

A PRACTITIONER'S GUIDE TO ESTIMATION OF RANDOM-COEFFICIENTS LOGIT MODELS OF DEMAND

AVIV NEVO

*University of California–Berkeley, Berkeley,
CA 94720-3880 and
NBER*

*Estimation of demand is at the heart of many recent studies that examine questions of market power, mergers, innovation, and valuation of new brands in differentiated-products markets. This paper focuses on one of the main methods for estimating demand for differentiated products: **random-coefficients logit models**. The paper carefully discusses the latest innovations in these methods with the hope of increasing the understanding, and therefore the trust among researchers who have never used them, and reducing the difficulty of their use, thereby aiding in realizing their full potential.*

1. INTRODUCTION

Estimation of demand has been a key part of many recent studies examining questions regarding market power, mergers, innovation, and valuation of new brands in differentiated-product industries.¹ This paper explains the random-coefficients (or mixed) logit methodology for estimating demand in differentiated-product markets using

An earlier version of this paper circulated under the title “A Research Assistant’s Guide to Random Coefficients Discrete Choice Model of Demand.” I wish to thank Steve Berry, Iain Cockburn, Bronwyn Hall, Ariel Pakes, various lecture and seminar participants, and an anonymous referee for comments, discussions, and suggestions. Financial support from the UC Berkeley Committee on Research Junior Faculty Grant is gratefully acknowledged.

1. Just to mention some examples, Bresnahan (1987) studies the 1955 price war in the automobile industry; Gasmi et al. (1992) empirically study collusive behavior in a soft-drink market; Hausman et al. (1994) study the beer industry; Berry et al. (1995, 1999) examine equilibrium in the automobile industry and its implications for voluntary trade restrictions; Goldberg (1995) uses estimates of the demand for automobiles to investigate trade policy issues; Hausman (1996) studies the welfare gains generated by a new brand of cereal; Berry et al. (1996) study hubs in the airline industry; Bresnahan et al. (1997) study rents from innovation in the computer industry; Nevo (2000a, b) examines price competition and mergers in the ready-to-eat cereal industry; Davis (1998) studies spatial competition in movie theaters; and Petrin (1999) studies the welfare gains from the introduction of the minivan.

market-level data. This methodology can be used to estimate the demand for a large number of products using market data and allowing for the endogeneity of price. While this method retains the benefits of alternative discrete-choice models, it produces more realistic demand elasticities. With better estimates of demand, we can, for example, better judge market power, simulate the effects of mergers, measure the benefits from new goods, or formulate innovation and competition policy. This paper carefully discusses the recent innovations in these methods with the intent of reducing the barriers to entry and increasing the trust in these methods among researchers who are not familiar with them.

Probably the most straightforward approach to specifying demand for a set of closely related but not identical products is to specify a system of demand equations, one for each product. Each equation specifies the demand for a product as a function of its own price, the price of other products, and other variables. An example of such a system is the linear expenditure model (Stone, 1954), in which quantities are linear functions of all prices. Subsequent work has focused on specifying the relation between prices and quantities in a way that is both flexible (i.e., allows for general substitution patterns) and consistent with economic theory.²

Estimating demand for differentiated products adds two additional nontrivial concerns. The first is the large number of products, and hence the large number of parameters to be estimated. Consider, for example, a constant-elasticity or log-log demand system, in which logarithms of quantities are linear functions of logarithms of all prices. Suppose we have 100 differentiated products; then without additional restrictions this implies estimating at least 10,000 parameters (100 demand equations, one for each product, with 100 prices in each). Even if we impose symmetry and adding up restrictions, implied by economic theory, the number of parameters will still be too large to estimate them. The problem becomes even harder if we want to allow for more general substitution patterns.

An additional problem, introduced when estimating demand for differentiated products, is the heterogeneity in consumer tastes: If all consumers are identical, then we would not observe the level of differentiation we see in the marketplace. One could assume that preferences are of the right form [the Gorman form: see Gorman (1959)], so that an aggregate, or average, consumer exists and has a demand function that satisfies the conditions specified by economic theory.³

2. Examples include the Rotterdam model (Theil, 1965; Barten, 1966), the translog model (Christensen et al., 1975), and the almost ideal demand system (Deaton and Muellbauer, 1980).

3. For an example of a representative consumer approach to demand for differentiated products see Dixit and Stiglitz (1977) or Spence (1976).

However, the required assumptions are strong and for many applications seem to be empirically false. The difference between an aggregate model and a model that explicitly reflects individual heterogeneity can have profound affects on economic and policy conclusions.

The logit demand model (McFadden, 1973)⁴ solves the dimensionality problem by projecting the products onto a space of characteristics, making the relevant size the dimension of this space and not the square of the number of products. A problem with this model is the strong implication of some of the assumptions made. Due to the restrictive way in which heterogeneity is modeled, substitution between products is driven completely by market shares and not by how similar the products are. Extensions of the basic logit model relax these restrictive assumptions, while maintaining the advantage of the logit model in dealing with the dimensionality problem. The essential idea is to explicitly model heterogeneity in the population and estimate the unknown parameters governing the distribution of this heterogeneity. These models have been estimated using both market- and individual-level data.⁵ The problem with the estimation is that it treats the regressors, including price, as exogenously determined. This is especially problematic when aggregate data is used to estimate the model.

This paper describes recent developments in methods for estimating random-coefficients discrete-choice models of demand [Berry, 1994; Berry et al., 1995 (henceforth BLP)]. The new method maintains the advantage of the logit model in handling a large number of products. It is superior to prior methods because (1) the model can be estimated using only market-level price and quantity data, (2) it deals with the endogeneity of prices, and (3) it produces demand elasticities that are more realistic—for example, cross-price elasticities are larger for products that are closer together in terms of their characteristics.

The rest of the paper is organized as follows. Section 2 describes a model that encompasses, with slight alterations, the models previously used in the literature. The focus is on the various modeling assumptions and their implications for estimation and the results. In Section 3 I discuss estimation, including the data required, an outline of the algorithm, and instrumental variables. Many of the nitty-gritty details of estimation are described in an appendix (available from

4. A related literature is the characteristics approach to demand, or the address approach (Lancaster, 1966, 1971; Quandt, 1968; Rosen, 1974). For a recent exposition of it and a proof of its equivalence to the discrete choice approach see Anderson et al.

5. For example, the generalized extreme-value model (McFadden, 1978) and the random-coefficients logit model (Cardell and Dunbar, 1980; Boyd and Mellman, 1980; Tardiff, 1980; Cardell, 1989; and references therein). The random-coefficients model is often called the hedonic demand model in this earlier literature; it should not be confused with the hedonic price model (Court, 1939; Griliches, 1961).

<http://elsa.berkeley.edu/~nevo>). Section 4 provides a brief example of the type of results the estimation can produce. Section 5 concludes and discusses various extensions and alternatives to the method described here.

2. THE MODEL

In this section I discuss the model with an emphasis on the various modeling assumptions and their implications. In the next section I discuss the estimation details. However, for now I want to stress two points. First, the method I discuss here uses (market-level) price and quantity data for each product, in a series of markets, to estimate the model. Some information regarding the distribution of consumer characteristics might be available, but a key benefit of this methodology is that we do not need to observe individual consumer purchase decisions to estimate the demand parameters.⁶

Second, the estimation allows prices to be correlated with the econometric error term. This will be modeled in the following way. A product will be defined by a set of characteristics. Producers and consumers are assumed to observe all product characteristics. The researcher, on the other hand, is assumed to observe only some of the product characteristics. Each product will be assumed to have a characteristic that influences demand but that either is not observed by the researcher or cannot be quantified into a variable that can be included in the analysis. Examples are provided below. The unobserved characteristics will be captured by the econometric error term. Since the producers know these characteristics and take them into account when setting prices, this introduces the econometric problem of endogenous prices.⁷ The contribution of the estimation method presented below is to transform the model in such a way that instrumental-variable methods can be used.

2.1 THE SETUP

Assume we observe $t = 1, \dots, T$ markets, each with $i = 1, \dots, I_t$ consumers. For each such market we observe aggregate quantities,

6. If individual decisions are observed, the method of analysis differs somewhat from the one presented here. For clarity of presentation I defer discussion of this case to Section 5.

7. The assumption that when setting prices firms take account of the unobserved (to the econometrician) characteristics is just one way to generate correlation between prices and these unobserved variables. For example, correlation can also result from the mechanics of the consumer's optimization problem (Kennan, 1989).

average prices, and product characteristics for J products.⁸ The definition of a market will depend on the structure of the data. BLP use annual automobile sales over a period of twenty years, and therefore define a market as the national market for year t , where $t = 1, \dots, 20$. On the other hand, Nevo (2000a) observes data in a cross section of cities over twenty quarters, and defines a market t as a city-quarter combination, with $t = 1, \dots, 1124$. Yet a different example is given by Das et al. (1994), who observe sales for different income groups, and define a market as the annual sales to consumers of a certain income level.

The indirect utility of consumer i from consuming product j in market t , $U(x_{jt}, \xi_{jt}, p_{jt}, \tau_i; \theta)$,⁹ is a function of observed and unobserved (by the researcher) product characteristics, x_{jt} and ξ_{jt} respectively; price, p_{jt} ; individual characteristics, τ_i ; and unknown parameters, θ . I focus on a particular specification of this indirect utility,¹⁰

$$u_{ijt} = \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt},$$

$$i = 1, \dots, I_t, \quad j = 1, \dots, J, \quad t = 1, \dots, T, \quad (1)$$

where y_i is the income of consumer i , p_{jt} is the price of product j in market t , x_{jt} is a K -dimensional (row) vector of observable characteristics of product j , ξ_{jt} is the unobserved (by the econometrician) product characteristic, and ε_{ijt} is a mean-zero stochastic term. Finally, α_i is consumer i 's marginal utility from income, and β_i is a K -dimensional (column) vector of individual-specific taste coefficients.

