PART TWO: Measurement of Market Power


Aviv Nevo

© 1997
Food Marketing Policy Center
Department of Agricultural and Resource Economics
University of Connecticut
and
Department of Resource Economics
University of Massachusetts, Amherst

Aviv Nevo  (nevo@husc.harvard.edu)

Department of Economics
Littauer Center
Harvard University
Cambridge, MA  02138

Keywords:  Market power, consumer demand, cereal industry

Copyright © 1997 by Food Marketing Policy Center, University of Connecticut. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Aviv Nevo

Introduction

High price cost margins, high advertising to sales ratios, and aggressive introduction of new brands have led previous research to conclude that the ready-to-eat (RTE) cereal industry is a classic example of a concentrated, differentiated products industry that has channeled away from price competition, thus, making pricing approximately collusive. In a related paper (Nevo 1997) I examine this claim. This paper presents the methodology used there, and discusses the implications for measuring market power.

Price cost margins (PCM) have been traditionally used in homogenous good markets to construct a Lerner Index, which measures the degree of price competition. However, in a market with many closely related products, high PCM are potentially a result of three causes. The first cause is the firm’s ability to differentiate its brands from those of its competition. The second is the portfolio effect; if two brands are perceived as substitutes, a firm producing both would charge a higher price than two separate manufacturers. Finally, the main players in the industry could potentially collude on prices. This paper lays out a method to estimate consistently the true economic price cost margins (PCM) in an industry, empirically distinguish between the three causes of these margins, and measure the degree of price competition.

The general strategy is to model demand as a function of product characteristics, consumer preferences, and unknown parameters. Following recent developments in techniques for estimating demand of closely related products (see Bresnahan 1987; Berry 1994; Berry, Levinsohn, and Pakes (henceforth BLP) 1995), and using data rarely available for academic research, I am able to estimate the unknown parameters. These estimates are used to compute the PCM implied by three hypothetical industry structures: (1) single product firms; (2) multi-product firms; and (3) one firm producing all brands. The markup in the first structure is due only to product differentiation. In the second case the markup also includes the portfolio effect. Finally, the last structure produces the markups based on joint ownership, thus bounding the effect of collusion.

The results given in Nevo (1997) suggest that the markups implied by the current industry structure, under a Nash-Bertrand pricing game, explain the observed PCM in the RTE cereal industry. If we take Nash-Bertrand prices as the competitive benchmark, then even with PCM higher than 40 percent we have to conclude that pricing in the RTE cereal industry is approximately competitive. High PCM are not due to tacit price collusion (as claimed by previous research), but to consumers’ willingness to pay for their favorite brand, and firms that account for substitution between their own brands when setting prices. To the extent that there is any market power in this industry, it is due to the firms’ ability to maintain a portfolio of differentiated products and influence the perceived quality of these products by means of advertising.
This paper follows the tradition that has been termed the “New Empirical IO” (Bresnahan 1989). This now, two decade old, “new” IO focuses on estimating PCM and marginal costs, without observing actual costs. The markups, and the implied marginal costs, are inferred from firm conduct. These methods have been widely used for homogenous goods industries (see Bresnahan 1989). However, when it comes to differentiated products, the task is much harder because of the large number of parameters to be estimated. To be more specific, we require an estimate of the demand system and the pattern of substitution between the goods. If, for example, we have 200 differentiated products (as in the RTE cereal industry), assuming constant elasticity demand curves, this implies estimating 40,000 price elasticities. Even if we impose restrictions implied by economic theory, the number of parameters will still be too high to estimate with any reasonable data set.

One solution to this problem is given by the discrete choice literature (see Cardell 1989, Berry 1994, BLP 1995). Here the dimensionality problem is solved by projecting the products onto some characteristics space, making the relevant dimension the dimension of this space and not the square number of products. The method outlined below follows this approach, but points out various extensions introduced in Nevo (1997).

