Vertical Contracts between Manufacturers and Retailers: An Empirical Analysis*

Sofia Berto Villas-Boas
(University of California, Berkeley)

July, 2002

ABSTRACT

This paper tests different models of vertical contracting between manufacturers and retailers in the supermarket industry. I estimate demand and use the estimates to compute price-cost margins for retailers and manufacturers under different supply models without observing wholesale prices. I then test which set of margins seems to be compatible with the margins obtained from direct estimates of cost and select the best among the non-nested competing models. The models considered are: (1) a double marginalization pricing model; (2) a vertically integrated model; and (3) a variety of alternative (strategic) supply scenarios, allowing for collusion, non-linear pricing and strategic behavior with respect to private label products. Using data on yogurt sold at several stores in a large urban area of the United States, I find that wholesale prices are close to marginal cost and that retailers have pricing power in the vertical chain. This is consistent with non-linear pricing by the manufacturers or with high bargaining power of the retailers.

Keywords: Market power, vertical contracts, multiple manufacturers and retailers, non-nested tests, yogurt local market, non-linear pricing, bargaining power.

* I am grateful to my advisors Richard Gilbert and Aviv Nevo for constant encouragement and invaluable guidance. I wish also to specially thank Severin Borenstein, Ken Chay, Bronwyn Hall, Rene Kamita, Michael Katz, Dan McFadden, Thomas Rothenberg, Paul Ruud, Miguel Villas-Boas, Catherine Wolfram and participants in the Industrial Organization and in the Econometrics seminars at UC Berkeley for their comments. Fellowship *PRAXIS BD/9128/96* from the Fundação para a Ciência e Tecnologia is gratefully acknowledged. Address: Department of Economics, University of California at Berkeley, 549 Evans Hall, Berkeley, CA 94720-3880; e-mail:sberto@econ.berkeley.edu.

1. Introduction

Vertical contracts between retailers and manufacturers are multidimensional agreements with terms that reflect the relative bargaining power of the parties involved and that are responses to moral hazard and adverse selection problems and to the need of risk sharing between the parties. There are several wide-ranging reasons why we should care about analyzing vertical contracts. First, vertical contracts may promote efficiency in the vertical channel. This efficiency is a result from the departure from the simple uniform pricing scheme that results in double marginalization. The problem of double marginalization arises when the only contractual instrument used is the wholesale price. As a consequence, the sum of profits for the manufacturer and retailer may be less than if they could have coordinated their decisions. Second, vertical contracts may impair competition through their horizontal effects on the upstream (manufacturer) and downstream (retail) markets by increasing the possibility for oligopolistic coordination (increasing market power) or by excluding rivals (and hence diminishing product variety and choices).[1] Finally, it has been recognized that retailers are not neutral pass-through intermediaries. When analyzing price dynamics in the economy as a whole it is therefore important to consider not only a model for manufacturer behavior but also for retail behavior (see, for example, Chevalier, Kashyap and Rossi, 2000).

While there is extended theoretical work on vertical contracts (for a survey, see Katz (1989)), there is a smaller set of empirical studies on these contracts. Vertical contracts are especially difficult to examine due to their infra-marginal components and due to limited data availability. In particular, wholesale price data are typically unavailable and retailers' and manufacturers' marginal costs are difficult to measure separately. This paper presents a method to analyze vertical contracting that does not require data on wholesale prices or marginal cost (of either the retailers or the manufacturers). This methodology is applied to a particular market but can easily be adapted to analyzing similar questions when limited data observability might be the case.

In this paper I focus on the efficiency implications of vertical contracting. In particular, I examine whether the contracting between manufacturers and retailers in the supermarket industry follows the double marginalization model or something more efficient. In the last few decades, arguments have been made that retailers have acquired greater bargaining power relative to manufacturers (Progressive Grocer, April 1992) suggesting a possible departure from the double marginalization model in this industry. Several reasons for this have been pointed out by industry participants and

[1] In general, U.S. antitrust authorities haven't paid attention to vertical contracts unless foreclosure became an issue, and most vertical arrangements are treated as per se legal. Foreclosure occurs when a vertical contract closes off some or all of a market to competitors thereby permitting the exercise of market power.

by researchers. The private labels that compete directly with the national brands (e.g., Narasimhan and Wilcox, 1998) provide a new bargaining tool for retailers when negotiating with manufacturers.[2] Another reason is the increase of concentration at the retail level. As a result retailers have market power which they can use to bargain more aggressively with the manufacturers.[3] An indication of retailer market power is the increased competition for shelf space implying that manufacturers have to pay retailers slotting allowances (e.g., Chu, 1992 and Shaffer, 1991) to get their products displayed.

The research plan of this paper is as follows: First, I estimate demand and use the estimates to compute price-cost margins for retailers and manufacturers under different supply models, without observing wholesale prices. I then test which set of computed price-cost margins is most compatible with the price-cost margins obtained by the difference of observed retail prices and estimated marginal cost. Testing between the supply models, however, cannot be done using the traditional procedures because the null hypotheses are not particular cases of some other null hypothesis, i.e. they are non-nested (Gourieroux and Monfort, 1994 and Vuong, 1989). Therefore, I apply the methodology proposed by Smith (1992) to select among the non-nested competing models estimated by Generalized Method of Moments.

The first supply model I consider is the double marginalization model. The implied price-cost margins are inefficient from the perspective of the joint profit of retailers and manufacturers. The second model I consider is vertically integrated pricing, which will maximize joint profits and therefore is the efficient outcome from the retailers and manufacturers point of view. The implied price-cost margins correspond to those of a vertically integrated monopolist who sells all the products in the market. I also examine intermediate scenarios incorporating the role of private labels, collusion and fixed fees in their design. In two of the models, either the retailers or the manufacturers are allowed to use non-linear pricing contracts (involving fixed fees). In another model, the retailers are assumed to behave as if they were vertically integrated with respect to the private labels. Finally, collusion at the manufacturer level or at the retailer level is examined.

In terms of the techniques used in this paper, the estimation of firm's (implied) price-cost margins without observing actual costs follows Bresnahan (1981, 1987) (see Bresnahan (1989) for a survey). The starting point is the estimation of a demand system and the elasticities of substitution between the different products. In the context of oligopoly markets with differentiated products,

[2]Retailers are able to sell products that can be purchased at a potentially lower wholesale price, that carry their store brand and are displayed next to the national brands. At a 1995 convention, Douglas Ivester, then-president and CEO of Coca Cola, called private labels *"parasites"* and said they were responsible for *"eroding category profits."*

[3]For example, see New York Times, November 13, 1998, page C1.

two problems may arise: the high dimensionality of elasticities to be estimated and the endogeneity of prices. Discrete choice modeling of demand (see, e.g., McFadden (1973,1984), Cardell (1989), Berry (1994), Berry, Levinsohn and Pakes (1995), for future reference BLP (1995), and Nevo (2001)) is a way to solve the dimensionality problem and here I follow this strategy.[4] I project consumer choices on a set of product characteristics, with smaller dimension than the number of products. The second problem I deal with is the fact that prices set by retailers and manufacturers can be correlated with unobserved product characteristics. As instruments, I use direct components of marginal cost, namely input prices, interacted with product-specific fixed effects. The intuition for interacting input prices with product dummies is to allow for each input to enter the production function of each product differently. This is a new approach to instrument for prices and, given the good first-stage fit, appears to generate robust results.

This paper empirically focuses on the yogurt market in a large Midwestern city. I choose this product category because it has substantial retail price variability that is not solely due to promotional retail activity, which in turn is important for input price changes to be reflected in changes of retail prices. Otherwise, input prices would be poor instruments for the retail prices. In fact, yogurt is one of the least promoted product categories sold by the retailers in the market being considered. Another reason for choosing this product has to do with the potential wholesale price variability. Yogurt has a short shelf life and has to be produced and delivered to retailers more often than, for example, detergents or cereal. Manufacturers can adjust wholesale prices to respond to significant marginal cost changes in the next yogurt shipments. Two additional advantages of looking at the yogurt market are the small number of manufacturers and the well-established private label brands (Frozen Food Digest, October 1995). At the retail level there are three large retailers competing directly with each other in the market defined by two zip code areas. This paper uses a panel data set of quantities sold and retail prices for 43 products over a period of 104 weeks collected by scanning technology at the three retailers in the market.

The results do not indicate double marginalization pricing in the vertical structure. The double marginalization model provides a statistically worse fit than the other models considered and is rejected against them. The supply model that fits the data best assumes that wholesale prices are close to marginal cost and that the retailers have pricing power in the vertical chain. This is consistent with high bargaining power of the retailers or with non-linear pricing by the manufacturers. In the optimal non-linear pricing contract, the manufacturer sets the marginal wholesale price close to the manufacturer's marginal cost in order for the retailer to have the right incentives when setting

[4] Two other possible solutions for the dimensionality problem are to assume the existence of a representative consumer (e.g., Dixit-Stiglitz, 1977) or to use a multistage demand model (as in Hausman, 1996).

the retail prices. Then the manufacturer transfers revenue from the retailers via a fixed fee or by selling the non-marginal units at higher wholesale prices.

The two main contributions of this paper are that, given demand assumptions, I am able to (1) estimate, without observing wholesale prices, the price-cost margins for all manufacturers and all retailers in a certain local market given different supply models; and (2) test the validity of each one of the models by comparing the computed margins with the price-cost margins estimated using components of marginal cost. Previous work, typically, does not model the retailers' decisions (for example, BLP (1995) and Nevo (2001)). In these papers the implied price cost margins are determined by the manufacturers and by maximizing the profits from the set of products that each of them sells. My results suggest that, at least for the market I study, the model that is more consistent with the data has retailers making the pricing decisions not the manufacturers. This model implies different price-cost margins, since the retailers and manufacturers will be maximizing their profits over a different set of products.

Several recent papers examine retailer and manufacturer vertical relationships in different industries. Bresnahan and Reiss (1985) analyze the relative margins of automobile manufacturers and their dealers given some stylized facts that dealer discounts are larger on more expensive models as are manufacturer margins. Corts (2001) studies the effect of vertical market structure on competition in the downstream market in the U.S. motion picture industry. Messinger and Narasimhan (1995) correlate the supermarket industry structure with accounting-based measures of performance from the early 1960s to the early 1990s. They claim that there is no evidence of a systematic shift of power from the manufacturer to the retailer during that period. Four other recent studies, closer to this paper, address vertical market power in the grocery channel. Chintagunta, Bonfrer and Song (2000) estimate the impact of the introduction of a private label by one retailer on the relative market power of the retailer and the manufacturers. They use a conduct parameter approach, where the conduct parameter is interpreted as a measure of deviations from the Bertrand pricing behavior. Kadiyali, Chintagunta and Vilcassim (2000) also use a conduct parameter approach to measure the share of profits to retailers and manufacturers. One key distinguishing feature of this paper relative to the two previous ones is that they use data on wholesale prices reported by the retailer. Sudhir (2001) studies competition among manufacturers under alternative assumptions of vertical interactions with one retailer. Finally, Villas-Boas and Zhao (2001) evaluate the degree of manufacturer competition and the retailer and manufacturers interactions in the ketchup market in a certain city, finding results consistent with the ones in this paper. One major modeling difference in this paper from all of the previous ones is that this is the first paper to consider multiple retailers when analyzing the vertical interactions between manufacturers and retailers.

The rest of this paper is organized as follows. The next section presents the model. Section 3 describes with more detail the method of estimation and the instruments. In section 4, I describe the yogurt market and the data being used. Finally, section 5 looks at the results, and section 6 presents conclusions and extensions indicating, in particular, how the methodology proposed here can be used in different settings.

2. The Model

The model consists of a demand side and of different alternative scenarios of vertical relationships between manufacturers and retailers. For each supply scenario the price-cost margins for the retailers and for the manufacturers are expressed solely as functions of the demand parameters.

2.1. Demand Side

Each period t the consumer chooses among N_t different products[5] sold by three different retailers. The indirect latent utility of consumer i from buying product j during week t is given by

$$U_{ijt} = d_j + x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{1}$$

where d_j are product (brand-store) fixed effects capturing time invariant product characteristics, x_{jt} are the observed product characteristics, p_{jt} is the price of product j, ξ_{jt} are the mean across consumers of unobserved (by the econometrician) product characteristics (or better, changes in unobserved product characteristics since a product dummy is included in equation (1)) and ϵ_{ijt} represents the distribution of consumer preferences about this mean with density $f(\epsilon)$. The random coefficients β_i are consumer taste parameters for the different product characteristics and the term α_i represents the marginal utility of price. They are allowed to vary across consumers in the following way

$$[\alpha_i, \beta_i]' = [\alpha, \beta]' + \Gamma\ D_i + \Upsilon\ v_i \tag{2}$$

where the variable D_i has observed consumer characteristics such as demographics, while unobserved consumer characteristics are picked up by v_i. The matrices of non-linear demand parameters to be estimated are Γ and Υ. Unobserved consumer characteristics v_i are assumed to be normally distributed $N(0, I)$, and the observed consumer characteristics D_i have an empirical distribution

[5] The same physical product sold at two different retailers is defined as two different products.

$\hat{\mathrm{F}}(\mathrm{D})$ from the demographic data. It may be useful to rewrite equation (1) as

$$U_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt} \tag{3}$$

where

$$\mu_{ijt} = [x_{jt}, p_{jt}](\Gamma\ D_i + \Upsilon\ v_i). \tag{4}$$

and the mean utility level δ_{jt} is defined as

$$\delta_{jt} = d_j + x_{jt}\beta - \alpha p_{jt} + \xi_{jt}. \tag{5}$$

Then μ_{ijt} in (3) is interpreted as the deviation from the mean Utility level, allowing for consumer heterogeneity.

Additionally an outside good is included in the model, called good zero, allowing for the possibility of consumer i not buying one of the N_t marketed goods. Its price is not set in response to the prices of the other N_t products. In the outside good I include yogurts sold by smaller retail stores or grocery stores not considered in the analysis and also yogurts of small manufacturers sold in the three retail stores studied. As usual, the mean utility of the outside good, δ_{0t}, is normalized to be constant over time and equal to zero.[6]

Given a measure M of the market size, assumed proportional to the population in the contiguous zip code areas where the stores are located, then the observed share of product j is given by $s_j = q_j/M$, where q_j are the units sold. [7]

In discrete choice models for demand, consumers purchase one unit of that product among all the possible products available at a certain time t that maximizes their indirect utility. Consumers in reality may purchase multiple products at the same time. By assuming a single unit purchase, this paper is implicitly saying that a consumer who purchases three units of yogurts equals three consumers (or the same consumer) buying one unit each time.[8] This assumption implies that the sequential multiple unit purchases are independent.[9] Consumer i would purchase one unit of good

[6]Without making any additional assumptions it would not be identified. The alternative would be to normalize any one of the N_t goods.

[7]In this case, q_j are the servings sold of yogurt. One serving corresponds to a cup of 6 ounces. Accordingly, p_j is the price per serving of product j.

[8]The studies that explicitly model multiple-discrete choices (e.g., Dubin and McFadden, 1984; Hanemann, 1984; Hausman, Leonard and McFadden, 1995; Hendel, 1999 and Dubé, 2001) need individual level data for estimation. Since this paper uses only market-level data the techniques proposed could not be directly applied here.

[9]If there exist correlation between the purchases of the same consumer, when the number of consumers is large, one may be underestimating the standard errors, but the estimates of the probabilities of purchases are still consistent.

j if and only if

$$U_{ijt} \geq U_{iht} \tag{6}$$

for all $h = 0, ..., N_t$. Let $Q_{jt} = \{(D_i, v_i, \epsilon_{it}) \mid U_{ijt} \geq U_{iht} \ \forall \ h = 0, ...N_t\}$ and $1(Q_{jt})$ be an indicator function that takes the value of one when Q_{jt} is true and zero otherwise. Then, the market share of product j during week t is given by the probability that good j is chosen, that is,

$$s_{jt} = \int 1(Q_{jt}) \ dF(\epsilon) \ dF(v) \ dF(D). \tag{7}$$

Depending on the distribution assumptions, equation (7) can be analytically integrated or not.

Beginning with the easiest case, if consumer heterogeneity enters only through the random shock ϵ_{ijt} (that is, not through the random coefficients in equation (4) assuming that both D and v are fixed) and ϵ_{ijt} is distributed i.i.d. with an extreme value type I density, $f(\epsilon) = e^{-e^{-\epsilon}}$, then the market share of good j has a closed expression given by

$$s_{jt} = \frac{e^{\delta_{jt}}}{e^{\delta_{0t}} + \sum_{k=1}^{N_t} e^{\delta_{kt}}} = \frac{e^{\delta_{jt}}}{1 + \sum_{k=1}^{N_t} e^{\delta_{kt}}}. \tag{8}$$

Equation (8) is the Multinomial Logit model.

Assuming still that ϵ_{ijt} is distributed i.i.d. extreme value, but now allowing for consumer heterogeneity to affect the taste parameters for the different product characteristics corresponds to the full random coefficients model or mixed Logit model described by equations (3), (4) and (5).[10] The market share of product j will no longer have a closed form expression. The computation of the integral in equation (7) is done by simulation noting that, conditional on the unobserved v_{it} and observed D_{it} consumer characteristics, or in other words, conditional on μ_{it}, the probability of purchase of a certain product has the Multinomial Logit expression

$$s_{jt}(\mu_{it}) = \frac{e^{\delta_{jt} + \mu_{ijt}}}{1 + \sum_{k=1}^{N_t} e^{\delta_{kt} + \mu_{ikt}}} = \frac{e^{\delta_{jt} + [x_{jt}, p_{jt}](\Gamma \ D_i + \Upsilon \ v_i)}}{1 + \sum_{k=1}^{N_t} e^{\delta_{kt} + [x_{kt}, p_{kt}](\Gamma \ D_i + \Upsilon \ v_i)}}. \tag{9}$$

The unconditional true probabilities are given by the integral over all possible values of the random taste terms μ_{it} of the conditional probabilities, namely,

$$s_{jt} = \int s_{jt}(\mu_{it}) f(\mu_{it}) d\mu_{it}. \tag{10}$$

[10]This is a very general model. As shown in McFadden and Train (2000), any discrete choice model derived from random utility maximization can be approximated, to any degree of accuracy, to a Mixed Logit.

2.2. Supply Side

In what follows, each supply model is described in detail and then solved to obtain an expression for both the retailer's and the manufacturer's implied price-cost margins just as a function of demand side parameters.[11]

2.2.1. Scenario 1: Double Marginalization Model

In this model manufacturers set their prices first, and retailers follow, setting the retail prices given the wholesale prices. The margins that result from their behavior correspond to the pure double marginalization price-cost margins in the context of linear pricing in oligopoly markets at the manufacturer and at the retail level.

Let there be N_r retailers and N_w manufacturers in the market. All players are assumed to behave as Nash-Bertrand competitors. To solve this vertical model one starts, as usual, by looking at the retailer's problem. Each retailer r's profit function in week t is given by

$$\pi_{rt} = \sum_{j \epsilon S_{rt}} \left[p_{jt} - p_{jt}^{w} - c_{jt}^{r} \right] s_{jt}(p) \tag{11}$$

where S_{rt} is the set of products sold by retailer r during week t, p_{jt}^{w} is the wholesale price he pays for product j, c_j^r is the retailer's marginal cost of product j and $s_{jt}(p)$ is the share of product j. The first order conditions, assuming a pure strategy Nash-equilibrium in prices, are

$$s_{jt} + \sum_{m \epsilon S_{rt}} \left[p_{mt} - p_{mt}^{w} - c_{mt}^{r} \right] \frac{\partial s_{mt}}{\partial p_{jt}} = 0 \quad \text{for} \quad j = 1, \cdots, N_t \tag{12}$$

where N_t is the number of products in the market.

Define T_r as the retailer's ownership matrix with the general element $T_r(i,j)$ equal to one when both products i and j are sold by the same retailer and zero otherwise. Let Δ_{rt} be the retailer's response matrix, containing the first derivatives of all the shares with respect to all retail prices, with element $(i,j) = \frac{\partial s_{jt}}{\partial p_{it}}$. Stacking up the first order conditions given by (12) for all N_t products and rearranging terms, the following vector expression for the retailers' implied price-cost margins just as a function of the demand side for each week t is obtained

$$p_t - p_t^w - c_t^r = -(T_r.\Delta_{rt})^{-1} s_t(p), \tag{13}$$

[11] For illustration purposes see in the Appendix the (Logit and random coefficients) expressions of the price-cost margins in a simplified model with only two retailers, two wholesalers and two products.

where $T_r.\Delta_{rt}$ is the element by element multiplication of the two matrices. If the equilibrium is unique, equation (13) implicitly defines the retail prices as a function of the wholesale prices.