Observed characteristics vary with the product being considered. BLP examine the demand for cars, and include as observed characteristics horsepower, size and air conditioning. In estimating demand for ready-to-eat cereal Nevo (2000a) observes calories, sodium, and fiber content. Unobserved characteristics, for example, can include the impact of unobserved promotional activity, unquantifiable factors (brand equity), or systematic shocks to demand. Depending on the structure of the data, some components of the unobserved characteristics can be captured by dummy variables. For example, we can model $\xi_{jt} = \xi_j + \xi_t + \Delta\xi_{jt}$ and capture ξ_j and ξ_t by brand- and market-specific dummy variables.

Implicit in the specification given by equation (1) are three things. First, this form of the indirect utility can be derived from a

8. For ease of exposition I have assumed that all products are offered to all consumers in all markets. The methods described below can easily deal with the case where the choice set differs between markets and also with different choice sets for different consumers.

9. This is sometimes called the conditional indirect utility, i.e., the indirect utility conditional on choosing this option.

10. The methods discussed here are general and with minor adjustments can deal with different functional forms.

quasilinear utility function, which is free of wealth effects. For some products (for example, ready-to-eat cereals) this is a reasonable assumption, but for other products (for example, cars) it is an unreasonable one. Including wealth effects alters the way the term $y_i - p_{jt}$ enters equation (1). For example, BLP build on a Cobb-Douglas utility function to derive an indirect utility that is a function of $\log(y_i - p_{jt})$. In principle, we can include $f(y_i - p_{jt})$, where $f(\cdot)$ is a flexible functional form (Petrin, 1999).

Second, equation (1) specifies that the unobserved characteristic, which among other things captures the elements of vertical product differentiation, is identical for all consumers. Since the coefficient on price is allowed to vary among individuals, this is consistent with the theoretical literature of vertical product differentiation. An alternative is to model the distribution of the valuation of the unobserved characteristics, as in Das et al. (1994). As long as we have not made any distributional assumptions on consumer-specific components (i.e., anything with subscript i), their model is not more general. Once we make such assumptions, their model has slightly different implications for some of the normalizations usually made. An exact discussion of these implications is beyond the scope of this paper.

Finally, the specification in equation (1) assumes that all consumers face the same product characteristics. In particular, all consumers are offered the same price. Depending on the data, if different consumers face different prices, using either a list or average transaction price will lead to measurement error bias. This just leads to another reason why prices might be correlated with the error term and motivates the instrumental-variable procedure discussed below.¹¹

The next component of the model describes how consumer preferences vary as a function of the individual characteristics, τ_i . In the context of equation (1) this amounts to modeling the distribution of consumer taste parameters. The individual characteristics consist of two components: demographics, which I refer to as *observed*, and additional characteristics, which I refer to as *unobserved*, denoted D_i and v_i respectively. Given that no individual data is observed, neither component of the individual characteristics is directly observed in the choice data set. The distinction between them is that even though we do not observe individual data, we know something about the distribution of the demographics, D_i , while for the additional characteristics, v_i , we have no such information. Examples of demographics are income, age, family size, race, and education. Examples of the type of

11. However, as noted by Berry (1994), the method proposed below can deal with measurement error only if the variable measured with error enters in a restrictive way. Namely, it only enters the part of utility that is common to all consumers, δ_j in equation (3) below.

information we might have is a large sample we can use to estimate some feature of the distribution (e.g., we could use Census data to estimate the mean and standard deviation of income). Alternatively, we might have a sample from the joint distribution of several demographic variables (e.g., the Current Population Survey might tell us about the joint distribution of income, education, and age in different cities in the US). The additional characteristics, v_i , might include things like whether the individual owns a dog, a characteristic that might be important in the decision of which car to buy, yet even very detailed survey data will usually not include this fact.

Formally, this will be modeled as

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \quad v_i \sim P_v^*(v), \quad D_i \sim \hat{P}_D^*(D), \quad (2)$$

where D_i is a $d \times 1$ vector of demographic variables, v_i captures the additional characteristics discussed in the previous paragraph, $P_v^*(\cdot)$ is a parametric distribution, $\hat{P}_D^*(\cdot)$ is either a nonparametric distribution known from other data sources or a parametric distribution with the parameters estimated elsewhere, Π is a $(K+1) \times d$ matrix of coefficients that measure how the taste characteristics vary with demographics, and Σ is a $(K+1) \times (K+1)$ matrix of parameters.¹² If we assume that $P_v^*(\cdot)$ is a standard multivariate normal distribution, as I do in the example below, then the matrix Σ allows each component of v_i to have a different variance and allows for correlation between these characteristics. For simplicity I assume that v_i and D_i are independent. Equation (2) assumes that demographics affect the distribution of the coefficients in a fairly restrictive linear way. For those coefficients that are most important to the analysis (e.g., the coefficients on price), relaxing the linearity assumption could have important implications [for example, see the results reported in Nevo (2000a, b)].

As we will see below, the way we model heterogeneity has strong implications for the results. The advantage of letting the taste parameters vary with the observed demographics, D_i is twofold. First, it allows us to include additional information, about the distribution of demographics, in the analysis. Furthermore, it reduces the reliance on parametric assumptions. Therefore, instead of letting a key element of the method, the distribution of the random coefficients, be determined by potentially arbitrary distributional assumptions, we bring in additional information.

12. To simplify notation I assume that all characteristics have random coefficients. This need not be the case. I return to this in the appendix, when I discuss the details of estimation.

The specification of the demand system is completed with the introduction of an *outside good*: the consumers may decide not to purchase any of the brands. Without this allowance, a homogenous price increase (relative to other sectors) of all the products does not change quantities purchased. The indirect utility from this outside option is

$$u_{i0t} = \alpha_i y_i + \xi_{0t} + \pi_0 D_i + \sigma_0 v_{i0} + \varepsilon_{i0t}.$$

The mean utility from the outside good, ξ_{0t} , is not identified (without either making more assumptions or normalizing one of the inside goods). Also, the coefficients π_0 and σ_0 are not identified separately from coefficients on an individual-specific constant term in equation (1). The standard practice is to set ξ_{0t} , π_0 , and σ_0 to zero, and since the term $\alpha_i y_i$ will eventually vanish (because it is common to all products), this is equivalent to normalizing the utility from the outside good to zero.

Let $\theta = (\theta_1, \theta_2)$ be a vector containing all the parameters of the model. The vector $\theta_1 = (\alpha, \beta)$ contains the *linear* parameters, and the vector $\theta_2 = (\Pi, \Sigma)$ the *nonlinear* parameters.¹³ Combining equations (1) and (2), we have

$$\begin{aligned} u_{ijt} &= \alpha_i y_i + \delta_{jt}(x_{jt}, p_{jt}, \xi_{jt}; \theta_1) + u_{ijt}(x_{jt}, p_{jt}, v_i, D_i; \theta_2) + \varepsilon_{ijt}, \\ \delta_{jt} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \quad u_{ijt} = [-p_{jt}, x_{jt}](\Pi D_i + \Sigma v_i), \end{aligned} \quad (3)$$

where $[-p_{jt}, x_{jt}]$ is a $1 \times (K + 1)$ (row) vector. The indirect utility is now expressed as a sum of three (or four) terms. The first term, $\alpha_i y_i$, is given only for consistency with equation (1) and will vanish, as we will see below. The second term, δ_{jt} , which is referred to as the mean utility, is common to all consumers. Finally, the last two terms, $\mu_{ijt} + \varepsilon_{ijt}$, represent a mean-zero heteroskedastic deviation from the mean utility that captures the effects of the random coefficients.

Consumers are assumed to purchase one unit of the good that gives the highest utility.¹⁴ Since in this model an individual is de-

13. The reasons for the names will become apparent below.

14. A comment is in place here about the realism of the assumption that consumers choose no more than one good. We know that many households own more than one car, that many of us buy more than one brand of cereal, and so forth. We note that even though many of us buy more than one brand at a time, less actually consume more than one at a time. Therefore, the discreteness of choice can be sometimes defended by defining the choice period appropriately. In some cases this will still not be enough, in which case the researcher has one of two options: either claim that the above model is an approximation, or reduced-form, to the true choice model, or model the choice of multiple products, or continuous quantities, explicitly [as in Dubin and McFadden (1984) or Hendel (1999)].

defined as a vector of demographics and product-specific shocks, $(D_i, v_i, \varepsilon_{i0t}, \dots, \varepsilon_{ijt})$, this implicitly defines the set of individual attributes that lead to the choice of good j . Formally, let this set be

$$A_{jt}(x_{\cdot t}, p_{\cdot t}, \delta_{\cdot t}; \theta_2) = \left\{ (D_i, v_i, \varepsilon_{i0t}, \dots, \varepsilon_{ijt}) \mid u_{ijt} \geq u_{ilt} \right. \\ \left. \forall l = 0, 1, \dots, J \right\},$$

where $x_{\cdot t} = (x_{1t}, \dots, x_{Jt})'$, $p_{\cdot t} = (p_{1t}, \dots, p_{Jt})'$, and $\delta_{\cdot t} = (\delta_{1t}, \dots, \delta_{Jt})'$ are observed characteristics, prices, and mean utilities of all brands, respectively. The set A_{jt} defines the individuals who choose brand j in market t . Assuming ties occur with zero probability, the market share of the j th product is just an integral over the mass of consumers in the region A_{jt} . Formally, it is given by

$$s_{jt}(x_{\cdot t}, p_{\cdot t}, \delta_{\cdot t}; \theta_2) = \int_{A_{jt}} dP^*(D, v, \varepsilon) \\ = \int_{A_{jt}} dP^*(\varepsilon \mid D, v) dP^*(v \mid D) dP_D^*(D) \\ = \int_{A_{jt}} dP_\varepsilon^*(\varepsilon) dP_v^*(v) d\hat{P}_D^*(D), \tag{4}$$

where $P^*(\cdot)$ denotes population distribution functions. The second equality is a direct application of Bayes' rule, while the last is a consequence of the independence assumptions previously made.