The estimation in this paper takes as given advertising and the brands offered. Therefore, the results can be considered as measuring the short-run market power, i.e., without taking account of the market power firms might have because of their ability to introduce new brands and change perception of existing brands. For the question I address here, short-run price competition, this is not important. However, one has to be careful in drawing policy conclusions based on these results, since both advertising and brand introduction would not stay fixed if the structure of the industry changed. A complete model of the industry, which could answer these policy questions, is a dynamic one that takes account of both brand introduction and advertising.

The paper is organized in the following manner. First, I lay out the empirical framework. Next, I outline the data requirements and the estimation methods.

The Empirical Framework

My general strategy is to model demand as a function of product characteristics and consumer preferences. Unknown demand parameters are estimated and used to compute the PCM implied by different models of conduct. Finally, I use external data on costs to choose between these different models.

Supply

Suppose there are \( F \) firms, each of which produces some subset, \( \mathcal{F}_f \), of the \( j = 1, \ldots, J \) different brands of RTE cereal. The profits of firm \( f \) are:

\[
\Pi_f = \sum_{j \in \mathcal{F}_f} (p_j - mc_j) M s_j(p) - C_j
\]

where \( s_j(p) \) is the market share of firm \( j \), \( M \) is the size of the market, and \( C_j \) are the fixed costs of production. Assuming: (1) the existence of a pure-strategy Nash-Bertrand equilibrium in prices and (2) that the prices that support it are strictly positive, the price \( p_j \) of any product \( j \) produced by firm \( f \) must satisfy the first order condition:

\[
s_j(p) + \sum_{r \neq j \in \mathcal{F}_f} (p_r - mc_r) \frac{\partial s_j(p)}{\partial p_j} = 0.
\]
This set of $J$ equations imply price-costs margins for each good. The markups can be solved for explicitly by defining $S_{p} = -\partial s_{j}/\partial p_{j}$, where $j, r = 1,...,J$:

$$\Omega_{p}^{*} = \begin{cases} 1, & \text{if } \exists f: \{r,j\} \in \mathcal{F}; \\ 0, & \text{otherwise} \end{cases}$$

and $\Omega_{p} = \Omega_{p}^{*} \cdot S_{p}$. In vector notation the first order conditions become:

$$s(p) - \Omega(p - mc) = 0.$$ 

Which implies a markup equation:

$$p - mc = \Omega^{-1}s(p).$$

Given estimates of the demand parameters, we are able to estimate PCM without observing actual costs and can distinguish between three different causes of the markups: the effect due to the differentiation of the products, the portfolio effect, and the effect of price collusion. This is done by evaluating the PCM in three hypothetical industry conduct models. First is single product firms, where the price of each brand is set by a profit maximizer that considers only the profits from that brand. Second is the current structure, i.e., multi-product firms that set the prices of all their products jointly. Therefore, if brands are substitutes, a multi-product firm will set a higher price than single product firms. Finally is joint profit maximization of all the brands, which corresponds to monopoly or perfect price collusion. Each of these structures is estimated by setting a different $\Omega^{*}$ matrix. The PCM in the first structure are due only to product differentiation. While the difference between the margins in the first two cases is due to the portfolio effect. The last case bounds the increase in the margins due to price collusion.

Once these margins are computed we choose the model of conduct that seems to best fit the observed PCM. Here, I suggest examining a Nash-Bertrand price equilibrium under different ownership structures. Alternatively, one might either assume a different model of conduct (for example Stackelberg price leadership), or estimate a conduct parameter (namely a parameter that tells where on the continuum between perfect competition and monopoly the industry is located). The latter can be estimated, if we have reliable cost data, by adding a cost equation that equates the PCM, that are a function of conduct parameters and demand elasticities, to the margins implied by cost data in order to choose the conduct parameter that best fits the data. However, one should be careful in giving structural interpretation to this parameter both for theoretical and identification reasons (Corts 1996).

