MEASURING MARKET POWER IN THE READY-TO-EAT CEREAL INDUSTRY

Aviv Nevo

Working Paper **6387**

NBER WORKING PAPER SERIES

MEASURING MARKET POWER IN THE READY-TO-EAT CEREAL INDUSTRY

Aviv Nevo

Working Paper 6387
http://www.nber.org/papers/w6387

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1998

This paper is based on various chapters of my 1997 Harvard University Ph.D. dissertation. Special thanks to my advisors, Gary Chamberlain, Zvi Griliches and Michael Whinston for guidance and support. I wish to thank Ronald Cotterill, the director of the Food Marketing Policy Center at the University of Connecticut, for sharing with me his data. I am grateful to Steve Berry, Ernst Berndt, Tim Bresnahan, David Cutler, Jerry Hausman, Igal Hendel, Kei Hirano, John Horn, Joanne McLean, Ariel Pakes, Sadek Wahba, Frank Wolak, Catherine Wolfram, and participants in several seminars in the US and Europe for comments and suggestions. Financial support from the Graduate School Fellowship Fund at Harvard University and the Alfred P. Sloan Doctoral Dissertation Fellowship Fund is gratefully acknowledged. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Measuring Market Power in the Ready-to-Eat
Cereal Industry
Aviv Nevo
NBER Working Paper No. 6387
January 1998
JEL Nos. L13, L66

ABSTRACT

The ready-to-eat cereal industry is characterized by high concentration, high price-cost margins, large advertising to sales ratios, and numerous introductions of new products. Previous researchers have concluded that the ready-to-eat cereal industry is a classic example of an industry with nearly collusive pricing behavior and intense non-price competition. This paper empirically examines this conclusion. In particular, I estimate price-cost margins, but more importantly I am able empirically to separate these margins into three parts: (1) that which is due to product differentiation; (2) that which is due to multi-product firm pricing; and (3) that due to potential price collusion. The results suggest that given the demand for different brands of cereal, the first two effects explain most of the observed price-cost markups. I conclude that prices in the industry are consistent with non-collusive pricing behavior, despite the high price-cost margins. Leading firms are able to maintain a portfolio of differentiated products, and influence the perceived quality of these products, and it is these two factors that lead to high price-cost margins.

Aviv Nevo
Department of Economics
549 Evans Hall #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and NBER
nevo@econ.berkeley.edu

1. INTRODUCTION

The ready-to-eat (RTE) cereal industry is characterized by high concentration, high price-cost margins, large advertising to sales ratios and aggressive introduction of new products. These facts lead Schmalensee (1978, pg. 315) to conclude that the RTE cereal industry is a classic example of a concentrated industry with differentiated products in which price competition is suppressed and rivalry is channeled into advertising and new product introduction. Furthermore, Schmalensee claims that, in game-theoretic terms, pricing conduct in the industry is approximately cooperative. Scherer (1982) agrees with this conclusion and argues: "...the cereal industry's conduct fits well the model of price competition-avoiding, non-price competition-prone oligopoly" (pg. 189).

This paper examines these conclusions regarding price competition in the RTE cereal industry. In particular, I estimate the true economic price cost margins (PCM) in the industry and empirically distinguish between three sources of these margins. The first explanation is the firms' 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 collude on prices.

My general strategy is to model demand as a function of product characteristics, heterogeneous consumer preferences, and unknown parameters. Extending recent developments in techniques for estimating demand of closely related products (see Bresnahan, 1987; Berry, 1994; Berry, Levinsohn, and Pakes, henceforth BLP, 1995), and data rarely available for academic research, I estimate the unknown demand parameters. These estimates are used to compute the PCM implied by three hypothetical industry structures: (1) single product firms; (2)

the current structure (i.e., a few firms with many brands each); 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 multi-product firm portfolio effect. Finally, the last structure produces the markups based on joint ownership, or full collusion.

The results suggest that the markups implied by the current industry structure, under a Nash-Bertrand pricing game, match the observed PCM. If we take Nash-Bertrand prices as the non-collusive benchmark, then even with PCM higher than 40% we have to conclude that pricing in the RTE cereal industry is approximately non-collusive. High PCM are not due to lack of price competition, but are due to consumers' willingness to pay for their favorite brand, and pricing decisions by firms that take into account 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 perceived quality of these products by means of advertising.

This paper follows the tradition of what has been termed the "New Empirical IO" (Bresnahan, 1989), which focuses on estimating PCM and marginal costs without observing actual costs. The markups, and the implied marginal costs, are inferred from firm conduct. For homogenous goods industries these methods have been widely used (for a survey see Bresnahan, 1989). However, in industries with 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), then assuming constant elasticity demand curves, 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 available data set.

One solution to this problem is given by the discrete choice literature (for example see McFadden, 1973, 1978, 1986; Cardell, 1989; Berry, 1994; or BLP). Here the dimensionality problem is solved by projecting the products onto a space of characteristics, making the relevant dimension the dimension of this space and not the square of the number of products.

In this paper I follow this approach and add to it in several ways. First, the method is applied to RTE cereal in which one might doubt the ability of observed product characteristics to explain utility. By adding a brand fixed effect, I control for unobserved quality for which previous work had to instrument. Potential difficulties with identifying all the parameters are solved using a minimum distance procedure, as in Chamberlain (1982). Second, the identifying assumptions used in this paper differ from those used in previous discrete choice models. Most previous work assumed brand characteristics are exogenous and identified demand parameters using this assumption, which is not consistent with a broader model in which brand characteristics are chosen by firms that account for consumer preferences. The identifying assumption used in this paper is consistent with this broader model. Third, I model heterogeneity as a function of the empirical non-parametric distribution of demographics, thus, partially relaxing the parametric assumptions used by BLP. Finally, I reduce the computation time substantially, which not only permits the more complex estimation problem in this paper, but also makes a Monte-Carlo study of the statistical properties of this class of estimators feasible.

An alternative to the model of demand followed here is given by Hausman, Leonard, and Zona (1994) and Hausman (1996), which demonstrate the use of a multilevel demand model to estimate demand for differentiated products. The essential idea is to use aggregation and separability assumptions to justify different levels of demand. 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, for example, between kids cereals 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 proportionally to the inverse of 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.

A comparison of the multilevel demand model and the method employed in this paper is beyond the scope of this paper. For a theoretical and empirical comparison the interested reader is referred to Nevo (1997a, Chapter 6). The importance, as far as this paper is considered, is that the conclusion reached here is not sensitive to the model of demand.

Although the demand estimates are consistent with strategic advertising and introduction of brand characteristics, these decisions are not modeled explicitly. Therefore, the results can be interpreted as measuring the short-run market power, i.e., without taking into account the market power firms might have because of their ability to introduce new brands and change perception of existing brands. For short-run price competition, which I examine in this paper, the lack of explicit modeling of advertising and brand introduction is not a problem. However, one has to be careful in drawing policy conclusions based on the 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 policy questions, would be a dynamic one that accounted for both brand introduction and advertising, and is the topic of separate work.

The paper is organized as follows. Section 2 gives a short description of the industry, both of the history and the current facts. In Section 3 I lay out the empirical model and discuss the implications of different modeling decisions. Section 4 describes the data, the estimation procedure, instruments, and the inclusion of brand fixed effects. Results for two models, and different sets of instruments are presented in Section 5. Section 6 concludes and outlines extensions.

2. THE READY-TO-EAT CEREAL INDUSTRY

2.1 *History and Background*[2]

The first ready-to-eat cold breakfast cereal was probably introduced by James Caleb Jackson in 1863, at his Jackson Sanatorium in Dansville, New York. The product, named Granula, was made of baked bran filled whole wheat flour, and had to be soaked in milk overnight. Almost a quarter of a century later in a public library in Cambridge, Massachusetts Henry Perky began his search for a cure to his digestive problem. In 1884 he presented his discovery, shredded wheat, at the World's Fair in Boston. Based in Worcester, Massachusetts and Niagara Falls, his company, the Natural Food Company, produced what became known as "Shredded Wheat." In 1928 this company was sold to the National Biscuit Company, now known as Nabisco, for thirty-five million dollars.

Dr. John Harvey Kellogg offered Perky $100,000 for his business, in 1984, but was turned down. Dr. Kellogg was the manager of the vegetarian Seventh-Day Adventist

[2]For a detailed non economic description of the evolution of the industry see Bruce and Crawford (1995); for an economic analysis see Scherer (1982).

Sanatorium in Battle Creek, Michigan. One of his biggest challenges was providing suitable breakfast food. His first attempt to solve the problem owed a great deal to the work of Dr. Jackson. Kellogg and his wife made an attempt at producing their own version of Jackson's Granula, and called it by the same name. This was the cause of the first of many industry courtroom battles. In 1881 Jackson sued Kellogg and won, after which Kellogg changed the name of his product to Granola.

In 1885 Kellogg began production of his patented flaked cereal called Granose. The flakes were an immediate hit. At this point, Kellogg set up the Sanitas Food Company, and hired his younger brother Will Keith Kellogg. In 1898 the two brothers developed the Sanitas Toasted Corn Flakes, and worked together for a few more years to improve it. However, the harmony between the brothers did not last for long.

In 1906 W.K. Kellogg took over the Battle Creek Toasted Corn Flake Company. Competition was fierce. By 1911, 107 brands of cornflakes were being packaged in Battle Creek. Kellogg turned to aggressive merchandising techniques to increase sales. He used methods like premiums, taste tests, and advertising campaigns. The battle between the brothers ended in court where they fought over the right to use the Kellogg name. W.K. Kellogg won and his Battle Creek Toasted Corn Flake Company later turned into the Kellogg Company, today the leading world producer of ready-to-eat cereals.

One of the patients of Dr. Kellogg was Charles William Post, who thought Dr. Kellogg's cereals were so good that in 1892, after leaving the sanitarium, he moved down the road and started producing his own cereals. His first big success was Grape Nuts cereal, which he merchandised using aggressive advertising methods.

Quaker Oats has its origins in the hot oatmeal market. During the beginning of the century, the Quaker company made several unsuccessful attempts to enter the ready-to-eat cold cereal market. In 1902 they funded research that eventually led to the introduction of puffed wheat and puffed rice cereals at the 1904 World's Fair in St. Louis. Thus, by the turn of the century, four of the main producers of ready-to-eat cereals had been established.

The last major producer, a Minneapolis based milling company, launched in 1924 two experimental ventures: a new product line of ready to eat cereals and a radio station. This company, later called General Mills, took advantage of the radio station to advertise its new product, dubbed Wheaties, and created the world's first singing commercial.

2.2 *The Industry Today*

Fueled by vitamin fortification, pre-sweetening and the surge of interest in natural cereals, the sales of RTE cereals grew steadily. In 1995 the U.S. market consumed approximately 3 billion pounds of cereal, grossing roughly $9 billion in sales.

All the major players in the industry were present early. Advertising and new brand introduction, the important dimensions of competition, were also used from the start. However, in its first few decades, the industry's structure changed dramatically: from a fragmented industry at the turn of the century, to one of the most concentrated US industries by the late 40's. Table 1a shows the volume (pounds sold) market shares starting in 1988. The top three firms dominate the market, and the top six firms can almost be considered to define the market. The concentration measures increase when we look at dollar sales market shares, as seen in Table 1b.

By comparing Tables 1a and 1b we can see a few of the changes the industry has undergone in the last decade. The market share of private labels (brands sold under a retailer name) increased substantially during 1989 and 1990. A parallel and related trend is the decline in the market share of Kellogg, especially if we look at the revenue shares, due mostly to the increase in sales of private labels that imitate Kellogg's national brands. General Mills has been able to maintain both its market share, but even more so its high prices, due to little private label entry into its "niches" (although there are generic versions of the different types of Cheerios, they do not seem to have caught on). The dynamics of prices can be seen directly in Table 1c. From 1988 to 1992, the prices for branded cereal increased by 36%, while private label prices increased only by 23%.

These trends of price increases, continue through 1993 and 1994, during which high price inflation was accompanied by an increased issue of coupons. However, this strategy seemed to have backfired, as the market shares of private labels only increased. 1995 saw a slowdown in inflation, a trend that continued into 1996. On April 16, 1996, Philip Morris (which owns the cereal lines of Post and Nabisco) announced a 20% price reduction and a decrease in the issue of coupons. This was followed by similar announcements by Kellogg, General Mills, and Quaker Oats.

The growth of private label market shares continued beyond the years described in Table 1a-b. In 1996 the private label share exceeded 10%, yet it is still much lower than the market shares of private labels in other breakfast food categories, which vary between 20 and 90 percent. The issue of private labels, entry of generic brands, and factors that determine the success or failure of generic brands is of great importance to understanding the industry. However, in this paper I am not able to consider these issues due to data limitations.

For economists the concentration of the industry is troublesome because the industry leaders have been consistently earning high profits.[3] This has drawn the attention of regulatory

[3] Fruhan (1979, chapter 1) ranked Kellogg's as 3 out of 1285 U.S. nonfinancial corporations in terms of profitability, while Mueller (1986) estimated Kellogg's long run equilibrium profits rate to be 120% above the

agencies to the practices in the industry. Perhaps the best known case was the anti-competitive complaint brought by the FTC against the top three manufacturers -- Kellogg, General Mills and Post -- in the 1970's.

