Product Variety in the U.S. Yogurt Industry

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Joseph Rossetti, M.S.

Department of Economics

The Ohio State University

2018

Dissertation Committee

Javier Donna, Advisor

Jason Blevins

Bruce Weinberg

Abstract

My research investigates the determinants of the variety of products offered in consumer goods industries. Firms in consumer goods industries must decide which set of product lines to offer. I model this as an entry-exit decision across a set of possible product markets. Understanding how the variety of products offered by firms is determined is important due to its large impacts on consumer welfare. Also, the effect of changes in market concentration, like mergers, on product variety, depends on the specific features of the industry: the elasticity of demand, types of product differentiation available to firms, and competition from regional or niche producers.In Chapter 2 I lay out the theoretical model of product choice that I will estimate in two steps in the remaining two chapters. I provide a brief discussion of the incentives in the model for firms to introduce a new product.

In Chapter 2, I estimate a model of product entry and exit in the U.S. yogurt industry from 2001-2011 using supermarket scanner data from the IRI Marketing Database. I use a two-step procedure. I first estimate yogurt industry demand and variable costs using the standard framework of Berry et al. (1995). I account for store level adoption of product lines, and brand advertising decisions in a reduced form. I also study the consumer welfare changes that have occurred in the U.S. Yogurt industry due to the adoption of several new products lines between 2001 and 2011. Previous work on changes in consumer welfare in consumer goods industries has focused on the introduction of one or two new products or brands. The U.S. Yogurt industry has seen the introduction of 16 new product lines in my sample period. In my data, controlling for price changes, the

largest direct benefit from a new product is \$3.6 million from the introduction of Greek yogurt, however, the direct benefits only account for less than 5% of the total increases in consumer welfare from new products. Price decreases after product introductions are the main source of consumer welfare gains.

In Chapter 3, I estimate the fixed costs of offering a product. Estimation of the fixed cost is complicated because firms can offer any subset of the potential product lines in the industry, but I only observe in the sample a small number of the possible combinations of products. I apply the pairwise maximum score estimator of Fox (2007), which provides consistent estimates in settings with large choice sets. I use the first stage estimates to compute firms' expected variable profits from offering alternative sets of products and choose the fixed costs parameters to maximize the number of times the model predicts that the firms' observed choices were optimal. In a counterfactual analysis, I find that after a merger the merged firm may have a preference, depending on its fixed costs, to offer more unique products than the competitive industry, but that the incentive is not significantly stronger than the incentives already present in the competitive industry. Consumer welfare is lower after a merger regardless of the number of products. Together this leads consumers to prefer the competitive market structure since they may expect similar levels of product variety and lower prices.

Acknowledgments

I would like to thank my advisor Javier Donna, as well as the other members of my dissertation committee Jason Blevins and Bruce Weinberg, for advice and support. I would also like to thank participants of the Ohio State Department of Economics Applied Microeconomics Lunch Seminar and the 44th Annual Conference of the European Association for Research in Industrial Economics for many helpful comments. Finally, I would like to thank IRI for providing access to the IRI Marketing Database. Any errors and views expressed in the paper are my own.

Lastly, I would like to thank my wife Samantha for her love and support over the long process of earning a Ph.D.

Vita

Biographical Information

2009 .... Fayetteville High School

2013 .... B.S.B.A. Economics, University of Arkansas

2013 .... B.S. Mathematics, University of Arkansas

2014 .... M.S. Economics, The Ohio State University

Fields of Study

Major Field: Economics

Table of Contents

List of Figures

List of Tables

1 Introduction to Product Choice in Differentiated Product Markets

1.1 Introduction

Product variety, the number of unique products offered by firms, determines price competition, profitability, and consumer welfare in an industry. A firm cannot remain competitive without adjusting the set of products it offers to keep up with consumer taste. For example, in the U.S. Yogurt industry in 2016 Greek yogurt accounted for half of the sales of all yogurts but only accounted for 21% of General Mills yogurt sales. In response, General Mills, the owner of Yoplait, has announced in 2016 it will make adjustments to two-thirds of its product offerings Kell (2016). When adopting new products to match changing consumer demand, firms pay direct costs for marketing, producing, and distributing the new products, and indirect costs as consumers substitute away from existing towards the new product and competitors alter their prices. Consumers benefit from the product directly when they prefer it to existing products and indirectly when its introduction results in lower prices for competing products. In order to better understand how market structure and product differentiation determine firms' incentives to adopt products in consumer goods industries, I will estimate a model of product entry and exit in the U.S. Yogurt industry from 2001 to 2011.