The above analysis takes as given the products offered. However, existing theories of the RTE cereal industry suggest that it is precisely the ability to introduce differentiated new products that accounts for the market power of the industry leaders (Schmalensee 1978). In this context, we can view this paper as measuring short-run market power, i.e., the market power under the current industry power and not what would happen under a different structure. We can think of the estimates as an empirical measure of the potential market power due to brand proliferation or advertising. A complete model of this industry, one which allows us to measure the long-run market power, would have to take the dynamics of new brand introduction and advertising into account. This is the goal of work in progress.

**Demand**

The exercise suggested in the previous section allows us to estimate PCM and separate them into different parts. However, it relies heavily on the ability to consistently estimate own and cross price elasticities. As previously pointed out, this is not an easy task in an industry with many closely related
products. In the analysis below I follow the discrete choice literature and solve the dimensionality problem by projecting the products onto a utility space, making the relevant dimension the dimension of this space and not the number of products.

Suppose there are \( i = 1, \ldots, I \) households. The indirect utility\(^4\) of household \( i \) from product \( j \) at time \( t \) in market \( t, t = 1, \ldots, T \) is given by:

\[
\begin{align*}
    u_{ijt} &= x_i \beta_i^* - \alpha_i p_j + \xi_j + \Delta \xi_j + \epsilon_{ijt} \\
    i &= 1, \ldots, I \quad j &= 1, \ldots, J, \quad t = 1, \ldots, T
\end{align*}
\]

where: \( x_i \) are observable product characteristics, \( p_j \) is the price of product \( j \) at market \( t \), \( \xi_j \) is the mean valuation, across markets, of the unobserved (by the econometrician) product characteristics, \( \Delta \xi_j \) is a market specific deviation from this mean, and \( \epsilon_{ijt} \) is a mean zero stochastic term. The coefficients \( (\alpha_i, \beta_i^*) \) are \( K+1 \) household specific coefficients. The \( T \) different markets can include different geographical locations (different cities), different periods, different economic markets (for example cereals sold in supermarket versus those sold to schools), or any combination of the above.

For cereal, examples of observed characteristics are: calories, sodium, and fiber content. Unobserved characteristics include a vertical component (at equal prices all consumers weakly prefer a national brand to a generic version), components that are consumer specific (for example taste and valuation of freshness), and the market specific effects of merchandizing (other than national advertising). If \( T > 1 \) then we can control for the vertical component, \( \xi_j \), by including a brand dummy in the regression. The consumer and market specific components are included in \( \xi_j \) and are left as “error terms,”\(^5\) which also includes \( \xi_j \) if a dummy is not included in the regression. I assume both firms and consumers observe all the product characteristics and take them into consideration when making decisions.

I model the distribution of the consumer’s taste parameter for the characteristics as multi-variate normal with a mean that is a function of demographic variables and parameters to be estimated, and a variance-covariance matrix to be estimated. Let \( \gamma_i = (\alpha_i, \beta_1^*, \ldots, \beta_K^*) \) and \( \gamma_j = (\alpha_j, \beta_1^{j*}, \ldots, \beta_{Kj}^{j*}) \), where \( K \) is the dimension of the observed characteristics vector, then:

\[
\gamma_i^* = \gamma + \Pi D_i + \Sigma v_i, \quad v_i \sim N(0, I_{K+1})
\]

where \( D_i \) is a \( d \times 1 \) vector of demographic variables, \( \Pi \) is a \( (K+1) \times d \) matrix of coefficients that measure how the taste characteristics vary with demographics, and \( \Sigma \) is a scaling matrix.\(^6\) Examples of demographic variables are income, age, and household size.

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 we get that 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:

\[
u_{0it} = \xi_{0i} + \pi_0 D_i + \sigma_0 v_{0i} + \epsilon_{0it}.
\]

The mean utility of the outside good is not identified (without either making more assumptions or normalizing one of the “inside” goods), thus, I normalize \( \xi_{0i} \) to zero. The coefficients \( \pi_0 \) and \( \sigma_0 \) are not identified separately from coefficients on a constant that are allowed to vary by household. I interpret the coefficients on this constant as utility parameters of the outside good.