The focus of that specific complaint was one of the industry's key characteristics: an enormous amount of brands. There are currently over 200 brands, even without counting negligible brands or market tests. The brand level market shares vary from 5% (Kellogg's Corn Flakes and General Mills' Cheerios) to 1%(the 25th brand) to less than 0.1%(the 100th brand). Not only are there many brands in the industry, but the rate at which new ones are introduced is high and has been increasing over time.

During the 80's the top six producers introduced 67 new major brands, out of which only 25 (37 percent) survived to 1993,[4] while in the period between 1950 and 1972, only 80 new brands were introduced.[5] Brand extensions (like General Mills Apple Cinnamon Cheerios) are generally considered both cheaper to introduce (due to economies of scale in advertising and existing expertise in production) and more likely to succeed. Surprisingly, very little successful entry into other producers' "niches" is observed. An example is lack of successful entry into the "Cheerios niche", which accounts for three of the top selling brands.

Competition by means of advertising was a characteristic of the industry since its early days. Today, television advertising expenditures by cereal producers are second only to expenditures by automobile manufacturers. Advertising to sales ratios are about 13 percent, compared to 2-4 percent in other food industries. In order for a new brand to succeed, its

---

mean return of U.S. industrial firms. Scherer (1982) reports the weighted average after-tax returns on the cereal division assets, for the industry leaders, was 19.8%, for 1958-1970.

[4]See Corts (1996a) Exhibit 5.

[5]See Schmalensee (1978, pg. 306) and Scherer(1982, Table 3).

introduction must be accompanied by a costly marketing campaign. Even existing, well established, brands maintain high advertising levels. This is surprising because one would like to think of cereal as an "experience" good, where advertising would have less of an informative effect.

Contrary to common belief, RTE cereals are quite complicated to produce. There are five basic methods used in the production of RTE cereals:[6] granulation, flaking, shredding, puffing and extrusion. The first four have been used since the beginning of the century, while the extrusion process, in which dough is pressed through a die to form the desired shape before baking, was introduced in 1941 with the introduction of General Mills Cheerios. Although the fundamentals of the production are simple and well known, these processes, especially extrusion, require production experience.

A typical plant will produce $400 million of output per year, employ 800 workers, and will require an initial investment of $300 million. Several brands are produced in a single location in order to exploit economies of scale in packaging. Table 2a presents estimates of the cost of production, computed from aggregate Census of Manufacturers SIC 2043. The second column presents the equivalent figures for the food sector as a whole (SIC 20). The gross price-average variable cost margin for the RTE cereal industry is 64.4%, compared to 26.5% for the aggregate food sector.

These margins are also supported by an accounting estimate of the price-marginal cost margins taken from Cotterill (1996), and presented in Table 2b. Here the estimated gross margin is 7 percentage points lower than before, which can be attributed to the fact that these

[6]For a more detailed discussion of production and distribution see Corts(1996a).

are marginal costs versus average costs.[7] The last column of the table presents the retail margins.

3. The Empirical Framework

My general strategy is to consider different models of supply conduct. For each model of supply, the pricing decision depends on brand level demand, which is modeled 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. I use additional information on costs to compute observed PCM and choose the conduct model that best fits these margins.

3.1 *Supply*

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

$$\Pi_f = \sum_{j \in \mathcal{F}_f} (p_j - mc_j) M s_j(p) - C_f$$

where $s_j(p)$ is the market share of brand j, which is a function of the prices of all brands, M is the size of the market, and C_f are the fixed cost of production. Assuming (1) the existence of a pure-strategy Bertrand-Nash 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

$$s_j(p) + \sum_{r \in \mathcal{F}_f} (p_r - mc_r) \frac{\partial s_r(p)}{\partial p_j} = 0.$$

[7]If we envision a U-shaped cost curve and believe that existing plants are producing at least the minimum efficient scale (see Scherer ,1982, for support), then marginal costs will be higher than average costs.

These set of J equations imply price-costs margins for each good. The markups can be solved for explicitly by defining $S_{jr} = -\partial s_r / \partial p_j \quad j,r = 1,...,J$,

$$\Omega^*_{jr} = \begin{cases} 1, & \text{if } \exists f\text{: } \{r,j\} \subset \mathscr{F}_f; \\ 0, & \text{otherwise} \end{cases}$$

and $\Omega_{jr} = \Omega^*_{jr} * S_{jr}$. In vector notation the first order conditions become

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

This implies a markup equation

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

Using estimates of the demand parameters, we can estimate PCM without observing actual costs, and we 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. The first structure is single product firms, in which the price of each brand is set by a profit maximizing agent that considers only the profits from that brand . A second is the current structure, in which multi-product firms 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. The final structure is joint profit maximization of all the brands, which corresponds to monopoly or perfect price collusion. Each of these is estimated by defining the ownership structure, $\mathscr{F}_f$, and ownership matrix, Ω^*.

PCM in the first structure arise only from product differentiation. The difference between the margins in the first two cases is due to the portfolio effect. The last structure 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 you 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 minimizing the difference between computed PCM, which are a function of conduct parameters and demand elasticities, and the margins implied by cost data. However, one should be careful in giving structural interpretation to this parameter both for theoretical and identification reasons (see Corts, 1996b).

3.2 *Demand*

The exercise suggested in the previous section allows us to estimate the PCM, and separate them into different parts. However, it relies on the ability to consistently estimate the 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 approach taken by the discrete choice literature and solve the dimensionality problem by projecting the products onto a characteristics space, making the relevant dimension the dimension of this space and not the number of products.

Suppose we observe t=1,...,T markets, each with i=1,...,I_t consumers. In the results below a market will be defined as a city-quarter combination. The conditional indirect utility [8] of consumer i from product j at market t is

$$u_{ijt} = x_j \beta_i^* - \alpha_i^* p_{jt} + \xi_j + \Delta\xi_{jt} + \epsilon_{ijt}, \qquad i=1,\ldots,I_t \quad j=1,\ldots,J, \quad t=1,\ldots,T \tag{2}$$

[8] This indirect utility form can be derived from a quasi-linear utility function.

where x_j are K observable product characteristics, p_{jt} is the price of product j in market t, ξ_j is the national mean valuation of the unobserved (by the econometrician) product characteristics, $\Delta\xi_{jt}$ is a city-quarter specific deviation from this mean, and ε_{ijt} is a mean zero stochastic term. Finally, $(\alpha_i^* \ \beta_i^*)$ are K+1 individual specific coefficients.

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 merchandising (other than national advertising). In the results presented below I control for the vertical component, ξ_j, by including a brand-specific brand dummy variables in the regressions. The consumer and market specific components are included in $\Delta\xi_{jt}$ and are left as "error terms".[9] I assume both firms and consumers observe all the product characteristics and take them into consideration when making decisions.

Ideally we would observe individual purchase decisions and transactions prices. The data used below consists of aggregate market (city-quarter) sales and average pre-manufacturer coupon transaction prices. Nevertheless, the outlined estimation procedure allows us to estimate individual utility parameters and their distribution in the population.

I model the distribution of consumers taste parameters for the characteristics as multi-variate normal (conditional on demographics) with a mean that is a function of demographic variables and parameters to be estimated, and a variance-covariance matrix to be estimated. Let $\gamma_i^* = (\alpha_i^*, \beta_{i1}^*, \ldots, \beta_{iK}^*)$ and $\gamma = (\alpha, \beta_1, \ldots, \beta_K)$, where K is the dimension of the observed characteristics vector; therefore,

---

[9]An alternative is to model the distribution of the valuation of the unobserved characteristics, as in Das, Olley and Pakes (1994). For a further discussion see Nevo (1997b).

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

where D_i is a $d\times1$ vector of demographic variables, Π is a $(K+1)\times d$ matrix of coefficients that measure how the taste characteristics vary with demographics, and Σ is a scaling matrix.[10] This specification allows the individual characteristics to consist of demographics that are observed and additional characteristics that are unobserved, denoted D_i and v_i respectively. In the results below observed demographic variables include log income and log age.

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

$$u_{i0t} = \xi_0 + \pi_0 D_i + \sigma_0 v_{i0} + \epsilon_{i0t} \ .$$

In the results given below 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 ξ_0 to zero. The coefficients, π_0 and σ_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. The vector $\theta_1=(\alpha,\beta)$ contains the linear parameters and the vector $\theta_2=(\Pi, \Sigma, \pi_0, \sigma_0)$ the non-linear parameters.[11] Thus, combining equations (2) and (3) we get

(4) $$\begin{gathered} u_{ijt} = \delta_{jt}(x_j, p_{jt}, \xi_j, \Delta\xi_{jt}; \theta_1) + \mu_{ijt}(x_j, p_{jt}, v_i, D_i; \theta_2) + \epsilon_{ijt} \\ \delta_{jt} = x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt}, \quad \mu_{ijt} = [p_{jt}, x_j] * (\Pi D_i + \Sigma v_i) \end{gathered}$$

---

[10]Alternatively, one could think of a composite "error" term, v_i^*, which is distributed $N(0,\Sigma^*)$ and Σ is the Cholesky factorization of Σ^*.

[11]The reasons for names will become apparent below.

where $[p_{jt}, x_j]$ is a $(K+1)\times 1$ vector. The utility is now expressed as the mean utility, represented by δ_{jt}, and a mean zero heteroskedastic deviation from that mean, $\mu_{ijt} + \varepsilon_{ijt}$, that captures the effects of the random coefficients. The estimation exploits this separation to (1) reduce the number of parameters that enter in a non-linear fashion and (2) generate linear moment conditions.

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

$$A_{jt}(x,p_{.t},\delta_{.t};\theta_2) = \{(D_i, v_i, \epsilon_{it}) \mid u_{ijt} \geq u_{ilt} \;\; \forall l=0,1,\ldots,J\}$$

where x are the characteristics of all brands, $p_{.t} = (p_{1t},\ldots,p_{Jt})'$ and $\delta_{.t} = (\delta_{1t},\ldots,\delta_{Jt})'$. Assuming ties occur with zero probability, the market share of the jth product, as a function of the mean utility levels of all the J+1 goods, given the parameters, is

$$\begin{aligned} s_{jt}(x,p,\delta_{.t};\theta_2) &= \int_{A_{jt}} dP^*(D,v,\varepsilon) = \int_{A_{jt}} dP^*(\varepsilon \mid D,v)\, dP^*(v \mid D)\, dP^*(D) \\ &= \int_{A_{jt}} dP^*(\varepsilon)\, dP^*(v)\, dP^*(D) \;\; , \end{aligned} \tag{5}$$

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 assumptions made in equations (2) and (3).

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

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

$$\text{(6)} \qquad u_{ijt} = x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt} + \epsilon_{ijt}, \quad i=1,\ldots,I_t \quad j=1,\ldots,J, \quad t=1,\ldots,T.$$

If ε_{ijt} is distributed i.i.d. with a Type I extreme value distribution, this is the well-known (Multinominal) Logit model. The brand market shares relative to the total market, including the outside good, are

$$\text{(7)} \qquad s_{jt} = \frac{\exp(x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt})}{1 + \sum_{k=1}^{J} \exp(x_k\beta - \alpha p_{kt} + \xi_k + \Delta\xi_{kt})} \, .$$

Although the model implied by (6) and the extreme value distribution assumption is appealing, due to its tractability, it restricts the own- and cross-price elasticities.[12] If the observed market shares are small, and because prices enters equation (6) linearly, then the own-price elasticities implied by equation (6) are roughly proportional to own-price; the lower the price the lower (in absolute value) the elasticity. From equation (1) we see that this implies that the marginal costs of a low-price cereal is lower than that of a more expensive cereal (not just in value, but as a percentage of price.) Not only is this an unreasonable assumption, but it is also a direct implication of the functional form. If price enters equation (6) in a log form the implied own-price elasticities would be roughly constant.

An additional problem, which has been stressed in the literature, is with the cross-price elasticities. 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

[12]See Nevo (1997b) for a detailed discussion.

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 towards other brands in proportion to market shares, regardless of characteristics.

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.

The Nested Logit model allows for somewhat more flexible substitution patterns, yet retains the computational simplicity of the Logit structure by implying a closed form expression for the integral in equation (5).. 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 model does not help with the problem of own-price elasticities. This is usually handled by assuming some "nice" functional form, yet does not solve the problem of having the elasticities be driven by the functional form assumption.

An additional problem with both the Logit and the Nested Logit models, which will be important in the data used below, is that all consumer heterogeneity enters the model through the additive separable random shock. The distribution of this shock is not a function of market specific variables. In the data set used below I find there is a big difference in market shares between cities. However, the only other variable that varies between cities 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 shock, and imposing this parameter to be identical across cities.