Looking now at the manufacturer, each of them maximizes his profit choosing the wholesale prices p^w, knowing that the retailers behave according to (13). The manufacturer's profit function is given by

$$\pi_{wt} = \sum_{j \epsilon S_{wt}} [p^w_{jt} - c^w_{jt}] \; s_{jt}(p(p^w)), \tag{14}$$

where S_{wt} is the set of products sold by manufacturer w during week t and c^w_{jt} is the marginal cost of the manufacturer that produces product j. The first-order conditions are, assuming again a pure strategy Nash-Equilibrium in the wholesale prices,

$$s_{jt} + \sum_{m \epsilon S_{wt}} [p^w_{mt} - c^w_{mt}] \frac{\partial s_{mt}}{\partial p^w_{jt}} = 0 \text{ for } \; j = 1, \cdots, N_t. \tag{15}$$

Let T_w be a matrix of ownership for the manufacturers, analogously defined as the matrix T_r above. In particular, element (j, m) of T_w is equal to one if manufacturer who sells product j also sells product m and is equal to zero otherwise. Let Δ_{wt} be the manufacturer's response matrix, with element $(j, m) = \frac{\partial s_{mt}}{\partial p^w_{jt}}$, containing the derivatives of the market shares of all products with respect to all wholesale prices. In other words, this matrix has the cross-price elasticities of the derived demand and the effect of cost pass-through. Collecting terms and solving for the manufacturers' implied price-cost margins yields

$$p^w_t - c^w_t = -(T_w.\Delta_{wt})^{-1} s_t(p). \tag{16}$$

To obtain Δ_{wt}, first note that $\Delta_{wt} = \Delta'_{pt}\Delta_{rt}$, where Δ_{pt} is a matrix of derivatives of all the retail prices with respect to all the wholesale prices. So all that is needed is to find expressions for, and compute, Δ_{pt} and pre-multiply Δ_{rt}, from the retailer's problem, by the transpose of Δ_{pt} to get the manufacturer's response matrix Δ_{wt}. From now on, the time subscript is dropped to simplify notation. To get the expression for Δ_p, let us start by totally differentiating for a given j equation (12) with respect to all prices $(dp_k, k = 1, \cdots, N)$ and a wholesale price p^w_f, with variation dp^w_f:

$$\sum_{k=1}^{N} \underbrace{[\frac{\partial s_j}{\partial p_k} + \sum_{i=1}^{N} (T_r(i,j) \frac{\partial^2 s_i}{\partial p_j \partial p_k} (p_i - p^w_i - c^r_i)) + T_r(k,j) \frac{\partial s_k}{\partial p_j}]}_{g(j,k)} \; dp_k \; - \; \underbrace{T_r(f,j) \frac{\partial s_f}{\partial p_j}}_{h(j,f)} dp^w_f = 0. \tag{17}$$

Putting all $j = 1,...N$ products together, let G be the matrix with general element $g(j,k)$ and let

H_f be the N dimensional vector with general element $h(j,f)$. Then

$$G\ dp - H_f\ dp_f^w = 0. \tag{18}$$

Solving for the derivatives of all prices with respect to the wholesale price f the f-th column of Δ_p is obtained:

$$\frac{dp}{dp_f^w} = G^{-1} H_f. \tag{19}$$

Stacking all N columns together, $\Delta_p = G^{-1}H$, which has the derivatives of all prices with respect to all wholesale prices. The general element of Δ_p is $(i,j) = \frac{\partial p_j}{\partial p_i^w}$.

Finally, the sum of the implied price-cost margins for the retailers and the manufacturers is by definition obtained by adding up (13) and (16)

$$p_t - c_t^r - c_t^w = -(T_r.\Delta_{rt})^{-1} s_t(p) - (T_w.\Delta_{wt})^{-1} s_t(p). \tag{20}$$

2.2.2. Scenario 2: Non-Linear Pricing Models

In a one manufacturer and one retailer case, in the classical nonlinear optimal (two-part tariff) pricing model, the manufacturer sets the wholesale price equal to marginal cost and lets the retailer be the residual claimant. Then the manufacturer is able to extract part or the full "monopoly" (or vertically integrated firm's) surplus in the form of a fixed fee that the retailer has to pay. Two-part tariffs are seen as optimal contracts whenever there is downstream market power in the retail market[12] and under fairly general market assumptions. Two-part tariff as the optimal contract design has been shown to be optimal in the simple double marginalization model where retailers follow manufacturers in a price setting game with a certain demand (Tirole, 1988, page 176), an uncertain demand (Rey and Tirole, 1986) or under asymmetric information (Tirole, 1988, page 177).[13] It is no longer true that the optimal two-part tariff in the context of multiple retailers yields marginal cost pricing by the manufacturers (Mathewson and Winter, 1984 and Schmalensee, 1981). However two-part tariffs are still optimal in the context of multiple manufacturers and a single retailer (Shaffer and O'Brien, 1997 and Tirole, 1988, page 180). In the one manufacturer, one retailer model, resale price maintenance implies that the manufacturer sets a wholesale price equal to the monopoly price and then imposes resale price at the monopoly price. The retailer makes zero profit, and the manufacturer recovers the monopoly profit.

[12] This is not to be confused with retailers having vertical power in the vertical structure. If there is retail market power, retailers impose an externality by adding a second margin to the wholesale margin.

[13] However, in the presence of uncertainty, two-part tariffs have poor properties in terms of risk sharing.

Scenario 2 considers the existence of these non-linear pricing contracts in the context of multiple retailers and multiple manufacturers. In particular, two sub-cases are considered to test the validity of two solutions to the double marginalization problem. In the first case, the wholesale margins are assumed to be zero, which means that the retailers have the pricing decision given that wholesale prices equal marginal costs. In the second case, the retail margins are assumed to be zero. Given that, it is as if the manufacturers were setting the final price consumers are facing, like in resale price maintenance. In both of these sub-cases, the profits of the vertical structure may be greater than the sum of the profits of retailers and manufacturers in the first scenario of double marginalization. The potential increase in the whole channel's profits is due to the elimination of the first or the second margin in the vertical structure and the fact that the retailers have some retail power, i.e., face downward sloping demand curves.

Case 1: "Wholesale margins are zero and retailers have pricing decisions." In this case, retailers maximize their profits, given that wholesale prices are equal to marginal costs. The manufacturers' implied price-cost margins are zero for all products. The implied price-cost margins for the retailers are given by equation (13) subject to $p_t^w = c_t^w$, that is,

$$p_t - c_t^r - c_t^w = -(T_r.\Delta_{rt})^{-1} s_t(p). \tag{21}$$

This means that the retailer gets from his optimization the profits corresponding to the downstream vertically integrated structure for each of the j products.

Case 2: "Zero retail margins and manufacturers have pricing decisions." In this case, the retail implied price-cost margins are zero for all products since the retailers add to the wholesale prices only the retail costs, i.e. $p_{jt} = p_{jt}^w + c_{jt}^r \quad \forall j$. This means that the manufacturers get from their optimization the profits corresponding to the downstream vertically integrated structure for each of the j products. The manufacturers' implied price-cost margins are given by

$$p_t^w - c_t^r - c_t^w = -(T_w.\Delta_{rt})^{-1} s_t(p). \tag{22}$$

It is worth noting that the implied price-cost margins in equation (22) are different from equation (21) because the retail ownership T_r differs from the manufacturer ownership T_w or, in other words, because the manufacturers and the retailers are maximizing their profits over a different set of products. In BLP (1995) and Nevo (2001) the (manufacturer) implied price-cost margins computed are given by expressions similar to (22) and the retailers' decisions are not modeled.

2.2.3. Scenario 3: The Hybrid Model

Each retailer behaves as a vertically integrated firm with respect to its own private label products and plays the vertical Nash-Bertrand game in the other products (the national brands). This scenario's implied price-cost margins have bits and pieces of the ones from scenario 1 and scenario 2's first case (for the expressions in a simple model, please refer to the appendix). In particular, the retail margins will be the same as in scenario 1 given by equation (13). However, the wholesale margins change: When vertically integrating into the upstream market, the retailers affect the price-cost margins of the national brands' manufacturers. By vertically integrating, the retailers eliminate the wholesale margins in the private labeled products, and the final retail price of the private labels falls. Demand for the products sold by the manufacturers of national brands changes (decreases), and consequently the national brand manufacturers need to adjust their wholesale prices. For this particular market, at the manufacturer level, the wholesale margins for the private label products are zero and thus not optimized over. The implied manufacturers' price-cost margins for the national brands are given by

$$p_t^w - c_t^r - c_t^w = -(T_w^*.\Delta_{wt}^*)^{-1} s_t^*(p), \tag{23}$$

where T_w^* is the manufacturers' ownership matrix as before without the rows and columns that correspond to the private label products in the sample. The expression for Δ_{wt}^* is equal to Δ_{wt} but without rows and columns of the private label products. In $s_t^*(p)$ are the shares of the national brands, namely $s_t(p)$ without rows for the private label products.

2.2.4. Scenario 4: Manufacturer Level Collusion Model

This scenario corresponds to manufacturers choosing wholesale prices that maximize the sum of the manufacturers' profits. Since manufacturers are assumed to be colluding, it is as if one single upstream firm owned the full set of products. Thus the manufacturers' ownership matrix T_w is a matrix full of ones, henceforth called T_1. Manufacturers' price-cost margins are given by equation (16) subject to $T_w = T_1$, which results in

$$p_t^w - c_t^w = -(T_1.\Delta_{wt})^{-1} s_t(p). \tag{24}$$

The implied price-cost margins of the retailers, which are assumed to set their retail prices given the wholesale prices, are given by

$$p_t - p_t^w - c_t^r = -(T_r.\Delta_{rt})^{-1}s_t(p). \tag{25}$$

Finally, the sum of the implied price-cost margins for the retailers and the manufacturers is obtained by adding up (24) and (25)

$$p_t - c_t^r - c_t^w = -(T_r.\Delta_{rt})^{-1}s_t(p) - (T_1.\Delta_{wt})^{-1}s_t(p). \tag{26}$$

2.2.5. Scenario 5: Retail Level Collusion Model

Assuming collusion at the retail level corresponds to assuming that $T_r = T_1$. Retail price-cost margins are given by

$$p_t - p_t^w - c_t^r = -(T_1.\Delta_{rt})^{-1}s_t(p), \tag{27}$$

while manufacturer price-cost margins have the following expression

$$p_t^w - c_t^w = -(T_w.\Delta_{wt})^{-1}s_t(p). \tag{28}$$

Finally, the sum of the implied price-cost margins for the retailers and the manufacturers is, by definition, obtained by adding up (27) and (28)

$$p_t - c_t^r - c_t^w = -(T_1.\Delta_{rt})^{-1}s_t(p) - (T_w.\Delta_{wt})^{-1}s_t(p). \tag{29}$$

2.2.6. Scenario 6: Monopolist Model

This last scenario examines the question of whether this industry is jointly profit maximizing. The way scenario 6 relates to scenario 2 is that one can view scenario 2 as asking the question of whether this industry has efficient pricing subject to horizontal constraints. Scenario 6 implies, in a world where there are several manufacturers and several retailers, that they not only vertically integrate but that they coordinate their horizontal pricing decisions (meaning that they are colluding). The fact that they cannot collude is what is meant by horizontal constraints in scenario 2.

In this present scenario, wholesale margins are zero. Furthermore, $T_r = T_w = T_1$. Consequently the implied price-cost margins of the full vertically and horizontally integrated structure are given by

$$p_t - c_t^r - c_t^w = -(T_1.\Delta_{rt})^{-1} s_t(p). \tag{30}$$

3. Estimation

3.1. Estimation of the Model using Two-Step Generalized Method of Moments

Given N_s random draws from the distribution of v and D, the integral in equation (10) is approximated by

$$s_{jt} = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{e^{\delta_{jt} + [x_{jt}, p_{jt}](\Gamma\ D_i + \Upsilon\ v_i)}}{1 + \sum_{k=1}^{N_t} e^{\delta_{kt} + [x_{kt}, p_{kt}](\Gamma\ D_i + \Upsilon\ v_i)}}. \tag{31}$$

When estimating demand the idea is to estimate the parameters that produce shares close to the observed ones. This procedure is non-linear in the demand parameters and suffers from the fact that prices are endogenous variables. The key step in estimating demand is to construct a demand side equation to be estimated linear in the parameters associated with the endogenous variables so that instrumental variables estimation (or, in other words, Generalized Method of Moments (GMM) estimation) can be directly applied. This follows from equating the estimated shares, given by (8) for the Logit model or by (31) for the full model, to the observed shares and solving for the mean utility. For the Logit model the mean utility can be recovered analytically, following Berry (1994)'s inversion technique, by $log(s_{jt}) - log(s_{0t}) = \delta_{jt}$. However, in the full model, solving for the mean utility has to be done numerically (see BLP, 1995).

Finally, once this inversion has been made, the equation to be estimated relates the mean utility from consuming a product to its price, the product characteristics x_{jt}, the constant product characteristics d_j and the changes in unobserved product characteristics that affect demand ξ_{jt}. This equation is given by

$$\delta_{jt} = d_j + x_{jt}\beta - \alpha p_{jt} + \xi_{jt} \tag{32}$$

or, by stacking all j products in (32), as

$$\delta(\Gamma, \Upsilon) = d + x\beta - \alpha p + u^d \tag{33}$$

where u^d is the demand side residual (i.e. the vector of ξ_{jt}). Let θ be the demand side parameters to be estimated. In the Logit case $\theta = \theta_L = (\alpha, \beta, d)$ and in the full model $\theta = (\theta_L, \Gamma, \Upsilon)$ where Γ and Υ are the non-linear parameters. For the Logit case, θ_L is obtained directly from estimating (33). In the full random coefficients model, θ is obtained by simulation following Nevo's (2000) estimation algorithm, where equation (33) enters in one of the steps.[14] Let me denote henceforth the scenario implied price-cost margins by $SIPCM$. Under each supply model separately, given θ, the price-cost margins implied by the different scenarios can be computed: $SIPCM_r(\theta)$ for the retailers and $SIPCM_w(\theta)$ for the manufacturers. The supply pricing equation is constructed by associating parameters to the implied margins. The intuition behind this approach is to compare actual price-cost margins with the price-cost margins implied by the different supply models. Actual price-cost margins are defined as the difference between retail price and the estimate of marginal cost of the vertical structure. The supply pricing equation to be estimated is then given by

$$p = f(c\ \gamma) + SIPCM_r(\theta)\ \lambda_r + SIPCM_w(\theta)\ \lambda_w + u^s, \tag{34}$$

where c is a matrix of cost side variables such as input prices, γ is a vector of coefficients associated with cost-side variables, λ_r and λ_w are N-dimensional vectors of coefficients associated with the implied price-cost margins of the retailers and the manufacturers, respectively, and u^s is the supply-side residual that contains unobserved components of marginal cost. The parameters λ_r and λ_w measure deviations from the retail and manufacturer behavior assumed in each of the underlying models. For example, for the double marginalization model, if the estimated coefficient λ_r is not significantly different from one I cannot reject that that retailer is behaving as a Nash-Bertrand player. It is assumed that $f(c\ \gamma) = e^{c\ \gamma}$ to ensure that all products have positive estimated marginal costs.[15]

Equations (33) and (34) define a system of two simultaneous vector equations to be estimated, where the demand side parameters θ enter the supply equation in a non-linear fashion. The estimation method consists of a two-step procedure.[16] The first step is estimation of demand by Generalized Method of Moments (GMM) and the second step is estimation of supply given the demand parameters previously estimated also by GMM. The use of consistent estimates in the first step will not cause in general a consistency problem in the second step. However, it will affect the covariance matrix of the second step estimators. Ignoring that effect can result in inconsistent

[14]The Logit smoothed accept-reject simulator for the market shares, given by equation (31), is used. This simulator is continuous in the data and in the parameters to be estimated, so gradient-based methods can be applied to estimate Γ and Υ.

[15]Results do not change, compared to the linear marginal cost assumption, in terms of the ranking of the different supply models. The products whose estimates of marginal cost are negative have statnottically significantively higher values for λ_w or λ_r compared to the exponential marginal cost estimates.

[16]Alternatively, both the demand and supply equations can be estimated jointly. This feasible and efficient two-step approach is used here due to its simplicity, which is particularly important when the random coefficients model is estimated. For one example of the two-step procedure, see the Heckman (1976) sample selection estimator.

standard errors estimates.[17] In the first step, the GMM function to be minimized with respect to the demand parameters is given by

$$u_d' Z_d W_d^{-1} Z_d' u_d, \tag{35}$$

where the wings of the quadratic form are the moments $u_d' Z_d$ (i.e., the residuals interacted with the instruments) and W_d is a consistent estimate of the variance-covariance matrix of the moments. Since W_d itself depends on a consistent estimate of the true parameters, a circularity problem emerges when minimizing (35). To solve this problem I assume that $W_d = Z_d' Z_d$, which results in estimating the demand parameters θ by Two Stage Least Squares (2SLS). This is optimal in the presence of homoscedastic errors. The 2SLS estimators are unbiased, consistent and asymptotically normally distributed even in the presence of heteroscedasticity.[18] Given θ^* the implied price-cost margins, $SIPCM_r(\theta^*)$ and $SIPCM_w(\theta^*)$ are computed. In the second step, given θ^*, the GMM function to be minimized with respect to the supply parameters is given by

$$u_s' Z_s W_s^{-1} Z_s' u_s, \tag{36}$$

where the wings of the quadratic form are the moments $u_s' Z_s$ (i.e., the residuals interacted with the instruments Z_s) and W_s is a consistent estimate of the variance-covariance matrix of the moments. The supply parameters $\theta_s = [\gamma\ \lambda_r\ \lambda_w]'$ in (34) are estimated by 3SLS. I iterate once more, using the 2SLS supply parameter estimates, θ_s^*, to compute the following consistent and efficient estimate of the variance covariance matrix of the moments $Z_s' u(\theta_s^*) u(\theta_s^*)' Z_s$ which in turn is used to compute a new estimate of the parameters θ_s^{**}. The use of an efficient weight matrix is needed for GMM hypothesis testing (Newey and McFadden, 1994) and for non-nested testing.[19]

By assuming exponential marginal costs, this implies that the marginal cost parameters, γ, are non-linear. Furthermore, in the random coefficient model of demand, some demand parameters are also non-linear, namely Γ and Υ. The way to estimate demand for the random coefficients model and supply for the exponential marginal cost case is to concentrate the objective function such that it will be only a function of the non-linear parameters. First note that from the first order conditions with respect to the linear parameters, the optimal vector of linear parameters can be expressed as a function of the non-linear parameters. Then substituting back into the GMM objective function

[17] This problem is present if and only if consistency of the first-step estimator affects consistency of the second-step estimator (Newey and McFadden, 1994).

[18] However, one needs to obtain an appropriate White (1980) estimate of the 2SLS estimators' variance covariance matrix.

[19] Further iterations between estimating θ and W, even though implying an additional computational effort (specially in the random coefficient demand estimation), do not seem to dominate in finite samples and are asymptotically equivalent (to a first order approximation) to the procedure used here (see Hansen, Heaton and Yaron, 1996).

it can be expressed as a function of the non-linear parameters alone. Optimization is then done with respect to the non-linear parameters. To ensure finding a global minimum, I start by using a gradient method (providing an analytical gradient) using different starting values of the non-linear parameters to find a minimum of the GMM objective function. Then I use that minimum as a starting value for the Nelder-Mead algorithm to check if the results coincide.[20]

Finally, robust standard errors of the parameters are obtained. The White (1980) estimate of the covariance matrix of the demand side parameters estimated by 2SLS, defined in the presence of heteroscedastic demand residuals, is, for the Logit model, given by

$$V_l = (X'Z_d(Z_d'Z_d)^{-1}Z_d'X)^{-1}X'Z_d(Z_d'Z_d)^{-1}(\sum u_{d,i}^2 Z_d Z_d')(Z_d'Z_d)^{-1}Z_d'X(X'Z_d(Z_d'Z_d)^{-1}Z_d'X)^{-1},$$

where X are all the regressors in equation (33). For the random coefficient model, the White (1980) estimate of the covariance matrix of the demand (both linear, $\theta_L = (\alpha, \beta)$ and non-linear Γ, Υ) parameters estimated by 2SLS is given by

$$\begin{aligned} V_{rc} = \quad & ([X\vdots F_{(\Gamma,\Upsilon)}]'Z_d(Z_d'Z_d)^{-1}Z_d'[X\vdots F_{(\Gamma,\Upsilon)}])^{-1}[X\vdots F_{(\Gamma,\Upsilon)}]'Z_d(Z_d'Z_d)^{-1}(\sum u_{d,i}^2 Z_d Z_d')(Z_d'Z_d)^{-1}Z_d'[X\vdots F_{(\Gamma,\Upsilon)}]\cdot \\ & ([X\vdots F_{(\Gamma,\Upsilon)}]'Z_d(Z_d'Z_d)^{-1}Z_d'[X\vdots F_{(\Gamma,\Upsilon)}])^{-1}, \end{aligned}$$

where $[X\vdots F_{(\Gamma,\Upsilon)}]$ is a matrix that has in columns the regressors X associated with the linear demand parameters and the gradient $F_{(\Gamma,\Upsilon)}$ of the mean utility with respect to the non-linear parameters Γ and Υ.