Let \( \theta = (\theta_1, \theta_2) \) be a vector containing all the parameters of the model, where \( \theta_1 = (\alpha, \beta) \), for reasons that will be apparent below, are the linear parameters, and \( \theta_2 = (\Pi, \Sigma, \pi_0, \sigma_0) \) are the non-linear parameters. Thus, combining equations (2) and (3) we get:

88
where \([p_{jt}, x_j] \) is a \((K+1) \times 1\) vector. The utility is now expressed as the mean utility, represented by \(\delta_{jt}\), and a mean zero heteroskedastic deviation from that mean, \(\mu_{ijt} + \epsilon_{ijt}\), that captures the effects of the random coefficients.

Consumers are assumed to purchase one unit of the good that gives the highest utility. This implicitly defines the set of unobserved variables that lead to the choice of good \(j\). Formally, let this set be:

\[
A_j(\delta_j; \theta_2) = \{ (D_t, v_t, \epsilon_{ijt}) | u_{ijt} > u_{ilh} \ \forall \ l = 0, 1, \ldots, J \}
\]

where \(\delta_j = (\delta_{1j}, \ldots, \delta_{Kj})'\). Assuming ties occur with zero probability, the market share of the \(j\)th product, as a function of the mean utility levels of all the \(J+1\) goods, given the parameters, is:

\[
s_j(x, p, \delta_j; \theta) = \int_{A_j} dP^*(D_t, v_t, \epsilon_{ijt}) = \int_{A_j} dP^*(\epsilon | D_t, v_t) dP^*(v | D_t) dP^*(D_t)
\]

(5)

\[
= \int_{A_j} dP^*(\epsilon | D_t, v_t) dP^*(v) dP^*(D_t).
\]

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 modeling assumption made in equation (3).

Given assumptions on the distribution of the unobserved variables, we can compute the integral in (5), either analytically or numerically. A straightforward identification strategy is to choose parameters that minimize the distance (in some metric) between the market shares predicted by (5) and the observed shares. For reasons given below, this is not the strategy used. Nevertheless, it serves as an intuitive guide to the identification scheme.

Possibly the simplest distributional assumptions one can make in order to solve (5) are those made by the classical discrete choice models: consumer heterogeneity enters the model only through the separable, additive shock, \(\epsilon_{ijt}\). In our model this implies \(\theta_2 = 0\), and therefore \(\beta_j = \beta_j^*\), \(\alpha_i = \alpha\) for all \(i\), and (2) becomes:

\[
u_{ijt} = x_j \beta_j - \alpha p_{jt} + \xi_j + \Delta \xi_j + \epsilon_{ijt}, \quad i = 1, \ldots, I, \ j = 1, \ldots, J, \ t = 1, \ldots, T.
\]

(6)

If \(\epsilon_{ijt}\) is distributed i.i.d. with a Type I extreme value distribution this is the well-known (conditional) Logit model. The brand market shares are:

\[
s_j = \frac{\exp (x_j \beta_j - \alpha p_{jt} + \xi_j + \Delta \xi_j)}{1 + \sum_{k=1}^{J} \exp (x_k \beta_j - \alpha p_{kt} + \xi_k + \Delta \xi_k)}.
\]

(7)