The full model has several advantages over the Logit and Nested Logit models. It allows for flexible own- and cross-price substitution patterns, which are not constrained by a priori segmentation of the market (yet at the same time can take advantage of this segmentation.) If the price of a brand goes up the number of consumers that substitute away from it is determined by the demographics of consumers that purchase that brand, and not solely by the functional form. A brand that is consumed by price-sensitive consumers will have a higher own-price elasticity. Furthermore, since the composite random shock, $\mu_{ijt} + \varepsilon_{ijt}$, is no longer independent of the product characteristics, consumers are more likely to switch to brands with similar characteristics, rather than to the most popular brand. So, for example, consumers that substitute away from a sweet cereal have a higher than average taste for sugar and are more likely to substitute to other sweet cereals.

By allowing the distribution of the shock term to depend on demographic variables we let it vary across cities and regions. Thus, individuals with similar characteristics will tend to have similar purchasing patterns and the variation between cities in the market shares does not need to be explained solely by the variation in prices.

The full model identifies the parameters of the distribution of tastes by trying to fit differences in market shares between brands, but more importantly, between the same brands in different markets. BLP assume a parametric functional form for this distribution of tastes. This paper, partially, relaxes this parametric form by using the non-parametric distribution of demographics which varies between cities.

Unfortunately, these advantages do not come without cost. Estimation of the model specified by equation (4) is not as simple as that of the Logit or Nested Logit models. There are two immediate problems. First, the integral given in equation (5) no longer has an analytic closed form (like that given in equation (7) in the Logit case). Furthermore, the computation of the integral in equation (5) is difficult. This is solved using the simulation technique, introduced by Pakes (1986), to compute the integral. Second, we now require information on the distribution of demographics in the population in order to compute the market shares. This is solved by sampling individuals from the CPS.

4. DATA AND ESTIMATION

4.1 *The Data*

The data required to consistently estimate the model previously described consists of the following variables: market shares in each market (in this application each city-quarter), prices in each market, brand characteristics, advertising, and information on the distribution of demographics.

Market shares and prices were obtained from the IRI Infoscan Data Base at the University of Connecticut.[13] These data were collected by Information Resources, Inc. (IRI), a marketing firm in Chicago, using scanning devices in a national random sample of supermarkets located in various size metropolitan areas and rural towns. Weekly data for UPC-coded products are drawn from a sample which represents the universe of supermarkets with annual sales of more than $2 million dollars, accounting for 82% of grocery sales in the US. In most cities the

[13] I am grateful to Ronald Cotterill, the director of the Food Marketing Center at the University of Connecticut, for making these data available.

sample covers more than 20% of the relevant population, and due to the importance of the sample to its costumers, IRI makes an effort to make the sample representative. This is confirmed by unpublished analysis conducted by the BLS.

In the Infoscan Data Base the data are aggregated by brand (for example different size boxes are considered one brand), city[14] and quarter. The data covers up to 65 different cities (the exact number increases over time), and ranges from the first quarter of 1988 to the last quarter of 1992. Three types of variables are collected: sales, demographics, and market. These are described in Table 3. In this paper I use only the sales data.

Market shares are defined by converting volume sales into number of servings sold,[15] and dividing by the total potential number of servings in a city in a quarter. This potential was assumed to be one serving per capita per day.[16] The outside good market share was defined as the residual between one and the sum of the observed market shares.

A price variable was created by dividing the dollar sales by the number of servings sold, and was deflated using a regional urban consumers CPI. The dollar sales reflect the price paid by consumers at the cashier, generating an average real per serving transaction price. However, the sales data does not account for any coupons used post purchase. If coupons are used evenly across brands this is not a problem; otherwise the results are potentially biased. One should keep in mind that the data are from a period when coupons were issued less frequently than they are today.

The variation in both prices and market shares is presented in Table 4a. The last three

---

[14]Most of IRI's definition of cities are similar, but not identical, to MSA's.

[15]This was done by using the serving weight suggested by the manufacturer, which are assumed correct (or at least proportional to the "true" serving weight).

[16]Therefore, the total market size is defined as τ*population*365/4, with τ is assumed equal to 1. Alternatively, τ can be estimated.

columns show the percentage of the variance due to brand, city, and quarter dummy variables.[17] Controlling for the variation between brands, most of the variation in market shares, and even more so in prices, is due to differences between cities. The variation in prices is due both to endogenous variation (i.e., variation correlated with demand shocks) and exogenous variation. Consistent estimation will have to separate these.

The Infoscan data was matched with a few other sources. First, advertising data was taken from the Leading National Advertising data base, which contains quarterly national advertising expenditures by brand collected from 10 media sources.[18] I used only the total of the 10 types of media.

Product characteristics were collected in local supermarkets by examining cereal boxes. This implicitly assumes that the characteristics have not changed since 1988. Although this is not exactly true, it seems a reasonable first approximation. Each cereal was classified into "mushy" or not, depending on its sogginess in milk.[19] There might be some measurement error in this classification. Table 4b provides statistics for the sample of 25 brands used below.

Information on the distribution of demographics was obtained by sampling individuals from the March Current Population Survey for each year. Individual income was obtained by dividing household income by the size of the household. The sample statistics are presented in Table 4c. The national averages are representative of Census statistics. However, the statistics on city level demographics do not exactly match the analogous statistics from Census data. This is due to the small sample, and will have to be accounted in the estimation as an additional

[17]To be precise, this is defined as the part of the variation explained by either the brand, city, or quarter dummy variables, divided by the total variation in the variable.

[18]The sources include: magazines, Sunday magazines, newspapers, outdoor, network television, spot television, syndicated television, cable networks, network radio and national spot radio.

[19]I wish to thank Sandy Black for suggesting this variable and helping me classify the various brands.

source of variation.

Finally, instrumental variables were constructed using two additional data sources. An average of wages paid in the supermarket sector in each city was constructed from the NBER CPS Monthly Earning Extracts. Estimates of city density were taken from the BLS, as were regional price indices.

4.2 *The Estimation Algorithm*

In this section I outline how the parameters of the models described in Section 3 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 to compute the implied "error term", for a given value of the unknown parameters, 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 instrumental variables is deferred to the next section.

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

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

where $s(\cdot)$ are the market shares given by equation (5) and S are the observed market shares. However, this approach will not be taken for several reasons. First, all the parameters enter the minimization problem in equation (8) in a non-linear fashion. In the application below the inclusion of brand and time dummy variables results in a large number of parameters and a costly non-linear minimization problem. 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 equation (8), while it is totally natural to do so in the method used below.

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

(9)
$$E\left[Z \cdot \omega(\theta^*)\right] = 0 \ ,$$

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

(10)
$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \ \omega(\theta)' Z A^{-1} Z' \omega(\theta)$$

where A 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 (9), is equal to zero, so we choose our estimate such that it sets the sample analog of the moments defined in equation (9), i.e. $Z'\hat{\omega}$, to zero. If there are more independent moment equations than parameters (i.e., dim(Z)>dim(θ)), we can not 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

$$s_{.t}\left(\delta_{.t}; \theta_2\right) = S_{.t} \ .$$

In some cases (for example, the Logit or Nested Logit models) this 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

(11)
$$\omega_{jt} = \delta_{jt}(S_{.t};\theta_2) - (x_j\beta + \alpha p_{jt}) \ .$$

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 non-linearly.

The intuition in this definition is as follows. For a given value of the non-linear parameters, θ_2, we compute what is the mean valuation, $\delta_{jt}(\cdot)$, 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, α and β. The estimator, defined by equation (10), is the one the minimizes the distance between these different predictions.

Usually,[20] the error term, as defined by equation (11), is the unobserved product characteristic, ξ_j. However, due to the richness of my data I am able to include brand-specific dummy variables as product characteristics. These dummy variables include both the mean quality index of observed characteristics, βx_j, and the unobserved characteristics, ξ_j. Thus, the error term is the city-quarter specific deviation from the main valuation, i.e., $\Delta\xi_{jt}$. The inclusion of brand dummy variables introduce 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,[21] this procedure simplifies to two-stage least squares. In the full random coefficients model, both

---

[20]See for example Berry (1994), BLP (1995), Berry, Carnall, and Spiller (1996), Bresnahan, Stern and Trajtenberg (1997).

[21]I.e., $A=Z'Z$, which is the "optimal" weight matrix under the assumption of homoscadestic errors.

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 (10) is then computed using a non-linear search. This search is simplified in two ways. First, we note that the first order conditions of the minimization problem defined in equation (10) 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 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.

Some details of the computation are given in Appendix A, for a detailed discussion see Nevo (1997b).

4.3 *Instruments*

The key identifying assumption in the algorithm previously given is equation (9), which requires a set of exogenous instrumental variables. The first set that comes to mind are the instrumental variables defined by ordinary least squares, namely the regressors (or more generally the derivative of the moment function with respect to the parameters.) In order to determine the validity of this assumption we examine the pricing decision. By equation (1), prices are a function of marginal costs and a markup term,

$$p_{jt} = mc_{jt} + f(\xi_{jt},\ldots) = (mc_j + f_j) + (\Delta mc_{jt} + \Delta f_{jt}) \quad . \tag{12}$$

This can be decomposed into a national component (that does not vary by city and quarter), and a deviation from this national mean. As pointed out, once brand dummy variables are included in the regression, the error term is the unobserved city-quarter specific deviation from the national time mean valuation of the brand. Since I assumed that players in the industry observe and account for this deviation, it will influence the market specific markup term and bias the estimate of price sensitivity, α, towards zero. Indeed, the results presented in the next section support this.

By examining the pricing decision given in equation (12) we can evaluate the instrumental variables used in much of the previous work (see footnote 20), which include the product characteristics (excluding price and other potentially endogenous variables), the sum over all the firm's other brand of the same characteristics (if the firm produces more than one product) and the sum over all the firm's competing brands characteristics. These variables attempt to proxy for the degree and closeness of competition the brand is facing. The key moment condition, defined by equation (9), is as reasonable (or unreasonable) here as it was in previous work.[22] However, for our purposes the problem with these instruments is much more fundamental: there is little variation in these instruments over time, and even less so between cities, while the inclusion of brand dummy variables results in most of the additional variation in a brand's price being between cities (see Table 4a), so one would expect these instrumental variables to be weak.[23] These variables were tried, but as expected they have little power and the results are not presented.

Two sets of instrumental variables were used in an attempt to separate the exogenous variation in prices (due to differences in marginal costs) and endogenous variation (due to

---

[22]These instrumental variables treat the "location" of the products, in the attribute space, as exogenous. Yet, a more general model has these "locations" determined endogenously.

[23]We recall that a valid IV must posses two properties: it must be uncorrelated with the error term, and it should be correlated with the regressor. A weak IV is one with the second requirement is just barely satisfied, namely the correlation with the regressor is low.

differences 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 cities (but are allowed to be correlated within a city).[24] Given this assumption, the prices of the brand in other cities are valid IV's; from equation (12) we see that 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 market specific valuation. Potentially, one could use the prices in all other cities and all quarters as instruments. I use regional quarterly average prices (excluding the city being instrumented).[25]

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 reduces the risk of cancer. This discovery will increase the unobserved valuation of all fiber intensive brands in all cities, and the independence assumption will be violated. However, the results below concentrate on well established brands for which it seems reasonable to assume there are less systematic demand shocks. Also, aggregate shocks to the cereal market will be captured by time dummy variables.

Suppose one believes that local advertising and promotions are coordinated across city borders, but within regions, and that these activities influence demand. Then the independence assumption will be violated for cities in the same region, and prices in cities in the same region will not be valid instrumental variables. However, given the size of the IRI "cities" (which in most cases are larger than MSA's) and the size of the Census regions this might be less of a problem. The size of the IRI city determines how far the activity has to go in order to cross city

---

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

[25]There is no claim made here with regards to the "optimality" of these IV's. A potentially interesting question might be are there other ways of weighting the information from different cities.

borders; the larger the city the smaller the chance of correlation with neighboring regions. Similarly, the larger the Census region the less likely is correlation with all cities in the region. Finally, the IRI data is used by the firms in the industry; therefore, it is not unlikely that they base their strategies on this geographic split.

Determining how plausible are these, and possibly other, situations is an empirical issue. I approach it by examining another set of instrumental variables that attempts to proxy for the marginal costs directly and compare the difference between the estimates implied by the different sets of IV's. The marginal costs include production (materials, labor and energy), packaging, and distribution costs. Direct production and packaging costs exhibit little variation, and are too small a percentage of marginal costs, to be correlated with prices. Also, except for small variations over time, a brand dummy variable, which is included as one of the regressors, proxies for these costs. The last component of marginal costs, distribution costs, includes: costs of transportation, shelf space, and labor. These are proxied by region dummy variables, which pick up transportation costs; city density, which is a proxy for the difference in the cost of space; and average city earning in the supermarket sector computed from the CPS Monthly Earning Files.

A consistent regional shock for certain brands will violate the validity of these IV's. If, for example, all western states value natural cereals more than east-coast states, region-specific dummy variables will be correlated with the error term (which is this valuation). However, in order for this argument to work the different valuation of brands has to be above and beyond what is explained by demographics and heterogeneity, since both are controlled for.