Given V_l or V_{rc}, that I shall call generally V_d, then the White (1980) covariance matrix of the supply side parameters can be computed. Since the price-cost margins, which are regressors in the supply equation, have embedded estimators (as in Newey and McFadden (1994)) from the first stage of estimating demand, this results in having to correct the covariance matrix of the supply parameters for the embedded aspect. The variance of the supply parameters estimated by 3SLS is then given by

$$V = \quad (X_s'Z_s(Z_s'Z_s)^{-1}(\sum u_{2SLS,s,i}^2 Z_s Z_s')(Z_s'Z_s)^{-1}X_s'Z_s)^{-1} + \text{E.C.},$$

[20]The Nelder-Mead (1965) simplex search method is a direct search method that does not use numerical or analytical gradients, as in quasi-Newton methods. It is generally less efficient and slower to converge than the gradient methods but is more robust specially if the objective function is highly discontinuous.

where $X_s = [\nabla_\gamma \vdots SIPCM(\theta)]$ is a matrix that contains in the first columns the gradient of marginal cost with respect to marginal cost parameters, ∇_γ, and in the next columns the price-cost margins, $SIPCM(\theta)$, of the model in question. The term E.C. is the embedded correction which is given by[21]

$$\text{E.C.} = BV_dB' - BA_N^{-1}A'(Z_d'Z_d)^{-1}\Omega_{hg} - \Omega_{gh}(Z_d'Z_d)^{-1}AA_N^{-1'}B', \tag{37}$$

where each of the elements in (37), are consistently estimated by

$$\begin{aligned}
\hat{A} &= Z_d'X_d \\
\hat{A_N} &= X_d'Z_d(Z_d'Z_d)^{-1}Z_d'X_d \\
\hat{B} &= Z_s'\tfrac{\partial X_s \theta_s}{\partial \theta} \\
\hat{\Omega_{hh}} &= Z_d'\hat{u}_d\hat{u}_d{}'Z_d \\
\hat{\Omega_{gg}} &= Z_s'\hat{u}_s\hat{u}_s{}'Z_s \\
\hat{\Omega_{gh}} &= Z_s'\hat{u}_s\hat{u}_d{}'Z_d \\
\hat{\Omega_{hg}} &= Z_d'\hat{u}_d\hat{u}_s{}'Z_s.
\end{aligned} \tag{38}$$

3.2. Instruments and Identification of the Model

3.2.1. The Instruments

The need to use instrumental variables in the estimation of demand results from the fact that when retailers decide retail prices, they take into account all the product characteristics, not only the ones that are observed, the x_{jt}, but also those characteristics that are not observed by the analyst, the ξ_{jt}. The retailers also take into account any changes in their products' characteristics when setting the retail prices. Since a product fixed effect is included, it will capture the product characteristics that are constant over time, both observed and unobserved. The econometric error that remains in ξ_{jt} will therefore only include the changes in unobserved product characteristics such as unobserved promotions, changes in shelf display and/or changes in unobserved consumer preferences. This implies that the prices in (33) are correlated with the changes in unobserved product characteristics affecting demand. Hence, to obtain a precise estimate of the price coefficients, those endogenous variables need to be instrumented for.

[21] For the derivation of that embedded correction term given a Logit demand and a random coefficients demand model, please see the supplement to this paper available at http://socrates.berkeley.edu/~villas/homepage.html. In the supplement I also provide suggestions for programming the computation of the embedded correction term.

When estimating the supply-side parameters, the λs, note that they are associated with price-cost margins. Price-cost margins depend on the response matrices Δ_r and Δ_w and the shares $s(p)$, which in turn depend on the vector of prices for all products. So, in order to consistently estimate the supply-side parameters in the supply equation (34), there is a need to find instruments for the endogenous price-cost margins.

Recall that a valid instrument has to satisfy two requirements. It has to be both uncorrelated with the residual and correlated with the endogenous variable that one is instrumenting for. The price decision takes into account cost-side variables, such as input prices. It is reasonable to assume that the prices of inputs are uncorrelated with the changes in unobserved product characteristics, the ξ_{jt}. For example, changes in shelf display are most likely not correlated with input prices such as the prices of milk and sugar. One problem is that there is not brand or product level variation for the prices of the inputs in the data. This can be solved by interacting those input prices with product dummy variables. The idea behind these instruments is that different products use different levels of inputs. Since the exact content is not observed, it is estimated this way. Thus these cost data multiplied by product fixed effects are the instruments for the endogenous retail prices. In particular, two sets of instruments are considered when estimating demand to examine the effects of the two alternative instrumental variables specifications. In the first specification, by allowing marginal cost of a given yogurt brand sold at two different retailers to vary, I interact 43 product dummies (where product is defined as brand-store) with the input prices. In the alternative specification, I assume that the marginal cost of the same brand sold at two different retailers is the same. This results in 21 brand dummies interacted with the input prices. One last note on why these instruments (input prices multiplied by product fixed effects) are valid. The residual in the demand equation has only the part that is not explained by "store-brand" level fixed effects. If I had not included brand dummies in the demand, I would have the problem that the instruments would be correlated with constant unobserved product characteristics. Since I account for those by estimating the brand-store fixed effects, I don't have that problem.

Consider now the supply side. Lagged implied price-cost margins are used as instruments for the implied price-cost margins. Several lagged specifications are used in an attempt to access the robustness of the results. The identifying assumption lies in the fact that, controlling for observed cost side determinants of price, past price-cost margins are uncorrelated with contemporaneous unobserved determinants of prices (components of marginal cost that is unobserved). BLP (1995) suggest product characteristics or functions of product characteristics as instruments for supply-side variables, however, given the fact that product characteristics are assumed to be constant during the sample period, this class of instruments cannot be used here.

3.2.2. Identification

There are three levels of identification that I have to worry about. There is first the level of how I am consistently estimating demand. The second level is how can I separate empirically the different models that I am comparing. And finally, how can I distinguish between models that I don't consider.

Starting with the third level of identification, I cannot identify models with the same marginal pricing outcome. For example, models that have different wholesale and retail margins but have the same total margins are not distinguishable.

Let me now discuss, given demand, how I identify the vertical supply models of pricing behavior.To summarize, one needs the slope of demand to be changing over time and the curvature of demand (that affects the derived demand of manufacturers) to be changing over time. Depending on the vertical supply model in question, manufacturers and retailers react differently to changes in the curvature and the slope of demand. Let me consider the case of constant (not varying with quantity) marginal costs.[22] If marginal cost is not varying with quantity, I can trace out whether the manufacturers (facing a shifting derived demand - or a changing curvature of demand) and the retailers (facing a shifting demand) behave according to Bertrand-Nash pricing behavior. Assuming that one cannot find enough exogenous variables to shift demand and change it's curvature over time, the idea is to look at changes in demand over time that are caused by random shocks from previous periods. Based on distributional assumptions of the demand and supply econometric error terms, namely that demand is serially correlated, i.e. that $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$ and that supply shocks are not, one can ask how changes in demand that are caused by past period's demand shocks, affect prices today. If marginal costs do not vary with quantity, then one can trace out whether the manufacturers (facing a shifting derived demand, or in other words, facing a shifting curvature of demand)) and the retailers (facing a shifting demand) behave according to the base model of double marginalization or to another model. In practice, I use the part of ϵ_{t-1} that is captured by lagged price-cost margins of retailers and manufacturers as the role that demand shifters play. In terms of the econometric exclusion restrictions, lagged margins which are correlated with ϵ_{t-1} and are excluded from the contemporaneous supply pricing equation can be used as instruments for the supply endogenous variables in equation (34). Several lagged specifications, up to p-lags, can be used if the supply residual is not serially correlated and if one assumes that ϵ_t is AR(p).

From what was argued above, it is therefore crucial to be able to estimate consistently a very

[22]The same argument holds for the homogeneous product's and for the differentiated product's cases as well as for the models that don't consider retail behavior explicitly.

flexible demand model with a flexible Hessian. To consistently estimate demand, the experiment is to ask consumers to choose between different products over time, where a product is perceived as a bundle of attributes (among which are prices). The source of identification is the relative price variation over time. Since prices are not randomly assigned, I use input price changes over time that are significant and exogenous to unobserved changes in product characteristics as instruments for prices. Another source of identification of demand and, in particular, of demand elasticities, is the fact that the choice set of consumers changed due to renovation of one of the stores during some weeks.

4. The Market and the Data

4.1. The Market

The analysis focuses on the yogurt market in a Midwestern metropolitan area, more precisely in two zip code areas.[23] Yogurt has to be consumed within twenty-eight days of its production, so its shelf life is short. In case there are some significant marginal cost changes, there exists, in principle, the possibility for the manufacturers to adjust the wholesale prices accordingly. Some wholesale price variability can therefore be expected. In the data, retail yogurt prices for each product over time are fairly variable. Furthermore, the retail price variability of yogurt does not seem to be uniquely related to promotional activities. This is particularly important since if price promotional aspects were the drivers of retail price variability then input price changes may not have been reflected in changes of retail prices. This could imply that input prices would be poor instruments for the retail prices. Figure 1 plots the price series for one of the large selling brands of yogurt. Temporary price reductions are characterized by a decrease in price during a number of successive weeks, and after that period price rises to its original level. Such a pattern is not present for the yogurt price series. Looking instead at the price series for a heavily promoted brand, such as a large selling soft drink brand, (see Figure 2) the pattern of price promotions is evident.

Yogurt is produced by a few leading national yogurt manufacturers: Dannon and General Mills, who together account for almost 62% of the total U.S. yogurt sales. Private label brands from retail stores are in third place with 15% of the market and Kraft comes next. All other manufacturers have individual shares of less than 2% (Frozen Food Digest, October 1995: 38). This industry is fairly concentrated at the manufacturer level. Therefore, in principle, it is interesting to confront

[23] For confidentiality reasons the city's name is not revealed, nor are the retail store names.

supply models of upstream price collusion (for example, scenario 4) with the data. Scenario 3 was inspired by the importance that private labels seem to have in the yogurt market.

One of the most important characteristics of the yogurt market is that yogurt sales are mostly driven by new product introductions.[24] In 1994, there were over one thousand new dairy product introductions and over one hundred yogurt introductions alone[25]. In terms of product variety, each store sells an average of 150 yogurts from seven manufacturers. Product variety together with successful advertising (influencing consumers' evaluations of the different products) can result in positive price-cost margins for the manufacturers due to product differentiation alone. This would be reflected in the estimates for the price-cost margins in the non-collusive supply scenarios considered in this paper.

At the retail level there is a small number of large retailers (or retail chains) competing directly with each other and who have jointly 75% of total sales to final consumers in the whole metropolitan area. All other retailers not considered have individual shares less than 5% (figures for 1992). Three retail stores are considered in the data, where store 1 is a smaller store than stores 2 and 3. The last two retail stores belong to two retail chains, while store 1 is unique in the whole metropolitan area. The retail stores in the data are located within less than two miles from one another, and in fact two retailers are located at both sides of a street intersection (see Figure 3). Some smaller grocery stores are located within the two zip code areas considered, but the closest large retail store nearby is located in a different zip area.[26]

4.2. Data

The analysis is done using a data set on retail prices, advertising, aggregate market shares[27] and product characteristics for 43 products produced by five manufacturers. In particular, the number of products are equal to 43 for all weeks but for the weeks during which retailer 2 closed due

[24] On the competitive effects of product line extensions in the U.S. yogurt market by the two leading manufacturers, see Kadiyali, Vilcassim and Chintagunta (1999). In another paper, Draganska and Jain (2000) use store-level data for the yogurt category to derive recommendations for effective product-line extension decisions based on what-if experiments.

[25] For the sample period and market considered in this analysis, there are five new product introductions.

[26] The other grocery stores and nearby retailers that could overlap with the three stores are therefore in the outside option of the model.

[27] The household sample is not used in this analysis because it does not seem to be representative. In fact, twenty five products with susbtantial market shares in the aggregate sample have zero market shares in the individual sample during more than eight weeks. Combining market shares by store there are again discrepancies when using the individual sample. Store one has a 21% combined market share in the individual sample while stores two and three have 18% and 68%, respectively. However in the aggregate data store one has only 6%, store two has 57% and store three has 37%.

to remodeling, when the number of products in the sample are 25. Information on consumer demographics, wages by state and input prices is also used.

The price, feature (advertising) and market share data come from an Information Resources Inc. (IRI) scanner data set that covers the purchases in three retail stores in a Midwestern urban area during 104 weeks.[28] Summary statistics for prices, feature, quantity sold and shares are presented in Table 1. Feature is a dummy variable that takes the value of one when the product was featured during that week. Table 1 also presents summary statistics on combined shares for all the products, combined shares for the products sold by each manufacturer and combined shares for the products sold by each retailer. The combined shares for the products analyzed are on average 34%. Quantity sold is defined as servings sold, where one serving corresponds to a 6-ounce yogurt cup. Price and servings sold series for the 43 products in the sample were obtained by aggregation.[29]

Market shares are defined by converting quantity sold to servings sold and then dividing by the total potential servings in the market. The potential market, in terms of servings, is assumed to be half of a serving per capita a week. Hence the potential market in terms of servings is equal to half of the resident population in the two zip code areas. This assumption is consistent with U.S. consumption patterns. According to U.S. Department of Agriculture, Americans consume on average 9 pounds of yogurt a year, which in terms of servings corresponds to approximately half of a serving per capita a week. Table 3 provides the average U.S. per capita consumption for the years 1991-94 and compares it with international patterns. The highest per capita consumption average for the countries considered is in Bulgaria with 3.39 servings per week[30] and the lowest in Russia with less than 25% of a serving per week. The estimates of the marginal utility of price are robust to small variations of the weekly per capita consumption assumption,[31] as can be seen in Table 4.

In the potential market, Dannon comes first in terms of local market shares of its products with an average of 17%. Next comes General Mills with 9%. The private labels come third with 4%. Kraft comes last among the products analyzed with 3%. Furthermore, in this local market, combined shares for the 10 products sold by retailer 1 are on average 2%, and for store 2, which sells 18 products, they are 20%. Store 3 has average combined shares of its 15 products of about 14%.

[28] I thank David Bell for letting me have access to this data set. See data Appendix for a detailed description of the data set.

[29] For a particular retailer, a product is defined such that when sold in different sizes would be aggregated as the same product. Also products with the same brand name and with price correlation close to one and with similar product characteristics were aggregated. For the list of the products in the sample, please refer to Table 2.

[30] The primary yogurt culturing bacteria, Lactobacillus bulgaricus, was named by 1908's Physiology and Medicine Nobel laureate Dr. Metchnikoff, in honor of the yogurt-loving Bulgarians.

[31] The ranking of the different supply models is also invariant to small changes in weekly per capita assumption.

The product characteristics data were collected by inspection of the label reads and for those products currently unavailable in any supermarket because they were discontinued, from manufacturers' descriptions. Table 5 describes the following product characteristics: calories, total fat, cholesterol, carbohydrates, dummy for vitamins, dummy for calcium above 30% daily value, Aspartame dummy, Fruit on the Bottom dummy, available in different sizes dummy and store dummy variables. These are most of the characteristics found to be more relevant to consumers when purchasing yogurt, according to Frozen Food Digest (1995), and also according to manufacturer's yearly market surveys and brand-name comparison articles (e.g., Nutrition Action, 1998, Center for Science in the Public Interest).

The cost data set is described in Table 6, with reference about the different sources. In order to get an idea of the input price variability, the weekly input price series normalized by the correspondent weekly average, are presented in Figures 4 and 5. The conclusion is that there is, for most of the cost data series, considerable time variation.

Using information on consumer demographics, such as family size, income and age, allows us to consider consumer heterogeneity in the taste parameters for the different products. A sample from the joint distribution of income and age of the resident population was obtained from the 1990 Census at the zip code level, for the zip codes of interest (see Table 7). The population in the market considered is about 76% white with median household income of about thirty thousand dollars a year and on average with 2.5 persons in the household.

5. Results

5.1. Demand Estimation

The Logit model for demand is considered first to get a feel for what is going on in the data. It also allows to compare and choose between two different instrumental variable specifications and illustrate the need to instrument for prices when estimating demand. Understanding the drawback of having poor substitution patterns, I then estimate a full random coefficients discrete choice model of demand for differentiated products.

5.1.1. Logit Demand

Table 8 presents the results from regressing the mean utility δ_j, which for the Logit case is given by $ln(s_{jt}) - ln(s_{0t})$, on prices and product dummy variables in equation (32) or, in vector notation, in equation (33). The second column displays the estimate of ordinary least squares for the mean price coefficient α and columns three and four have estimates of α for two different instrumental variables (IV) specifications, using input prices as instruments for the prices. In the first IV specification, assuming that marginal cost for the same product sold at different retailers is different, prices are instrumented by input prices interacted with 43 product dummy variables. In the second IV specification, prices are instrumented by input prices interacted with 21 product dummies. This last specification corresponds to assuming that marginal cost for the same product sold at different retailers is constant. Regarding the need to instrument for prices, the Hausman (1978) test for exogeneity suggests that there is a gain from using instrumental variables versus ordinary least squares when estimating demand. The last columns of Table 8 present the results from including feature, i.e., from the regressing the mean utility δ_j on price, feature dummy variable and product dummy variables. The coefficient of feature is not significant for OLS as well as for the two IV specifications, and the effects of including feature on the price coefficients and on the product characteristic coefficients are insignificant both statistically and economically. Furthermore, I cannot reject the exogeneity of the feature variable.

The first stage R-squared and F-Statistic of both IV specifications are high, suggesting that the demand instruments have some power.[32] Estimates of first-stage coefficients have in general the expected positive sign and are significant for plastic, sugar and milk. Estimates for the average effect of strawberry price on the price of yogurt are positive and in general significant for fruit yogurts. Coefficients for the wages in the states where plants of the different products are located are significant and positive.

To choose between the two different specifications, the assumption of constant marginal cost (across same "physical" product sold in different retailers) is tested. In particular, I test the assumption that the coefficients associated with the same input for the same "physical" product are equal to each other. This is a much stronger assumption than what is needed but in case of not being able to reject it this makes the choice for the specification that assumes constant marginal cost (Specification 2) stronger. This assumption cannot be rejected, so I choose to proceed with the second specification.[33]

[32] First-stage results for both are available at http://socrates.berkeley.edu/~villas/homepage.html. In the webpage mentioned, the first part of the table assumes that marginal cost for the same product sold at different retailers is different, the second part assumes the opposite.

[33] The estimate of the Wald test statistic is 160.9, which is less than 326.3, the 95% the critical value for a chi-square with 286 degrees of freedom. The critical value C for a chi-square with large R degrees of freedom for the significance level a can be approximated (Greene, 1997, p.70) as $C \approx 0.5[\Phi^{-1}(a) + \sqrt{2R-1}]^2$.

5.1.2. Random Coefficients Demand

Results from estimating equation (33) for the full model are presented in Table 9 considering consumer heterogeneity by allowing the coefficients on price, calories, calcium and store-specific dummy variables to vary across consumers as a function of their income, their age and other unobserved consumer characteristics. Interpreting the estimates, the mean price coefficient is similar to the Logit case, however, unobservable characteristics in the population seem to affect it significantly. The coefficients associated with the store dummies are to be interpreted as relative to the smaller store 1. For example, unobservable characteristics in the population do not seem to explain why people choose stores 2 and 3 over store 1. In fact, older people seem to significantly prefer store 1 over both the two other stores, given the negative and significant coefficient associated to the interaction between the store dummies and age of the population. The preferences for the larger stores 2 and 3 rise with an increase in income. Higher calcium content seems to be preferred by older consumers. In terms of the mean price coefficient it is similar to the Logit estimate for the mean of the marginal utility of price. From the coefficient on the interaction of price with income one interprets that consumers with higher income are less price sensitive. Age does not significantly seem to affect the mean price sensitivity. The estimates for the interactions of demographics with the constant term (that captures consumers' valuation for the outside option) suggest that older consumers and consumers with less income are less likely to buy yogurt.