Firms can change the variety of products they offer in two ways[1]. The first way is a

[1] Quelch and Kenny (1994) discusses this distinction in depth terms of business strategy and I borrow

product line extension. Product line extensions in the yogurt industry would include: changing branding, packaging, and package size, as well as adding a new flavor (e.g. adding a strawberry low-fat yogurt where only vanilla low-fat was offered before). The second way is to adopt or introduce a new product line. Examples of product line introductions in the yogurt industry during my sample include the introduction of drink, greek, and probiotic yogurts. In addition to these new categories of yogurt, firms also adopt new combinations of existing categories. For example, after the introduction of probiotic yogurts, a firm could choose to combine probiotic yogurt with the existing category of lite (diet) yogurt to create probiotic lite yogurt. The distinction between an extension and a new product line is relevant to firms for two reasons. First, new product lines may cost more than line extensions. They require changes to production and shipping technology, and while consumers are relatively well informed about the attributes of a product line extensions, new marketing expenses may be required to inform consumers about a new product line. Second, product lines tend to be priced together. The extension may raise the value of the product line allowing a higher price to be charged, but this price will be the same across products in the line. A new product line will usually be offered at a new price point, which reflects its cost and demand by consumers for the new type of product.

Product line adoption lends itself to a simple model of product entry and exit. Product lines are different from each other in discrete ways, particularly in the yogurt industry, and each is offered at its own price point. I think of an industry as a group of closely related product markets–one product market per possible product line. In each period, firms choose which markets they will offer products in from among this set. A standard model of entry and exit would predict that firms would enter until the expanding supply lowers the price enough that the marginal firm would not be able to cover their fixed costs. If the product markets were independent then this intuition would be correct, and standard methods for estimating entry and exit games could be applied directly.

their definition.

Product markets within an industry are not independent, however, and as prices fall in one market, demand will rise for that product and fall for other products.

The inherent interdependence between product markets means that I will need to adapt existing models of entry and exit in two ways. First, I will need to estimate the demand system for the yogurt industry in order to estimate how firms would expect demand curves in each product market to respond when they enter or exit products. Second, since firms can enter products in any subset of the product markets in the industry the number of possible actions is large, and this will pose a challenge for estimation. In later chapters, I will take up these issues and estimate demand and fixed costs of offering products. In the current chapter, I will review the literature on product choice in Section 1.1.1, lay out my model of product choice in Section 1.2, and then in Section 1.3 I discuss the incentives for firms to introduce products in the model and how they may relate to market structure.

1.1.1 Literature Review

There is a considerable literature in economics and related business fields about the design and choice of products, but relatively few papers that study product variety empirically using entry and exit models. I will focus on reviewing the literature on product entry and exit since my model will not say anything about product design or business strategy; my model focuses on estimating fixed costs using the revealed preference approach that is common in the entry and exit literature. Firms choose best responses, and I look for parameters that rationalize their behavior. First, I will review the literature on modeling the entry and exit of products. Then, I will discuss the related literature on firm entry and exit from which the modeling strategy is drawn. The model that I estimate adds additional assumptions to the standard product entry and exit model to achieve point rather than partial identification of the fixed costs, and complements the literature on product entry and exit by studying an industry where product differentiation is more difficult and competition stronger than in previously studied settings.

One of the earliest empirical investigations into the relationship between product variety and market structure was Connor (1981), which examined a cross section of food industries and reported a positive correlation between market concentration and new product introduction. A structural model of product choice was estimated by Draganska et al. (2009) in the ice cream industry. They consider the entry and exit of flavors of vanilla ice cream. This limits the number of potential products to 3 potential products. When Draganska et al. (2009) was written, game estimation relied on solving for equilibria as the inner loop in a nested fixed point algorithm. I update their work on product variety in consumer goods industries by taking advantage of advances in estimation that avoid solving for equilibria. Their model assumes firms have private information about their fixed costs of offering flavors, and firms play a Bayes Nash Equilibrium in a two stage game where they first choose which products to offer and then set prices in a second stage. The authors find that in counterfactuals where firms merge, the variety of products offered increases or decreases depending on market conditions. By focusing on the adoption of product lines, I also study a different type of product variety. The yogurt industry has experienced the introduction of several new product lines, e.g. greek yogurt, and drink yogurt, that are different from existing products in a way that flavors of vanilla ice cream are not. Other empirical work has investigated product line extensions in the yogurt industry: Kadiyali et al. (1999) and Draganska and Jain (2005). Generally, product line extensions appear to increase market power of the firm offering the extension and appear to benefit consumers more than firms–firms would prefer a world without product line extensions since the proliferation of extensions in equilibrium is costly.