Given assumptions on the distribution of the (unobserved) individual attributes, we can compute the integral in equation (4), either analytically or numerically. Therefore, for a given set of parameters equation (4) predicts of the market share of each product in each market, as a function of product characteristics, prices, and unknown parameters. One possible estimation strategy is to choose parameters that minimize the distance (in some metric) between the market shares predicted by equation (4) and the observed shares. This estimation strategy will yield estimates of the parameters that determine the distribution of individual attributes, but it does not account for the correlation between prices and the unobserved product characteristics. The method proposed by Berry (1994) and BLP, which is presented in detail in the Section 3, accounts for this correlation.

2.2 DISTRIBUTIONAL ASSUMPTIONS

The assumptions on the distribution of individual attributes made in order to compute the integral in equation (4) have important implications for the own- and cross-price elasticities of demand. In this section I discuss some possible assumptions and their implications.

Possibly the simplest distributional assumption one can make in order to evaluate the integral in equation (4) is that consumer heterogeneity enters the model only through the separable additive random shock, ε_{ijt} . In our model this implies $\theta_2 = 0$, or $\beta_i = \beta$ and $\alpha_i = \alpha$ for all i , and equation (1) becomes

$$u_{ijt} = \alpha(y_i - p_{jt}) + x_{jt}\beta + \xi_{jt} + \varepsilon_{ijt},$$

$$i = 1, \dots, I_t, \quad j = 1, \dots, J, \quad t = 1, \dots, T. \quad (5)$$

At this point, before we specify the distribution of ε_{ijt} , the model described by equation (5) is as general as the model given in equation (1).¹⁵ Once we assume that ε_{ijt} is i.i.d., then the implied substitution patterns are severely restricted, as we will see below. If we also assume that ε_{ijt} are distributed according to a Type I extreme-value distribution, this is the (aggregate) logit model. The market share of brand j in market t , defined by equation (4), is

$$s_{jt} = \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^J \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}. \quad (6)$$

Note that income drops out of this equation, since it is common to all options.

Although the model implied by equation (5) and the extreme-value distribution assumption is appealing due to its tractability, it restricts the substitution patterns to depend only on the market shares. The price elasticities of the market shares defined by equation (6) are

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k, \\ \alpha p_{kt} s_{kt} & \text{otherwise.} \end{cases}$$

There are two problems with these elasticities. First, since in most cases the market shares are small, the factor $\alpha(1 - s_{jt})$ is nearly constant; hence, the own-price elasticities are proportional to own price. Therefore, the lower the price, the lower the elasticity (in absolute value), which implies that a standard pricing model predicts a higher markup for the lower-priced brands. This is possible only if the marginal cost of a cheaper brand is lower (not just in absolute value, but as a percentage of price) than that of a more expensive product. For some products this will not be true. Note that this problem is a direct implication of the functional form in price. If, for example, indirect utility was a function of the logarithm of price, rather than price, then the implied elasticity would be roughly constant. In other

15. To see this compare equation (5) with equation (3).

words, the functional form directly determines the patterns of own-price elasticity.

An additional problem, which has been stressed in the literature, is with the cross-price elasticities. For example, in the context of RTE cereals the cross-price elasticities imply that if Quaker CapN Crunch (a children's cereal) and Post Grape Nuts (a wholesome simple nutrition cereal) have similar market shares, then the substitution from General Mills Lucky Charms (a children's cereal) toward either of them will be the same. Intuitively, if the price of one children's cereal goes up, we would expect more consumers to substitute to another children's cereal than to a nutrition cereal. Yet, the logit model restricts consumers to substitute towards other brands in proportion to market shares, regardless of characteristics.

The problem in the cross-price elasticities comes from the i.i.d. structure of the random shock. In order to understand why this is the case, examine equation (3). A consumer will choose a product either because the mean utility from the product, δ_{ijt} , is high or because the consumer-specific shock, $\mu_{ijt} + \varepsilon_{ijt}$, is high. The distinction becomes important when we consider a change in the environment. Consider, for example, the increase in the price of Lucky Charms discussed in the previous paragraph. For some consumers who previously consumed Lucky Charms, the utility from this product decreases enough so that the utility from what was the second choice is now higher. In the logit model different consumers will have different rankings of the products, but this difference is due only to the i.i.d. shock. Therefore, the proportion of these consumers who rank each brand as their second choice is equal to the average in the population, which is just the market share of the each product.

In order to get around this problem we need the shocks to utility to be correlated across brands. By generating correlation we predict that the second choice of consumers that decide to no longer buy Lucky Charms will be different than that of the average consumer. In particular, they will be more likely to buy a product with a shock that was positively correlated to Lucky Charms, for example CapN Crunch. As we can see in equation (3), this correlation can be generated either through the additive separable term ε_{ijt} or through the term μ_{ijt} , which captures the effect the demographics, D_i and v_i . Appropriately defining the distributions of either of these terms can yield the exact same results. The difference is only in modeling convenience. I now consider models of the two types.

Models are available that induce correlation among options by allowing ε_{ijt} to be correlated across products rather than independently distributed, are available (see the generalized extreme-value model, McFadden, 1978). One such example is the nested logit model,

in which all brands are grouped into predetermined exhaustive and mutually exclusive sets, and ε_{ijt} is decomposed into an i.i.d. shock plus a group-specific component.¹⁶ This implies that correlation between brands within a group is higher than across groups; thus, in the example given above, if the price of Lucky Charms goes up, consumers are more likely to rank CapN Crunch as their second choice. Therefore, consumers that currently consume Lucky Charms are more likely to substitute towards Grape Nuts than the average consumer. Within the group the substitution is still driven by market shares, i.e., if some children's cereals are closer substitutes for Lucky Charms than others, this will not be captured by the simple grouping.¹⁷ The main advantage of the nested logit model is that, like the logit model, it implies a closed form for the integral in equation (4). As we will see in Section 3, this simplifies the computation.

This nested logit can fit into the model described by equation (1) in one of two ways: by assuming a certain distribution of ε_{ijt} , as was motivated in the previous paragraph, or by assuming one of the characteristics of the product is a segment-specific dummy variable and assuming a particular distribution on the random coefficient of that characteristic. Given that we have shown that the model described by equations (1) and (2) can also be described by equation (3), it should not be surprising that these two ways of describing the nested logit are equivalent. Cardell (1997) shows the distributional assumptions required for this equivalence to hold.

The nested logit model allows for somewhat more flexible substitution patterns. However, in many cases the a priori division of products into groups, and the assumption of i.i.d. shocks within a group, will not be reasonable, either because the division of segments is not clear or because the segmentation does not fully account for the substitution patterns. Furthermore, the nested logit does not help with the problem of own-price elasticities. This is usually handled by assuming some "nice" functional form (i.e., yield patterns that are consistent with some prior), but that does not solve the problem of having the elasticities driven by the functional-form assumption.

In some industries the segmentation of the market will be multilayered. For example, computers can be divided into branded versus generic and into frontier versus nonfrontier technology. It turns

16. For a formal presentation of the nested logit model in the context of the model presented here, see Berry (1994) or Stern (1995).

17. Of course, one does not have to stop at one level of nesting. For example, we could group all children's cereals into family-acceptable and not acceptable. For an example of such grouping for automobiles see Goldberg (1995).

out that in the nested logit specification the order of the nests matters.¹⁸ For this reason Bresnahan et al. (1997) build on a the general extreme-value model (McFadden 1978) to construct what they call the principles-of-differentiation general extreme-value (PD GEV) model of demand for computers. In their model they are able to use two dimensions of differentiation, without ordering them. With the exception of dealing with the problem of ordering the nests, this model retains all the advantages and disadvantages of the nested logit. In particular it implies a closed-form expression for the integral in equation (4).

In principle one could consider estimating an unrestricted variance-covariance matrix of the shock, ε_{ijt} . This, however, reintroduces the dimensionality problem discussed in the Introduction.¹⁹ If in the full model, described by equations (1) and (2), we maintain the i.i.d. extreme-value distribution assumption on ε_{ijt} . Correlation between choices is obtained through the term μ_{ijt} . The correlation will be a function of both product and consumer characteristics: the correlation will be between products with similar characteristics, and consumers with similar demographics will have similar rankings of products and therefore similar substitution patterns. Therefore, rather than having to estimate a large number of parameters, corresponding to an unrestricted variance-covariance matrix, we only have to estimate a smaller number.

The price elasticities of the market shares, s_{jt} , defined by equation (4) are

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i s_{ijt} (1 - s_{ijt}) d\hat{P}_D^*(D) dP_v^*(v) & \text{if } j = k, \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i s_{ijt} s_{ikt} d\hat{P}_D^*(D) dP_v^*(v) & \text{otherwise,} \end{cases}$$

where $s_{ijt} = \exp(\delta_{jt} + \mu_{ijt}) / [1 + \sum_{k=1}^K \exp(\delta_{kt} + \mu_{ikt})]$ is the probability of individual i purchasing product j . Now the own-price elasticity will not necessary be driven by the functional form. The partial derivative of the market shares will no longer be determined by a single parameter, α . Instead, each individual will have a different price sensitivity, which will be averaged to a mean price sensitivity using the individual specific probabilities of purchase as weights. The price sensitivity

18. So, for example, classifying computers first into branded/nonbranded and then into frontier/nonfrontier technology implies different substitution patterns than classifying first into frontier/nonfrontier technology and then into branded/nonbranded, even if the classification of products does not change.

19. See Hausman and Wise (1978) for an example of such a model with a small number of products.

will be different for different brands. So if, for example, consumers of Kellogg's Corn Flakes have high price sensitivity, then the own-price elasticity of Kellogg's Corn Flakes will be high despite the low prices and the fact that prices enter linearly. Therefore, substitution patterns are not driven by functional form, but by the differences in the price sensitivity, or the marginal utility from income, between consumers that purchase the various products.

The full model also allows for flexible substitution patterns, which are not constrained by a priori segmentation of the market (yet at the same time can take advantage of this segmentation by including a segment dummy variable as a product characteristic). The composite random shock, $\mu_{ijt} + \varepsilon_{ijt}$, is then no longer independent of the product and consumer characteristics. Thus, if the price of a brand goes up, consumers are more likely to switch to brands with similar characteristics rather than to the most popular brand.