Product fixed effects are estimated, the d_js, capturing the part of the mean utility level that is constant over time and associated with product characteristics that don't change in the sample period, improving the fit of the model. However, given that the product characteristics used are time invariant, the estimates of their coefficients cannot be directly obtained when product fixed effects are included. It is easy to see that the coefficient β of a constant product characteristic x_j is not identified, i.e. it is indistinguishable from the coefficient d_j. Nevertheless, the taste coefficients for the product characteristics can be indirectly obtained by estimating the following regression $\hat{d} = X\beta + \zeta$ using Generalized Least Squares, where $\hat{d}$ are the estimated product fixed effects and assuming $E[\zeta|X] = 0$. Table 9 displays the estimates of the consumer taste parameters for the different product characteristics. The R-squared of 0.78 suggests that the fit obtained is good. All coefficients are statistically significant, and the signs are, in general, consistent with consumer surveys. For example, for the average consumer, calories and Aspartame (an artificial sweetener that was not 100% FDA approved) have a negative marginal utility. The availability of different sizes of yogurt for a certain brand is, on average, positively valued, as is less sugar content, the most calcium content and the possibility of trying new yogurt flavors. Finally, on average, consumers tend to prefer stores 2 and 3 over the smaller store 1.

Additional specifications are presented in Table 10. Column 3 presents the results for the GMM estimation of the full model while column 4 presents the NLLS estimates of the full model. The coefficients change considerably as do the estimated price-cost margins. Column 5 presents the estimates from the specification that sets the unobserved shocks v_i to zero for all the product characteristics. Comparing columns 3 and 5, the estimates are essentially unchanged, as are the average estimated price-cost margins for the different scenarios considered and also the ranking of the different supply models. This suggests that the heterogeneity is driven by demographics and not by random shocks. Finally, columns 6 through 8 present the results for a full model of demand including feature: Column 6 is equal to 3 but adds feature, column 7 has the NLLS results with feature and, finally, column 8 does not allow for random shocks v_i and includes feature. Comparing the estimates of columns 6 and 7, there is still a considerable change when instrumenting for prices after including feature. The heterogeneity is mostly explained by demographics since the estimates from column 6 are similar to those in column 8. The coefficients for feature are overall statistically insignificant and comparing column 3 with column 6, there is not a significant effect on the estimates, on the price-cost margins and also on the ranking of the models from including feature.

5.2. Elasticities and Price-Cost Margins

5.2.1. Elasticities and Price-Cost Margins: Logit Demand

Given the demand parameters one can compute the implied price-cost margins for the different supply models being considered. Summary statistics for the price-cost margin estimates, given a Logit demand model, are provided in the top part of Table 13. In each line are the price-cost margins for the different models. For the models that estimate both retail and wholesale margins those are added up to have an estimate of the whole vertical margin for each product. Comparing the whole vertical margins for the different models, one notes that for four of the models there are some products, during some weeks that exhibit estimated price-cost margins greater than 100%. This happens in particular for the double marginalization model, for the hybrid model, for the wholesale (scenario 4) and for the retail (scenario 5) collusion models being considered. This is a first indicator of a possible problem in those models fitting the data. Whether this problem is going to be statistically significant is tested in the next sections. When retailers decide the prices (scenario 2, case 1) the price-cost margins estimated are, on average, slightly higher than the price-cost margins that result when manufacturers decide the prices (scenario 2, case 2). Finally, the monopolist model in scenario 6 predicts, on average, larger margins than the previous two models.

Additionally, columns 3 through 11 of Table 2 present the average estimated price-cost margins by product for the retailers and the wholesalers for each different scenario. One fact that is evident is that private labels have the largest estimated price-cost margins under all scenarios considered. This is an implication of the Logit demand specification. The elasticities of the market shares with respect to prices are, for the Logit, given by

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

The lower the price, the lower the elasticity (in absolute value). Since relative price-cost margins are negatively related to elasticities, the lower in absolute value are the elasticities, the higher the price-cost margins. Private labels with lower prices than national brands exhibit therefore, not surprisingly, the highest implied price-cost margins. One last limitation of the Logit demand specification are the implied cross-price elasticities. Products with similar market shares and prices will have similar cross-price elasticities. For the estimated own and cross elasticities for the Logit model, see columns three and four of Table 12. The elasticities vary by brand, where the mean of the distribution of own-price elasticities is -8.01 with a standard deviation of 0.03. In terms of cross-price elasticities, they are on average 0.026 with a standard deviation of 0.02.

5.2.2. Elasticities and Price-Cost Margins: Random Coefficients Demand

In the full model, the above described and other limitations in terms of elasticities disappear. On the one hand, own price elasticities are no longer uniquely driven by functional form specifications, such as above. In particular, the marginal utility of price α will now vary by product, in the sense that it is obtained as the average of all the price sensitivities for all the consumers of that particular product. On other hand cross-price substitution patterns are richer. The elasticities for the full model are given by

$$\frac{\partial s_{jt}}{\partial p_{kt}} \frac{p_{kt}}{s_{jt}} = \begin{cases} \frac{p_{jt}}{s_{jt}} \int \alpha_i s_{jt,i}(1 - s_{jt,i}) \, dF(v) \, dF(D) & \text{if j=k} \\ -\frac{p_{kt}}{s_{jt}} \int \alpha_i s_{jt,i} s_{kt,i} \, dF(v) \, dF(D) & \text{otherwise,} \end{cases}$$

where

$$s_{ijt} = \frac{e^{\delta_{jt} + \mu_{ijt}}}{1 + \sum_{k=1}^{N_t} e^{\delta_{kt} + \mu_{ikt}}}$$

is now the individual probability of i purchasing product j during week t. In particular α_i is the marginal utility of price for consumer i. The integral over the unobserved and observed consumer demographics is simulated by drawing N_s random pairs (v, D). The simulated elasticities are then given by

$$\frac{\partial s_{jt}}{\partial p_{kt}}\frac{p_{kt}}{s_{jt}} = \begin{cases} \frac{p_{jt}}{s_{jt}}\sum_{n=1}^{N_s}\alpha_n s_{jt,n}(1-s_{jt,n}) & \text{if j=k} \\ -\frac{p_{kt}}{s_{jt}}\sum_{n=1}^{N_s}\alpha_n s_{jt,n}s_{kt,n} & \text{otherwise,} \end{cases}$$

Table 12 presents the own price elasticities as well as the mean and the standard deviations of the cross-price elasticities for the random coefficients demand model. In terms of own price elasticities, they are on average slightly lower than the ones estimated in the Logit demand model. The mean of the own price elasticities is now -7.5 with a standard deviation of 0.13. In terms of the cross-price elasticities, their summary statistics are presented in the last columns of Table 12. They vary significantly by product, ranging, on average, from 0.001 to 0.194. Due to its large dimension, detailed information on estimated cross-price elasticities for the 43 products in the sample is not presented. Let me however summarize that, overall, the results seem reasonable and intuitive. As an illustrative example, Dannon Lowfat Plain Yogurt sold in store 3 is less sensitive to a change in price of Dannon Lowfat Plain Yogurt sold at any of the two other stores (elasticities of 0.005 and 0.01) than to a change in the price of Dannon Light Vanilla Yogurt sold at the same store (elasticity of 0.023). Also, when looking within the same store, the effect on Dannon Light fruit yogurt from changes in the price of Dannon Classic Flavor Fruit yogurt (elasticity of 0.222) seems to be larger than the effect from changes in the price of Dannon Lowfat plain yogurt (elasticity of 0.063). These yogurts are maybe used for different purposes (plain yogurt is sometimes used for cooking) and therefore purchased by consumers with different characteristics.

Overall, the products seem to be less sensitive to changes in prices of products in other stores than to changes in prices in the same store. To summarize this, Table 11 presents the mean cross-price elasticities for products within a store and contrasts it with the mean cross-price elasticities between products of different stores. If one defines a measure of the relative substitution, as the mean cross-price elasticities within the same store divided by the mean cross-price elasticities across stores, on average, the substitution within store is four times the substitution across stores. In particular, breaking up this analysis by store, it is interesting to verify that the smaller store 1 seems to have the most loyal customers, versus the larger stores. Nevertheless, the substitution within stores 2 and 3 is still larger than across stores.

As a diagnostic of how far the results are from the Logit demand model, the variance of the cross-price elasticities in the random coefficients model is computed. The last columns of Table 12 present the mean and the standard deviations of the cross-price elasticities with respect to a certain price (which should be zero, according to Logit assumptions). For all products, the standard deviation, relative to the mean, is fairly large and the Logit restrictions seem less reasonable, especially for products of store 1.

The estimated price-cost margins differ by retail store. Retail store 1, the smallest retailer in the sample, exhibits the largest variability of price-cost margins across time for all scenarios estimated and also the largest average price-cost margins. Summary statistics for the price-cost margin estimates in the random coefficients demand specification are presented in the bottom part of Table 13. In each line are the price-cost margins for the different models. For the models that estimate both retail and wholesale margins, those are added up to have an estimate of the whole vertical margin for each product. Comparing the total vertical margins for the different models, one notes again that for some of the models there are certain products, during some weeks, that exhibit estimated price-cost margins greater than 100%.

5.3. Supply Estimation

To estimate the supply parameters, based on the assumption that the supply residuals are uncorrelated over time, lagged implied price-cost margins are used as instruments for the implied price-cost margins. This is a reasonable assumption since the null hypothesis of autocorrelation of the supply residuals is rejected for all the estimated supply scenarios. Furthermore, the results are robust to different lagged specifications. The first stage R-squared is large, of about 0.7 for all the supply scenarios estimated. When comparing the seven scenarios, an informal diagnostic for the validity of each model is done first, by looking at each of the λs individually and testing for them to be equal to one. These null hypotheses imply that, given the assumptions for demand, the price-cost margins estimated under the scenario in question are consistent with the price-cost margins obtained from supply-side estimates.[34] Results are presented in the last column of Table 14. The price-cost margins implied by the double marginalization model are the ones that seem less consistent with the data, since the null hypothesis of the parameter associated with each individual margin being one is rejected in 86% of the cases. Looking at more efficient contracting solutions between manufacturers and retailers, I reject the null hypothesis of each individual margin being consistent with the manufacturer having the pricing decision and zero retail margins 84% of the times. In contrast, the retail pricing decision and zero wholesale margins hypothesis is only rejected 24% of the times. The hybrid model is rejected 81% of the times. Both collusion models are rejected more than 80% of the times, and this result can be explained maybe by the fact that there exists product differentiation in

[34] The purpose and interpretation of the supply parameters here is different from the Conduct Parameter (CP) Models where, for some values of the estimate of the CP, inferences (subject to Corts' (1999) cautions) are drawn about market power in a certain industry.

this market. This is because it is more difficult to coordinate on a price (see Scherer, 1980) and also to penalize (and hence to sustain collusion) in the context of differentiated products.[35] Finally, the efficient pricing model is only rejected 33% of the times. As a preliminary conclusion, there seems to be informal evidence that the contracting solution may follow something between model 2, case 1 and model 6, that is between zero manufacturer margin and retail pricing decisions and the fully efficient model. This would mean that not only retailers are deciding prices but the margins could be higher than the ones predicted by Nash-pricing behavior of the retailers. The formal ranking of the models follows in the next section.

Looking at each of the supply models separately, one can formally test if the data are consistent with each one. This is done by testing the restrictions imposed by each of the supply models using the GMM analog of the Likelihood ratio statistic (Newey and West, 1987), which is proportional to the minimized restricted GMM objective function minus the minimized unrestricted GMM objective function. The statistic is defined as a distance metric DM, which is distributed Chi-square with degrees of freedom equal to the number of restrictions. Estimates for the distance metric statistics and critical values are presented in the second and third columns of Table 14. A large DM leads to rejection of the null hypothesis, which indicates that the data do not seem to support the scenario in question. All models cannot be individually rejected.

5.4. Ranking of Supply Models: Non-nested Hypothesis Testing

This paper follows a "menu approach," as in Bresnahan (1987) or in Gasmi, Laffont and Vuong (1992), presenting different models of supply. To determine which model fits the data best, one needs to be able to make comparisons between the results from different models. This cannot be done in a nested way using the traditional testing methods, since most of the null hypotheses are not particular cases of some other null hypothesis. This means that the hypotheses are non-nested (Gourieroux and Monfort, 1994). When selecting the best among the seven scenarios, one can identify three particular cases of comparisons between each pair of competing models (following Vuong (1987)):

[35] Firms may in fact want to choose minimum differentiation to support collusion, as in Friedman and Thisse (1993).

	Model 2.1	Model 2.2	Model 3	Model 4	Model 5	Model 6
Model 1	AEV	SNN	SNN	OVL	OVL	SNN
Model 2.1	–	SNN	SNN	AEV	SNN	SNN
Model 2.2	–	–	SNN	SNN	SNN	SNN
Model 3	–	–	–	SNN	SNN	SNN
Model 4	–	–	–	–	SNN	SNN
Model 5	–	–	–	–	–	AEV

Classification of pair-wise comparisons, following Vuong (1989).

Model 1. Double Marginalization pricing; Model 2.1 Zero wholesale margin; Model 2.2 Zero retail margin; Model 3. Hybrid Model; Model 4. Wholesale collusion; Model 5. Retail collusion; Model 6. Monopolist.

First, there is the case where the two competing models are strictly non-nested (SNN in the above table). Second, there are cases of overlapping models, where there are common explanatory variables and different additional explanatory variables (OVL in the above table). Third, and finally, there are pairs of overlapping models with the only difference that one of the models has an additional set of explanatory variables (called AEV in the above table).

In the non-nested literature there are two general approaches to compare two competing models: either to assume that one of the models is true and to contrast it with other models (as in Cox, 1961) or to relax the assumption that one of the models contains the true law generating the observations and to verify if a particular null model can predict the relevant features of the competing alternative model, even if neither models is true (the encompassing tests, as in Mizou and Richard, 1986). The Cox-test generalizes the likelihood ratio procedure to non-nested hypotheses. The intuition behind the Cox (one sided) tests is as follows: A significantly positive test statistic leads to the rejection of the null model because the alternative model is performing too well for the null model to be regarded as true. One problem with Cox-type tests is that they are not symmetric. If a model H_g is rejected against model H_h it can also happen that H_h is rejected against H_g. In this paper, that means that when confronting seven alternatives one has to make the whole forty-two comparisons. Then, models that escape rejection against any alternative are the best models to be selected.

Related to the model comparison approaches described above are model selection tests (as in Vuong (1989) in the context of likelihood ratio tests). There the idea is to assume as null hypothesis that the two models are equivalent, and, given the estimate of the test statistic, one can conclude that either one of the two models is better (statistically dominating the other) or that one cannot discriminate between the two competing models given the data. The advantage of these model selection tests, in contrast to Cox-type tests, is that they are symmetric and directional. So, when confronting seven alternative models, only twenty-one comparisons have to be made.[36]

Smith (1992) extends the likelihood ratio based non-nested tests, presenting Cox-type and encompassing tests in the context of GMM, in particular for the cases of competing Instrumental Variables regression models with different regressors and differing instrument validity assumptions. The tests used in this paper are based on Smith (1992) and Vuong (1989) to confront the different supply models and to rank them.[37]

As in Smith (1992), the Cox-type statistic for the SNN case is constructed by examining the behavior under H_h of the difference of the estimated GMM criterion functions under H_h and the alternative H_g. It can be shown that this statistic is distributed asymptotically normal. Normalized, standardized and compared to a standard normal critical value, a large positive statistic in this one-sided test leads to a rejection of the null model H_h against H_g.

The testing procedure for the case of overlapping models (OVL and AEV) is more complicated than for the strictly non-nested cases. A two-step procedure (Vuong, 1989) has to be applied where in the first step one tests whether both models can or cannot be discriminated. If they can, one proceeds as in the SNN case. If they cannot, then the Cox-type test statistic no longer has a standard normal limiting distribution. It can be shown (Vuong, 1989) that it converges to a weighted sum of Chi-square distribution[38].

Table 15 presents the estimates for the test statistics for pair-wise comparisons of all models, given a random coefficients demand specification. In each row is the (null) model being tested and in each column is the alternative being used to test it. If the alternative model is performing too well, then the null model is rejected by a large and significant test statistic. After doing the pair-wise comparison, the model that assumes zero wholesale margin and in which retailers have pricing decisions provides the best fit. It outperforms other models at 5% significance since all elements of the column correspondent to this model (labeled 2.1) are larger than the critical value. This leads to a rejection of the null models in each row against model 2.1. The best model also escapes rejection against any alternative specified since all the elements in the row correspondent to model 2.1 are less than the critical value for all alternatives considered. The "second place" model seems to be model 6, the monopolist model. It is only rejected by the best model and by the model that assumes that there are no retail margins and that manufacturers are setting the prices.

[36] These directional and symmetric tests have been derived in the context of likelihood ratio tests. Extending them to GMM is the object of a companion project still in progress.

[37] For expressions of the test statistics, their asymptotic distributions and other details, please see Part III.

[38] See Dubin and Rivers (1989) for subroutines to compute the cumulative distribution of a weighted sum of Chi-square distribution.

For the Logit demand specification, the ranking in terms of the two best models is invariant. The ranking of the different models is also invariant for the additional demand specifications that are presented in Table 10, which sheds some evidence for the robustness of the results with respect to demand-side assumptions.

6. Discussion

The purpose of this paper is to present a method to analyze vertical contracts. Alternative models of competing manufacturers' and retailers' decision-making are used to determine whether contracting in the supermarket industry follows the double marginalization pricing model or whether more efficient contracting solutions are used. This paper extends the literature in analyzing vertical contracts in as much as it considers multiple retailers and does not require the need to observe data on wholesale prices. The approach, given demand estimates, is to compute price-cost margins for retailers and manufacturers implied by alternative vertical contracting models and to confront those with price-cost margins obtained from direct estimates of cost. In the more efficient contracts considered, via vertical integration, collusion or bargaining power, the double marginalization externality imposed by the retailers disappears. Consequently the sum of retailers' and manufacturers' profits may increase.[39]

For the market I study, the results rule out double marginalization. In particular, they suggest that, on the margin, manufacturers are pricing at marginal cost and that retail prices are the unconstrained profit maximizing prices. This result is consistent with several scenarios. For example, this result is consistent with non-linear pricing by manufacturers, via quantity discounts or two-part tariff contracts. In the optimal non-linear pricing contract, the manufacturer sets the marginal wholesale price close to the manufacturer's marginal cost for the retailer to have the right incentives when setting the retail prices. Then the manufacturer extracts revenue from the retailers via a fixed fee or by selling the non-marginal units at higher wholesale prices. The existence of quantity discounts is common practice in this industry while anecdotal evidence suggests that retail supermarkets do not often pay fixed fees to their manufacturers, and if they do, these fees are not close to the retail profits. Instead, there seem to be substantial fees paid by the manufacturers to the retailers (the so-called slotting allowances). The non-existence of (or the small) fixed fees from the retailers to the manufacturers could be explained by the fact that there are multiple manufacturers

[39] In certain cases, profits may decrease and the manufacturers may not choose the vertically integrated solution, as e.g. in Mc Guire and Staelin (1983) and Coughlan and Wernerfelt (1989).

in this market with whom the retailers can bargain more aggressively the fixed fee down, threatening to buy from another manufacturer. This result is also consistent with high bargaining power of the retailers that are able to force the wholesale prices down to marginal cost. The bottom line is that, without information on fixed fees, the above theoretical and anecdotal predictions cannot be tested and one cannot formally identify which interpretation of the results applies.

Another implication of the results relates to the pricing decision-makers in a particular industry. In the related literature, traditionally, the retailers' pricing decisions have been assumed away. For the market I study, this model is outperformed by the alternative model of retailers having the pricing decisions. Estimating the price-cost margins under the assumption that manufacturers are setting the prices and retailers are neutral pass-through intermediaries, when in fact retailers are deciding the prices, could lead to bias and affect the conclusions when accessing market power in a certain industry. Furthermore, the bias is expected to be more serious the more the sets of products that retailers sell and the sets of products that manufacturers sell do not coincide. More broadly, and since retailers may not be a neutral pass-through intermediaries, when analyzing price dynamics in the economy as a whole, retail behavior and retail market conditions should also be considered in addition to manufacturer behavior.