Although the model implied by (6) and the extreme value distribution assumption is appealing, due to its tractability, it restricts the substitution patterns to depend only on the market shares. In the context of RTE cereals this implies that if, for example, Quaker CapN Crunch (a kids’ cereal) and Post Grape Nuts (a wholesome simple nutrition cereal) have similar market shares then the substitution from General Mills Lucky Charms (a kids’ cereal) toward either of them will be the same. Intuitively, if the price of one kids’ cereal goes up we would expect more consumers to substitute to another kids’ cereal than to a nutrition cereal. Yet, the Logit model restricts consumers to substitute more towards the more popular brands (the ones with the higher market shares).
The Nested Logit (McFadden 1978) is a slightly more complex model in which the i.i.d. extreme value assumption is replaced with a variance components structure. All brands are grouped into exhaustive and mutually exclusive sets. A consumer has a common shock to all the products in a set, so she is more likely to substitute to other products in the group. Cardell (1991) shows the distributional assumptions required in order to retain the extreme value distribution on the composite term. In the context of our model this model implies a closed form expression for (5).

The Nested Logit model allows for somewhat more flexible substitution patterns, yet retains the computational simplicity of the Logit structure. 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.

An additional problem with both the Logit and the Nested Logit models is that all consumer heterogeneity enters the model through the error term. The distribution of this error term is not a function of market specific variables. In the data used in Nevo (1997), I find there is large between market variation. However, the only variable that varies between markets is the price variable, which is not always enough to explain the variation in market shares. This is especially problematic for the Nested Logit model since we are trying to estimate a parameter of the distribution of the error and imposing this parameter to be identical across markets.

The full model described in this section has two advantages over the Logit and Nested Logit models. First, it 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). The composite error terms, \( \mu_{ij} + \epsilon_{ijt} \), are no longer independent of the product and individual characteristics. Thus, if the price of a brand goes up consumers are more likely to switch to brands with similar characteristics, rather then to the most popular brand. Also, households with similar characteristics will tend to have similar purchasing patterns. Second, by allowing the distribution of the error term to depend on demographic variables we let the distribution of the error term vary across markets. Thus the variation between markets in the market shares does not need to be explained solely by the variation in prices.

Unfortunately, these advantages do not come without cost. Estimation of equation (4) is not as simple as that of the Logit or Nested Logit models. There are two immediate problems. First, the integral in equation (5) no longer has an analytic closed form (like equation (7) in the Logit case). Furthermore, the computation of the integral in (5) is difficult. This is solved using the simulation technique, introduced by Pakes (1986), to compute the integral. Second, we now require individual demographics in order to compute the market shares, yet the data described below consists of aggregate data. This is solved by incorporating additional demographic information into the estimation algorithm.

**Data and Estimation**

**Data Requirements**

The data required to consistently estimate the model previously described consists of the following variables: brand level market shares and prices in each market, brand characteristics, and information on the distribution of demographics.

In Nevo (1997) market shares and prices were obtained from the IRI Infoscan Data Base at the University of Connecticut. However, less detailed data, which is widely available, will also do. Product characteristics were collected in local supermarkets by examining cereal boxes. Information on the distribution of household demographics was obtained by sampling households from the March Current Population Survey for each year.
### The Estimation Algorithm

In this section I outline how the parameters of the models described in the previous section can be consistently estimated, using the data described in the previous section. I follow the suggestion of Berry (1994) to construct a GMM estimator. Essentially the idea is, for a given value of the unknown parameters, to compute the implied “error term” and interact it with instruments, thus forming 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 section I discuss what the error term is, how it can be computed, and some computational details. Discussion of the instruments is deferred to the next section.

As previously pointed out a straightforward approach to estimation is to solve:

$$
\min_{\theta} \left\| s(x, p, \delta; \theta) - S \right\|
$$

where $s(\cdot)$ are the market shares given by equation (5) and $S$ are the observed market shares. However, this approach is not taken for several reasons. First, all the parameters enter (8) in a non-linear fashion. In certain applications the number of parameters is large and a non-linear minimization problem costly. The estimation procedure suggested by Berry (1994), which is used below, avoids this problem by transforming the minimization problem such that some (or all) of the parameters enter the objective function linearly. Furthermore, it is much harder to think of identifying assumptions in the context of (8), while it is natural to do so in the method used below.