Unfortunately, these advantages do not come without cost. Estimation of the model specified in equation (3) is not as simple as that of the logit, nested logit, or GEV models. There are two immediate problems. First, equation (4) no longer has an analytic closed form [like that given in equation (6) for the Logit case]. Furthermore, the computation of the integral in equation (4) is difficult. This problem is solved using simulation methods, as described below. **Second, we now require information about the distribution of consumer heterogeneity in order to compute the market shares.** This could come in the form of a parametric assumption on the functional form of the distribution or by using additional data sources. Demographics of the population, for example, can be obtained by sampling from the CPS.

3. ESTIMATION

3.1 THE DATA

The market-level data required to consistently estimate the model previously described consists of the following variables: market shares and prices in each market, and brand characteristics. In addition information on the distribution of demographics, \hat{P}_D^* ,²⁰ as are marketing mix variables (such as advertising expenditures or the availability of coupons or promotional activity). In principle, some of the parameters of the model are identified even with data on one

20. Recall that we divided the demographic variables into two types. The first were those variables for which we had some information regarding the distribution. If such information is not available we are left with only the second type, i.e., variables for which we assume a parametric distribution.

market. However, it is highly recommended to gather data on several markets with variation in relative prices of the products and/or products offered.

Market shares are defined using a quantity variable, which depends on the context and should be determined by the specifics of the problem. BLP use the number of automobiles sold, while Nevo (2000a, b) converts pounds of cereal into servings. Probably the most important consideration in choosing the quantity variable is the need to define a market share for the outside good. This share will rarely be observed directly, and will usually be defined as the total size of the market minus the shares of the inside goods. The total size of the market is assumed according to the context. So, for example, Nevo (2000a, b) assumes the size of the market for ready-to-eat cereal to be one serving of cereal per capita per day. Bresnahan et al. (1997), in estimating demand for computers, take the potential market to be the total number of office-based employees.

In general I found the following rules useful when defining the market size. You want to make sure to define the market large enough to allow for a nonzero share of the outside good. When looking at historical data one can use eventual growth to learn about the potential market size. One should check the sensitivity of the results to the market definition; if the results are sensitive, consider an alternative. There are two parts to defining the market size: choosing the variable to which the market size is proportional, and choosing the proportionality factor. For example, one can assume that the market size is proportional to the size of the population with the proportionality factor equal to a constant factor, which can be estimated (Berry et al., 1996). From my own (somewhat limited) experience, getting the right variable from which to make things proportional is the harder, and more important, component of this process.

An important part of any data set required to implement the models described in Section 2 consists of the product characteristics. These can include physical product characteristics and market segmentation information. They can be collected from manufacturer's descriptions of the product, the trade press, or the researcher's prior. In collecting product characteristics we recall the two roles they play in the analysis: explaining the mean utility level $\delta(\cdot)$ in equation (3), and driving the substitution patterns through the term $\mu(\cdot)$ in equation (3). Ideally, these two roles should be kept separate. If the number of markets is large enough relative to the number of products, the mean utility can be explained by including product dummy variables in the regression. These variables will absorb any product characteristics that are constant across markets. A discussion of the issues arising from including brand dummy variables is given below.

Relying on product dummy variables to guide substitution patterns is equivalent to estimating an unrestricted variance-covariance matrix of the random shock ε_{ijt} in equation (1). Both imply estimating $J(J - 1)/2$ parameters. Since part of our original motivation was to reduce the number of parameters to be estimated, this is usually not a feasible option. The substitution patterns are explained by the product characteristics, and in deciding which attributes to collect the researcher should keep this in mind.

The last component of the data is information regarding the demographics of the consumers in different markets. Unlike market shares, prices, or product characteristics, this estimation can proceed without demographic information. In this case the estimation will rely on assumed distributional assumptions rather than empirical distributions. The Current Population Survey (CPS) is a good, widely available source for demographic information.

3.2 IDENTIFICATION

This section discusses, informally, some of the identification issues. There are several layers to the argument. First, I discuss how in general a discrete choice model helps us identify substitution patterns, using aggregate data from (potentially) a small number of markets. Second, I ask what in the data helps us distinguish between different discrete choice models—for example, how we can distinguish between the logit and the random-coefficients logit.

A useful starting point is to ask how one would approach the problem of estimating price elasticities if a controlled experiment could be conducted. The answer is to expose different consumers to randomly assigned prices and record their purchasing patterns. Furthermore, one could relate these purchasing patterns to individual characteristics. If individual purchases could not be observed, the experiment could still be run by comparing the different aggregate purchases of different groups. Once again, in principle, these patterns could be related to the difference in individual characteristics between the groups.

There are two potential problems with mapping the data described in the previous section into the data that arises from the ideal controlled experiment, described in the previous paragraph. First, prices are not randomly assigned; rather, they are set by profit-maximizing firms that take into account information that, due to inferior knowledge, the researcher has to include in the error term. This problem can be solved, at least in principle, by using instrumental variables.

The second, somewhat more conceptual difficulty arises because discrete choice models, for example the logit model, can be estimated

using data from just one market; hence, we are not mimicking the experiment previously described. Instead, in our experiment we ask consumers to choose between products, which are perceived as bundles of attributes. We then reveal the preferences for these attributes, one of which is price. The data from each market should not be seen as one observation of purchases when faced with a particular price vector; rather, it is an observation on the relative likelihood of purchasing J different bundles of attributes. The discrete choice model ties these probabilities to a utility model that allows us to compute price elasticities. The identifying power of this experiment increases as more markets are included with variation both in the characteristics of products and in the choice set. The same (informal) identification argument holds for the nested logit, GEV, and random-coefficients models, which are generalized forms of the logit model.

There are two caveats to the informal argument previously given. If one wants to tie demographic variables to observed purchases [i.e., allow for ΠD_i in equation (2)], several markets, with variation in the distribution of demographics, have to be observed. Second, if not all the product characteristics are observed and these unobserved attributes are correlated with some of the observed characteristics, then we are faced with an endogeneity problem. The problem can be solved by using instrumental variables, but we note that the formal requirements from these instrumental variables depend on what we believe goes into the error term. In particular, if brand-specific dummy variables are included, we will need the instrumental variables to satisfy different requirements. I return to this point in Section 3.4.

A different question is: What makes the random-coefficients logit respond differently to product characteristics? In other words, what pins down the substitution patterns? The answer goes back to the difference in the predictions of the two models and can be best explained with an example. Suppose we observe three products: A, B, and C. Products A and B are very similar in their characteristics, while products B and C have the same market shares. Suppose we observe market shares and prices in two periods, and suppose the only change is that the price of product A increases. The logit model predicts that the market shares of both products B and C should increase by the same amount. On the other hand, the random-coefficients logit allows for the possibility that the market share of product B, the one more similar to product A, will increase by more. By observing the actual relative change in the market shares of products B and C we can distinguish between the two models. Furthermore, the degree of change will allow us to identify the parameters that govern the distribution of the random coefficients.

This argument suggests that having data from more markets helps identify the parameters that govern the distribution of the random coefficients. Furthermore, observing the change in market shares as new products enter or as characteristics of existing products change provides variation that is helpful in the estimation.

3.3 THE ESTIMATION ALGORITHM

In this subsection I outline how the parameters of the models described in Section 2 can be consistently estimated using the data described in Section 3.1. Following Berry's (1994) suggestion, a GMM estimator is constructed. Given a value of the unknown parameters, the implied error term is computed and interacted with the instruments to form the GMM objective function. Next, a search is performed over all the possible parameter values to find those values that minimize the objective function. In this subsection I discuss what the error term is, how it can be computed, and some computational details. Discussion of the instrumental variables is deferred to the next section.

As previously pointed out, a straightforward approach to the estimation is to solve

$$\text{Min}_{\theta} \|s(x, p, \delta(x, p, \xi; \theta_1); \theta_2) - S\|, \quad (7)$$

where $s(\cdot)$ are the market shares given by equation (4), and S are the observed market shares. However, this approach is usually not taken, for several reasons. First, all the parameters enter the minimization in equation (7) in a nonlinear fashion. In some applications the inclusion of brand and time dummy variables results in a large number of parameters and a costly nonlinear minimization problem. The estimation procedure suggested by Berry (1994), which is described below, avoids this problem by transforming the minimization problem so that some (or all) of the parameters enter the objective function linearly.

Fundamentally, though, the main contribution of the estimation method proposed by Berry (1994) is that it allows one to deal with correlation between the (structural) error term and prices (or other variables that influence demand). As we saw in Section 2.1, there are several variables that are unobserved by the researcher. These include the individual-level characteristics, denoted $(D_i, v_i, \varepsilon_i)$, as well as the unobserved product characteristics, ξ_i . As we saw in equation (4), the unobserved individual attributes $(D_i, v_i, \varepsilon_i)$ were integrated over. Therefore, the econometric error term will be the unobserved product characteristics, ξ_i . Since it is likely that prices are correlated with this term, the econometric estimation will have to take account of this. The

standard nonlinear simultaneous-equations model [see, for example, Amemiya (1985, Chapter 8)] allows both parameters and variables to enter in a nonlinear way, but requires a separable additive error term. Equation (7) does not meet this requirement. The estimation method proposed by Berry (1994), and described below, shows how to adapt the model described in the previous section to fit into the standard (linear or) nonlinear simultaneous-equations model.

Formally, let $Z = [z_1, \dots, z_M]$ be a set of instruments such that

$$E[Z_m \omega(\theta^*)] = 0, \quad m = 1, \dots, M, \quad (8)$$

where ω , a function of the model parameters, is an error term defined below, and θ^* denotes the "true" values of the parameters. The GMM estimate is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \omega(\theta)' Z \Phi^{-1} Z' \omega(\theta), \quad (9)$$

where Φ is a consistent estimate of $E[Z' \omega \omega' Z]$. The logic driving this estimate is simple enough. At the true parameter value, θ^* , the population moment, defined by equation (8), is equal to zero. So we choose our estimate such that it sets the sample analog of the moments defined in equation (8), i.e., $Z' \hat{\omega}$, to zero. If there are more independent moment equations than parameters [i.e., $\dim(Z) > \dim(\theta)$], we cannot set all the sample analogs exactly to zero and will have to set them as close to zero as possible. The weight matrix, Φ^{-1} , defines the metric by which we measure how close to zero we are. By using the inverse of the variance-covariance matrix of the moments, we give less weight to those moments (equations) that have a higher variance.