Future research considers the fact that looking at just one category may be restrictive since manufacturers, retailers and consumers make their pricing and purchase decisions in the context of multiple categories.[40] Given that consumers purchase a basket of goods during a shopping trip, a multiple category demand may be a more realistic framework to consider (see Part IV). In terms of pricing decisions, the fact that one manufacturer sells products in different product categories affects not only its pricing strategy but may possibly benefit its bargaining flexibility with the retailers. Also retailers use strategic category pricing to drive consumers into the store and increase sales.

Finally, and to motivate future empirical research on vertical contracts, I illustrate how the methodology proposed in this paper can be applied to address two questions. First, given the estimates of demand and a model of a pre- and post-vertical merger supply behavior, one can predict whether a potential vertical merger affects horizontal competition in the upstream and downstream markets involved.[41] Second, and related to pass-through effects of foreign trade policy, given the estimates of demand in a certain country for a particular good that involves a vertical trading supply model across different countries, one can analyze the effect of an increase of a

[40]For the retailers analyzed, yogurt sales represent on average only 2% of total retail sales in contrast to the two largest dollar sales categories: soft drinks 17 % and cereal 12 %.

[41]See Manuszak (2001) for an analysis on how upstream (horizontal) mergers affect market power in the retail gasoline markets.

tariff (depreciation of the exchange rate) on domestic or foreign margins. Trade policy makers are particularly interested in who absorbs most of the effects of a particular trade policy (foreign margins or domestic margins). That is in turn determined by the vertical relationships between domestic and foreign upstream or downstream firms.[42]

The question is under what conditions, given having only data on aggregate quantities and prices in a particular industry, can one estimate if the pricing behavior of manufacturers and retailers is consistent with a certain model of pricing behavior previously defined by the researcher. The special case when wholesale prices are not observed is considered. This part discusses the identification of vertical supply models when not only manufacturer pricing behavior but also retail pricing behavior in a certain industry is considered. In order to identify a vertical supply model one needs demand to be changing over time and the curvature of demand (that affects the derived demand of manufacturers) to be changing over time. It is therefore very important to have consistently estimated a very flexible demand model with very flexible second derivatives. Depending on the vertical supply model in question, manufacturers and retailers react differently to changes in the curvature of demand. If marginal cost is not varying with quantity, I can trace out whether the manufacturers (facing a shifting derived demand - or a changing curvature of demand) and the retailers (facing a shifting demand) behave according to Bertrand-Nash pricing behavior (the double marginalization model) or to another pricing bahavior model.

Villas-Boas (2001) selects the best among a menu of models of vertical contracts between manufacturers and retailers in a local (yogurt) market.[43] There, the fit of a finite set of models, that may be plausible explanations for the vertical pricing behavior in the market, is compared by constructing non-nested tests. Villas-Boas (2001) assumes that marginal cost is constant. A very flexible demand model is then estimated and it is assumed that demand is serially correlated. In this fashion, the demand retailers are facing and the derived demand manufacturers are facing are changing over time due to past period demand shocks. Given that supply does not seem to be significantly serially correlated, past periods price-cost margins, that are correlated with past period demand shocks, play the role of the exogenous demand and curvature shifters in identifying the vertical models.

[42] For example, if import prices do not rise as much as the dollar depreciation (i.e. the pass-through effect is less than one), then foreign profit margins are being diminished (see, for example, Feenstra (1989)).

[43] Gasmi, Laffont and Vuong (1992) use also a menu approach, where different models of conduct for soft-drink manufacturers are compared.

7. References

Berry, Steven T., 1994. "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, No. 2, pp.242-262.

Berry, S., J.Levinsohn and A.Pakes, 1995. "Automobile Prices in Market Equilibrium," *Econometrica*, 63, No. 4, pp.841-890.

Bresnahan, T., 1981. "Departures from Marginal-Cost Pricing in the American Automobile Industry," *Journal of Econometrics*, 17, pp.201-227.

Bresnahan, T. and P. C. Reiss, 1985. "Dealer and Manufacturer Margins," *RAND Journal of Economics*, 26, No. 2, pp.253-268.

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

Bresnahan, T., 1989. "Empirical Studies of Industries with Market Power," in Schmalansee, R., R. D. Willig eds., *Handbook of Industrial Organization, Volume II*, Amsterdam: North Holland, pp.1011-1057.

Cardell, N. Scott,1997. "Variance Components Structures for the Extreme Value and Logistic Distributions With Application to Models of Heterogeneity," *Econometric Theory*, 13 (2), pp.185-213.

Chevalier, J.A., A. K. Kashyap and P.E. Rossi. "Why don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data," *NBER Working Paper*, No. W7981.

Chintagunta, P., A. Bonfrer and I. Song, 2000. "Studying the Impact of Store Brand Entry: An application of a Random Coefficients Logit Model with Aggregate Data," *working paper* , University of Chicago.

Chu, W., 1992. "Demand Signaling and Screening in Channels of Distribution," *Marketing Science*, 11, No. 4, pp.327-347.

Corts, K. S., 1999. "Conduct Parameters and the Measurement of Market Power," *Journal of Econometrics*, 88, pp.227-250.

Corts, K. S., 2001. "The Strategic Effects of Vertical Market Structure: Common Agency and Divisionalization in the U. S. Motion Picture Industry," *Journal of Economics and Management Strategy*, 10, No. 4, pp.509-528.

Coughlan, A.T. and B. Wernerfelt, 1989. "On Credible Delegation by Oligopolists: A Discussion of Distribution Channel Management," *Management Science*, 35, No. 2, pp.226-239.

Cox, D.R., 1961. "Tests of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics*, 1, pp.105-123.

Dixit, A. and J. E. Stiglitz, 1977. "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, pp.297-308.

Draganska, M. and D. Jain, 2000. "Product-Line Length and Competitive Pricing," *working paper*, Northwestern University.

Dubé, J.P., 2001. "Multiple Discreteness and Product Differentiation: Strategy and Demand for Carbonated Softdrinks," *working paper*, University of Chicago.

Dubin, J. A. and D. McFadden, 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, No. 2, pp.345-362.

Dubin, J. and D. Rivers, 1986. "Statistical Software Tools," Pasadena: *California Institute of Technology*.

Feenstra, R.C., 1989. "Pass-through of tariffs and exchange rates," *Journal of International Economics*, 27, pp.25-45.

Friedman, J.W. and J-F Thisse, 1993. "Partial Collusion Fosters Minimum Product Differentiation," *RAND Journal of Economics*, 24, No. 4, pp.631-645.

Gasmi, F., J.J. Laffont and Q. Vuong, 1992. "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics and Management Strategy*, 1, pp.277-311.

Gourieroux, C. and A. Monfort, 1994. "Testing Non-nested Hypotheses," in R.F.Engle and D. L. McFadden, eds., *Handbook of Econometrics, Volume IV*, Amsterdam: North Holland, pp.2583-2637.

Greene, W.H., 1997. "Econometric Analysis," 3rd Edition, New Jersey, Prentice-Hall.

Hanemann, W. M., 1984. "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, No. 3, pp.541-561.

Hansen, L.P., J. Heaton and A. Yaron, 1996. "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14 (3), pp.262-280.

Hausman, J., 1978. "Specification Tests in Econometrics," *Econometrica*, 46, pp.1251-1271.

Hausman, J., 1983. "Specification and Estimation of Simultaneous Equations Models," in Z. Griliches and M. Intilligator, eds., *Handbook of Econometrics, Volume I*, Amsterdam: North-Holland, sezione 7.

Hausman, J., G. Leonard and D. McFadden, 1995. "A Utility Consistent, Combined Discrete Choice and Count Data Model," *Journal of Public Economics*, 56 (1), pp.1-30.

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.

Heckman, J.J., 1976. “The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models,” *Annals of Economic and Social Measurement*, 5, pp.475-492.

Hendel, I., 1999. “Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns,” *Review of Economic Studies*, 66, pp.423-446.

Kadiyali, V., P. Chintagunta and N. Vilcassim, 2000. “Manufacturer-Retailer Channel Interactions and Implications for Channel Power: An empirical investigation of pricing in a local market,” *Marketing Science*,v.19 n.2 , pp.127-148.

Kadiyali, V., N. Vilcassim and P. Chintagunta, 1999. “Product line extensions and competitive market interactions: An empirical Analysis,” *Journal of Econometrics*, 88, pp.339-363.

Katz, M., 1989. “Vertical Contractual Relations,” in Schmalansee, R., R. D. Willig eds., *Handbook of Industrial Organization, Volume I*, Amsterdam: North Holland, pp.655-721.

Manuszak, M. D., 2001. “The Impact of Upstream Mergers on Retail Gasoline Markets,” *working paper*, Carnegie Mellon University.

Mathewson, G. F. and R. A. Winter, 1984. “An Economic Theory of Vertical Restraints,” *Rand Journal of Economics*, 15, No. 1, pp.27-38.

McFadden, D., 1973. “Conditional Logit Analysis of Qualitative Choice Behavior,” *Frontiers of Econometrics*, P. Zarembka, eds., New York, Academic Press, pp.105-142.

McFadden, D., 1984. “Econometric Analysis of Qualitative Response Models,” in Z. Griliches and M. Intilligator, eds., *Handbook of Econometrics, Volume II*, Amsterdam: North-Holland, pp.1396-1456.

McFadden, D. and K. Train, 2000. "Mixed MNL Models of Discrete Response," *Journal of Applied Econometrics*, 15, No. 5, pp.447-470.

McGuire, T.W. and R. Staelin, 1983. "An Industry Analysis of Downstream Vertical Integration," *Marketing Science*, 2, No. 2, pp.161-191.

Messinger, P. R. and C. Narasimhan, 1995. " Has Power Shifted in the Grocery Channel?," *Marketing Science*, Vol. 14, No. 2, pp.189-223.

Mizon, G. and J-F. Richard, 1986. "The Encompassing Principle and its Application to Non-nested Hypotheses," *Econometrica*, 54, pp.657-678.

Narasimhan, C. and R.T. Wilcox, 1998. "Private Labels and the Channel Relationship: A Cross-Category Analysis," *Journal of Business*, 71, No. 4, pp.573-600.

Nevo, A. 1998. "Identification of the Oligopoly Solution Concept in a Differentiated-Products Industry", *Economics Letters*, 59, pp.391-395.

Nevo, A., 2000. " A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9, No. 4, pp.513-548.

Nevo, A., 2001. " Measuring Market Power in the Ready-To-Eat Cereal Industry ," *Econometrica*, 69, No. 2, pp.307-342.

Newey W. and D. McFadden, 1994."Large Sample Estimation and Hypothesis Testing," in R.F. Engle and D. McFadden, eds., *Handbook of Econometrics, Volume IV*,Amsterdam: North-Holland, pp.2211-2245.

Newey W. and K. West, 1987. " Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, pp.777-787.

Rey, P. and J. Tirole, 1986. "The Logic of Vertical Restraints," *American Economic Review*, 76, pp.921-939.

Scherer, F., 1980. "Industrial Market Structure and Economic Performance," second edition. Chicago: Rand-McNally.

Schmalensee, R., 1981. "Monopolistic Two-Part Pricing Arrangements," *RAND Journal of Economics*, 12, No. 2, pp.445-466.

Shaffer, G., 1991. "Slotting Allowances and Resale Price Maintenance: A Comparison of Facilitating Practices," *Rand Journal of Economics*, 22, No. 1, pp.120-135.

Shaffer, G. and D. P. O'Brien, 1997. "Nonlinear Supply Contracts, Exclusive Dealing, and Equilibrium Market Foreclosure," *Journal of Economics & Management Strategy*, 6, pp.755-785.

Smith, R. J., 1992. "Non-nested Tests for Competing Models Estimated by Generalized Method of Moments," *Econometrica*, 60, No. 4, pp.973-980.

Sudhir, K., 2001. "Structural Analysis of Manufacturer Pricing in the Presence of a Strategic Retailer,"*Marketing Science*, 20, No. 3, pp.244-264.

Tirole, J., 1988. *The Theory of Industrial Organization*, Cambridge: The MIT Press.

Villas-Boas, J.M. and Y. Zao, 2000. "The Ketchup Marketplace: Retailer, Manufacturers and Individual Consumers," *working paper*, University of California, Berkeley.

Vuong, Q.H., 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, No. 2, pp.307-333.

White, H. 1980. "A Heteroskedasticity-consistent covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, No. 4, pp.817-838.

8. Appendix 1: Solution for a Simple 2 by 2 by 2 Model for Supply

This appendix presents the expressions for the retailers' and manufacturers' price-cost margins in a "simple" model for the supply side with two manufacturers who sell two products to two retailers. This is done given a Logit demand model and then given the random coefficients model for demand. In particular, I focus on the supply model of double marginalization since the price-cost margins for the other supply scenarios considered in this paper can be derived from the formulae of the double marginalization price-cost margins by changing the ownership matrices accordingly. Expressions for the crucial matrices for obtaining the retailers' and the manufacturers' price-cost margins are provided. These crucial matrices are:

1. The retail and manufacturer ownership matrices (T_r and T_w), with general elements $T_r(j,k) = 1$ if both products j and k are sold at the same retailer (and zero otherwise) and $T_w(j,k) = 1$ if both products j and k are produced by the same manufacturer (and zero otherwise).

2. The retailer's response matrix Δ_r containing the derivatives of the shares with respect to all retail prices.

3. The manufacturer's response matrix Δ_w containing the derivatives of the shares with respect to all wholesale prices.

The third matrix is very complicated to obtain, while the others are very straightforward. A procedure to obtain Δ_w is described and intermediary expressions handy for computer coding are provided.

The rest of this appendix is organized as follows. The ownership matrices are derived first and then the expressions for the price-cost margins in the double marginalization model are obtained. A short illustration is provided next on how all the price-cost margins in the other supply models considered are derived. Finally, expressions for the response matrices Δ_r and Δ_w are given.

8.1. The Ownership Matrices

The case of two manufacturers a and b producing one good each, which they sell to two retailers 1 and 2 is considered. Furthermore, and for the purpose of illustrating the price-cost margins for scenario 3, it is assumed that the good produced by manufacturer a is a private label of retailer 1. This implies, given the definition of products in this paper, that there are three products in this model. Let product a_1 be the private label produced by manufacturer a and sold to retailer 1. Product b_1 is produced by manufacturer b and sold to retailer 1 and product b_2 is produced by manufacturer b and sold to retailer 2. This implies that the retailer's (T_r) and the wholesaler's (T_w)

ownership matrices are given by

$$T_r = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } T_w = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

8.2. Double Marginalization Model

In this model the manufacturers set their prices first and then the retailers follow. Each one of the two retailers maximizes his profit function given by

$$\pi_{rt} = \sum_{j \epsilon S_{rt}} \left[p_{jt} - p_{jt}^w - c_{jt}^r \right] s_{jt}(p) \quad \text{for} \quad r = 1, 2 \tag{39}$$

where S_{rt} is the set of products sold by retailer r in week t, p_{jt} is the retail price, p_{jt}^w is the wholesale price he pays, c_j^r is the retailer's marginal cost and $s_{jt}(p)$ is the product's market share. The first order conditions, assuming a pure strategy Nash-equilibrium in prices, are

$$s_{jt} + \sum_{m=a_1,b_1,b_2} T_r(m,j) \left[p_{mt} - p_{mt}^w - c_{mt}^r \right] \frac{\partial s_{mt}}{\partial p_{jt}} = 0 \quad \text{for} \quad j = a_1, b_1, b_2 \tag{40}$$

or in matrix notation

$$\begin{bmatrix} \frac{\partial s_{a1}}{\partial p_{a1}} & \frac{\partial s_{b1}}{\partial p_{a1}} & 0 \\ \frac{\partial s_{a1}}{\partial p_{b1}} & \frac{\partial s_{b1}}{\partial p_{b1}} & 0 \\ 0 & 0 & \frac{\partial s_{b2}}{\partial p_{b2}} \end{bmatrix} \begin{bmatrix} p_{a1} - p_{a1}^w - c_{a1}^r \\ p_{b1} - p_{b1}^w - c_{b1}^r \\ p_{b2} - p_{b2}^w - c_{b2}^r \end{bmatrix} = - \begin{bmatrix} s_{a1} \\ s_{b1} \\ s_{b2} \end{bmatrix}.$$

Define $[A * B]$ as the element by element multiplication of the two matrices of the same dimensions A and B. Then using the ownership matrix notation and solving for the price-cost margins yields

$$\begin{bmatrix} p_{a1} - p_{a1}^w - c_{a1}^r \\ p_{b1} - p_{b1}^w - c_{b1}^r \\ p_{b2} - p_{b2}^w - c_{b2}^r \end{bmatrix} = - \left[\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \frac{\partial s_{a1}}{\partial p_{a1}} & \frac{\partial s_{b1}}{\partial p_{a1}} & \frac{\partial s_{b2}}{\partial p_{a1}} \\ \frac{\partial s_{a1}}{\partial p_{b1}} & \frac{\partial s_{b1}}{\partial p_{b1}} & \frac{\partial s_{b2}}{\partial p_{b1}} \\ \frac{\partial s_{a1}}{\partial p_{b2}} & \frac{\partial s_{b1}}{\partial p_{b2}} & \frac{\partial s_{b2}}{\partial p_{b2}} \end{bmatrix} \right]^{-1} \begin{bmatrix} s_{a1} \\ s_{b1} \\ s_{b2} \end{bmatrix}. \tag{41}$$

Writing the price-cost margins for all products in vector notation gives the implied price-cost margins for the retailers

$$p - p^w - c^r = -[T_r * \Delta_r]^{-1} s(p), \tag{42}$$

which is a system of three implicit functions that expresses the three retail prices as a function of the wholesale prices.

Looking now at the manufacturers, they maximize their profits choosing the wholesale prices p^w knowing that the retailers behave according to (42). The first-order conditions are, assuming again pure strategy Nash-Equilibrium in the wholesale prices,

$$\begin{bmatrix} \frac{\partial s_{a1}}{\partial p^w_{a1}} & 0 & 0 \\ 0 & \frac{\partial s_{b1}}{\partial p^w_{b1}} & \frac{\partial s_{b2}}{\partial p^w_{b1}} \\ 0 & \frac{\partial s_{b1}}{\partial p^w_{b2}} & \frac{\partial s_{b2}}{\partial p^w_{b2}} \end{bmatrix} \begin{bmatrix} p^w_{a1} - c^w_{a1} \\ p^w_{b1} - c^w_{b1} \\ p^w_{b2} - c^w_{b2} \end{bmatrix} = - \begin{bmatrix} s_{a1} \\ s_{b1} \\ s_{b2} \end{bmatrix}. \tag{43}$$

Solving for the price-cost margins of the manufacturers and using the ownership matrix notation yields

$$\begin{bmatrix} p^w_{a1} - c^w_{a1} \\ p^w_{b1} - c^w_{b1} \\ p^w_{b2} - c^w_{b2} \end{bmatrix} = - \left[\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{\partial s_{a1}}{\partial p^w_{a1}} & \frac{\partial s_{b1}}{\partial p^w_{a1}} & \frac{\partial s_{b2}}{\partial p^w_{a1}} \\ \frac{\partial s_{a1}}{\partial p^w_{b1}} & \frac{\partial s_{b1}}{\partial p^w_{b1}} & \frac{\partial s_{b2}}{\partial p^w_{b1}} \\ \frac{\partial s_{a1}}{\partial p^w_{b2}} & \frac{\partial s_{b1}}{\partial p^w_{b2}} & \frac{\partial s_{b2}}{\partial p^w_{b2}} \end{bmatrix} \right]^{-1} \begin{bmatrix} s_{a1} \\ s_{b1} \\ s_{b2} \end{bmatrix} \tag{44}$$

or short

$$(p^w_t - c^w_t) = -[T_w * \Delta_w]^{-1} s_t(p), \tag{45}$$

where c^w is the manufacturer's marginal cost.

8.3. The Other Supply Models

The price-cost margins of the other supply models considered can be obtained from equations (41) and (44) by changing the ownership matrices accordingly. First, for the model that assumes zero manufacturer margins and retailers having the pricing decisions (Scenario 2 Case 1), the price-cost margins for the manufacturers are zero for all products, since $p^w = c^w$ and the price-cost margins for the retailers are given by (41).

Second, for the model that assumes zero retail margin and in which the manufacturers have the pricing decision (Scenario 2 Case 2), the price-cost margins for the retailers are zero for all products since $p = c^r + p^w$. The price-cost margins for the manufacturers are given by (41) with the only change that the ownership matrix in the expression is now T_w and not T_r.