4.4 *Brand-Specific Dummy Variables*

A main difference between this work and previous work is the inclusion of brand-specific dummy variables as part of the product characteristics. 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 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 were not to include a brand fixed effect.

Furthermore, a major motivation (see 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_j. A brand-specific dummy variable captures the characteristics that do not vary by market, namely, $x_j\beta + \xi_j$. Therefore, the correlation between prices and the unobserved quality is fully accounted for and does not require an instrument. In order to introduce brand-specific dummy variable we require observations on more than one market. However, even without these 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 quality. Rather, it is market specific deviations from this unobserved national 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 difficultly 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 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 all enter as part of the 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 problem seems meager.

A more serious objection to the use of brand dummy variables is that 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 $d=(d_1, \ldots, d_J)'$ denote the $J\times 1$ vector of brand dummy coefficients, X be the $J\times K$ ($K<J$) matrix of product characteristics, and $\xi=(\xi_1, \ldots, \xi_J)'$ be the $J\times 1$ vector of unobserved product qualities. Then from (2)

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

The estimates of β and ξ are

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

where $\hat{d}$, is the vector of coefficients estimated from the procedure described in the previous section, and Ω is the variance-covariance matrix of these estimates. The coefficients on the brand dummy variables provide an "unrestricted" estimate of the mean utility. The minimum distance estimates project these estimates onto a lower dimensional space, which is implied by a "restricted" model that sets ξ to zero. Chamberlain provides a χ^2 test to evaluate these restrictions.

5. RESULTS

I provide two sets of results. First, I examine the Logit model which, as previously pointed out, imposes restrictions on the pattern of substitution. These results are given as a baseline comparison. They also easily demonstrate the importance of instrumenting for the endogenous variables and allow the comparison of different sets of instrumental variables. Next, I present results from the full random coefficients model described in Section 3.

The results presented were computed using the 25 brands with highest national market shares in the last quarter of 1992. For all, except one, there are 1124 observations (i.e., they are present in all quarters and all cities). The exception is Post Honey Bunches of Oats, which appears in the data only in the first quarter of 1989. The sample is given in Table 4. The combined market share of the brands in the sample varies between 43 and 62 percent of the total volume of cereal sold in each city and quarter. If we look at national shares, the combined market share varies between 55 and 60 percent.

5.1 *Logit Results*

As pointed out in Section 3 the Logit model specified by equation (6) yields restrictive and unrealistic substitution patterns, and therefore is inadequate for measuring market power. Nevertheless, due to its computational simplicity it is a useful tool in getting a feel for the data. In this section I use the Logit model to examine: (1) the importance of instrumenting for price; and (2) the effects of the different sets of instrumental variables discussed in the previous section.

Table 6 displays the results obtained by regressing $\ln(s_{jt})$ - $\ln(s_{0t})$ on prices, advertising expenditures, brand and time dummy variables. In the first column I report the results of ordinary least squares regression. The coefficient on price and the implied own price elasticities are relatively low. Since the Logit demand structure does not impose a constant elasticity, the estimates imply a different elasticity for each brand-city-quarter combination. Some statistics of the distribution of the elasticities are shown in the bottom of column (i). The low elasticities and the high number of inelastic demands seem to support the claim previously made that OLS estimates are downward biased.

Column (ii) of Table 6 uses the average regional prices as instrumental variables in a two stage least squares regression. These IV's are valid under the assumptions given in the previous section. Not surprisingly, the coefficient on price increases and all demands are elastic. Column (iii) uses a different set of IV's: the proxies for city level marginal costs. The coefficient on price is similar in the two regressions.

The similarity between the estimates of the price coefficient continues to hold when we introduce demographics into the regression equation. Columns (iv)-(v) present the results from the previous two sets of IV's, while column (vi) presents an estimation using both sets of instruments jointly. Potentially, more demographics could be added to the regression, but since the purpose of the Logit model is mainly descriptive this is done only for the full model.

The first stage R-squared and F-statistic for all the instrumental variable regressions are high, suggesting (although not promising) that the IV's have some power. The tests of over-identification are all rejected, suggesting the identifying assumptions are not valid. However, it is unclear whether the large number of observations is the reason for the rejection[26] or that the

[26]It is well known that with a large enough sample a χ^2 test will reject essentially any model.

IV's are not valid.

The addition of demographics increases the absolute value of the price coefficient; thus, increasing, in absolute value, the price elasticity. As we recall from the previous section, if there are regional demand shocks then both sets of IV's are not valid. City specific valuations might be a function of demographics, and if demographics are correlated within a region these valuations will be correlated. Under this story, adding demographics eliminates the omitted variable bias and improves the over identification test statistic.

The important thing to take from these results is the similarity between estimates using the two sets of IV's, and the importance of controlling for demographics and heterogeneity. The similarity between the coefficients does not promise the two sets of IV's will produce identical coefficients in different models or that these are valid IV's. However, I believe that with proper control for demographics and heterogeneity, as in the full model, these are valid IV's.

In the full random coefficients model, the coefficients on demographics capture the change in the value of the outside option as a function of demographics. If we take the same interpretation here, the results in Table 6 suggest the value of cereals increases with income, while age and household size seem to be non-significant.

The regression also includes advertising, which has a statistically significant coefficient. The mean elasticity with respect to advertising is approximately 0.06, which seems low: a Dorfman-Steiner condition requires advertising elasticities to be an order of magnitude higher. This is probably a result of measurement error in advertising data. Non linear effects in advertising were also tested for and were found to be non-significant.

The price-cost margins implied by the estimates are given in the first column of Table 10. A discussion of these results is deferred to later.

5.2 *Results from the Full Model*

The estimates of the full model are based on equation (4) and were computed using the procedure described in Section 4. Predicted market shares are computed using equation (5) and are based on the empirical distribution of demographics (as sampled from the March CPS for each year), independent normal distributions (for v), and Type I extreme value (for ε). Coefficients on price, calories from fat, sugar, fiber, a dummy variable that measures if the cereal gets soggy in milk, and segment specific dummy variables were allowed to vary between individuals as a function of demographics. The demographics considered are log income, log income squared, log age, and a dummy variable if the individual's age is less than 16 years . These demographics were obtained by sampling individuals from the March CPS for each year by cities (as defined by the IRI data). All demographic variables were expressed as deviations from the mean.[27]

Before discussing the results, recall that the main advantage of the full model is allowing for a distribution in the population of the marginal utility from the product characteristics. The results below give the mean of this distribution and measures of heterogeneity which include the influence of demographic variables, and a standard deviation that captures any additional heterogeneity not expressed by the demographics.

The results are presented in Tables 7a-b. Different specifications are displayed in different columns. The first column of each specification displays the parameter estimates, while the second presents the standard error. The first set of rows give the means of the taste parameters, α and β. Coefficients on price and advertising are estimated with the GMM procedure, while the coefficients on the physical characteristics come from a Minimum Distance

[27]Since the regression includes a constant this has no effect on the results, just as in the linear regression case. However, now the mean of the distribution can be read directly from the estimation results.

regression of the GMM brand dummy coefficients on product characteristics. The next three groups of rows give the parameters that measure heterogeneity in the population: standard deviations, interaction with income and interaction with age. Table 7b presents results for an additional specification. The coefficients are now organized somewhat differently.

The first specification limits the number of variables that interact with demographics; only the constant and price are allowed to vary as a function of individual demographics. In the second, sugar and the sogginess dummy variables are also allowed to vary with demographics. The third column displays the results from a specification where all physical characteristics are allowed to vary. Finally, additional characteristics and demographics are included in the specification presented in Table 7b. For expository simplicity, I focus on the second specification.

The discussion below is organized in the following fashion. I start by discussing the estimates of the means of the distribution of marginal utilities, the standard deviations, and the influence of demographics on these distributions. Measures of fit for the Minimum Distance procedure are presented, followed by a discussion of the effects of advertising. Next, I discuss the distribution of price sensitivity and the substitution patterns. I conclude by discussing the difference between the different specifications.

The means of the distribution of marginal utilities, β's, are estimated by a Minimum Distance procedure. This amounts to choosing estimates, $\hat{\beta}$, that minimize the difference between the coefficients on the brand dummy variables estimated in the main GMM regression and the predicted values, $X\hat{\beta}$, using the variance-covariance matrix of the estimates as a weight matrix. Except for *CAL from FAT*, all coefficients are statistically significant. For the average consumer, sugar has positive marginal utility, as does fiber. The coefficient on *SUGAR* is consistent with the finding of Stanley and Tschirhart (1991) and with evidence they bring from the U.S.D.A. Food Consumption Survey. Stanley and Tschirhart find a negative coefficient on *FIBER* and attribute this finding to the taste component of fiber dominating the nutrition component. Here I find that it is the nutrition component that dominates, which can be attributed to the taste of fiber-rich cereals improving from the time they conducted their study, or to a increased importance of nutrition. As expected, sogginess in milk is valued negatively by the average consumer.

The estimates of standard deviations of the taste parameters are all non-significant at conventional significance levels, while most interactions with demographics are significant. Marginal utility from sugar decreases with age and income, although these coefficients are not statistically significant (a fact that might be due to the log functional form). Marginal valuation of sogginess increases with age and income. In other words, adults are less sensitive to the crispness of a cereal as are wealthier (and possibly more health conscious) consumers. The distribution of the *MUSHY* coefficient can be seen in Figure 1; most of the consumers value sogginess in a negative way, but approximately 30% of consumers actually prefer a mushy cereal. The economic significance of the various hedonic variables can be seen in Table 8, and will be discussed below.

The ability of the observed characteristics to fit the coefficient on the brand dummy variables is measured in several ways. In the unrestricted version of the model the estimates of the brand dummy variables include $X_j\beta + \xi_j$. A restricted version has $\xi_j = 0$. This leads to a χ^2 test as in Chamberlain (1982). The results of this test are presented in the bottom of Table 7. Since the brand dummy variables are estimated very precisely (due to the large number of observations) it is not surprising that the restricted model is rejected. This motivates the modeling of unobserved quality, here as in previous work. A direct measure of fit is given by the amount of variation in the estimates of the brand dummy variables explained by the variation in observed characteristics, which can be measured in two ways: by using the original or transformed regression.[28] The first measure is more adequate for our purposes but is not constrained to be positive (it will, however, be no greater than one.) Both measures are provided at the bottom of the table. The un-weighted R-square suggests that although the unobserved quality is important, as expressed by the χ^2 test, the fit obtained by the observed characteristics is good.

The mean sensitivity to advertising is captured by the advertising coefficient, which is significant and slightly higher than that presented in Table 6. Like the Logit results, the implied elasticity, given in Table 8, seems to be too low, although much more in line with what a model of advertising would predict.

The mean price coefficient is of the same order of magnitude as those presented in Table 6. However, the implied elasticities and margins are different, as discussed below. Coefficients on the interaction of price with demographics are statistically significant. The estimate of the standard deviation is not statistically significant, suggesting that most of the heterogeneity is explained by the demographics (an issue we shall return to below). Consumers with above average income tend to be less price sensitive as do older consumers.

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. Part of the

---

[28]Just to be sure,

$$R_w^2 = 1 - \frac{(y - X\hat{\beta})'\Omega^{-1}(y - X\hat{\beta})}{(y-\bar{y})'\Omega^{-1}(y-\bar{y})}, \text{ and } R_{uw}^2 = 1 - \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{(y-\bar{y})'(y-\bar{y})}$$

where $\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ and Ω is the variance-covariance matrix of the coefficients on dummy variables.

tail of the distribution has individuals with positive price sensitivity -- the higher the price the higher their utility. This might be attributed to the model's inability to precisely estimate the tails of the distribution, or to a rare but existing economic phenomena. The percent of positive price coefficients is given in the last row of the table. This phenomenon is one of the main motivations to explore the additional specifications given in columns (iii) and (iv), which are discussed below.

Table 8 presents the demand elasticities with respect to the continuous attributes and price. The top part of the table shows some descriptive statistics for the distribution of these elasticities in the population, while the bottom part presents these elasticities for the brands present in the sample in the last quarter of 1992 in Boston. For each attribute the left column gives the value of that attribute and the right gives the demand elasticity.

The elasticities with respect to *SUGAR* illustrate the importance of considering both the mean and the standard deviation of the distribution of tastes. Consider, for example, the case of Quaker 100% Natural and General Mills Raisin Nut. Both have similar quantities of sugar per serving (14 and 16 grams) and similar market shares; therefore, in the Logit demand, which does not allow for a distribution of tastes, they would have similar demand elasticities with respect to sugar. In the demand system considered here this is not necessarily true. Indeed, we see in Table 9 that the estimated demand elasticities of these two brands is different, which is a result of the model predicting that consumers who purchase them are different. A somewhat disturbing pattern appears in the own-price elasticities, displayed in column (i): an almost linear relation exists between the estimated elasticity and the price. I discuss this in detail below.

Valuation of the outside good is measured by the constant term. Older consumers and consumers with above average income are less likely to buy cereal, and value the outside option higher. The distribution of the utility from the outside option can be seen in Figure 3.