Following Berry (1994), the error term is not defined as the difference between the observed and predicted market shares, as in equation (7); rather it is defined as the structural error, ξ_{jt} . The advantage of working with a structural error is that the link to economic theory is tighter, allowing us to think of economic theories that would justify various instrumental variables.

In order to use equation (9) we need to express the error term as an explicit function of the parameters of the model and the data. The key insight, which can be seen in equation (3), is that the error term, ξ_{jt} , only enters the mean utility level, $\delta(\cdot)$. Furthermore, the mean utility level is a linear function of ξ_{jt} ; thus, in order to obtain an expression for the error term we need to express the mean utility as a

linear function of the variables and parameters of the model. In order to do this, we solve for each market the implicit system of equations

$$s(\delta_{.t}; \theta_2) = S_{.t}, \quad t = 1, \dots, T, \tag{10}$$

where $s(\cdot)$ are the market shares given by equation (4), and S are the observed market shares.

The intuition for why we want to do this is given below. In solving this system of equations we have two steps. First, we need a way to compute the left-hand side of equation (10), which is defined by equation (4). For some special cases of the general model (e.g., logit, nested logit, and PD GEV) the market-share equation has an analytic formula. For the full random-coefficients model the integral defining the market shares has to be computed by simulation. There are several ways to do this. Probably the most common is to approximate the integral given by equation (4) by

$$\begin{aligned} & s_{jt}(p_{.t}, x_{.t}, \delta_{.t}, P_{ns}; \theta_2) \\ &= \frac{1}{ns} \sum_{i=1}^{ns} s_{jti} = \frac{1}{ns} \sum_{i=1}^{ns} \\ & \times \frac{\exp[\delta_{jt} + \sum_{k=1}^K x_{jt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id})]}{1 + \sum_{m=1}^J \exp[\delta_{mt} + \sum_{k=1}^K x_{mt}^k (\sigma_k v_i^k + \pi_{k1} D_{i1} + \dots + \pi_{kd} D_{id})]}, \end{aligned} \tag{11}$$

where (v_i^1, \dots, v_i^K) and (D_{i1}, \dots, D_{id}) , $i = 1, \dots, ns$, are draws from $\hat{P}_v^*(v)$ and $P_D^*(D)$, respectively, while x_{jt}^k , $k = 1, \dots, K$, are the variables that have random slope coefficients. Note that we use the extreme-value distribution $P_\varepsilon^*(\varepsilon)$, to integrate the ε 's analytically. Issues regarding sampling from P_v^*v and $\hat{P}_D^*(D)$, alternative methods to approximate the market shares, and their advantages are discussed in detail in the appendix (available from <http://elsa.berkeley.edu/~nevo>).

Second, using the computation of the market share, we invert the system of equations. For the simplest special case, the logit model, this inversion can be computed analytically by $\delta_{jt} = \ln S_{jt} - \ln S_{0t}$, where S_{0t} is the market share of the outside good. Note that it is the observed market shares that enter this equation. This inversion can also be computed analytically in the nested logit model (Berry, 1994) and the PD GEV model (Bresnahan et al., 1997).

For the full random-coefficients model the system of equations (10) is nonlinear and is solved numerically. It can be solved by using

the contraction mapping suggested by BLP (see there for a proof of convergence), which amounts to computing the series

$$\delta_t^{h+1} = \delta_t^h + \ln S_t - \ln S(p_{\cdot t}, x_{\cdot t}, \delta_t^h, P_{ns}; \theta_2),$$

$$t = 1, \dots, T, \quad h = 0, \dots, H, \tag{12}$$

where $s(\cdot)$ are the predicted market shares computed in the first step, H is the smallest integer such that $\|\delta_t^H - \delta_t^{H-1}\|$ is smaller than some tolerance level, and δ_t^H is the approximation to δ_t .

Once the inversion has been computed, either analytically or numerically, the error term is defined as

$$\omega_{jt} = \delta_{jt}(S_{\cdot t}; \theta_2) - (x_{jt}\beta + \alpha p_{jt}) \equiv \xi_{jt}. \tag{13}$$

Note that it is the observed market shares, S , that enter this equation. Also, we can now see the reason for distinguishing between θ_1 and θ_2 : θ_1 enters this term, and the GMM objective, in a linear fashion, while θ_2 enters nonlinearly.

The intuition to the definition is as follows. For given values of the nonlinear parameters θ_2 , we solve for the mean utility levels $\delta_{\cdot}(\cdot)$, that set the predicted market shares equal to the observed market shares. We define the residual as the difference between this valuation and the one predicted by the linear parameters α and β . The estimator, defined by equation (9), is the one that minimizes the distance between these different predictions.

Usually,²¹ the error term, as defined by equation (13), is the unobserved product characteristic, ξ_{jt} . However, if enough markets are observed, then brand-specific dummy variables can be included as product characteristics. The coefficients on these dummy variables capture both the mean quality of observed characteristics that do not vary over markets, βx_j , and the overall mean of the unobserved characteristics, ξ_j . Thus, the error term is the market-specific deviation from the main valuation, i.e., $\Delta \xi_{jt} \equiv \xi_{jt} - \xi_j$. The inclusion of brand dummy variables introduces a challenge in estimating the taste parameters, β , which is dealt with below.

In the logit and nested logit models, with the appropriate choice of a weight matrix,²² this procedure simplifies to two-stage least squares. In the full random-coefficients model, both the computation of the market shares and the inversion in order to get $\delta_{jt}(\cdot)$ have to be done numerically. The value of the estimate in equation (9) is then computed using a nonlinear search. This search is simplified by noting

21. See for example Berry (1994), BLP, Berry et al. (1996), and Bresnahan et al. (1997).

22. That is, $\Phi = Z'Z$, which is the "optimal" weight matrix under the assumption of homoskedastic errors.

that the first-order conditions of the minimization problem defined in equation (9) with respect to θ_1 are linear in these parameters. Therefore, these linear parameters can be solved for (as a function of the other parameters) and plugged into the rest of the first-order conditions, limiting the nonlinear search to the nonlinear parameters only.

The details of the computation are given in the appendix.

3.4 INSTRUMENTS

The identifying assumption in the algorithm previously given is equation (8), which requires a set of exogenous instrumental variables. As is the case with many hard problems, there is no global solution that applies to all industries and data sets. Listed below are some of the solutions offered in the literature. A precise discussion of how appropriate each set of assumptions has to be done on a case-by-case basis, but several advantages and problems are mentioned below.

The first set of variables that comes to mind are the instrumental variables defined by ordinary (or nonlinear) least squares, namely the regressors (or more generally the derivative of the moment function with respect to the parameters). As previously discussed, there are several reasons why these are invalid. For example, a variety of differentiated-products pricing models predict that prices are a function of marginal cost and a markup term. The markup term is a function of the unobserved product characteristic, which is also the error term in the demand equation. Therefore, prices will be correlated with the error term, and the estimate of the price sensitivity will be biased.

A standard place to start the search for demand-side instrumental variables is to look for variables that shift cost and are uncorrelated with the demand shock. These are the textbook instrumental variables, which work quite well when estimating demand for homogeneous products. The problem with the approach is that we rarely observe cost data fine enough that the cost shifters will vary by brand. A restricted version of this approach uses whatever cost information is available in combination with some restrictions on the demand specification [for example, see the cost variables used in Nevo (2000a)]. Even the restricted version is rarely feasible, due to lack of any cost data.

The most popular identifying assumption used to deal with the above endogeneity problem is to assume that the location of products in the characteristics space is exogenous, or at least determined prior to the revelation of the consumers' valuation of the unobserved product characteristics. This assumption can be combined with a specific model of competition and functional-form assumptions to generate an implicit set of instrumental variables (as in Bresnahan, 1981, 1987).

BLP derive a slightly more explicit set of instrumental variables, which build on a similar economic assumption. They use the observed product characteristics (excluding price and other potentially endogenous variables), the sums of the values of the same characteristics of other products offered by that firm (if the firm produces more than one product), and the sums of the values of the same characteristics of products offered by other firms.²³

Instrumental variables of this type have been quite successful in the study of many industries, including automobiles, computers, and pharmaceutical drugs. One advantage of this approach is that the instrumental variables vary by brand. The main problem is that in some cases the assumption that observed characteristics are uncorrelated with the unobserved components is not valid. One example is when certain types of products are better characterized by observed attributes. Another example is if the time required to change the observed characteristics is short and therefore changes in characteristics could be reacting to the same sort of shocks as prices. Finally, once a brand dummy variable is introduced, a problem arises with these instrumental variables: unless there is variation in the products offered in different markets, there is no variation between markets in these instruments.

The last set of instrumental variables I discuss here was introduced by Hausman et al. (1994) and Hausman (1996) and was used in the context of the model described here by Nevo (2000a, b). The essential ideal is to exploit the panel structure of the data. This argument is best demonstrated by an example. Nevo (2000a, b) observes quantities and prices for ready-to-eat cereal in a cross section of cities over twenty quarters. Following Hausman (1996), the identifying assumption made is that, controlling for brand-specific intercepts and demographics, the city-specific valuations of the product, $\Delta\xi_{jt} = \xi_{jt} - \xi_j$, are independent across cities but are allowed to be correlated within a city over time. Given this assumption, the prices of the brand in other cities are valid instruments; prices of brand j in two cities will be correlated due to the common marginal cost, but due to the independence assumption will be uncorrelated with the market-specific valuation of the product.