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

$$
E[Z \cdot \omega(\theta^*)] = 0
$$

where $\omega$ is an “error term” defined below, which is a function of the model parameters, and $\theta^*$ denotes the “true” value of these parameters. The GMM estimate is:

$$
\theta = \arg\min_{\theta} \omega(\theta) Z A^{-1} Z^T \omega(\theta)
$$

where $A$ is a consistent estimate of $E[Z \omega(\theta) Z^T]$. The logic driving this estimate is simple enough. At the true parameter value, $\theta^*$, the population moment defined by (9) is equal to zero so we choose our estimate such that it sets the sample analog of (9), i.e. $Z^T \omega$, to zero. If there are more independent moment equations than parameters (i.e., $M = \dim(Z) > \dim(\theta)$), 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, $A$, defines the metric by which we measure how close to zero we are. By using 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, rather it is obtained by inverting the market share function to obtain the vector of mean valuations that equates the observed market shares to the predicted shares. This is done by solving, for each market, the implicit system of equations:

$$
s_j(x, p, \delta; \theta_j) = S_j.
$$

In some cases (for example, the Logit model and one level Nested Logit) this system can be solved analytically. However, for the full model suggested above this has to be done numerically. Once this inversion has been done, either analytically or numerically, the “error term” is defined as:
Note that it is the observed market shares, $S$, that enter this equation. Also, we can now see the reason for distinguishing between $\theta_1$ and $\theta_2$: $\theta_1$ enters this term, and the GMM objective, in a linear fashion, while $\theta_2$ enters non-linearly.

The intuition in this definition is as follows. For a given value of the non-linear parameters, $\theta_2$, we compute what is the mean valuation, $\delta_j(\cdot)$, that would make the predicted market share equal to the observed market share. We define the residual as the difference between this valuation and the one “predicted” by the linear parameters, $\alpha$ and $\beta$. The estimator, defined by (10), is the one that minimizes the distance between these different predictions.

Usually, in these types of models, the error term, as defined by (11) is the unobserved product characteristic, $\xi_j$. In Nevo (1997) I am able to include a brand dummy as a regressor. This dummy captures both the mean quality index of observed characteristics, $\beta x$, and the unobserved characteristics, $\xi_j$. Thus, the error term is the market specific deviation from the mean valuation, i.e., $\Delta \xi_{jt}$. The inclusion of a brand dummy introduces a challenge in estimating the taste parameters, $\beta$, which can be solved using a minimum distance procedure (see Nevo 1997, Chapter 3.4).

In the Logit and Nested Logit models, with the appropriate choice of a weight matrix, 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 $\delta_j(\cdot)$, have to be done numerically. The value of the estimate in (10) is then computed using a non-linear search. This search is simplified in two ways. First, we note that first order conditions of (10) with respect to $\theta_1$ 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 non-linear search to only the non-linear parameters.

Second, the results produced below were computed using a Quasi-Newton method with a user supplied gradient. This was found to work much faster than the Nelder-Mead non-derivative "simplex" search method used by BLP. The details of the computation and a step-by-step description of the algorithm are given in Nevo (1997, Chapter 4).

**Instruments**

The key identifying assumption, in the algorithm previously given, is equation (9) which requires a set of exogenous instruments. The first set that comes to mind are the instruments defined by ordinary least squares, namely the regressors. As pointed out by Berry (1994), these instruments are not valid, because we assumed the firm knows the unobserved (to us) quality and takes account of it when setting prices.

Different instruments have been used in the literature and their validity depends on the data, the model, and the specifics of the industry. Much of the previous work (see footnote 9), used instruments which include the number of competing products and the sum over competition of product characteristics. These instruments attempt to proxy for the degree and closeness of competition the brand is facing. These instruments treat the “location” of the products, in the attribute space, as exogenous. Yet, a more general model has these “locations” determined endogenously, and therefore we might doubt the validity of them as instruments.