As noted above, all the estimates of the standard deviations are statistically non-significant, suggesting that the heterogeneity in the coefficients is mostly explained by the included demographics. A measure of the relative importance of the demographics and the 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 .95 for the results in column (ii). Therefore, unlike previous work (see Rossi, McCulloch, and Allenby 1996), it seems as if the demographics explain most of the variation in the coefficients.

Different columns of Table 7 present the estimates from specifications that differ in the coefficients allowed to vary between individuals. The coefficients in the different columns vary slightly, with the price coefficients exhibiting the largest change. However, we note that these price coefficients are not directly comparable yet the implied PCM are similar, as discussed below. At the bottom of the table the value of the GMM objective function is provided. The test of over-identifying restrictions is satisfied. Note that these values cannot be used to test the restrictions between the columns because the weight matrix is not held constant.

As previously noted a major motivation for exploring the specifications in columns (iii) and (iv) was the high dispersion in price sensitivities presented in Figure 2. The specification given in column (ii) reduces this dispersion by not allowing the model to change the price sensitivity as a function of demographics. As in the results of column (ii) the estimates of the standard deviations are low, implying little heterogeneity in the price sensitivity. This suggests that the own-price elasticities would be similar to those implied by the Logit model and therefore suffer from the same troubling trend pointed out in Section 2. The sample of estimated own-price elasticities, presented in Table 8, supports this conclusion: the estimated own-price elasticities are nearly linear in price.

The estimates presented in column (iv) solve both these problems. On one hand, the distribution of the price sensitivity is almost completely negative, while on the hand it is quite disperse unlike the distribution implied by the results of column (iii). Note that this is achieved mainly by freeing the restrictive linear form in which log of income influenced the price coefficient; once we allow log income to enter in a non-linear fashion, by introducing a quadratic term, we achieve a reasonable distribution of price sensitivity, as seen in Figure 4. Furthermore, due to the high level of heterogeneity in price sensitivity the elasticities computed from the results in column (iv), which are presented in Table 9, do not exhibit the unrealistic pattern present in Table 8.

Table 9 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 j with respect to a change in the price of i. The figures are median the median of the elasticities of brands in the sample over all quarters and all cities, and are computed using the results of Table 7b. The results are intuitive. For example, Lucky Charms, a kids cereal, is most sensitive to a change in the price of Corn Pops and Froot Loops, also kids cereals. At the same time it is least sensitive to a change in the price of cereals like Corn Flakes, Total or Wheaties, all cereals aimed at different market segments. These substitution patterns are persistent across the table. The elasticities are of the same order of magnitude as those given by Hausman (1996), which were computed using different data and a different demand system suggesting that the conclusions drawn in the next section are not sensitive to the specific assumptions and data employed here.

An additional 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 restricts all these elasticities 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 ration of the maximum to the minimum cross-price elasticity (the Logit implies a ratio of one.) This ratio varies from 21 (Corn Flakes) to 3.5 (Shredded Wheat). Not only does this tell us the results have overcome the Logit restrictions, but also for which brands the characteristics do not seem strong enough to overcome the restrictions. Thus, suggesting which additional characteristics we might want to add.

Finally, the bottom row of Table 9 presents the elasticity of the share of the outside good with respect to the price of the "inside" goods. By comparing the ratio of these elasticities to the average in each column we see the relative importance of the outside good to that specific brand. So for example elasticity of the outside good is higher with respect to a change in the price of Kellogg's Corn Flakes than with respect to the price of Froot Loops. Not only os it higher in absolute terms, but more importantly it is higher as a ration of the average cross-price elasticity in that column.[29] Once again this is an intuitive result. Private label versions of Kellogg's Corn Flakes are available and have higher shares than generic versions of Froot Loops. All these generic products are included in the outside good and therefore it should not be surprising that the outside good is more sensitive to the price of Corn Flakes.

5.4 *Price-Cost Margins*

Given the demand parameters estimated in the previous sections, we can use equation (1)

[29] Comparing the absolute value of the elasticities across columns is somewhat meaningless, since in each column the absolute price change is different. In order to compare across columns semi-elasticities, or the percent change in market share due to say a 10 cents change in price, need to be computed.

to compute PCM for different conduct models. As explained in Section 3.1, I compute PCM for three hypothetical industry structures, thus placing bounds on the importance of the different causes for PCM.

Our theory is one of the markups charged by the manufacturers. However, the observed prices are the pre-coupons retail prices, which include a retail margin. As in much of the previous work, I assume that the retail margin is just a fixed percent of the wholesale price and can be considered as an additional cost to the manufacturers. This assumption is supported by practices in the industry, but more importantly by no observed correlation between the price and the structure of the retail market in different cities.

Table 10 presents the median PCM for the Logit model in the first column, and the full model in the next three columns. The different rows present the PCM that different models of pricing conduct predict, taking as given the brands offered and advertising. This means that these estimates should not be used to evaluate policy changes. For example, the PCM given in the first row are not a prediction of the PCM that would prevail if the industry was indeed divided into single product firms. Not only would the cost structure of such firms change, but also the decisions with respect to advertising and new brand introduction. However, these estimates can be used to distinguish between different models of pricing conduct.

Although the price coefficients from the full model, given in Table 7, are of similar order of magnitude to those for the Logit model, given in Table 6, the implied markups are different. Since the full model does a better job of estimating the cross price elasticities, it is not surprising that the difference increases as we go from single, to multi-product firms, and then to joint ownership.

Finally, we note that despite the differences in the different specifications in Table 7 they all imply similar price cost margins. Similar PCM are also implied by a variety of specifications examined, that due to space constraints are not presented here. As a matter of fact, the PCM implied by these different specifications is the most robust estimate of all the parameters. Furthermore, essentially the same PCM are implied by the estimates of Hausman (1996), which uses both a different model of demand and different data, and by Nevo (1997a Chapter 6), which uses the same data as in this paper but a different method for estimating demand.

In order to determine which model of conduct fits the industry, we need to compare the PCM computed assuming different models of conduct to actual margins. Unfortunately, we do not observe actual margins and will have to use accounting estimates. Such an estimate is given in Table 2b. The relative comparison for our purposes is the gross retail margin, estimated at 46.0%. Note, that this margin does not include promotional costs, some of which can be argued to be marginal costs (for example, coupons). For the conclusions made below this makes my estimate a conservative one.

These estimates are supported by Census data (presented in Table 2a), which as we saw are slightly higher because they are average variable costs and can therefore be considered an upper bound to PCM. A lower bound on the margins is the margin between the price of national brands and the corresponding private labels. Using data from Wongtrakool (1994), these margins are approximately 31%. Prices of private labels will be higher than marginal costs for several reasons. First, they also potentially include a markup term, but lower than the national brands. Second, the private label manufacturers might have different marginal costs, most likely higher. For these reasons this margin is only a lower bound on PCM.

After examining the margins computed in Table 2b, the bounds on these estimates, and the PCM in Table 10, we are left to conclude that a multi-product Nash-Bertrand equilibrium in prices is consistent with observed PCM. Furthermore, if there was any significant price collusion, as suggested by previous work, the observed margins would be much closer to those implied by joint ownership, and therefore much higher.

6. Conclusions and Extensions

This paper estimated a brand level demand system for RTE cereal. The estimated elasticities were used to compute price cost margins that would prevail under different conduct models. A Nash-Bertrand pricing game, played between multi-product firms (as the firms in the industry are), was found to be consistent with the observed price cost margins. Furthermore, it seems that if any significant price collusion existed, the observed margins would be much higher. 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%, prices in the industry are not a result of collusive behavior.

There is no claim in this paper regarding the competitiveness of the industry (which is different from non-collusive prices). The high observed PCM are due to the firms' ability to maintain a portfolio of differentiated brands and influence the perceived quality of these brands by means of advertising. In order to make claims regarding the competitiveness of the industry, or evaluate suggested policies, one has to deal with the dynamic issues of brand introduction and advertising, which were taken as given in this paper. Such a model is the subject of separate 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 comparison of this method to the one employed here is beyond the scope of this paper. The interested reader is referred to Nevo (1997a, Chapter 6).

References

Berry, S. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242-262.

Berry, S., M. Carnall, and P. Spiller (1996), "Airline Hubs: Costs and Markups and the Implications of Consumer Heterogeneity," National Bureau of Economic Research, Working paper no. 5561 (also available at http://www.econ.yale.edu/~steveb/imdex.html)

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

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

Bresnahan, T. (1989), "Empirical Methods for Industries with Market Power," in R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization*, Vol. II, Amsterdam: North-Holland.

Bresnahan, T., 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.

Broydon, C.G. (1970), "The Convergence of Double-rank Minimization Algorithms," *J. Inst. Maths. Applics.*, Vol. 6, 76-90.

Bruce, S., and B. Crawford (1995), *Cerealizing America*, Boston: Faber and Faber.

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.

Cardell, N.S. (1991), "Variance Components Structures for the Extreme Value and Logistic

Distributions," mimeo, Washington State University.

Chamberlain, G. (1982), "Multi Variate Regression Models for Panel Data," *Journal of Econometrics*, 18(1), 5-46.

Corts, K.S. (1996a), "The Ready-to-eat Breakfast Cereal Industry in 1994 (A)," Harvard Business School case number N9-795-191.

Corts, K.S. (1996b), "Conduct Parameters and the Measurement of Market Power: The Impossibility of Astructural Inference," Harvard Business School Working Paper #96-036.

Cotterill, R.W. (1996), "High Cereal Prices and the Prospects for Relief by Expansion of Private Label and Antitrust Enforcement," Testimony offered at the Congressional Forum on the Performance of the Cereal Industry, Washington D.C., March 12.

Das, S., S. Olley, and A. Pakes (1994), "Evolution of Brand Qualities of Consumer Electronics in the U.S.," mimeo.

Deaton, A., and J. Muellbauer (1980), "An Almost Ideal Demand System," *American Economic Review,* 70, 312-326.

Flecther, R. (1970), " A New Approach to Variable Metric Algorithms," *Computer Journal*, Vol. 13, 317-322.

Fruhan, W.H. (1979), *Financial Strategy: Studies in the Creation, Transfer, and Destruction of Shareholder Value*, Homewood, IL: Irwin.

Goldfarb, D. (1970), "A Family of Variable Metric Updates Derived by Variational Means," *Mathematics of Computing*, Vol. 24, 23-26.

Hausman, J. (1983), "Specification and Estimation of Simultaneous Equations Models,"in Z. Griliches and M. Intiligator, eds., *Handbook of Econometrics*, Amsterdam: North Holland.

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.

Hausman, J., G. Leonard, and J.D. Zona (1994), "Competitive Analysis with Differentiated Products," *Annales D'Economie et de Statistique*, 34, 159-80.

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

McFadden, D. (1978), "Modeling the Choice of Residential Location," in A. Karlgvist, et al., eds., *Spatial Interaction Theory and Planning Models*, Amsterdam: North-Holland.

McFadden, D. (1986), "Econometric Analysis of Qualitative Response Models," in Z. Griliches and M. Intilligator, eds., *Handbook of Econometrics, Volume III*, Amsterdam: North-Holland.

Mueller, D.C. (1986), *Profits in the Long Run*, Cambridge: Cambridge University Press.

Nelder, J. A., and R. Mead (1965), "A Simplex Method for Function Minimization," *Computer Journal*, Vol. 7, 308-313.

Nevo, A. (1997a), *Demand for Ready-to-Eat Cereal and Its Implications for Price Competition, Merger Analysis, and Valuation of New Goods*, Ph.D. Dissertation, Harvard University.

Nevo, A. (1997b), "A Research Assistant's Guide to Random Coefficients Discrete Choice Models of Demand," University of California at Berkeley, mimeo.

Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding A European Patent Stocks," *Econometrica*, 54, 755-784.

Rossi, P., R.E. McCulloch, and G.M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15(4), 321-40.

Scherer, F., M. (1982), "The Breakfast Cereal Industry," in W. Adams, ed., *The Structure of American Industry*, New York: Macmillian.

Schmalensee, R. (1978), "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry," *Bell Journal of Economics*, 9, 305-327.

Shanno, D. F., (1970), "Conditioning on Quasi-Newton Methods for Function Minimization," *Mathematics of Computing*, Vol. 24, 647-656.

Simon, C. P. , and L. Blume (1994), *Mathematics for Economists*, New York: W. W. Norton & Company.

Stanley, L. R., and J. Tschirhart (1991), "Hedonic Prices for a Nondurable Good: The Case of Breakfast Cereals," *The Review of Economics and Statistics*, 73(3), 537-541.

Wongtrakool, B. (1994), "An Assessment of the "Cereal Killers": Private Labels in the Ready-to-eat Cereal Industry," Senior Thesis, Harvard University.

Appendix A: Computational Details

In this appendix I lay out some of the computational details of the estimation algorithm for the Logit and full random coefficients models. The intuition of the method is the same for both models and they differ only in computational details. For a more detailed discussion see Nevo (1997b).