There are several plausible situations in which the independence assumption will not hold. Suppose there is a national (or regional) demand shock, for example, discovery that fiber may reduce the risk

23. Just to be sure, suppose the product has two characteristics: horsepower (HP) and size (S), and assume there are two firms producing three products each. Then we have six instrumental variables: The values of HP and S for each product, the sum of HP and S for the firm's other two products, and the sum of HP and S for the three products produced by the competition.

of cancer. This discovery will increase the unobserved valuation of all fiber-intensive cereal brands in all cities, and the independence assumption will be violated. Alternatively, suppose one believes that local advertising and promotions are coordinated across city borders and that these activities influence demand. Then the independence assumption will be violated.

The extent to which the assumptions needed to support any of the above instrumental variables are valid in any given situation is an empirical issue. Resolving this issue beyond any reasonable doubt is difficult and requires comparing results from several sets of instrumental variables, combing additional data sources, and using the researcher's knowledge of the industry.

3.5 BRAND-SPECIFIC DUMMY VARIABLES

As previously pointed out, I believe that brand-specific fixed effects should be used whenever possible. There are at least two good reasons to include these dummy variables. First, in any case where we are unsure that the observed characteristics capture the true factors that determine utility, fixed effects should be included in order to improve the fit of the model. We note that this helps fit the mean utility level δ_j while substitution patterns are driven by observed characteristics (either physical characteristics or market segmentation), as is the case if we do not include a brand fixed effect.

Furthermore, the major motivation (Berry, 1994) for the estimation scheme previously described is the need to instrument for the correlation between prices and the unobserved quality of the product, ξ_{jt} . A brand-specific dummy variable captures the characteristics that do not vary by market and the product-specific mean of unobserved components, namely, $x_j\beta + \xi_j$. Therefore, the correlation between prices and the brand-specific mean of unobserved quality is fully accounted for and does not require an instrument. In order to introduce a brand dummy variable we require observations on more than one market. However, even without brand dummy variables, fitting the model using observations from a single market is difficult (see BLP, footnote 30).

Once brand dummy variables are introduced, the error term is no longer the unobserved characteristics. Rather, it is the market-specific deviation from this unobserved mean. This additional variance was not introduced by the dummy variables; it is present in all models that use observations from more than one market. The use of brand dummy variables forces the researcher to discuss this additional variance explicitly.

There are two potential objections to the use of brand dummy variables. First, as previously mentioned, a major difficulty in estimating demand in differentiated product markets is that the number of parameters increases proportionally to the square of the number of products. The main motivation for the use of discrete choice models was to reduce this dimensionality problem. Does the introduction of parameters that increase in proportion to the number of brands defeat the whole purpose? No. The number of parameters increases only with J (the number of brands) and not J^2 . Furthermore, the brand dummy variables are linear parameters and do not increase the computational difficulty. If the number of brands is large, the size of the design matrix might be problematic, but given the computing power required to run the full model, this is unlikely to be a serious difficulty.

A more serious objection to the use of brand dummy variables is that the taste coefficients β cannot be identified. Fortunately, this is not true. The taste parameters can be retrieved by using a minimum-distance procedure (as in Chamberlain, 1982). Let $d = (d_1, \dots, d_j)'$ denote the $J \times 1$ vector of brand dummy coefficients, X be the $J \times K$ ($K < J$) matrix of product characteristics that are fixed across markets, and $\xi = (\xi_1, \dots, \xi_j)'$ be the $J \times 1$ vector of unobserved product qualities. Then from equation (1),

$$d = X\beta + \xi.$$

If we assume that $E[\xi | X] = 0$,²⁴ then the estimates of β and ξ are

$$\hat{\beta} = (X'V_d^{-1}X)^{-1}X'V_d^{-1}\hat{d}, \quad \hat{\xi} = \hat{d} - X\hat{\beta},$$

where \hat{d} is the vector of coefficients estimated from the procedure described in the previous section, and V_d is the variance-covariance matrix of these estimates. This is simply a GLS regression where the independent variable consists of the estimated brand effects, estimated using the GMM procedure previously described and the full sample. The number of "observations" in this regression is the number of brands. The correlation in the values of the dependent variable is treated by weighting the regression by the estimated covariance matrix, V_d , which is the estimate of this correlation. The coefficients on the brand dummy variables provide an unrestricted estimate of the mean utility. The minimum-distance procedure project these estimate onto a lower-dimensional space, which is implied by a restricted

24. Note that this is the assumption required to justify the use of observed product characteristics as instrumental variables. Here, however, this assumption is only used to recover the taste parameters. If one is unwilling to make it, the price sensitivity can still be recovered using the other assumptions discussed in the previous section.

model that sets ξ to zero. Chamberlain (1982) provides a chi-square test to evaluate these restrictions.

4. AN APPLICATION

In this section I briefly present the type of estimates one can obtain from the random-coefficients Logit model. The data used for this demonstration were motivated by real scanner data (from the ready-to-eat industry). However, the data is not real and should not be used for any analysis. The focus is on providing a data set that can easily be used to learn the method. Therefore, I estimate a somewhat restricted version of the model (with a limited amount of data). For a more detailed and realistic use of the models presented here see either BLP or Nevo (2000a, b). The data set used to generate the results and the Matlab code used to perform the computation, is available from <http://elsa.berkeley.edu/~nevo>.

The data used for the analysis below consists of quantity and prices for 24 brands of a differentiated product in 47 cities over 2 quarters. The data was generated from a model of demand and supply.²⁵ The marginal cost and the parameters required to simulate this model were motivated by the estimates of Nevo (2000b). I use two product characteristics: Sugar, which measures sugar content, and Mushy, a dummy variable equal to one if the product gets soggy in milk. Demographics were drawn from the Current Population Survey. They include the log of income (Income), the log of income squared, (Income Sq), Age, and Child, a dummy variable equal to one if the individual is less than sixteen. The unobserved demographics, v_i , were drawn from a standard normal distribution. For each market I draw 20 individuals [i.e., $ns = 20$ in equation (11)].

The results of the estimation can be found in Table I. The means of the distribution of marginal utilities (β 's) are estimated by a minimum-distance procedure described above and presented in the first column. The results suggest that for the average consumer more sugar increases the utility from the product. Estimates of heterogeneity around these means are presented in the next few columns. The column labeled "Standard Deviations" captures the effects of the unobserved demographics. The effects are insignificant, both economically and statistically. I return to this below. The last four columns present the effect of demographics on the slope parameters. The point estimates are economically significant; I return to statistical significance below. The estimates suggest that while the average consumer

25. The demand model of Section 2 was used. Supply was modeled using a standard multiproduct-firms differentiated-product Bertrand model.

TABLE I.
RESULTS FROM THE FULL MODEL

Variable	Means β	Standard Deviations σ	Interactions with Demographic Variables			
			Income	IncomeSq	Age	Child
Price	-32.433 (7.743)	1.848 (1.075)	16.598 (172.334)	-0.659 (8.955)	—	11.625 (5.207)
Constant	-1.841 ^a (0.258)	0.377 (0.129)	3.089 (1.213)	—	1.186 (1.016)	—
Sugar	0.148 ^a (0.258)	0.004 (0.012)	-0.193 (0.005)	—	0.029 (0.036)	—
Mushy	0.788 ^a (0.013)	0.081 (0.205)	1.468 (0.697)	—	-1.514 (1.103)	—
GMM objective (Degrees of freedom)			14.9 (7)			

Based on 2256 observations. Except where noted, parameters are GMM estimates. All regressions include brand and time dummy variables. Asymptotically robust standard errors are given in parentheses.

^aEstimates from a minimum-distance procedure.

might like a soggy cereal, the marginal valuation of soggyiness decreases with age and income. In other words, adults are less sensitive to the crispness of a cereal, as are the wealthier consumers. The distribution of the Mushy coefficient can be seen in Figure 1. Most of the consumers value soggyiness in a positive way, but approximately 31% of consumers actually prefer a crunchy cereal.

The mean price coefficient is negative. Coefficients on the interaction of price with demographics are economically significant, while the estimate of the standard deviation suggests that most of the heterogeneity is explained by the demographics (an issue we shall return to below). Children and above-average-income consumers tend to be less price-sensitive. The distribution of the individual price sensitivity can be seen in Figure 2. It does not seem to be normal, which is a result of the empirical distribution of demographics. In principle, the tail of the distribution can reach positive values—implying that the higher the price, the higher the utility. However, this is not the case for these results.

Most of the coefficients are not statistically significant. This is due for the most part to the simplifications I made in order to make this example more accessible. If the focus were on the real use of these estimates, their efficiency could be greatly improved by using more data, increasing the number of simulations (ns), improving the simulation methods (see the appendix on the web), or adding a supply side (see Section 5). For now I focus on the economic significance.