Nevo (1997) uses two sets of instruments in an attempt to separate between the exogenous variation in prices between markets (due to difference in marginal costs) and endogenous variation (due to difference in unobserved valuation). First, I exploit the panel structure of the data. The identifying assumption is that, controlling for brand specific means and demographics, city specific valuations are independent across markets (but are allowed to be correlated within a market). Second, I examine another set of instruments that attempts to proxy for the marginal costs directly.
Conclusions and Extensions

This paper reviews how one could estimate a brand level demand system for many closely related products and use the estimated elasticities to compute price cost margins that prevail under different conduct models. Selection between these models is done based on observed PCM. This method is used in a related paper (Nevo 1997). The results suggest that a Nash-Bertrand pricing game, played between multi-product firms, is consistent with observed price cost margins in the RTE cereal industry. The implication of such a result is that if we are willing to accept Nash-Bertrand as a benchmark of non-collusive pricing, we are left to conclude, unlike previous work, that even with PCM greater than 40 percent, prices in the industry are competitive.

A more general point is that a Lerner Index type measure is not appropriate for measuring price competition, and price collusion, in differentiated products markets. This paper proposes a general method of obtaining such a measure. However, a word of caution is needed. The estimates using the methods proposed here should be used very carefully for policy analysis. We assumed that advertising and brand introduction are exogenous, yet, in most cases when we want to evaluate policy this will not hold (for example in merger cases). In order to make claims regarding the long-run competitiveness of an industry, as opposed to short-run market power, one has to deal with the dynamic, long-run, issues of brand introduction and advertising, which were taken as given in this paper. Such a model is the subject of future work.

The demand elasticities presented above were obtained by following the discrete choice literature in order to solve the problem of estimating many elasticity parameters. One would like to know the sensitivity of the conclusion to the method used. An alternative method is the multi-level demand model (for example, see Hausman 1996), which estimates different levels of demand. A topic of ongoing research, joint with Ronald Cotterill and Li Yu Ma, is the relative performance of these alternatives.

Notes

1 Aviv Nevo is a Ph.D. candidate, Department of Economics, Harvard University. This paper was written while visiting at the Food Marketing Policy Center at the University of Connecticut, and is based on ongoing research presented in Nevo (1997). The author wishes to thank Ronald Cotterill for sharing his data and knowledge of the cereal industry.

2 For example, see Schmalensee (1978: 315): “Overall, the observed pattern of conduct in the RTE cereal market seems consistent with received doctrine about highly concentrated industries with differentiated products: price competition was suppressed and rivalry was channeled into advertising and new product introduction. In game-theoretic terms while pricing conduct may have been approximately cooperative, . . . .” Also, Scherer (1982: 189) claims: “. . . the cereal industry’s conduct fits well the model of price competition-avoiding, non-price competition-prone oligopoly.”

3 Caplin and Nalebuff (1991) provide a set of conditions that promise the existence of equilibrium in the case of single product firms. Their results do not easily generalize to the multi-product firm case. However, I am able to check whether my final estimates are consistent with the existence of an equilibrium.

4 This indirect utility form can be derived from a quasi-linear utility function.

5 An alternative is to model the distribution of the valuation of the unobserved characteristics, as in Das et al. (1994).

6 Alternatively, one could think of a composite “error” term, $\epsilon_i$, which is distributed $N(0, \Sigma^*)$ and $\Sigma$ is the Cholesky factorization of $\Sigma^*$.

7 See BLP, page 846, for a detailed discussion.
I am grateful to Ronald Cotterill, the Director of the Food Marketing Center at the University of Connecticut, for making these data available.

See for example Berry (1994), BLP (1995), Berry et al. (1994), and Bresnahan et al. (1997).

In other words, \( A = Z'Z \), which is the “optimal” weight matrix under the assumption of homoskedastic errors.

This assumption is similar to the one made in Hausman (1996), although our setups differ substantially.

References