Recent work on product entry and exit has also focused on consumer durables. By studying the yogurt industry I update the recent product entry and exit literature to include perishable consumer goods. Fan and Yang (2016) study the variety of products offered in the U.S. cellphone industry. They estimate demand for models of cellphones using a random coefficients logit model estimated from aggregated data. They then use this model to back out a model of marginal cost and compute firm's variable profits as a function of which products they offer. They assume that firms choose products according to a Nash Equilibrium and use the Nash Equilibrium necessary conditions to partially identify the fixed costs of offering different models of cellphones. They find that increases in market concentration have a negative impact on the variety of products offered by firms. I find the opposite: the monopolist often prefers adding products to the set offered by the competitive industry. One explanation for this difference is that it relates to the type of product differentiation possible in each industry. Fan and Yang (2016) model differentiation using a single index of product quality determined by underlying product characteristics, and these characteristics are often continuous variables (for example screen size). In the yogurt industry, I model product differentiation as a function of membership in discrete product categories: a yogurt is either Greek or not Greek, and either marketed toward kids or not. This makes differences between products discrete. A single index of quality in this setting would have only a finite number of points available for firms to occupy. Thus it is easier in the context of cellphones to produce a new close substitute to rival products, but in the yogurt industry potential new products will be less likely to be close substitutes. For firms with more market power in the cellphone industry, close substitutes generate too much cannibalization of sales to be profitable, but in the yogurt industry there is less cannibalization of sales to deter the monopolist from adding products.

The estimation strategy outlined above was first proposed by Eizenberg (2014) in studying product variety in the personal computer industry, and I follow a similar strategy adding an assumption when estimating fixed costs in order to apply maximum score and get point estimates. I depart from this strategy which is used in Fan and Yang (2016) by applying pairwise maximum score rather than partial identification. Both methods attempt to find parameters of the fixed costs function that make the observed choices best responses, but maximum score imposes the additional assumption of exchangeability on the unobservables. Under this added assumption point identification is possible. Besides contributing an empirical strategy used in several papers on product entry, Eizenberg

(2014) reports an interesting distributional issue that arises because there are price sensitive consumers who are not well served by the industry because manufacturers over provide expensive cutting edge PCs instead of older cheaper models. Another paper that considers product variety in a durable goods market is Wollmann (2014), which investigates product variety in the commercial vehicles market (cargo vans and light trucks). That paper partially identifies fixed costs in a dynamic model of product choice, and finds that after a merger product entry by un-merged firms can make up for the welfare losses from the merger.

A notable early paper combining product and firm entry is Mazzeo (2002), which extended the literature on firm entry and exit that began with Bresnahan and Reiss (1991) to allow firms to choose first whether to enter, and then to choose a market segment in which to compete. Draganska et al. (2009) can be viewed as an extension of this literature to the case where firms are multi-product. Recently, Cilberto et al. (2015) have extended the firm entry and exit literature significantly by estimating demand, marginal cost, and fixed costs simultaneously, rather than estimating demand and marginal cost in a first step and then using computed differences in variable profits to identify fixed costs as in Eizenberg (2014); Fan and Yang (2016); Wollmann (2014) and this paper. They partially identify the entire system using the N.E. necessary conditions. This allows them to consider correlation between unobserved shifters of demand, marginal cost, and fixed costs. They find that this correlation is important in the U.S. airline industry and leads to under-estimates of mark ups when demand is estimated separately. This correlation may be more important in the airline industry relative to the yogurt industry and other consumer goods industries because of how markets are defined. In the airline industry markets are usually defined as origin-destination airport pairs. Therefore, airlines can often easily enter or exit a market in response to current period shocks to demand or variable costs since they are often already flying in and out of at least one of the airports that makes up the market. Further, the computational complexity of the Cilberto et al. (2015) estimator would increase dramatically if firms were choosing from a large set of products to offer rather than making a binary entry and exit decision.