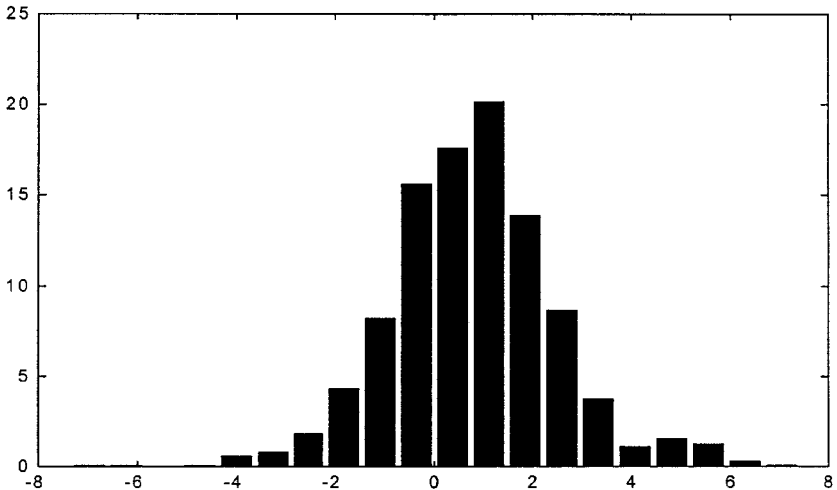


FIGURE 1. FREQUENCY DISTRIBUTION OF TASTE FOR SOGGINESS

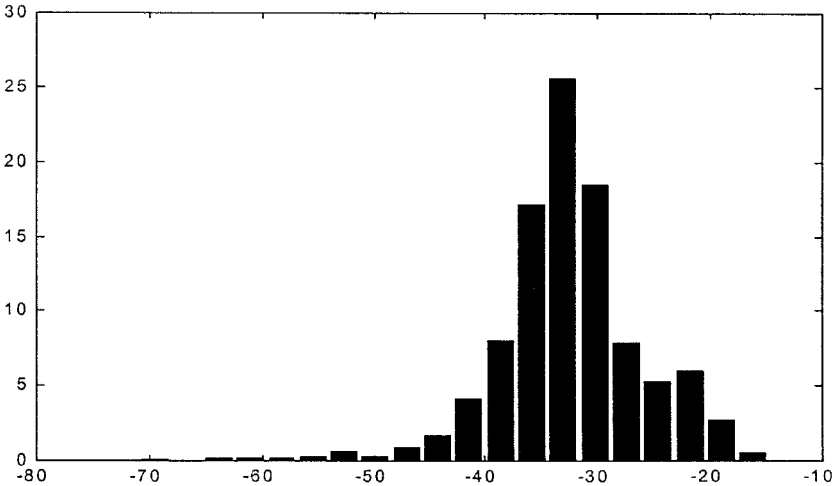


FIGURE 2. FREQUENCY DISTRIBUTION OF PRICE COEFFICIENT

As noted above, all the estimates of the standard deviations are economically insignificant,²⁶ suggesting that the heterogeneity in the coefficients is mostly explained by the included demographics. A measure of the relative importance of the demographics and random shocks can be obtained from the ratios of the variance explained by the demographics to the total variation in the distribution of the estimated coefficients; these are over 90%. This result is somewhat at odds with previous work.²⁷ The results here do not suggest that observed demographics can explain all heterogeneity; they only suggest that the data rejects the assumed normal distribution.

An alternative explanation for this result has to do with the structure of the data used here. Unlike other work (for example, BLP), by construction I have no variation across markets in the choice set. As I mentioned in Section 3.2, this sort of variation in the choice set helps identify the variance of the random shocks. This explanation, however, does not explain why the point estimates are low (as opposed to the standard errors being high) and why the effect of demographic variables is significant.

Table II presents a sample of estimated own- and cross-price elasticities. Each entry i, j , where i indexes row and j column, gives the elasticity of brand i with respect to a change in the price of j . Since the model does not imply a constant elasticity, this matrix will depend

26. Unlike the interactions with demographics, even after taking measures to improve the efficiency of the estimates, they will still stay statistically insignificant.

27. Rossi et al. (1996) find that using previous purchasing history helps explain heterogeneity above and beyond what is explained by demographics alone. Berry et al. (1998) reach a similar conclusion using second-choice data.

TABLE II.
MEDIAN OWN- AND CROSS-PRICE ELASTICITIES

Brand	Characteristics		Elasticity											
	Sugar	Mushy	Brand	Brand	Brand	Brand	Brand	Brand	Brand	Brand	Brand	Brand	Brand	Brand
			1	2	3	4	5	14	15	19	24			
1	2	1	-2.1689	0.0562	0.2429	0.1161	0.0645	0.0596	0.0611	0.0759	0.2375			
2	18	1	0.1119	-3.4505	0.1302	0.0924	0.1905	0.2305	0.2606	0.4380	0.1165			
3	4	1	0.0674	0.2715	-2.9208	0.1083	0.0709	0.0627	0.0944	0.0944	0.2237			
4	3	0	0.1868	0.0736	0.1501	-3.2709	0.1470	0.1557	0.1564	0.1037	0.2670			
5	12	0	0.0965	0.1540	0.0947	0.1326	-5.5116	0.2852	0.3085	0.2988	0.1715			
6	14	0	0.0699	0.1642	0.0698	0.1123	0.2518	0.3628	0.3971	0.3513	0.1369			
7	3	1	0.2828	0.0606	0.2390	0.1035	0.0648	0.0563	0.0865	0.0865	0.2225			
8	4	0	0.1743	0.0789	0.1476	0.1622	0.1524	0.1620	0.1656	0.1180	0.2555			
9	14	0	0.0735	0.1704	0.0717	0.1128	0.2596	0.3586	0.3966	0.3443	0.1333			
10	1	0	0.2109	0.0643	0.1682	0.1665	0.1224	0.1199	0.1298	0.0817	0.2809			
11	11	0	0.0967	0.1556	0.0890	0.1377	0.2426	0.3322	0.3427	0.2899	0.1769			
12	4	1	0.2711	0.0691	0.2339	0.1121	0.0735	0.0662	0.0704	0.0985	0.2260			
13	3	1	0.0597	0.2827	0.2390	0.0983	0.0632	0.0554	0.0582	0.0866	0.2189			
14	13	0	0.0834	0.1825	0.0786	0.1247	0.2697	-3.8393	0.3922	0.3380	0.1600			
15	13	0	0.0889	0.1763	0.0814	0.1256	0.2491	0.3542	-4.3982	0.3383	0.1582			
16	16	1	0.1430	0.2282	0.1523	0.0873	0.1447	0.1442	0.1702	0.3270	0.1264			
17	10	0	0.1138	0.1399	0.1005	0.1466	0.2172	0.2867	0.2912	0.2498	0.1875			
18	3	0	0.1900	0.0717	0.1542	0.1656	0.1362	0.1508	0.1578	0.1026	0.2660			
19	20	1	0.0944	0.2691	0.1130	0.0771	0.2163	0.2618	0.3107	-3.3335	0.1013			
20	7	0	0.1420	0.0999	0.1270	0.1546	0.1791	0.2074	0.2080	0.1628	0.2251			
21	14	0	0.0773	0.1706	0.0741	0.1160	0.2534	0.3573	0.3794	0.3530	0.1416			
22	6	0	0.1500	0.0950	0.1283	0.1602	0.1797	0.2096	0.2085	0.1530	0.2311			
23	12	0	0.0854	0.1561	0.0815	0.1307	0.2559	0.3484	0.3753	0.3154	0.1608			
24	0	0	0.2249	0.0548	0.1731	0.1619	0.1110	0.1044	0.1114	0.0733	-4.1215			
Outside good			0.0303	0.2433	0.0396	0.0552	0.2279	0.3349	0.3716	0.3896	0.0399			

Cell entries i, j , where i indexes row and j column, give the percent change in market share of brand i with a one-percent change in price of j . Each entry represents the median of the elasticities from the 94 markets.

on the values of the variables used to evaluate it. Rather than choosing a particular value (say the average, or a value at a particular market), I present the median of each entry over the 94 markets in the sample. The results demonstrate how the substitution patterns are determined in this model. Products with similar characteristics will have larger substitution patterns, all else equal. For example, brands 14 and 15 have identical observed characteristics, and therefore their cross-price elasticities are essentially identical.

A diagnostic of how far the results are from the restrictive form imposed by the logit model is given by examining the variation in the cross-price elasticities in each column. As discussed in Section 2, the logit model restricts all elasticities within a column to be equal. Therefore, an indicator of how well the model has overcome these restrictions is to examine the variation in the estimated elasticities. One such measure is given by examining the ratio of the maximum to the minimum cross-price elasticity within a column (the logit model implies that all cross-price elasticities within a column are equal and therefore a ratio of one). This ratio varies from 9 to 3. Not only does this tell us the results have overcome the logit restrictions, but more importantly it suggests for which brands the characteristics do not seem strong enough to overcome the restrictions. This test therefore suggests which characteristics we might want to add.²⁸

5. CONCLUDING REMARKS

This paper has carefully discussed recent developments in methods of estimating random-coefficients (mixed) logit models. The emphasis was on simplifying the exposition, and as a result several possible extensions were not discussed. I briefly mention these now.

5.1 SUPPLY SIDE

In the above presentation the supply side was used only in order to motivate the instrumental variables; it was not fully specified and estimated. In some cases we will want to fully specify a supply relationship and estimate it jointly with the demand-side equations (for example, see BLP). This fits into the above model easily by adding moment conditions to the GMM objective function. The increase in computational and programming complexity is small for standard static supply-side models. As usual, estimating demand and supply

28. A formal specification test of the logit model [in the spirit of Hausman and McFadden (1984)] is the test of the hypothesis that all the nonlinear parameters are jointly zero. This hypothesis is easily rejected.

jointly has the advantage of increasing the efficiency of the estimates, at the cost of requiring more structure. The cost and benefits are specific to each application and data set.

5.2 CONSUMER-LEVEL DATA

This paper has assumed that the researcher does not observe the purchase decisions of individuals. There are many cases where this is not true. In cases where only consumer data is observed, usually estimation is conducted using either maximum likelihood or the simulated method of moments [for recent examples and details see Goldberg, (1995), Rossi et al. (1996), McFadden and Train (2000), or Shum (1999)]. The method discussed here can be applied in such cases by using the consumer-level data to estimate the mean utility level δ_{jt} . The estimated mean utility levels can now be treated in a similar way to the mean utility levels computed from the inversion of the aggregate market shares. Care has to be taken when computing the standard errors, since the mean utility levels are now measured with error.

In most studies that use consumer-level data, the correlation between the regressors and the error term, which was the main motivation for the method discussed here, is usually ignored [one notable exception is Villas-Boas and Winer (1999)]. This correlation might still be present, for at least two reasons. First, even though consumers take prices and other product characteristics as given, their optimal choice from a menu of offerings could imply that econometric endogeneity might still exist (Kennan, 1989). Second, unless enough control variables are included, common unobserved characteristics, ξ_{jt} , could still bias the estimates. The method proposed here could, in principle, deal with the latter problem.