Third, for the Hybrid model (Scenario 3), retailer 1 is vertically integrated with respect to its own private label a_1, therefore $p^w_{a1} = c^w_{a1}$. Compared to scenario 1, this has no impact in the price-cost margins of the competing retailer 2 and of the other brands sold by retailer 1. This is due to the fact that the private label brand a_1 is not sold at the other retailer. Therefore the retail price-cost margins are given again by (41). On the contrary, by vertically integrating into the upstream industry, the retailer affects the price-cost margins of the other manufacturers since p^w_{a1} is not optimized over. Price-cost margins for the manufacturers are obtained from (44) by ignoring all rows and columns that correspond to the first product a_1. This yields

$$\begin{bmatrix} p^w_{b1} - c^w_{b1} \\ p^w_{b2} - c^w_{b2} \end{bmatrix} = - \left[\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{\partial s_{b1}}{\partial p^w_{b1}} & \frac{\partial s_{b1}}{\partial p^w_{b2}} \\ \frac{\partial s_{b1}}{\partial p^w_{b2}} & \frac{\partial s_{b2}}{\partial p^w_{b2}} \end{bmatrix} \right]^{-1} \begin{bmatrix} s_{b1} \\ s_{b2} \end{bmatrix}.$$

In terms of the economic intuition, by vertically integrating into the upstream market the retailer eliminated the double margin in its own product. So manufacturers of other products see the final retail price of product $a1$ fall. Accordingly, demand for the products that they produce changes and they need to adjust their own wholesale prices, since the ones that they were charging before are no longer optimal.

Fourth, in the manufacturer collusion model (Scenario 4), if manufacturers are colluding, the relevant ownership matrix for the manufacturers is T_1, an ownership matrix full of ones. It is as if a single manufacturer was deciding the wholesale prices for all the products. For the retailers, the relevant ownership matrix is still T_r. The price-cost margins for retailers are the same as in scenario 1, given by (41). The price-cost margins for the manufacturers are given by (44) but substituting T_w by T_1, that is,

$$\begin{bmatrix} p^w_{a1} - c^w_{a1} \\ p^w_{b1} - c^w_{b1} \\ p^w_{b2} - c^w_{b2} \end{bmatrix} = - \left[\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} \frac{\partial s_{a1}}{\partial p^w_{a1}} & \frac{\partial s_{b1}}{\partial p^w_{a1}} & \frac{\partial s_{b2}}{\partial p^w_{a1}} \\ \frac{\partial s_{a1}}{\partial p^w_{b1}} & \frac{\partial s_{b1}}{\partial p^w_{b1}} & \frac{\partial s_{b2}}{\partial p^w_{b1}} \\ \frac{\partial s_{a1}}{\partial p^w_{b2}} & \frac{\partial s_{b1}}{\partial p^w_{b2}} & \frac{\partial s_{b2}}{\partial p^w_{b2}} \end{bmatrix} \right]^{-1} \begin{bmatrix} s_{a1} \\ s_{b1} \\ s_{b2} \end{bmatrix}.$$

Fifth, in the retail collusion model (Scenario 5), the price-cost margins for the retailers are given by (41) changing the ownership matrix to T_1 since they are the ones colluding now. The price-cost margins for manufacturers are the same as in scenario 1, given by equation (44).

Sixth and finally, in the monopolist model (Scenario 6), the relevant ownership matrix is T_1, a matrix full of ones, since joint profit maximization is done. The sum of retailers' and wholesalers' price-cost margins are given by (41) with the unitary ownership matrix instead of T_r.

8.4. The Response Matrix Δ_r

Δ_r is a matrix with derivatives of all shares with respect to all retail prices, with general element $\Delta_r(j,k) = \frac{\delta s_k}{\delta p_j}$. For the Logit model, its expression is straightforward:

$$\Delta_r = \begin{bmatrix} \alpha s_{a1}(1-s_{a1}) & -\alpha s_{a1}s_{b1} & -\alpha s_{a1}s_{b2} \\ -\alpha s_{b1}s_{a1} & \alpha s_{b1}(1-s_{b1}) & -\alpha s_{b1}s_{b2} \\ -\alpha s_{b2}s_{a1} & -\alpha s_{b2}s_{b1} & \alpha(1-s_{b2})s_{b2} \end{bmatrix}.$$

For the random coefficients model, Δ_r is computed by simulation. Given N_s random draws of unobserved and observed consumer characteristics,

$$\Delta_r = \begin{bmatrix} \sum_{n=1}^{N_s} \alpha_n s_{a1,n}(1-s_{a1,n}) & -\sum_{n=1}^{N_s} \alpha_n s_{a1,n}s_{b1,n} & -\sum_{n=1}^{N_s} \alpha_n s_{a1,n}s_{b2,n} \\ -\sum_{n=1}^{N_s} \alpha_n s_{b1,n}s_{a1,n} & \sum_{n=1}^{N_s} \alpha_n s_{b1,n}(1-s_{b1,n}) & -\sum_{n=1}^{N_s} \alpha_n s_{b1,n}s_{b2,n} \\ -\sum_{n=1}^{N_s} \alpha_n s_{b2,n}s_{a1,n} & -\sum_{n=1}^{N_s} \alpha_n s_{b2,n}s_{b1,n} & \sum_{n=1}^{N_s} \alpha_n (1-s_{b2,n})s_{b2,n} \end{bmatrix},$$

The term α_n is the marginal utility of price for a certain consumer n and is given by

$$\alpha_n = \alpha + \Gamma_\alpha D_n + \Upsilon_\alpha v_n$$

where α is the mean across consumers of the marginal utility of price, D_n are the observed characteristics of consumer n, v_n are the unobserved characteristics of consumer n, Γ_α is the first column of Γ and Υ_α is the first element of Υ. See also equations (2) and (4) in the paper for the heterogeneity specification.

8.5. The Response Matrix Δ_w

This matrix Δ_w contains the derivatives of all shares with respect to all wholesale prices and has the general element $\Delta_w(j,k) = \frac{\delta s_k}{\delta p_j^w}$. An expression for this matrix is very complicated even in the Logit case. After noting that $\Delta_w = \Delta_p' \Delta_r$, where Δ_p is a matrix of derivatives of all the retail prices with respect to all the wholesale prices, one needs only to find an expression for Δ_p to get Δ_w. Let us start by totally differentiating, for a given j, equation (40) with respect to all prices $(dp_{a1}, dp_{b1}, dp_{b2})$ and a wholesale price p_f^w, with variation dp_f^w:

$$\sum_{k=a_1,b_1,b_2} \underbrace{[\frac{\partial s_j}{\partial p_k} + \sum_{i=a_1,b_1,b_2} (T_r(i,j)\frac{\partial^2 s_i}{\partial p_j \partial p_k}(p_i - p_i^w - c_i^r)) + T_r(k,j)\frac{\partial s_k}{\partial p_j}]}_{g(j,k)} \; dp_k \; - \; \underbrace{T_r(f,j)\frac{\partial s_f}{\partial p_j}}_{h(j,f)} dp_f^w = 0. \qquad (46)$$

Putting all j products together, let G be the matrix with general element $g(j,k)$ and let H_f be the J dimensional vector with general element $h(j,f)$. Then

$$G \; dp - H_f \; dp_f^w = 0. \qquad (47)$$

Solving for the derivatives of all prices with respect to the wholesale price f the f-th column of Δ_p is obtained:

$$\frac{dp}{dp_f^w} = G^{-1} H_f. \tag{48}$$

Stacking all N columns together, $\Delta_p = G^{-1}H$, which has the derivatives of all prices with respect to all wholesale prices, with general element $(i, j) = \frac{\partial p_j}{\partial p_i^w}$.

The only expression in (46) that needs to be obtained is the (3 by 3 by 3) Hessian, a matrix of second derivatives of the shares with respect to all retail prices. For the Logit case the Hessian is given by

$$\frac{\partial^2 s_{it}}{\partial p_{jt} \partial p_{kt}} = \begin{cases} \alpha^2(1-2s_{it})s_{it}(1-s_{it}) & \text{if } i=j=k, \\ -2\alpha^2 s_{it}s_{jt}s_{kt} & \text{if } i \neq j \text{ and } j \neq k \text{ and } i \neq k, \\ \alpha^2 s_{it}s_{jt}(2s_{jt}-1) & \text{if } i \neq j \text{ and } j = k, \\ \alpha^2 s_{it}s_{kt}(2s_{it}-1) & \text{if } i \neq k \text{ and } i = j, \\ \alpha^2 s_{it}s_{jt}(2s_{it}-1) & \text{otherwise.} \end{cases} \tag{49}$$

The Hessian for the random coefficients models is obtained by simulation. Given N_s random draws of unobserved and observed consumer characteristics, the Hessian is computed as

$$\frac{\partial^2 s_{it}}{\partial p_{jt} \partial p_{kt}} = \begin{cases} \sum_{n=1}^{N_s} \alpha_n^2(1-2s_{it,n})s_{it,n}(1-s_{it,n}) & \text{if } i=j=k, \\ -2\sum_{n=1}^{N_s} \alpha_n^2 s_{it,n}s_{jt,n}s_{kt,n} & \text{if } i \neq j \text{ and } j \neq k \text{ and } i \neq k, \\ \sum_{n=1}^{N_s} \alpha_n^2 s_{it,n}s_{jt,n}(2s_{jt,n}-1) & \text{if } i \neq j \text{ and } j = k, \\ \sum_{n=1}^{N_s} \alpha_n^2 s_{it,n}s_{kt,n}(2s_{it,n}-1) & \text{if } i \neq k \text{ and } i = j, \\ \sum_{n=1}^{N_s} \alpha_n^2 s_{it,n}s_{jt,n}(2s_{it,n}-1) & \text{otherwise.} \end{cases}$$

9. Appendix 2: Ranking of Supply Models: Non-nested Hypothesis Testing

The tests presented here have the purpose of comparing different supply models and ranking them. For the strictly non-nested comparisons, I derive (based on Smith, 1992) Cox-type tests to examine the behavior of the difference of the GMM criterion functions for two competing models under one of the competing hypotheses. The testing procedure for the case of overlapping models is more complicated than for the strictly non-nested cases. A two-step procedure as proposed in Vuong (1989) is described here.

The basic framework is as follows. Consider two competing regression models

$$
\begin{aligned}
M_g: \quad & y = X_g\beta + u_g \\
M_h: \quad & y = X_h\gamma + u_h
\end{aligned}
$$

where, in the model, y is a vector of the observations on prices for all products during all weeks (the dependent variable), X_g and X_h are (T by n_g) and (T by n_h) matrices of the supply equation regressor variables, respectively, and β and γ are parameters to be estimated. In particular X_g (resp. X_h) includes input prices and price-cost margins implied by model M_g (resp. M_h) for all retailers and all manufacturers. Since the price-cost margins in X_g (resp. X_h) are correlated with the unobserved determinants of price u_g (resp. u_h) there are $k_g > n_g$ (resp. $k_h > n_h$) instrumental variables Z_g (resp. Z_h) used to estimate the parameters. Thus, $E[T^{-1}Z_g'u_g] = 0$ (resp. $E[T^{-1}Z_h'u_h] = 0$). Let me define the two sample moments $g_T(\beta) \equiv T^{-1}Z_g'u_g$ and $h_T(\gamma) \equiv T^{-1}Z_h'u_h$ and the asymptotic variances $V_g \equiv \lim_{T\to\infty} \mathrm{Var}_h[T^{1/2}g_T(\beta_0)]$ and $V_h \equiv \lim_{T\to\infty} \mathrm{Var}_g[T^{1/2}h_T(\gamma_0)]$ where $Var_j(.)$ means variance taken with respect to the density under M_j for $j = g, h$. Let $\hat{V}_g$ (resp. $\hat{V}_h$) be consistent estimators of the asymptotic variances of the two sample moments. The optimal GMM estimator for β, $\hat{\beta}_T$, minimizes $g_T(\beta)'\hat{V}_g^{-1}g_T(\beta)$ and satisfies $\hat{G}_T'\hat{V}_g^{-1}\hat{g}_T = 0$ where $\hat{g}_T \equiv g_T(\hat{\beta}_T)$ and $\hat{G}_T \equiv \nabla_\beta' g_T(\hat{\beta}_T)$. Similarly, the optimal GMM estimator $\hat{\gamma}$ minimizes $h_T(\gamma)'\hat{V}_h^{-1}h_T(\gamma)$ and satisfies $\hat{H}_T'\hat{V}_h^{-1}\hat{h}_T = 0$ where $\hat{h}_T \equiv h_T(\hat{\gamma}_T)$ and $\hat{H}_T \equiv \nabla_\gamma' h_T(\hat{\gamma}_T)$.

Let us assume from now on that the null hypothesis is H_h and the alternative hypothesis is H_g defined as

$$
\begin{aligned}
H_g: \quad & y = X_g\beta + u_g \quad \text{and} \quad \beta_m^0 = 1 \\
H_h: \quad & y = X_h\gamma + u_h \quad \text{and} \quad \gamma_m^0 = 1
\end{aligned}
$$

where $\gamma = (\gamma_m' \mid \gamma_c')'$ and $\beta = (\beta_m' \mid \beta_c')'$. In particular, γ_c (resp. β_c) are parameters associated with cost side variables and γ_m (resp. β_m) are vectors of parameters associated with the price-cost margins implied by model M_g (resp. M_h). Under H_h

$$
T^{1/2}h_T(\gamma^0) \xrightarrow{D} N_{k_h}(0, V_h). \tag{50}
$$

Consequently,

$$
T^{1/2}\hat{h}_T = M_h\ T^{1/2}h_T(\gamma^0) + o_P(1), \tag{51}
$$

where M_h is defined as usual $M_h \equiv I_{k_h} - H(H'V_h^{-1}H)^{-1}H'V_h^{-1}$, $H \equiv E_h[\nabla_\gamma' h_T(\gamma^0)]$ is the probability limit taken under H_h of $\hat{H}_T$ previously defined.

For the strictly non-nested (SNN) comparisons, I use Cox-type tests to examine the behavior of the difference of the GMM criterion functions for two competing models under one of the competing hypotheses. As in Smith (1992), the Cox-type statistic is constructed by examining the behavior under H_h of the difference of the estimated GMM criterion functions under H_h and the alternative H_g

$$T^{1/2}[\hat{g}_T'\hat{V}_g^{-1}\hat{g}_T - \hat{h}_T'\hat{V}_h^{-1}\hat{h}_T]. \tag{52}$$

The plim of (52) under H_h simplifies to

$$2\ \hat{g}_T'\hat{V}_g^{-1}T^{1/2}[\hat{g}_T - \text{plim}_(\hat{g}_T)] + o_p(1). \tag{53}$$

I need a consistent estimator for $\text{plim}_h(\hat{g})$ that I define as

$$\text{plim}_(\hat{g}_T) = \hat{g}_T - \hat{A}_T\hat{h}_T \tag{54}$$

where $\hat{A}_T$ is a (k_g by k_h) matrix such that $\text{plim}(\hat{A}_T) = A_h$ is non-null. The idea behind choosing a certain $\hat{A}_T$ is to construct a common statistical framework. I use $\hat{A}_T = Z_g'Z_h(Z_h'Z_h)^{-1}$, but other A_T could have been used (see Smith, 1992). Finally the Cox-type GMM statistic is defined as

$$\hat{g}_T'\hat{V}_g^{-1}\hat{A}_T T^{1/2}\hat{h}_T \tag{55}$$

and is distributed asymptotically normal with variance w^2. The asymptotic variance w^2 is estimated by

$$\hat{w}^2 = [\hat{g}_T - \hat{A}_T\hat{h}_T]'\hat{V}_g^{-1}\hat{A}_T\hat{M}_h\hat{V}_h\hat{M}_h'\hat{A}_T'\hat{V}_g^{-1}[\hat{g}_T - \hat{A}_T\hat{h}_T] \tag{56}$$

where $\hat{M}_h = I_{k_h} - \hat{H}(\hat{H}'\hat{V}_h^{-1}\hat{H})^{-1}\hat{H}'\hat{V}_h^{-1}$ and $\hat{H}_T = Z_h'X_h$. Normalized, standardized and compared to a standard normal critical value, a large positive statistic in this one-sided test leads to a rejection of the null model H_h against H_g.

The testing procedure for the case of overlapping models (OVL and AEV) is more complicated than for the strictly non-nested cases. A two-step procedure (Vuong, 1989) has to be applied where in the first step one tests whether both models can or cannot be discriminated. This first step is called the variance test for discrimination. The null hypothesis that the models cannot be discriminated is equivalent to the null hypothesis that the variance ω^2 of the Cox-type statistic (which is the difference of the GMM criterion functions under H_h) is equal to zero. The variance statistic given $\omega^2 = 0$ is distributed weighted Chi-square or central Chi-square, depending on whether the "information matrix equivalencies"(see name for GMM context) hold. If they do then the limiting distribution is a central Chi-square with degrees of freedom equal to the sum of number of moments in H_h, say k_h, and H_g, say k_g, minus twice the rank of the k_h by k_g matrix of expectation of the product of the first derivatives of the GMM criterion functions of the two competing models, GMM_h and GMM_g: $N\omega^2 \xrightarrow{D} \chi^2_{(k_h+k_g-2rank(B))}$ where $B = E[\frac{\partial GMM_h}{\partial \gamma}.\frac{\partial GMM_g}{\partial \beta'}]$. Note that in the nested context where H_h is the larger model and H_g is the restrictive model, if the larger model is correctly specified then the limiting distribution is $\chi^2_{(k_h-k_g)}$. If in the first step the models can be discriminated, I proceed as in the SNN case. If they can't, then the Cox-type test statistic has a weighted sum of Chi-square distribution.

10. Appendix 3: Tables and Figures

Description	Mean	Median	Standard Deviation	Max	Min	Brand Variation	Week Variation
Prices (cents per serving)	49	48	9.2	72	24	68.3%	2.4%
Feature (=1 if featured)	0.03	0	0.15	1	0	10.8%	5.3%
Servings sold (1 serving=6 ounces)	246	132	393.3	9538	1	43.6%	4.1%
Share of product within market (%)	0.8	0.4	1.3	32	0.03	43.6%	4.1%
Combined Shares of products (%)	34	37	12.7	75	12		
Combined Shares by Manufacturer (%)	Mean	Median	Standard Deviation	Max	Min		
Dannon	16.8	16.4	7.6	50.0	4.7		
General Mills	8.8	9.0	3.6	31.1	4		
Private Label of Retailer 2	4.1	3.3	4.2	38.5	0.6		
Kraft	3.4	3.1	1.6	13.6	1.1		
Private Label of Retailer 3	1.3	1.2	0.5	3.7	0.6		
Combined Shares by Retailer (%)	Mean	Median	Standard Deviation	Max	Min		
Retailer 1	2.3	2.3	1.0	9.2	1		
Retailer 2	19.8	20.5	9.2	57.6	1.2		
Retailer 3	13.6	13.5	3.4	24.3	6.7		

Table 1: Prices, Feature, Servings Sold and Market Shares of Products in Sample: Summary Statistics.
Source: IRI.