There are essentially four steps (plus an initial step) to follow in computing the estimates:

(0) prepare the data including draws from the distribution of v and D (only for full model);

(1) for a given value of θ_2 and δ, compute the market shares implied by equation (5) (only for full model);

(2) for a given θ_2, compute the vector δ that equates the market shares computed in Step 1 to the observed shares;

(3) for a given θ, compute the error term (as a function of the mean valuation computed in Step 2), interact it with the instruments, and compute the value of the objective function given in equation (10);

(4) search for the value of θ that minimizes the objective function computed in Step 3.

Step 0: For the full model draw ns draws from the distribution of v and the demographics, D. In the results presented above v was drawn from a multi-variate normal, and the demographics were drawn by sampling from the CPS.

Step 1: For a given value of θ_2 and δ, compute the predicted market shares given by the integral in equation (5). For the full model this integral has to be computed numerically. In the results above $P^*(\epsilon)$ is assumed to be Type 1 extreme value; therefore, the integral in equation (5) can be approximated by

$$s_j(p,x,\delta,P_{ns};\theta_2) = \frac{1}{ns}\sum_{i=1}^{ns} s_{ji} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\exp\left(\delta_j + \sum_{k=1}^{K} x_{jk}(\sigma_k v_{ik} + \pi_{k1} D_{i1} + \ldots + \pi_{kd} D_{id})\right)}{1 + \sum_{m=1}^{J} \exp\left(\delta_m + \sum_{k=1}^{K} x_{mk}(\sigma_k v_{ik} + \pi_{k1} D_{i1} + \ldots + \pi_{kd} D_{id})\right)} ,$$

where $(v_{i1},\ldots,v_{iK})$ and $(D_{i1},\ldots D_{id})$, $i=1,\ldots,ns$, are the draws from step (0).

Step 2: Compute the vector δ that equates the market shares computed in Step 1 to the observed shares (see Berry 1994 for proof of existence and uniqueness). For the Logit model this can be computed analytically by $\delta_j = \ln(s_j) - \ln(s_0)$. For the full model I use the contraction mapping suggested by BLP (see there for proof of convergence), which amounts to computing

$$\delta^{t+1} = \delta^t + \ln(S) - \ln\left(s(p,x,\delta^t,P_{ns};\theta_2)\right), \qquad t=0,\ldots,T,$$

where T is the smallest integer such that $\|\delta^T - \delta^{T-1}\|$ is smaller than some tolerance level, and δ^T is the approximation to δ. Convergence can be reached faster by choosing a good starting value, δ_0.

Step 3: For both the Logit and the full model the error term is

$$\omega = \delta - (x\beta - \alpha p) \ .$$

This is interacted in a straight forward way with the instrument matrix to form the moment conditions, which are used to compute the objective function given in equation (10).

Step 4: Search for the value of θ that minimizes the objective function. For the Logit model this can be done analytically (it is an IV regression). For the full model we need to perform a non-linear search over θ. This search can be sped up by using the first order conditions, with respect to θ_1, to express θ_1 as a function of θ_2, i.e.,

$$\hat{\theta}_1 = (X'ZA^{-1}Z'X)^{-1}X'ZA^{-1}Z'\delta(\hat{\theta}_2)$$

where X is a the design matrix and Z is the instrument matrix. Now, the non-linear search can be limited to θ_2.

Two search methods were explored. The Nelder-Mead (1965) non derivative "simplex" search method, and a quasi-Newton method with an analytic gradient. The latter used an Hessian updating method (developed by Broydon, 1970; Flecther, 1970; Goldfarb, 1970; and Shanno, 1970) and a mixed quadratic and cubic line search procedure. The first was more robust but was much slower to converge, while the latter was two orders of magnitude faster, yet was sensitive to starting values; due to the non-linear objective function, if the initial values were extremely poor the algorithm would reach regions where the objective was not defined. This was especially true if the variables were scaled differently. The recommended practice is to start with the non derivative method and switch to the gradient method.

The Jacobian of the function computed in Step 2 can be computed in the following manner. The mean valuations of the J brands in each market are implicitly defined by the following system of J equations

$$s_j(\delta_1,\ldots,\delta_J,\theta_2;x,p,P_{ns}) = S_j, \quad j=1,\ldots,J.$$

By the Implicit Function Theorem (see Simon and Blume, 1994, Theorem 15.7 pg 355) the derivatives of the mean value with respect to the parameters are

$$\textbf{(13)} \qquad D\delta = \begin{pmatrix} \frac{\partial\delta_1}{\partial\theta_{21}} & \cdots & \frac{\partial\delta_1}{\partial\theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\delta_J}{\partial\theta_{21}} & \cdots & \frac{\partial\delta_J}{\partial\theta_{2L}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial s_1}{\partial\delta_1} & \cdots & \frac{\partial s_1}{\partial\delta_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_J}{\partial\delta_1} & \cdots & \frac{\partial s_J}{\partial\delta_J} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial s_1}{\partial\theta_{21}} & \cdots & \frac{\partial s_1}{\partial\theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_J}{\partial\theta_{21}} & \cdots & \frac{\partial s_J}{\partial\theta_{2L}} \end{pmatrix},$$

where θ_{2i}, $i=1,\ldots,L$ denotes the i's element of the vector θ_2, which contains the non-linear parameters of the model. The share function defined in this paper is

$$s_j = \frac{1}{ns}\sum_{i=1}^{ns} s_{ji} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\exp\left(\delta_j + \sum_{k=1}^{K} x_{jk}(\sigma_k v_{ik} + \pi_{k1}D_{i1} + \ldots + \pi_{kd}D_{id})\right)}{1+\sum_{m=1}^{J}\exp\left(\delta_m + \sum_{k=1}^{K} x_{mk}(\sigma_k v_{ik} + \pi_{k1}D_{i1} + \ldots + \pi_{kd}D_{id})\right)}.$$

Therefore, the derivatives are

$$\frac{\partial s_j}{\partial \delta_j} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\partial s_{ji}}{\partial \delta_j} = \frac{1}{ns}\sum_{i=1}^{ns} s_{ji}(1 - s_{ji})$$

$$\frac{\partial s_j}{\partial \delta_m} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\partial s_{ji}}{\partial \delta_m} = -\frac{1}{ns}\sum_{i=1}^{ns} s_{ji}s_{mi}$$

$$\frac{\partial s_j}{\partial \sigma_k} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\partial s_{ji}}{\partial \sigma_k} = \frac{1}{ns}\sum_{i=1}^{ns} s_{ji}\left(x_{jk}v_{ik} - \sum_{m=1}^{J} x_{mk}v_{ik}s_{mi}\right) = \frac{1}{ns}\sum_{i=1}^{ns} v_{ik}s_{ji}\left(x_{jk} - \sum_{m=1}^{J} x_{mk}s_{mi}\right)$$

$$\frac{\partial s_j}{\partial \pi_{kd}} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\partial s_{ji}}{\partial \pi_{kd}} = \frac{1}{ns}\sum_{i=1}^{ns} s_{ji}\left(x_{jk}D_{id} - \sum_{m=1}^{J} x_{mk}D_{id}s_{mi}\right) = \frac{1}{ns}\sum_{i=1}^{ns} D_{id}s_{ji}\left(x_{jk} - \sum_{m=1}^{J} x_{mk}s_{mi}\right)$$

Substituting this back into equation (13), we obtain the Jacobian of the function computed in Step 2.

The gradient of the objective function is

$$2 * D\delta' * Z * A^{-1} * Z' * \omega \ .$$

TABLE 1A
VOLUME MARKET SHARES

	88Q1	88Q4	89Q4	90Q4	91Q4	92Q4
Kellogg	41.39	39.91	38.49	37.86	37.48	33.70
General Mills	22.04	22.30	23.60	23.82	25.33	26.83
Post	11.80	10.30	9.45	10.96	11.37	11.31
Quaker Oats	9.93	9.00	8.29	7.66	7.00	7.40
Ralston	4.86	6.37	7.65	6.60	5.45	5.18
Nabisco	5.32	6.01	4.46	3.75	2.95	3.11
C3	75.23	72.51	71.54	72.64	74.18	71.84
C6	95.34	93.89	91.94	90.65	89.58	87.53
Private Label	3.33	3.75	4.63	6.29	7.13	7.60

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE 1B
SALES MARKET SHARES

	88Q1	88Q4	89Q4	90Q4	91Q4	92Q4
Kellogg	41.23	39.84	38.16	37.26	36.00	32.57
General Mills	24.73	25.16	26.86	27.70	29.66	31.39
Post	11.15	9.87	8.76	10.56	11.19	10.97
Quaker Oats	9.46	8.54	8.26	7.30	6.89	6.94
Ralston	5.12	6.61	7.80	7.03	5.78	5.45
Nabisco	5.00	5.59	4.13	3.58	3.04	3.33
C3	77.11	74.87	73.78	75.52	76.85	74.93
C6	96.69	95.61	93.97	93.43	92.56	90.65
Private Label	2.12	2.30	2.81	3.74	4.28	4.55

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE 1C
PRICE INDEXES (88Q1=100)

	88Q4	89Q4	90Q4	91Q4	92Q4
CPI	103.9	108.7	115.3	118.8	122.3
All RTE Cereals	107.9	118.6	124.6	128.6	133.0
Private Label	103.0	111.0	114.8	119.4	123.2
K Corn Flakes	109.6	114.7	117.6	117.4	121.2
K Raisin Bran	104.9	113.0	121.0	119.9	123.3
K Frosted Flakes	107.4	118.5	122.3	118.4	122.3
K Rice Krispies	106.0	115.8	122.0	121.1	117.8
K Frosted Mini Wheats	103.9	113.4	121.5	123.8	126.3
K Froot Loops	109.1	119.6	126.3	132.3	131.2
K Special K	107.7	115.3	123.0	126.5	133.8
K Crispix	107.0	120.3	127.1	121.2	129.4
K Corn Pops	105.7	117.1	129.0	135.0	135.5
GM Cheerios	109.7	119.8	130.6	134.1	137.9
GM Honey Nut Cheerios	107.5	119.9	128.1	133.7	135.6
GM Wheaties	110.5	120.9	127.9	137.1	144.3
GM Total	108.7	116.2	124.0	134.2	137.5
GM Lucky Charms	108.8	121.0	127.9	135.4	140.6
GM Trix	109.3	121.9	132.0	139.6	142.2
GM Raisin Nut	106.5	117.5	127.7	134.9	139.1
GM Cinn Toast Crunch	108.4	125.0	123.3	130.3	139.6
GM Kix	105.0	122.8	132.1	133.7	139.7
Post Raisin Bran	107.2	116.4	115.7	120.9	123.1
Post Grape Nuts	107.7	112.5	125.5	131.2	133.9
Q 100% Natural	104.6	116.9	121.3	124.4	130.1
Q Life	111.0	126.1	136.3	144.4	137.3
Q Cap N Crunch	104.8	115.0	120.0	131.6	129.6
N Shredded Wheat	108.0	113.3	123.5	146.4	159.3

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center, and BLS series.

TABLE 2A
AGGREGATE ESTIMATES OF PRODUCTION COSTS

	RTE Cereal (SIC 2043)		All Food Industries (SIC 20)	
Item	M$	% of value	M$	% of value
Value of Shipments	8,211	100.0	371,246	100.0
Materials	2,179	26.5	235,306	63.4
Labor	677	8.2	32,840	8.8
Energy	76	0.9	4,882	1.3
Gross Margin		64.4		26.5

Source: Annual Survey of Manufacturers 1988-1991.

TABLE 2B
DETAILED ESTIMATES OF PRODUCTION COSTS

Item	$/lb	% of mfr price	% of retail price
Manufacturer Price	2.40	100.0	80.0
Manufacturing Cost:	1.02	42.5	34.0
Grain	0.16	6.7	5.3
Other Ingredients	0.20	8.3	6.7
Packaging	0.28	11.7	9.3
Labor	0.15	6.3	5.0
Manufacturing Costs (net of capital costs)*	0.23	9.6	7.6
Gross Margin		57.5	46.0
Marketing Expenses:	0.90	37.5	30.0
Advertising	0.31	13.0	10.3
Consumer Promo (mfr coupons)	0.35	14.5	11.7
Trade Promo (retail in-store)	0.24	10.0	8.0
Operating Profits	0.48	20.0	16.0

*Capital costs were computed from ASM data.

Source: Cotterill (1996) reporting from estimates in CS First Boston Reports "Kellogg Company," New York, October 25,1994. Similar estimates include Corts (1996a).