Potentially, one could observe both consumer and aggregate data. In such cases the analysis proposed here could be enriched. Petrin (1999) observes, in addition to the aggregate market shares of automobile models, the probability of purchase by consumers of different demographic groups. He uses this information in the form of additional moment restriction (thus forcing the estimated probabilities of purchase to predict the observed probabilities). Although technically somewhat different, the idea is similar to using multiple observations on the same product in different markets (i.e., different demographic groups). Berry et al. (1998) generalize this strategy by fitting three sets of moments to their sample counterparts: (1) the market shares, as above, (2) the covariance of the product characteristics and the observed demographics, and (3) the covariance of first and second choice (they have a survey that describes what the consumer's second choice was). As Berry et al. (1998) point out, the algorithm they use is

very similar to the one they introduced in BLP, which was the basis for the discussion above.

5.3 ALTERNATIVE METHODS

An alternative to the discrete-choice methods discussed here is a multilevel demand model. The essential idea is to use aggregation and separability assumptions to justify different levels of demand [see Gorman (1959, 1971), or Deaton and Muellbauer (1980) and references therein]. Originally these methods were developed to deal with demand for broad categories like food, clothing, and shelter. Recently, however, they have been adapted to demand for differentiated products [see Hausman et al. (1994) or Hausman (1996)]. The top level is the overall demand for the product category (for example, RTE cereal). Intermediate levels of the demand system model substitution between various market segments (e.g., between children's and natural cereals). The bottom level is the choice of a brand within a segment. Each level of the demand system can be estimated using a flexible functional form. This segmentation of the market reduces the number of parameters in inverse proportion to the number of segments. Therefore, with either a small number of brands or a large number of (a priori) reasonable segments, this method can use flexible functional forms [for example, the almost ideal demand system of Deaton and Muellbauer (1980)] to give good first-order approximations to any demand system. However, as the number of brands in each segment increases beyond a handful, this method becomes less feasible. For a comparison between the methods described below and these multilevel models see Nevo (1997, Chapter 6).

5.4 DYNAMICS

The model presented here is static. However, it has close links to several dynamic models. The first class of dynamic models are models of dynamic firm behavior. The links are twofold. The model used here can feed into the dynamic model as in Pakes and McGuire (1994). On the other hand, the dynamic model can be used to characterize the endogenous choice of product characteristics, therefore supplying more general identifying conditions.

An alternative class of dynamic models examine demand-side dynamics (Erdem and Keane, 1996; Ackerberg, 1996). These models generalize the demand model described here and are estimated using consumer-level data. Although in principle these models could also be estimated using aggregate (high-frequency) data, consumer-level data is better suited for the task.

5.5 INSTRUMENTS AND ADDITIONAL APPLICATIONS

As was mentioned in Sections 3.2–3.4, the identification of parameters in these models relies heavily on having an adequate set of exogenous instrumental variables. Finding such instrumental variables is crucial for any consistent estimation of demand parameters, and in models of demand for differentiated products this problem is further complicated by the fact that cost data are rarely observed and proxies for cost will rarely exhibit much cross-brand variation. Some of the solutions available in the literature have been presented, yet all suffer from potential drawbacks. It is important not to get carried away in the technical fireworks and to remember this most basic, yet very difficult, identification problem.

This paper has surveyed some of the growing literature that uses the methods described here. The scope of application and potential of use are far from exhausted. Of course, there are many more potential applications within the study of industrial economics, both in studying new industries and in answering different questions. However, the full scope of these methods is not limited to industrial organization. It is my hope that this paper will facilitate further application of these methods.

REFERENCES

- Akerberg, D., 1996, "Empirically Distinguishing Informative and Prestige Effects of Advertising," Mimeo, Boston University.
- Anderson, S., A. de Palma, and J.F. Thisse, 1992, *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: The MIT Press.
- Amemiya, T., 1985, *Advanced Econometrics*, Cambridge, MA: Harvard University Press.
- Barten, A.P., 1966, "Theorie en Empirie van een Volledig Stelsel van Vraagvergelijkingen," Doctoral Dissertation, Rotterdam: University of Rotterdam.
- Berry, S., 1994, "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.
- , J. Levinsohn, and A. Pakes, 1995, "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.
- , M. Carnall, and P. Spiller, 1996, "Airline Hubs: Costs and Markups and the Implications of Consumer Heterogeneity," Working Paper No. 5561, National Bureau of Economic Research.
- , J. Levinsohn, and A. Pakes, 1998, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," Working Paper No. 6481, National Bureau of Economic Research; also available at <http://www.econ.yale.edu/~steveb>.
- , —, and —, 1999, "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," *American Economic Review*, 89 (3), 400–430.
- Boyd, H.J. and R.E. Mellman, 1980, "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis," *Transportation Research*, 14A, 367–378.

- 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.
- , S. Stern, and M. Trajtenberg 1997, "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's" *RAND Journal of Economics*, 28, S17-S44.
- Cardell, N.S., 1989, "Extensions of the Multinomial Logit: The Hedonic Demand Model, The Non-independent Logit Model, and the Ranked Logit Model," Ph.D. Dissertation, Harvard University.
- , 1997, "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13 (2), 185-213.
- , and F. Dunbar, 1980, "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research*, 14A, 423-434.
- Chamberlain, G., 1982, "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18 (1), 5-46.
- Christensen, L.R., D.W. Jorgenson, and L.J. Lau, 1975, "Transcendental Logarithmic Utility Functions," *American Economic Review*, 65, 367-383.
- Court, A.T., 1939, "Hedonic Price Indexes with Automotive Examples," in Anon., *The Dynamics of Automobile Demand*, New York: General Motors.
- Das, S., S. Olley, and A. Pakes, 1994, "Evolution of Brand Qualities of Consumer Electronics in the U.S.," Mimeo, Yale University.
- Davis, P., 1998, "Spatial Competition in Retail Markets: Movie Theaters," Mimeo, Yale University.
- Deaton, A., and J. Muellbauer, 1980, "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326.
- Dixit, A., and J.E. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297-308.
- Dubin, J. and D. McFadden, 1984, "An Econometric Analysis of Residential Electric Appliance Holding and Consumption," *Econometrica*, 52, 345-362.
- Erdem, T. and M. Keane, 1996, "Decision-making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1-20.
- Gasmi, F., J.J. Laffont, and Q. Vuong, 1992, "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics & Strategy*, 1 (2), 277-311.
- Goldberg, P., 1995, "Product Differentiation and Oligopoly in International Markets: The Case of the Automobile Industry," *Econometrica*, 63, 891-951.
- Gorman, W.M., 1959, "Separable Utility and Aggregation," *Econometrica*, 27, 469-481.
- , 1971, Lecture Notes, Mimeo, London School of Economics.
- Griliches, Z., 1961, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, hearing before the Joint Economic Committee of the U.S. Congress, 173-76, Pt. 1, 87th Cong., 1st sess. Reprinted in Z. Griliches, ed., 1971, *Price Indexes and Quality Change: Studies in New Methods in Measurement*, Cambridge, MA: Harvard University Press.
- Hausman, J., 1996, "Valuation of New Goods under Perfect and Imperfect Competition," in T. Bresnahan and R. Gordon, eds., *The Economics of New Goods*, Studies in Income and Wealth, Vol. 58, Chicago: National Bureau of Economic Research.
- , and D. McFadden, 1984, "Specification Tests for the Multinomial Logit Model," *Econometrica*, 52 (5), 1219-1240.

- , and D. Wise, 1978, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 49, 403–426.
- , G. Leonard, and J.D. Zona, 1994, "Competitive Analysis with Differentiated Products," *Annales d'Economie et de Statistique*, 34, 159–180
- Hendel, I., 1999, "Estimating Multiple Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 66, 423–446.
- Lancaster, K., 1966, "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132–157.
- , 1971, *Consumer Demand: A New Approach*, New York: Columbia University Press.
- Kennan, J., 1989, "Simultaneous Equations Bias in Disaggregated Econometric Models," *Review of Economic Studies*, 56, 151–156.
- McFadden, D., 1973, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed., *Frontiers of Econometrics*, New York: Academic Press.
- , 1978, "Modeling the Choice of Residential Location," in A. Karlqvist et al., eds., *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland.
- , and K. Train, 2000, "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, forthcoming; available from <http://emlab.berkeley.edu/~train>.
- Nevo, A., 1997, "Demand for Ready-to-Eat Cereal and Its Implications for Price Competition, Merger Analysis, and Valuation of New Goods." Ph.D. Dissertation, Harvard University.
- , 2000a, "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, forthcoming; also available from <http://emlab.berkeley.edu/~nevo>.
- , 2000b, "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry," *Rand Journal Economics*, forthcoming, 31 (Autumn).
- Pakes, A. and P. McGuire, 1994, "Computation of Markov Perfect Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *Rand Journal of Economics*, 25 (4), 555–589.
- Petrin, A., 1999, "Quantifying the Benefits of New Products: The Case of the Minivan," Mimeo, University of Chicago; available from <http://gsbwww.uchicago.edu/fac/amil.petrin>.
- Quandt, R.E., 1968, "Estimation of Model Splits," *Transportation Research*, 2, 41–50.
- Rosen, S., 1974, "Hedonic Prices and Implicit Markets," *Journal of Political Economy*, 82, 34–55.
- Rossi, P., R.E. McCulloch, and G.M. Allenby, 1996, "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321–340.
- Shum, M., 1999, "Advertising and Switching Behavior in the Breakfast Cereal Market," Mimeo, University of Toronto; available from <http://www.chass.utoronto.ca/eco/eco.html>.
- Spence, M., 1976, "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, 43, 217–235.
- Stern, S., 1995, "Product Demand in Pharmaceutical Markets," Mimeo, Stanford University.
- Stone, J., 1954, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *Economic Journal*, 64, 511–527.
- Tardiff, T.J., 1980, "Vehicle Choice Models: Review of Previous Studies and Directions for Further Research," *Transportation Research*, 14A, 327–335.
- Theil, H., 1965, "The Information Approach to Demand Analysis," *Econometrica*, 6, 375–380.
- Villas-Boas, M. and R. Winer, 1999, "Endogeneity in Brand Choice Models," *Management Science*, 45, 1324–1338.