Product ID #	Manufacturer	Retailer	Product Name	Price Mean	Price Std
1	Kraft	1	Breyer Light Fruit Yogurt	38.94	4.73
2	Dannon	2	Dannon Light Vanilla Yogurt	47.62	3.48
3	Dannon	3	Dannon Light Vanilla Yogurt	42.06	3.07
4	Dannon	1	Dannon Lowfat Plain Yogurt	52.56	3.97
5	Dannon	2	Dannon Lowfat Plain Yogurt	48.19	4.75
6	Dannon	3	Dannon Lowfat Plain Yogurt	46.90	2.48
7	Dannon	1	Dannon Light Fruit Yogurt	57.87	5.01
8	Dannon	2	Dannon Light Fruit Yogurt	54.69	5.09
9	Dannon	3	Dannon Light Fruit Yogurt	47.08	2.33
10	Dannon	2	Dannon Nonfat Plain Yogurt	48.69	4.54
11	Dannon	3	Dannon Nonfat Plain Yogurt	46.56	2.58
12	Dannon	1	Dannon Classic Flavor Fruit Yogurt	52.50	5.53
13	Dannon	2	Dannon Classic Flavor Fruit Yogurt	53.68	7.55
14	Dannon	3	Dannon Classic Flavor Fruit Yogurt	46.96	3.29
15	Dannon	1	Dannon Classic Flavor Vanilla Yogurt	53.31	3.27
16	Dannon	2	Dannon Classic Flavor Vanilla Yogurt	48.82	4.68
17	Dannon	3	Dannon Classic Flavor Vanilla Yogurt	46.38	3.04
18	Dannon	1	Dannon Fruit on the Bottom Yogurt	51.12	6.48
19	Dannon	2	Dannon Fruit on the Bottom Yogurt	53.18	6.47
20	Dannon	3	Dannon Fruit on the Bottom Yogurt	47.31	2.41
21	Store 2	2	Private Label 2 Lowfat Fruit Yogurt	52.17	7.43
22	Store 2	2	Private Label 2 Lowfat Plain Yogurt	30.76	2.00
23	Store 2	2	Private Label 2 Lowfat Vanilla Yogurt	30.13	0.87
24	Store 2	2	Private Label 2 Nonfat Fruit Yogurt	54.63	7.29
25	Store 2	2	Private Label 2 Nonfat Plain Yogurt	54.82	7.35
26	Store 3	3	Private Label 3 Lowfat Fruit Yogurt	35.83	1.01
27	Store 3	3	Private Label 3 Lowfat Plain Yogurt	30.52	2.07
28	Kraft	1	Light N'Lively Nonfat Fruit Yogurt	48.40	4.52
29	Kraft	2	Light N'Lively Nonfat Fruit Yogurt	46.93	4.71
30	Kraft	3	Light N'Lively Nonfat Fruit Yogurt	46.44	3.24
31	Kraft	1	Light N'Lively Lowfat Fruit Yogurt	49.38	4.28
32	Kraft	2	Light N'Lively Lowfat Fruit Yogurt	46.67	5.04
33	Kraft	3	Light N'Lively Lowfat Fruit Yogurt	45.23	4.26
34	General Mills	2	Yoplait Custard Style Lowfat Fruit Yogurt	60.69	5.86
35	General Mills	3	Yoplait Custard Style Lowfat Fruit Yogurt	57.52	4.77
36	General Mills	2	Yoplait Custard Style Lowfat Vanilla Yogurt	63.54	6.58
37	General Mills	3	Yoplait Custard Style Lowfat Vanilla Yogurt	57.06	5.48
38	General Mills	1	Yoplait Fruit Yogurt	57.69	9.47
39	General Mills	2	Yoplait Fruit Yogurt	58.67	4.73
40	General Mills	3	Yoplait Fruit Yogurt	52.62	4.67
41	General Mills	1	Yoplait Light Fruit Yogurt	52.10	10.65
42	General Mills	2	Yoplait Light Fruit Yogurt	56.21	5.61
43	General Mills	3	Yoplait Light Fruit Yogurt	49.15	4.10

Table 2: Information about the 43 Products in Sample - Prices.

Price in cents per serving. One serving is equivalent to 6 ounces of yogurt. Product ID #: First column has the product identification number. Source: IRI.

Product ID #	PCM$^{(*)}$ S-1,R	PCM$^{(*)}$ S-1,W	PCM$^{(*)}$ S-2,C 1	PCM$^{(*)}$ S-2,C 2	PCM$^{(*)}$ S-3,R	PCM$^{(*)}$ S-3,W	PCM$^{(*)}$ S-4,R	PCM$^{(*)}$ S-4,W	PCM$^{(*)}$ S-5,R	PCM$^{(*)}$ S-5,W	PCM$^{(*)}$ S-6
1	33%	33%	33%	33%	33%	33%	33%	41%	39%	34%	39%
2	29%	30%	29%	28%	29%	30%	29%	36%	31%	32%	31%
3	32%	34%	32%	32%	32%	33%	32%	39%	35%	36%	35%
4	24%	26%	24%	26%	24%	26%	24%	30%	28%	29%	28%
5	29%	30%	29%	28%	29%	29%	29%	35%	31%	32%	31%
6	28%	30%	28%	29%	28%	30%	28%	35%	31%	32%	31%
7	22%	24%	22%	23%	22%	24%	22%	27%	26%	26%	26%
8	25%	26%	25%	25%	25%	26%	25%	31%	27%	28%	27%
9	28%	30%	28%	29%	28%	30%	28%	35%	31%	32%	31%
10	28%	30%	28%	28%	28%	29%	28%	35%	31%	31%	31%
11	29%	30%	29%	29%	29%	30%	29%	35%	32%	32%	32%
12	24%	27%	24%	26%	24%	27%	24%	30%	29%	29%	29%
13	26%	28%	26%	26%	26%	27%	26%	33%	29%	29%	29%
14	28%	30%	28%	29%	28%	30%	28%	35%	32%	32%	32%
15	24%	26%	24%	25%	24%	26%	24%	29%	28%	28%	28%
16	28%	30%	28%	28%	28%	29%	28%	35%	31%	31%	31%
17	29%	31%	29%	29%	29%	30%	29%	35%	32%	33%	32%
18	25%	27%	25%	27%	25%	27%	25%	31%	29%	30%	29%
19	26%	27%	26%	26%	26%	27%	26%	32%	28%	29%	28%
20	28%	30%	28%	29%	28%	30%	28%	35%	31%	32%	31%
21	27%	25%	27%	25%	27%	0%	27%	33%	29%	25%	29%
22	44%	42%	44%	41%	44%	0%	44%	55%	48%	42%	48%
23	45%	43%	45%	42%	45%	0%	45%	56%	49%	43%	49%
24	26%	24%	26%	24%	26%	0%	26%	32%	28%	24%	28%
25	26%	24%	26%	24%	26%	0%	26%	32%	28%	24%	28%
26	37%	35%	37%	35%	37%	0%	37%	46%	41%	35%	41%
27	44%	41%	44%	41%	44%	0%	44%	54%	48%	42%	48%
28	26%	27%	26%	26%	26%	27%	26%	32%	31%	27%	31%
29	30%	28%	30%	27%	30%	27%	30%	37%	32%	28%	32%
30	29%	28%	29%	27%	29%	27%	29%	35%	32%	28%	32%
31	26%	26%	26%	26%	26%	26%	26%	32%	30%	26%	30%
32	30%	28%	30%	28%	30%	27%	30%	37%	32%	28%	32%
33	30%	28%	30%	28%	30%	28%	30%	36%	33%	29%	33%
34	23%	22%	23%	22%	23%	22%	23%	28%	25%	23%	25%
35	23%	23%	23%	23%	23%	23%	23%	29%	26%	24%	26%
36	22%	21%	22%	21%	22%	21%	22%	27%	24%	22%	24%
37	24%	24%	24%	23%	24%	23%	24%	29%	26%	24%	26%
38	23%	23%	23%	23%	23%	23%	23%	28%	26%	25%	26%
39	23%	23%	23%	22%	23%	22%	23%	29%	25%	23%	25%
40	25%	25%	25%	25%	25%	25%	25%	31%	28%	26%	28%
41	25%	26%	25%	26%	25%	26%	25%	31%	30%	28%	30%
42	24%	24%	24%	23%	24%	23%	24%	30%	27%	25%	27%
43	27%	27%	27%	27%	27%	27%	27%	33%	30%	28%	30%

Table 2: (cont.) Information about the 43 Products in Sample - Price-Cost Margins (PCM).
(*): PCM= Average Price-cost Margin across weeks. Price-Cost Margin=$(p-c)/p$, where p is price and c is marginal cost. S=Scenario; R=retail; W=wholesale. **S-1**: Double Marginalization; **S-2, C 1**: Wholesale Price at marginal cost, Retailer pricing decisions; **S-2, C 2**: Zero Retail Margin, Manufacturer pricing decisions; **S-3**: Hybrid model; **S-4**: Wholesale collusion; **S-5**: Retail collusion; **S-6**: Monopolist. Source: My calculations, Logit Demand. For product names correspondent to each product identification number (**Product ID #**) please refer to previous page in the fourth column of Table 2.

	USA	Germany	France	Bulgaria	Russia	Canada
$(*)$	0.48	1.24	1.92	3.39	0.22	0.34
$(**)$	4.7	11	17	30	2	3.1

Table 3: US Per Capita Consumption compared to selected Countries.
Row **(*)**: Per capita weekly consumption in servings (1 serving = 6 ounces). Row **(**)**: Per capita yearly consumption in kg. 1990's. Sources **(**)**: USA: Economic Research Service, USDA, Statistical Series, Food Consumption Prices & Expenditures 1991-94. Figures for Germany, France, Bulgaria and Russia- Boston Consulting Group , 1998. Canada: Statistics Canada, Matrix 5666. Per capita consumption of Dairy Products. Conversion **(*)**: 1kg = 1000/170 servings. 1 year = 52 weeks.

	Specification 1				Specification 2			
Servings per capita per week	0.25	0.5	0.75	1	0.25	0.5	0.75	1
α	-8.89	-7.14	-7.01	-6.99	-9.31	-8.42	-8.17	-8.04
	(0.72)	(0.59)	(0.58)	(0.59)	(0.81)	(0.64)	(0.63)	(0.63)

Table 4: Sensitivity Analysis to Yogurt weekly per capita consumption assumption. Dependent variable in all columns 2 through 9 is $ln(s_{jt}) - ln(s_{0t})$. Regressions include brand dummy variables. 4310 observations. White standard errors are in parenthesis. α: Estimate of marginal utility of price. Instrumental Variable Specification 1 assumes that marginal cost of the same product sold at different retailers is different, while Specification 2 assumes that it is constant.

Description	Mean	Median	Std	Max	Min
Store 2 (=1 if product from store 2)	0.42	–	–	–	–
Store 3 (=1 if product from store 3)	0.35	–	–	–	–
Total Calories	170.6	150	59.15	253	0
Fat Calories	16.4	20	14.45	39	0
Total Fat (g)	1.94	2.5	1.66	4.6	0
Cholesterol (mg)	11.67	13	5.62	20	5
Total Carbohydrates (g)	29.32	25	11.32	48	13
Dietary Fiber (g)	0.09	0	0.29	1	0
Sugar (g)	25.2	22	10.89	42	10
Protein (g)	8.67	9	2.55	13	4
Vitamin (=1 if product has vitamin A or C)	0.51	–	–	–	–
Calcium (=1 if product has more calcium than 30% Daily Value)	0.63	–	–	–	–
Aspartame (=1 if product has Aspartame)	0.11	–	–	–	–
Fruit on the Bottom (=1 if yes)	0.09	–	–	–	–
Available in Different Sizes (=1 is yes)	0.53	–	–	–	–
New Fruit Flavors (=1 if yes)	0.26	–	–	–	–

Table 5: Characteristics of Products in Sample.
Source: Yogurt cups' label-reads.

Description	Mean	Median	Std	Max	Min
Citric Acid ($/Lb)	1.9	1.3	0.84	3	1.23
Plastic (cents/Lb)	32.6	33	3.26	3.8	27
Sugar (cents/Lb)	9	8.6	1.14	14.4	8.2
Non-fat Grade A milk ($/Lb)	1	1.1	0.08	1.2	0.86
Whey Protein ($/Lb)	0.5	0.5	0.09	0.6	0.31
Corn ($/Bushel)	2.3	2.3	0.16	2.5	1.98
Strawberry ($/CWT)	0.8	0.7	0.29	1.4	0.35
Wages Ohio (weekly earnings/number hours a week - $/hour)	11.2	11	0.56	12.6	10.4
Location of plant for Dannon Yogurts: Minster OH.					
Wages Illinois (weekly earnings/number hours a week - $/hour)	12.1	12.1	0.3	12.8	11.5
Location of plant for Breyers, Light N'Lively (Kraft): Moleena, IL; location of plant for Private Label of Store 3 and location of the three retailers.					
Wages Michigan (weekly earnings/number hours a week - $/hour)	12	11.8	0.61	14.4	10.9
Location of plant for Yoplait Yogurts: Kalamazoo, MI.					
Wages Oregon (weekly earnings/number hours a week - $/hour)	12.9	13	0.37	13.8	12.1
Location of plant for Private Label of Store 2: Clackamas, OR.					
Interest Rate (Federal Funds Effective Rate - %)	4	3.7	1	6.3	2.9
Interest Rate (Commercial Paper 3 months - %)	4.1	3.9	0.96	6.2	3.1

Table 6: Input Prices.

Sources: Citric Acid (Chemical Week); Plastic (Chemical Marketing Reporter); Sugar (Coffee, Sugar and Cocoa Exchange); Non-fat Grade A milk, Whey protein (Cheese Market News, US. Dep. Agriculture); Corn, Strawberry (National Agriculture Statistics Service, US. Dep. Agriculture); Wages (CPS Annual Earning File - NBER 50); Interest Rates (Federal Reserve).

ZIP Area 1				Age						
				< 25	$25-34$	$35-44$	$45-54$	$55-64$	$65-74$	> 75
Income	Less	than	$5,000	190	355	263	161	194	188	83
	$5,000	to	$9,999	114	281	210	154	148	407	411
	$10,000	to	$14,999	85	263	197	213	155	424	298
	$15,000	to	$24,999	201	735	551	327	386	567	285
	$25,000	to	$34,999	160	943	751	407	568	424	146
	$35,000	to	$49,999	90	1016	1109	816	652	325	93
	$50,000	to	$74,999	30	483	926	878	609	166	82
	$75,000	to	$99,999	6	37	212	271	156	74	7
	$100,000	or	more	0	46	71	142	97	17	19
ZIP Area 2				Age						
				< 25	$25-34$	$35-44$	$45-54$	$55-64$	$65-74$	> 75
Income	Less	than	$5,000	41	157	167	163	118	149	193
	$5,000	to	$9,999	84	74	128	171	177	481	756
	$10,000	to	$14,999	106	253	251	160	227	479	627
	$15,000	to	$24,999	162	1126	697	540	725	764	542
	$25,000	to	$34,999	230	1049	939	525	485	628	305
	$35,000	to	$49,999	127	1258	1255	826	748	494	279
	$50,000	to	$74,999	62	699	1032	816	705	278	110
	$75,000	to	$99,999	32	201	228	262	242	89	72
	$100,000	or	more	0	15	93	108	84	31	36
Two ZIPS				Age						
				< 25	$25-34$	$35-44$	$45-54$	$55-64$	$65-74$	> 75
Income	Less	than	$5,000	231	512	430	324	312	337	276
	$5,000	to	$9,999	198	355	338	325	325	888	1167
	$10,000	to	$14,999	191	516	448	373	382	903	925
	$15,000	to	$24,999	363	1861	1248	867	1111	1331	827
	$25,000	to	$34,999	390	1992	1690	932	1053	1052	451
	$35,000	to	$49,999	217	2274	2364	1642	1400	819	372
	$50,000	to	$74,999	92	1182	1958	1694	1314	444	192
	$75,000	to	$99,999	38	238	440	533	398	163	79
	$100,000	or	more	0	61	164	250	181	48	55

Table 7: Demographics by Zip Code Areas

Age of Householder by Household Income. Source: 1990 Census at the Zip Code level.

Variable	No Feature OLS	IV1	IV2	With Feature OLS	IV1	IV2	IV1$^{(*)}$	IV2$^{(*)}$
Price	-5.54	-7.14	-8.42	-5.27	-7.00	-8.29	-6.96	-8.10
	(0.34)	(0.59)	(0.64)	(0.35)	(0.63)	(0.66)	(0.37)	(0.69)
Feature				0.32	0.16	0.04	0.36	0.60
				(0.22)	(0.14)	(0.14)	(0.36)	(0.47)
Measures of Fit								
R^2	0.72			0.72				
Price Exogeneity Test		10.58	28.38		11.17	28.20	10.30	22.64
Feature Exogeneity Test							0.005	0.35
95% critical value		(3.84)	(3.84)		(3.84)	(3.84)	(3.84)	(3.84)
Test of Overidentification		1139	728		1138	727	1137	2173
95% critical value		(613)	(312)		(613)	(312)	(618)	(311)
First Stage								
R^2		0.78	0.74		0.81	0.76		
F-Statistic		13.64	12.95		14.61	13.18		
Wald Test: cost coefficients =0		5134	2106		5157	2173		
95% critical value		(615)	(311)		(615)	(311)		

Table 8: Results from Logit Demand.

Dependent variable in all columns is $ln(s_{jt}) - ln(s_{0t})$. Regressions include brand dummy variables. 4310 observations. White standard errors are in parenthesis. Instrumental Variables (IV1) for prices in this column are input prices multiplied by 43 product dummy variables, assuming that marginal cost differs for the same product sold at different retailers (Specification 1). Instrumental Variables (IV2) for prices in this column are input prices multiplied by 21 product dummy variables, assuming that marginal cost for the same product sold at different retailers is constant (Specification 2). IV(*): Specifications that also instrument for feature are in these last two columns. Source: My calculations.

Variable	Mean in population	Interaction with		
		Unobservables	Income	Age
Constant*	−74.262	−0.270	−1.823	10.811
	(9.600)	(0.186)	(0.410)	(2.436)
Price	−7.884	1.0116	3.212	0.091
	(0.975)	(0.351)	(1.355)	(0.072)
Store 2 (=1 if product from store 2)*	55.846	0.375	2.009	−10.283
	(9.786)	(0.283)	(0.654)	(2.479)
Store 3 (=1 if product from store 3)*	58.851	0.813	1.263	−9.885
	(9.343)	(0.518)	(0.413)	(2.202)
Total Calories*	−0.162	0.002	0.002	−0.006
	(0.011)	(0.002)	(0.002)	(0.007)
Fat Calories*	−2.719			
	(0.078)			
Cholesterol (mg)*	−0.056			
	(0.012)			
Total Carbohydrates (g)*	2.237			
	(0.065)			
Dietary Fiber (g)*	0.769			
	(0.096)			
Sugar (g)*	−1.201			
	(0.036)			
Protein (g)*	0.441			
	(0.027)			
Vitamin (=1 if product has vitamin A or C)*	0.619			
	(0.057)			
Calcium (=1 if more than 30% Daily Value)*	5.314	0.190	0.246	0.236
	(0.946)	(0.218)	(0.212)	(0.086)
Aspartame (=1 if product has Aspartame)*	−5.719			
	(0.206)			
Fruit on the Bottom (=1 yes)*	−4.044			
	(0.146)			
Available in Different Sizes (=1 is yes)*	4.651			
	(0.127)			
New Fruit Flavors (=1 if yes)*	−13.339			
	(0.394)			
GMM	566.83			
R^2 of GLS regression	0.56			
Weighted R^2 of GLS regression	0.78			

Table 9: Results from the Random Coefficients Model of Demand.
(*) were obtained from a (**GLS**) regression of estimated product dummy variables on product characteristics, with 43 observations. **Std**: Standard errors are in parenthesis. Source: My calculations.