TABLE 3
DESCRIPTION OF IRI INFOSCAN DATA BASE

Name	Description
Sales data:	
sales	dollar sales
volume	volume (pounds) sales
units	number of units (of any size) sold
anymerch	percent volume sold with either display, any feature, or more than 5% price reduction
feature	percent volume sold any feature
display	percent volume sold any display
preduc	weighted average price reduction
Demographics:	
pop	total city population
numHH	number of households in city
medinc	median income
lowinc	percent of households under $10K
highinc	percent of households over $50K
hisp	percent Hispanic
medage	median family age
medsize	median household size
Market data:	
supgroc	supermarket/grocery store sales ratio
grocCR4	city grocery CR4
ACV	the sum of all commodity volume (ACV) sold by stores carrying the product over the ACV of all stores in the city
avg ACV	weekly dollar sales of the product over the ACV of stores selling at least one unit

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE 4A
PRICES AND MARKET SHARES OF BRANDS IN SAMPLE

Description	Mean	Median	Std	Min	Max	Brand Variation	City Variation	Quarter Variation
Prices (¢ per serving)	19.4	18.9	4.8	7.6	40.9	88.4%	5.3%	1.6%
Advertising (M$ per quarter)	3.56	3.04	2.03	0	9.95	66.2%	--	1.8%
Share within Cereal Market (%)	2.2	1.6	1.6	0.1	11.6	82.3%	0.5%	0%

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE 4B
CHARACTERISTICS OF BRANDS IN SAMPLE

Description	Mean	Median	Std	Min	Max
Calories	137.6	120	36.32	110	220
Fat Calories	12.4	10	13.93	0	60
Sodium (% RDA)	8.7	9	4.15	0	15
Total Carbohydrate (% RDA)	10	9	2.85	7	16
Fiber (% RDA)	9.5	5	9.38	0	31
Sugar (g)	8.4	7	5.98	0	20
# of vita & min with at least 10% RDA	9.4	10	2.52	3	14
Max % RDA	35.8	25	20.45	15	100
Average % RDA (for vit & min with >10%RDA)	24.6	23.9	11.38	11.7	76.1
Mushy (=1 if cereal gets soggy in milk)	0.35	--	--	0	1
Serving weight (g)	35.1	30	9.81	25	58

Source: Cereal boxes.

TABLE 4C
DEMOGRAPHIC VARIABLES

Description	Mean	Median	Std	Min	Max
Income	13,083	10,475	11,182	14	275,372
Age	29.99	28	23.14	1	90

Source: samples from the CPS.

TABLE 5
BRANDS IN SAMPLE

	Manufacturer	Brand name	Segment
1	General Mills	Cheerios	All Family/ Basic
2	Kellogg	Frosted Flakes	Kids Family Acceptable
3	Kellogg	Corn Flakes	All Family/ Basic
4	Kellogg	Rice Krispies	All Family/ Basic
5	Kellogg	Raisin Bran	Taste Enhanced Wholesome
6	General Mills	Honey Nut Cheerios	Kids Family Acceptable
7	Kellogg	Frosted Mini Wheats	Simple Health/ Nutrition
8	Quaker	Cap N Crunch	Kids
9	Post	Grape Nuts	Simple Health/ Nutrition
10	General Mills	Total	Simple Health/ Nutrition
11	Post	Raisin Bran	Taste Enhanced Wholesome
12	General Mills	Kix	Kids
13	General Mills	Lucky Charms	Kids
14	Kellogg	Fruit Loops	Kids
15	General Mills	Wheaties	All Family/ Basic
16	Kellogg	Special K	Simple Health/ Nutrition
17	Quaker	Life	Kids
18	General Mills	Cinnamon Toast Crunch	Kids
19	Nabisco	Spoon Size Shredded Wheat	Simple Health/ Nutrition
20	Kellogg	Corn Pops	Kids
21	General Mills	Trix	Kids
22	Post	Honey Bunches of Oats	Taste Enhanced Wholesome
23	General Mills	Raisin Nut	Taste Enhanced Wholesome
24	Kellogg	Crispix	All Family/ Basic
25	Quaker	100% Natural	Taste Enhanced Wholesome

TABLE 6
RESULTS FROM LOGIT DEMAND
(27,862 OBSERVATIONS)

	OLS	IV				
Variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Price	-7.26 (0.16)	-17.57 (0.50)	-18.46 (0.51)	-22.56 (0.52)	-22.08 (0.54)	-23.37 (0.48)
Advertising	0.026 (0.002)	0.020 (0.002)	0.019 (0.002)	0.018 (0.002)	0.020 (0.002)	0.018 (0.002)
log median income	--	--	--	1.06 (0.02)	1.11 (0.02)	1.12 (0.02)
log of median age	--	--	--	-0.063 (0.061)	-0.04 (0.061)	-0.007 (0.062)
median HH size	--	--	--	-0.053 (0.031)	-0.044 (0.032)	-0.038 (0.032)
Measures of fit:[a]	0.72	168.49 (30.14)	160.39 (16.92)	81.15 (30.14)	67.10 (16.92)	82.95 (42.56)
First Stage:						
R^2	--	0.91	0.91	0.91	0.91	0.91
F-statistic	--	3796	4902	3736	4720	3646
Instruments:						
Avg regional prices	--	✓	--	✓	--	✓
Cost proxies	--	--	✓	--	✓	✓
Own Price Elasticity:						
Mean	-1.39	-3.38	-3.56	-4.35	-4.25	-4.50
Std.	0.35	0.85	0.89	1.09	1.07	1.13
Median	-1.36	-3.30	-3.47	-4.24	-4.15	-4.39
% of Inelastic Demands (+/- 2 s.e.'s)	10.0% (7.4-13.6%)	0	0	0	0	0

Dependant variable is $\ln(s_{jt})$-$\ln(s_{0t})$. All regressions include time and brand dummy variables.
[a]Adjusted R^2 for the OLS regression, and a test of over identification for the IV regressions (Hausman, 1983) with the 0.95 critical values in parenthesis.

TABLE 7A
RESULTS FROM THE FULL MODEL
(27,862 OBSERVATIONS)

	Variable	(i) Estimate	(i) Std Error	(ii) Estimate	(ii) Std Error	(iii) Estimate	(iii) Std Error
Means (β's)	Price	-16.712	1.145	-11.671	1.927	-15.753	1.123
	Advertising	0.031	0.003	0.044	0.007	0.033	0.006
	Constant[a]	-2.718	0.068	-1.917	0.114	-1.853	0.097
	Cal from Fat[a]	0.011	0.007	-0.010	0.009	0.072	0.003
	Sugar[a]	0.011	0.001	0.021	0.002	0.012	0.009
	Mushy[a]	-0.687	0.047	-2.207	0.523	-3.193	0.764
	Fiber[a]	-0.014	0.008	0.014	0.001	0.006	0.005
Standard Deviations (σ's)	Price	1.570	1.043	0.805	2.278	0.257	4.030
	Constant	0.209	0.169	0.150	0.450	0.035	0.450
	Cal from Fat	0.011	0.013	0.013	0.028	0.004	0.033
	Sugar	0.003	0.020	0.008	0.040	0.010	0.060
	Mushy	0.261	0.445	0.926	0.498	1.221	0.964
	Fiber	0.005	0.013	0.004	0.027	0.010	0.033
Interaction with Income	Price	10.701	1.602	8.405	3.641	--	--
	Constant	-1.320	0.295	-1.450	0.766	-0.880	0.495
	Cal from Fat	--		--		0.021	0.029
	Sugar	--		--		-0.107	0.060
	Mushy	--		2.223	0.533	1.930	0.604
	Fiber	--		--		0.009	0.046
Interaction with Age	Price	8.749	1.284	6.026	2.846	--	--
	Constant	-1.744	0.256	-1.630	0.567	-0.156	0.429
	Cal from Fat	--		--		0.028	0.027
	Sugar	--		-0.005	0.036	-0.106	0.034
	Mushy	--		3.717	0.708	4.663	1.156
	Fiber	--		--		0.004	0.030
GMM Objective		9.82		3.36		4.53	
(degrees of freedom)		(17)		(14)		(11)	
MD χ^2		11015.7		3083.5		4086.2	
MD un-weighted R^2		-0.78		0.79		-2.01	
MD weighted R^2		0.11		0.16		0.28	
% of Price Coefficients >0		12.8		11.7		0	

Except where noted, parameters are GMM estimates.
All regressions include brand and time dummy variables.
[a]Estimates from a minimum distance procedure.

TABLE 7B
RESULTS FROM THE FULL MODEL
(27,862 OBSERVATIONS)

(iv)						
Variable	Means (β's)	Standard Deviations (σ's)	Interactions with Demographic Variables: Income	Income Sq	Age	Child
Price	-27.198	2.453	315.894	-18.200	-	7.634
	(5.248)	(2.978)	(110.385)	(5.914)		(2.238)
Advertising	0.020	-	-	-	-	-
	(0.005)					
Constant	-3.592[a]	0.330	5.482	-	0.204	-
	(0.138)	(0.609)	(1.504)		(0.341)	
Cal from Fat	0.011[a]	0.016	-	-	-	-
	(0.001)	(0.028)				
Sugar	0.059[a]	0.016	-0.251	-	0.051	-
	(0.006)	(0.059)	(0.091)		(0.034)	
Mushy	0.565[a]	0.244	1.265	-	0.809	-
	(0.052)	(0.623)	(0.737)		(0.385)	
Fiber	0.015[a]	0.002	-	-	-	-0.110
	(0.003)	(0.036)				(0.0513)
All-family	0.781[a]	0.1330	-	-	-	
	(0.075)	(1.365)				
Kids	1.021[a]	2.031	-	-	-	
	(0.168)	(0.448)				
Adults	1.972[a]	0.247	--	--	-	
	(0.186)	(1.636)				
GMM Objective			5.05			
(degrees of freedom)			(8)			
MD χ^2			3472.3			
MD un-weighted R^2			-0.27			
MD weighted R^2			0.19			
% of Price Coefficients >0			0.7			

Except where noted, parameters are GMM estimates. All regressions include brand and time dummy variables. Robust standard errors are given in parenthesis.
[a]Estimates from a minimum distance procedure.

TABLE 8
ESTIMATED DEMAND ELASTICITIES
(BASED ON TABLE 7 COLUMN III)

	Price		Cal from Fat		Sugar		Mushy		Fiber		Advertising	
Descriptive Statistics:[a]												
median		-2.88		0.69		0.09		--		0.03		0.1
mean		-2.95		0.91		0.13		--		0.06		0.1
std		0.79		1.08		0.14		--		0.07		0.1
Boston 1992Q4:[b]												
N Shredded Wheat	27	-4.230	5	0.36	0	0	0	--	21	0.2	0.1	0.00
GM Cheerios	19	-2.769	15	1.05	1	0	0	--	11	0	6.2	0.19
K Rice Krispies	13	-1.945	0	0.00	3	0	0	--	4	0	7.3	0.23
GM Kix	21	-3.284	5	0.36	3	0	0	--	3	0	3.9	0.13
K Crispix	19	-2.969	0	0.00	4	0	0	--	4	0	2.9	0.09
Q Life	16	-2.421	15	1.10	6	0	0	--	8	0	2.0	0.07
P Grape Nuts	24	-3.681	10	0.73	7	0	0	--	21	0.2	5.2	0.17
GM Cinnamon Toast Crunch	21	-3.276	30	2.28	10	0.1	0	--	4	0	2.9	0.09
GM Honey Nut Cheerios	18	-2.723	10	0.70	11	0.1	0	--	5	0	3.8	0.12
K Frosted Mini Wheats	28	-4.301	10	0.73	12	0.2	0	--	24	0.2	4.8	0.15
Q CapNCrunch	14	-2.081	15	1.08	12	0.2	0	--	3	0	1.5	0.05
GM Lucky Charms	20	-3.125	10	0.72	13	0.2	0	--	5	0	2.5	0.08
GM Trix	24	-3.752	15	1.10	13	0.2	0	--	0	0	3.1	0.10
K Corn Pops	21	-3.348	0	0.00	14	0.2	0	--	0	0	2.9	0.09
K Froot Loops	19	-2.933	10	0.72	14	0.2	0	--	4	0	2.1	0.07
Q 100% Natural	26	-4.036	60	4.85	14	0.2	0	--	15	0.1	1.4	0.04
K Frosted Flakes	15	-2.039	0	0.00	13	0.3	1	--	0	0	7.8	0.23
K Corn Flakes	10	-1.394	0	0.00	2	0	1	--	4	0	6.7	0.19
K Special K	21	-3.087	0	0.00	3	0	1	--	4	0	2.1	0.07
GM Total	23	-3.310	5	0.31	3	0	1	--	0	0	3.2	0.10
GM Wheaties	15	-2.279	10	0.66	4	0	1	--	12	0.2	0.4	0.01
P Honey Bunches of Oats	17	-2.590	25	1.75	7	0.1	1	--	6	0	3.5	0.11
GM Raisin Nut	33	-5.050	40	2.99	16	0.4	1	--	20	0.3	0.6	0.02
K Raisin Bran	17	-2.328	10	0.64	18	0.4	1	--	28	0.3	5.2	0.15
P Raisin Bran	16	-2.417	10	0.68	20	0.5	1	--	31	0.4	4.8	0.15

For each variable the left column displays the value of the attribute/price (in cents, calories, grams, ==1 if mushy, grams, and million of dollars, respectively) and the right column displays the elasticity w.r.t. to that attribute.

[a]Descriptive statistics of elasticities in all quarters and all cities.

[b]A sample of elasticities computed for the last quarter of 1992 in Boston.

