi(\mu))$. This is the value of δ_j that would explain a market share for good j of s_j if all the other market shares (other than the outside good) were equal to zero.

Unfortunately, $f(\delta)$ is increasing in δ_j without bound. Berry (1994) does, however, show how to establish the existence of a value, $\bar{\delta}$, such that if any element of δ is greater than $\bar{\delta}$, then there is some k such that $s_k(\delta)>s_k$. The vector with each element equal to $\bar{\delta}$ then satisfies the requirements of $\bar{x}$ in Assumption (3), for if $s_k(\delta)>s_k$, then $f_k(\delta)<\delta_k$.

Berry (1994) shows that an appropriate $\bar{\delta}$ is found as follows. For product j, define $\bar{\delta}_j$ as the value of $-\delta_j$ that would explain the market share of the outside good, s_0, when $\delta_o=0$ and all the other $\delta_k=-\infty$. Then set $\bar{\delta}>\max_j \bar{\delta}_j$.

REFERENCES

ANDERSON, S., A. DE PALMA, AND F. THISSE (1989): "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach," *Review of Economics Studies*, 56, 21–35.

BEN-AKIVA, M. (1973): "Structure of Passenger Travel Demand Models," Ph.D. Dissertation, Department of Civil Engineering, MIT.

BERKOVEC, J. (1985): "New Car Sales and Used Car Stocks: A Model of the Automobile Market," *RAND Journal of Economics*, 16, 195–214.

BERKOVEC, J., AND J. RUST (1985): "A Nested Logit Model of Automobile Holdings for One Vehicle Households," *Transportation Research*, 19B, 275–285.

BERRY, S. (1991): "Airport Presence as Product Differentiation," *American Economic Review, Papers and Proceedings*, 80, 394–399.

——— (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1993): "Automobile Prices in Market Equilibrium," NBER Working Paper #4264.

——— (1994): "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," mimeo.

BERRY, S., AND A. PAKES (1993): "Some Applications and Limitations of Recent Advances in Empirical Industrial Organization: Merger Analysis," *American Economic Review, Papers and Proceedings*, 83, 247–252.

——— (in process): "The Limit Properties of an Estimator for Differentiated Products Demand Systems as the Number of Products Grows Large," mimeo, Yale University.

BRESNAHAN, T. (1981): "Departures from Marginal-Cost Pricing in the American Automobile Industry," *Journal of Econometrics*, 17, 201–227.

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

CAPLIN, A., AND B. NALEBUFF (1991): "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 59, 26–61.

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 33, 305–334.

DEBREU, G. (1960): "Review of R. D. Luce Individual Choice Behavior," *American Economic Review*, 50, 186–188.

ERICSON, R., AND A. PAKES (1995): "Markov Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53–82.

FEENSTRA, R., AND J. LEVINSOHN (1995): "Estimating Markups and Market Conduct with Multidimensional Product Attributes," *Review of Economic Studies*, 62, 19–52.

GOLDBERG, P. (1993): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," Mimeo, Princeton University.

GRILICHES, Z. (1971): "Hedonic Price Indices for Automobiles: an Econometric Analysis of Quality Change," in *Price Indices and Quality Change*, ed. by Z. Griliches. Cambridge: Harvard University Press.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HASEMAN, J., AND L. KUPER (1979): "Analysis of Dichotomous Response Data from Certain Toxicological Experiments," *Biometrics*, 35, 281–293.

HAUSMAN, J. (1975): "Project Independence Report: An Appraisal of U.S. Energy Needs Up to 1985," *BELL Journal of Economics*, 6, 517–551.

HOTZ, V. J., AND R. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

KINGMAN, J. F. C. (1978): "The 1977 Wald Memorial Lectures," *The Annals of Probability*, 6, 183–197.

LANCASTER, K. J. (1971): *Consumer Demand: A New Approach*. New York: Columbia University Press.

LEVINSOHN, J. (1988): "Empirics of Taxes on Differentiated Products: The Case of Tariffs in the U.S. Automobile Industry," in *Trade Policy Issues and Empirical Analysis*, ed. by R. Baldwin. Chicago: University of Chicago Press.

MANSKI, C. (1983): "Analysis of Equilibrium Automobile Holdings in Israel with Aggregate Discrete Choice Models," *Transportation Research*, 17B, 373–389.

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers of Econometrics*, ed. by P. Zarembka. New York: Academic Press.

——— (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlvist, et al. Amsterdam: North Holland, 75–96.

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

——— (1986): "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics, Volume III*, ed. by Z. Griliches and M. Intriligator. Amsterdam: North Holland.

McFadden, D., A. Talvitie, and Associates (1977): *Demand Model Estimation and Validation*. Berkeley: The Institute of Transportation Studies, University of California, Berkeley and Irvine.
Morgan, M. (1990): *The History of Econometric Ideas*. New York: Cambridge University Press.
Morrison, S., and C. Winston (1989): "Enhancing the Performance of the Deregulated Air Transportation System," *Brookings Papers on Economic Activity: Microeconomics*, 2, 61–112.
Nelder, J., and R. Mead (1965): "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–313.
Newey, W. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–839.
Ohta, M., and Z. Griliches (1976): "Automobile Prices Revisited: Extensions of the Hedonic Hypothesis," in *Household Production and Consumption*, ed. by N. Terlecky. New York: NBER.
Pakes, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.
——— (1994): "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in *Advances in Econometrics: The Sixth World Congress of the Econometric Society, Vol. II*, ed. by J.-J. Laffont and C. Sims. New York: Cambridge University Press, pp. 171–260.
Pakes, A., S. Berry, and J. Levinsohn (1993): "Applications and Limitations of Some Recent Advances in Empirical Industrial Organization: Price Indexes and the Analysis of Environmental Change," *American Economic Review, Papers and Proceedings*, 83, 240–246.
Pakes, A., and P. McGuire (1994): "Computation of Markov Perfect Nash Equilibria I: Numerical Implications of a Dynamic Product Model," *RAND Journal of Economics*, 25, 555–589.
Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.
Perloff, J., and S. Salop (1985): "Equilibrium with Product Differentiation," *Review of Economic Studies*, 52, 107–120.
Robinson (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.
Rubinstein, R. (1981): *Simulation and the Monte Carlo Method*. New York: Wiley.
Sattinger, M. (1984): "Value of an Additional Firm in Monopolistic Competition," *Review of Economic Studies*, 49, 3–13.
Schmalensee, R. (1985): "Econometric Diagnosis of Competitive Localization," *International Journal of Industrial Organization*, 57–70.
Shaked, A., and J. Sutton (1982): "Relaxing Price Competition Through Product Differentiation," *Review of Economic Studies*, 49, 3–13.
Suits, D. (1958): "The Demand for Automobiles in the United States," *Review of Economics and Statistics*, 40, 273–280.
Toder, E., with N. Cardell and E. Burton (1978): *Trade Policy and the U.S. Automobile Industry*. New York: Praeger.
Train, K. (1986): *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. Cambridge: MIT Press.
Williams, D. (1982): "Extra-binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144–148.
Working, E. (1926): "What Do Statistical Demand Curves Show?" *Quarterly Journal of Economics*, 41, 212–235.