	Variable	No Feature GMM Est. (s.e)	NLLS Est. (s.e)	$v_i = 0$ Est. (s.e)	With Feature GMM Est. (s.e)	NLLS Est. (s.e)	$v_i = 0$ Est. (s.e)
Mean	Constant	−74.26(9.6)	−8.07(1.14)	−64.8(8.1)	−2.09(0.45)	−1.91(0.28)	−1.5(0.41)
	Price	−7.88(0.9)	−5.59(0.34)	−8.0(0.88)	−7.61(1.0)	−5.32(0.36)	−7.83(0.9)
	Feature				0.44(0.29)	0.35(0.12)	0.28(0.21)
	Store 2	55.9(9.8)	0.92(0.88)	52.1(8.5)	12.3(0.76)	1.76(0.05)	12.3(0.6)
	Store 3	58.9(9.3)	−2.29(0.83)	47.9(8.29)	13.2(0.87)	2.82(0.06)	13.1(0.8)
	Calories	−0.2(0.01)	−0.22(0.01)	−0.18(0.01)	−0.2(0.01)	−0.07(0.01)	−0.01(0.01)
	Calcium	5.31(0.9)	−0.57(0.68)	5.12(0.80)	2.36(0.6)	3.4(0.07)	2.61(0.23)
Std dev	Constant	0.27(0.18)	0.09(0.04)		0.26(0.19)	0.08(0.04)	
	Price	1.01(0.4)	0.47(0.13)		1.01(0.35)	0.48(0.12)	
	Store 2	0.4(0.3)	0.07(0.08)		0.37(0.28)	0.06(0.08)	
	Store 3	0.8(0.5)	0.13(0.06)		0.82(0.51)	0.12(0.06)	
	Calories	0(0.002)	0(0.003)		0(0.0015)	0(0.0003)	
	Calcium	0.2(0.22)	0.16(0.05)		0.19(0.22)	0.16(0.05)	
Interact	Constant	−1.82(0.41)	−0.2(0.15)	−1.5(0.31)	−1.85(0.42)	−0.23(0.16)	−1.5(0.31)
With	Price	3.21(1.36)	1.23(0.45)	2.97(1.37)	3.18(1.40)	1.24(0.55)	2.80(1.34)
Income	Store 2	2.0(0.65)	−0.05(0.1)	1.76(0.61)	2.01(0.65)	−0.02(0.1)	1.75(0.61)
	Store 3	1.26(0.4)	0.14(0.1)	1.25(0.36)	1.28(0.42)	0.17(0.09)	1.24(0.36)
	Calories	0(0.002)	0(0.001)	0(0.001)	0(0.002)	0(0.0006)	0(0.001)
	Calcium	0.25(0.2)	0.08(0.08)	0.27(0.15)	0.26(0.21)	0.08(0.08)	0.27(0.16)
Interact	Constant	10.81(2.4)	−0.57(0.33)	10.43(2.1)	10.74(2.4)	−0.52(0.33)	10.4(2.1)
With	Price	0.09(0.07)	0.13(0.11)	0.12(0.13)	0.09(0.06)	0.08(0.08)	0.09(0.05)
Age	Store 2	−10.3(2.5)	0.37(0.25)	−9.91(2.2)	−10.15(2.5)	0.34(0.24)	−9.88(2.18)
	Store 3	−9.9(2.2)	0.38(0.23)	−10.1(1.9)	−9.76(2.2)	0.36(0.22)	−10.05(1.9)
	Calories	−0.01(0.01)	0(0.001)	−0.01(0.01)	−0.01(0.01)	0(0.002)	0(0.01)
	Calcium	0.24(0.09)	0.4(0.19)	0.15(0.69)	0.21(0.86)	0.38(0.18)	0.27(0.7)
DoubleMg							
PCMw		31.6%	41.6%	30.6%	32.8%	29.1%	30.4%
PCMr		34.5%	41.5%	32.8%	35.8%	28.8%	31.7%
Monopolist		39.2%	47.1%	38.3%	40.6%	32.8%	35.5%
First Stage							
R2		0.74		0.74	0.76		0.76
Wald (cr.val)		2166(312)		2166(312)	2173(311)		2173(311)
GMM/NLLS		566.8	2157	644.7	565.1	2146	644.1
R2 min. dist.		0.56	0.68	0.69	0.83	0.65	0.81

Table 10: Additional Specifications - Random Coefficient Model of Demand.
Estimates (Est.) and standard errors in parenthesis (s.e) for different specifications in each column. Column 3 presents the GMM estimates without feature, column 4 the NLLS for the same specification and column 5 presents the GMM estimates with $v_i = 0$. Columns 6 through 8 are analogous to 3 through 5 but include feature. Source: My calculations.

	Mean Cross Price Elasticities		Relative Substitution
	Same Store Products	Other Store Products	
	(a)	(b)	(a)/(b)
Average across Products	0.070	0.020	3.559
Average by Store	(c)	(d)	(c)/(d)
Store 1	0.062	0.007	9.397
Store 2	0.036	0.019	1.868
Store 3	0.091	0.029	3.185

Table 11: Relative Substitution for products within and across stores.

Product		Logit Demand		Random Coefficients Demand			
		Own Price	Cross-Price Elasticities	Own Price			
ID #	Store	Elasticity	(varying price of product in row)	Elasticity	Mean (a)	Std (b)	(b)/(a)
1	1	−8.028	0.012	−7.631	0.036	0.044	1.227
2	2	−8.023	0.018	−7.530	0.022	0.011	0.512
3	3	−8.022	0.019	−7.570	0.029	0.018	0.606
4	1	−8.038	0.002	−7.681	0.006	0.008	1.257
5	2	−8.026	0.014	−7.515	0.013	0.007	0.548
6	3	−8.033	0.007	−7.526	0.012	0.008	0.625
7	1	−8.031	0.009	−7.568	0.028	0.034	1.238
8	2	−7.911	0.129	−7.267	0.134	0.066	0.492
9	3	−7.925	0.115	−7.196	0.194	0.122	0.626
10	2	−8.003	0.037	−7.470	0.039	0.019	0.490
11	3	−8.018	0.022	−7.500	0.039	0.024	0.620
12	1	−8.037	0.003	−7.681	0.008	0.009	1.208
13	2	−8.016	0.024	−7.447	0.026	0.017	0.646
14	3	−8.020	0.020	−7.496	0.028	0.016	0.557
15	1	−8.037	0.003	−7.675	0.007	0.009	1.220
16	2	−8.021	0.020	−7.525	0.018	0.012	0.687
17	3	−8.027	0.013	−7.505	0.021	0.012	0.561
18	1	−8.036	0.004	−7.645	0.010	0.012	1.220
19	2	−7.984	0.056	−7.400	0.059	0.037	0.627
20	3	−7.996	0.044	−7.427	0.060	0.034	0.562
21	2	−7.983	0.057	−7.387	0.053	0.033	0.629
22	2	−8.032	0.008	−7.675	0.008	0.004	0.542
23	2	−8.033	0.007	−7.730	0.007	0.005	0.664
24	2	−8.015	0.025	−7.419	0.034	0.016	0.468
25	2	−8.040	0.001	−7.476	0.001	0.000	0.480
26	3	−7.994	0.046	−7.519	0.066	0.036	0.550
27	3	−8.036	0.004	−7.692	0.006	0.003	0.583
28	1	−8.032	0.008	−7.657	0.024	0.029	1.246
29	2	−8.023	0.017	−7.520	0.018	0.006	0.321
30	3	−8.029	0.011	−7.528	0.020	0.015	0.752
31	1	−8.031	0.010	−7.584	0.021	0.027	1.260
32	2	−8.010	0.030	−7.502	0.032	0.014	0.448
33	3	−8.020	0.020	−7.498	0.037	0.028	0.753
34	2	−8.013	0.028	−7.391	0.026	0.017	0.635
35	3	−8.014	0.026	−7.356	0.042	0.028	0.664
36	2	−8.037	0.003	−7.432	0.004	0.002	0.628
37	3	−8.037	0.003	−7.457	0.006	0.004	0.708
38	1	−8.032	0.008	−7.525	0.025	0.032	1.245
39	2	−7.977	0.063	−7.333	0.072	0.043	0.592
40	3	−7.989	0.051	−7.290	0.088	0.062	0.708
41	1	−8.036	0.005	−7.661	0.020	0.027	1.335
42	2	−7.991	0.049	−7.382	0.049	0.023	0.464
43	3	−7.984	0.057	−7.318	0.092	0.071	0.769
Average					0.036	0.033	0.923

Table 12: Diagnostic of Logit Assumption.
Mean (column (a)) and Standard deviations (column (b)) of the cross-price elasticities for the different products under a random coefficients demand specification. For names equivalent to the Product Identification Numbers **ID#** see Table 2.

Description	Mean	Median	Std	Min	Max
Given a Logit Demand					
PCM Model 1: Double Marginalization - Wholesale Margin (%)	28.3	26.7	6.2	17.7	73.5
PCM Model 1: Double Marginalization - Retail Margin (%)	28.0	27.5	6.5	17.4	62.4
PCM Model 1: Retail+Wholesale Margin (%)	56.3	54.6	12.6	35.1	135.4
PCM Model 2 Case 1 : Zero Wholesale Margin, Retailer Decision (%)	28.0	27.5	6.5	17.4	62.4
PCM Model 2 Case 2: Zero Retail Margin, Manufacturer Decision (%)	27.6	26.4	5.9	17.6	61.9
PCM Model 3: Hybrid Model - Wholesale Margin (%)	22.7	25.8	10.8	0	72.9
PCM Model 3: Hybrid Model - Retail Margin (%)	28.0	27.5	6.5	17.4	62.4
PCM Model 3: Retail+Wholesale Margin (%)	50.8	50.8	11.4	21.8	134.8
PCM Model 4: Wholesale Collusion - Wholesale Margin (%)	34.5	33.0	8.4	20.1	88.0
PCM Model 4: Wholesale Collusion - Retail Margin (%)	28.0	27.5	6.5	17.4	62.4
PCM Model 4: Retail+Wholesale Margin (%)	62.5	60.4	14.9	37.5	149.8
PCM Model 5: Retail Collusion - Wholesale Margin (%)	29.5	28.1	6.4	18.1	78.2
PCM Model 5: Retail Collusion - Retail Margin (%)	31.2	30.2	7.1	19.5	67.5
PCM Model 5: Retail+Wholesale Margin (%)	60.6	57.9	13.2	37.7	145.6
PCM Model 6: Monopolist (%)	31.2	30.2	7.1	19.5	67.5
Given a Random Coefficients Demand					
PCM Model 1: Double Marginalization - Wholesale Margin (%)	31.6	29.8	8.4	16.6	131.3
PCM Model 1: Double Marginalization - Retail Margin (%)	34.5	32.0	10.0	17.8	103.5
PCM Model 1: Retail+Wholesale Margin (%)	66.2	62.5	17.3	34.4	216.8
PCM Model 2 Case 1 : Zero Wholesale Margin, Retailer Decision (%)	34.5	32.0	10.0	17.8	103.5
PCM Model 2 Case 2: Zero Retail Margin, Manufacturer Decision (%)	30.7	29.2	6.9	16.9	80.9
PCM Model 3: Hybrid Model - Wholesale Margin (%)	24.6	8.1	48.5	0	527.8
PCM Model 3: Hybrid Model - Retail Margin (%)	34.5	32.0	10.0	17.8	103.5
PCM Model 3: Retail+Wholesale Margin (%)	59.11	46.2	49.2	17.8	562.2
PCM Model 4: Wholesale Collusion - Wholesale Margin (%)	51.6	42.0	32.4	19.9	409.1
PCM Model 4: Wholesale Collusion - Retail Margin (%)	34.5	32.0	10.0	17.8	103.5
PCM Model 4: Retail+Wholesale Margin (%)	86.1	75.4	39.8	35.7	481.5
PCM Model 5: Retail Collusion - Wholesale Margin (%)	39.9	31.0	9.1	16.9	136.8
PCM Model 5: Retail Collusion - Retail Margin (%)	39.2	36.2	12.4	18.6	138.1
PCM Model 5: Retail+Wholesale Margin (%)	72.1	67.7	19.7	35.8	226.1
PCM Model 6: Monopolist (%)	39.2	36.2	12.4	18.6	138.1

Table 13: Price-Cost Margins (PCM) by Scenario.

PCM=$(p-c)/p$ where p is price and c is marginal cost. Std: Standard deviation. Source: My calculations.

Model	DM$^{(*)}$	Critical value (**C**)	Cases of Rejection$^{(**)}$
1. Double Marginalization	9.08	103.06	86%
2.1 Zero Wholesale Margin, Retailer Pricing decisions	18.24	59.02	24%
2.2 Zero Retail Margin, Manufacturer Pricing decisions	20.10	59.02	84%
3. Hybrid Model	6.52	95.36	81%
4. Wholesale Collusion	16.28	103.06	85%
5. Retail Collusion	7.26	103.06	83%
6. Monopolist	14.29	59.02	33%

Table 14: Validity of Different Supply Models.
(*): Distance Metric (DM) Tests for Validity of Different Supply Models based on Generalized Method of Moments (GMM) Estimation. $DM = R[GMM^r(\hat{\theta}^r_T) - GMM^u(\hat{\theta}^u_T)]$ which is distributed Chi-square with degrees of freedom equal to the number of restrictions R. GMM_r is the estimated GMM criterion function for the restricted model and GMM_u is the estimated GMM criterion function for the unrestricted model. Critical values C for a Chi-square χ^2_R with large degrees of freedom R for the significance level a can be approximated (Greene, 1997, p.70) as follows: $C \approx 0.5[\Phi^{-1}(a) + \sqrt{2R-1}]^2$. Significance level above is 10 %. Null hypothesis is that all the coefficients associated with the price-cost margins are equal to one. **(**)**: Percentage cases of rejection from testing individually if each product exhibits price-cost margins consistent with the ones implied by the model in each row (parametrically this is done by testing individually whether each λ is equal to one). Source: My calculations.

	Alternative Models						
H_0 Model	1	2.1	2.2	3	4	5	6
1. Double Marginalization	–	2.04	1.39	0.13	1.82	2.61	1.93
2.1 Wholesale Price at marginal cost	1.00	–	1.29	0.14	0.61	0.71	1.42
2.2 Zero Retail Margin	1.18	3.87	–	0.18	0.60	0.80	3.31
3. Hybrid model	0.14	2.29	1.03	–	0.23	0.21	2.10
4. Wholesale Collusion	0.41	2.28	0.18	0.06	–	0.41	2.29
5. Retail Collusion	2.10	2.94	0.59	0.13	1.18	–	2.13
6. Monopolist	1.06	4.24	2.56	0.14	0.61	0.86	–

Table 15: Pair-wise statistics to determine which model most adequately explains the data.
In each row is the null model being tested and in each column the alternative model being used to test the null model. Source: My calculations. Non-nested Cox-type test statistic for strictly non-nested hypothesis (SNN) and for overlapping models (OVE and AEV) that can be discriminated in the Vuong (1989) two step procedure are distributed standard normal. One-sided test at 5% with critical value 1.65. Random coefficients demand.

Figure 1

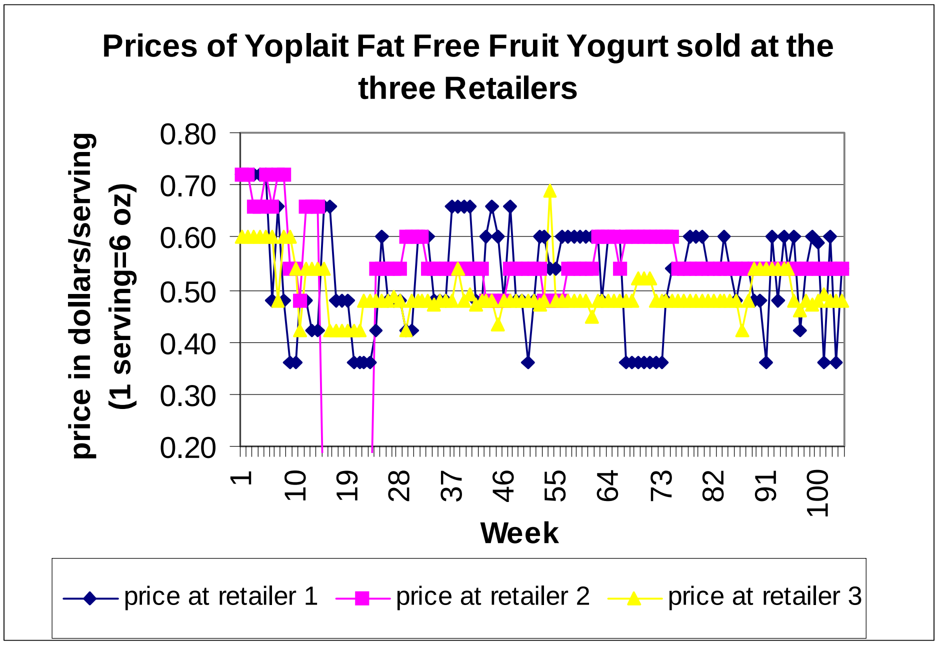


Figure 2

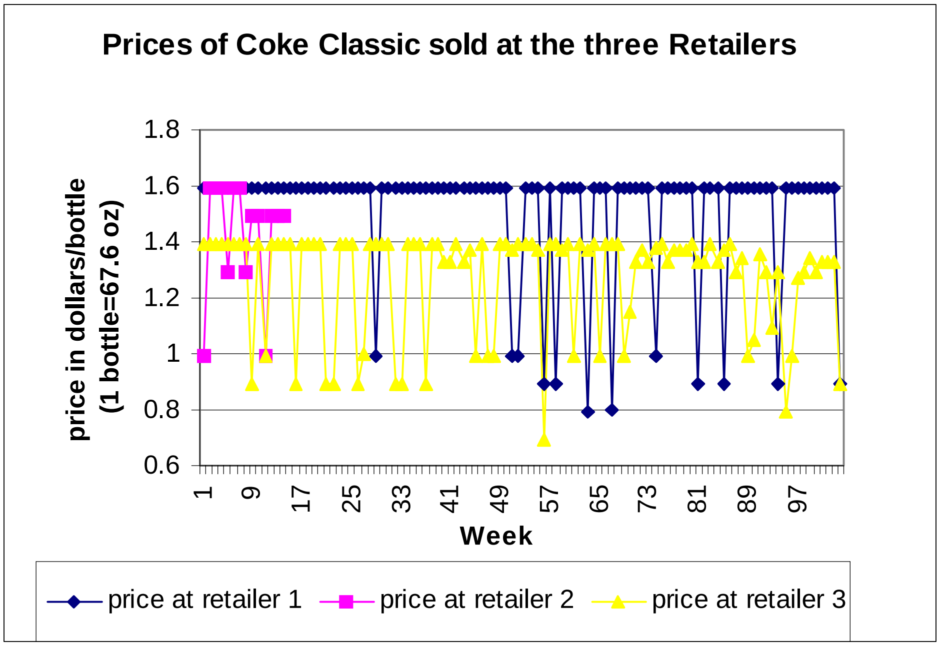
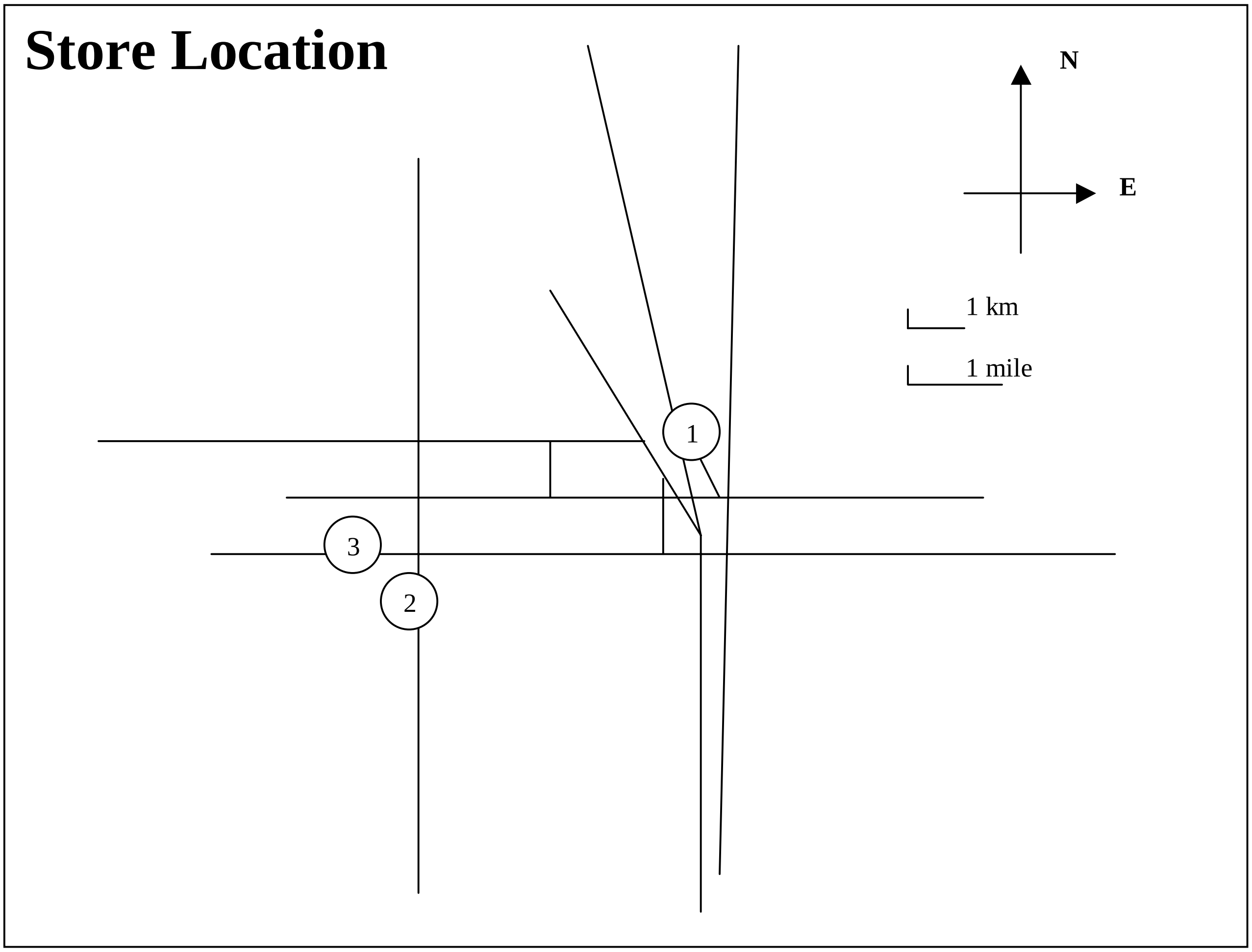


Figure 3



Figures 4 and 5