TABLE 9

MEDIAN OWN AND CROSS-PRICE ELASTICITIES

(BASED ON TABLE 7 COLUMN IV)

#	Brand	Corn Flakes	Frosted Flakes	Rice Krispies	Froot Loops	Cheerios	Total	Lucky Charms	P Raisin Bran	CapN Crunch	Shredded Wheat
1	K Corn Flakes	-3.387	0.213	0.198	0.014	0.203	0.097	0.012	0.013	0.038	0.028
2	K Raisin Bran	0.036	0.046	0.079	0.043	0.146	0.041	0.037	0.057	0.051	0.040
3	K Frosted Flakes	0.151	-3.139	0.106	0.069	0.130	0.078	0.061	0.013	0.138	0.023
4	K Rice Krispies	0.196	0.144	-3.246	0.031	0.242	0.087	0.026	0.031	0.055	0.047
5	K Frosted Mini Wheats	0.014	0.024	0.053	0.044	0.105	0.028	0.038	0.055	0.045	0.033
6	K Froot Loops	0.019	0.131	0.042	-2.338	0.073	0.025	0.107	0.028	0.149	0.020
7	K Special K	0.114	0.124	0.105	0.021	0.152	0.151	0.019	0.021	0.035	0.034
8	K Crispix	0.077	0.086	0.114	0.034	0.181	0.085	0.030	0.037	0.049	0.043
9	K Corn Pops	0.013	0.108	0.034	0.113	0.058	0.025	0.097	0.024	0.127	0.016
10	GM Cheerios	0.128	0.112	0.153	0.034	-3.682	0.084	0.030	0.037	0.056	0.051
11	GM Honey Nut Cheerios	0.033	0.192	0.058	0.123	0.095	0.033	0.107	0.027	0.162	0.024
12	GM Wheaties	0.244	0.169	0.176	0.025	0.242	0.112	0.021	0.026	0.050	0.044
13	GM Total	0.096	0.108	0.086	0.018	0.130	-2.865	0.016	0.016	0.029	0.029
14	GM Lucky Charms	0.019	0.130	0.041	0.124	0.073	0.026	-2.541	0.028	0.147	0.020
15	GM Trix	0.012	0.103	0.031	0.109	0.056	0.026	0.095	0.023	0.122	0.016
16	GM Raisin Nut	0.013	0.025	0.041	0.035	0.090	0.039	0.031	0.046	0.037	0.027
17	GM Cinnamon Toast Crunch	0.026	0.163	0.049	0.119	0.090	0.034	0.102	0.026	0.151	0.022
18	GM Kix	0.050	0.279	0.070	0.101	0.106	0.056	0.088	0.020	0.149	0.025
19	P Raisin Bran	0.027	0.037	0.068	0.044	0.127	0.034	0.038	-2.502	0.049	0.036
20	P Grape Nuts	0.037	0.049	0.089	0.042	0.166	0.049	0.037	0.052	0.052	0.047
21	P Honey Bunches of Oats	0.100	0.098	0.104	0.022	0.172	0.108	0.020	0.024	0.038	0.033
22	Q 100% Natural	0.013	0.021	0.047	0.042	0.104	0.028	0.036	0.052	0.046	0.029
23	Q Life	0.077	0.328	0.091	0.115	0.139	0.045	0.096	0.023	0.182	0.029
24	Q CapNCrunch	0.043	0.217	0.065	0.124	0.102	0.033	0.106	0.026	-2.277	0.024
25	N Shredded Wheat	0.076	0.082	0.126	0.038	0.212	0.074	0.034	0.044	0.054	-4.281
26	Outside good	0.141	0.078	0.084	0.022	0.105	0.041	0.018	0.021	0.033	0.021

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.

TABLE 10
MEDIAN MARGINS

	Logit (Table 6 column ii)	Full Model (Table 7 columns ii-iv)		
Single Product Firms	35.8%	35.2%	36.8%	35.8%
Current Ownership of 25 Brands	38.2%	40.1%	43.7%	42.2%
Joint Ownership of 25 Brands	44.6%	64.7%	74.6%	72.6%
Current Ownership of All Brands	39.8%	--	--	--
Monopoly/ Perfect Price Collusion	57.8%	--	--	--

The margins are defined as (p-mc)/p.

FIGURE 1
FREQUENCY DISTRIBUTION OF TASTE FOR SOGGINESS
(BASED ON TABLE 7 COLUMN II)

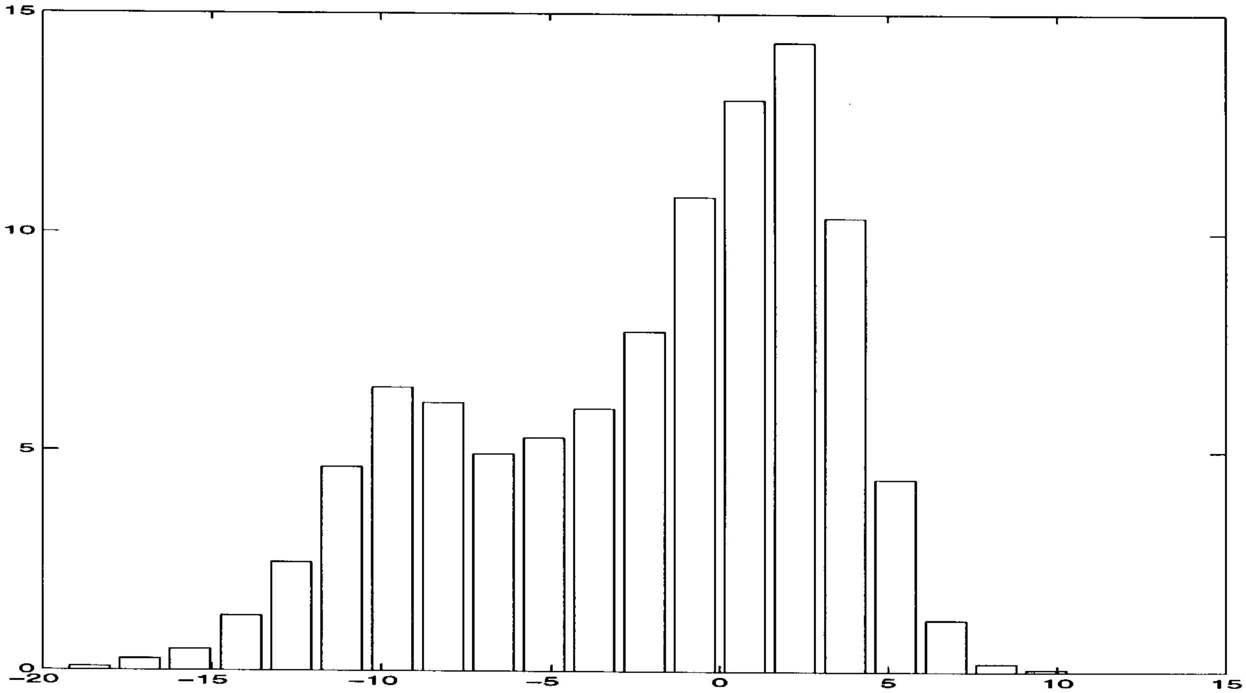

FIGURE 2
FREQUENCY DISTRIBUTION OF PRICE COEFFICIENT
(BASED ON TABLE 7 COLUMN II)

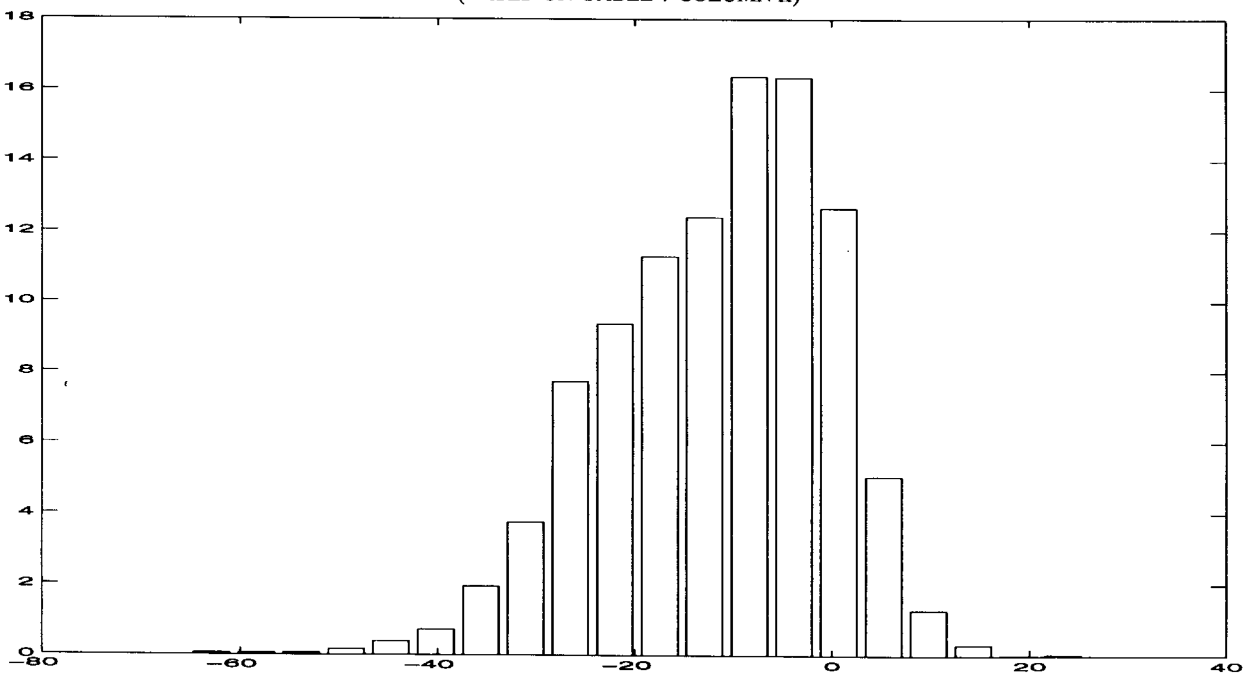

FIGURE 3
FREQUENCY DISTRIBUTION OF CONSTANT TERM
(BASED ON TABLE 7 COLUMN II)

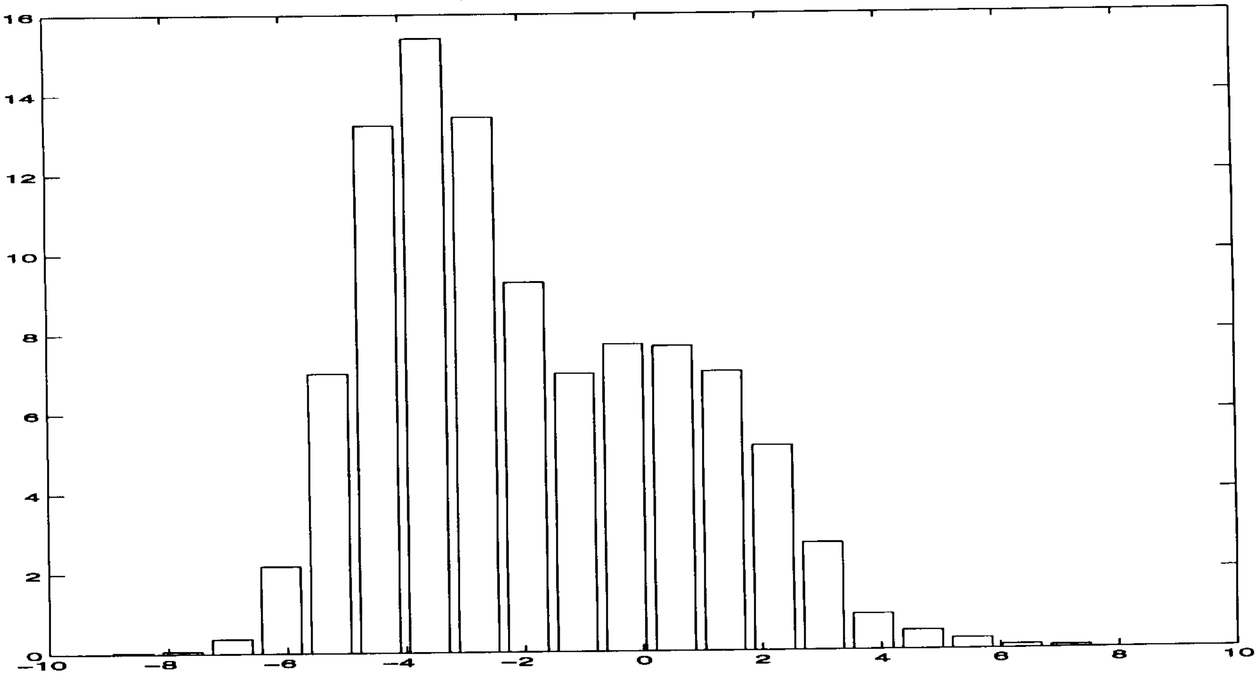

FIGURE 4
FREQUENCY DISTRIBUTION OF PRICE COEFFICIENT
(BASED OF TABLE 7 COLUMN IV)