The theoretical literature on product choice is broad. I will focus on reviewing a strand of papers on product choice that is closely related to the model I estimate in three ways: demand is based on a logit model, product choice is a discrete entry and exit decision for each product, and prices are endogenous and determined after products are chosen. The seminal paper in this strand of the literature is Anderson and de Palma (1992), presents a two-stage game where firms choose a set of products and then determine prices. The demand is nested logit with no outside option, and the product's qualities are identical. They consider the case where the brands form the nests (so that consumers can be viewed as first choosing among the brands, then choosing a product from their preferred brand). Their setting makes it so that firms charge the same price for all products, and the number of products offered by the firm is all that matters for product choice. They show that the number of products is below the social optimum for a given number of firms. They note that this was in contrast to existing work which tended to point to over entry due to business stealing in other settings. In their setting firms internalize some of the business stealing, because they offer multiple products, reducing the incentive for over-entry. In their model additional firms cause existing firms to decrease the number of products they offer. This is because price competition increases with both the number of products and the number of firms so that firms accommodate the entrant by decreasing the number of products they offer as a substitute for lowering prices. The internalization of business stealing and price competition effects identified in Anderson and de Palma (1992) remain in the model that I consider, although the conclusions of Anderson and de Palma (1992) do not necessarily hold since I do no maintain many of their simplifying assumptions (for example the qualities of products will not be equal).

Aydin and Ryan (2000) and Besbes and Saure (2016) generalize the model in Anderson and de Palma (1992) by allowing for products to have different qualities and costs so that product identities matter not just the total number of products. They also do not adopt the nested demand where consumers select brands and then products, but Besbes and Saure (2016) does discuss an extension to the nested case. While products are no longer offered at equal prices, both papers show that the equilibrium markups of each firm's products will be equal. This equality of markups is a general property of the logit model related to the independence of irrelevant alternatives assumption since this implies all products compete equally with each other. This property means that if the firm adds a product that provides more average utility to consumers it will increase its equilibrium markup (charging a higher markup or prices on all products). The equal markups in the more general case are similar to the equal prices in Anderson and de Palma (1992)[2]. In the discussion in Section 1.3 of the incremental profits from adding a product, this will be referred to as the own price effect (which has a direct effect via the change in prices, and an indirect effect as consumers respond to the price increase). The own price effect in the random coefficients model of demand cannot be easily characterized since the markups of products are no longer equal. Both Aydin and Ryan (2000) and Besbes and Saure (2016) find that firms competing in an oligopoly market have a strong incentive to offer products; both papers present cases where firms offer all available products up until fixed costs or constraints make added products too costly or infeasible to add. Besbes and Saure (2016) find that with exogenous prices the monopolist will offer fewer products than the competitive market, and because markups are equal when prices are endogenous the results for the exogenous prices case generalize to the endogenous case. The literature summarized above highlights the role that the functional form assumptions about demand may play in determining predictions about the incentives of firms. Not only does the multinomial logit model impose unrealistic structure on demand, but it also imposes considerable structure on markups and the firm's incentive to introduce products. To address this concern I do not use the multinomial logit, but instead use the random coefficients logit model.

[2]There is another parallel between the extended model and the earlier model of Anderson and de Palma (1992): Besbes and Saure (2016) show that with the multinomial logit model the choice of products can be reduced to the choice of by the firm of the expected value (or inclusive value) to the consumer of purchasing exclusively from that firm. This reduction is similar to the simplification in Anderson and de Palma (1992) that reduces product choice by firms to a choice of the number of products to offer (with equal qualities increasing the number of products increases the inclusive value).

1.2 Product Choice Game

Firms compete in a two-stage game. In the first stage, firms select the products that they plan to offer. Denote the firms choice in stage one by a^f a J-vector of product activity indicators (each element a_j^f is 1 or 0 depending on whether firm f offers product j), and let $\mathcal{A}$ be the set of all distinct a^f. Based on its choice of a^f the firm can compute its fixed costs $C(a^f, \eta^f)$, which are also a function of an unobserved firm-specific state $\eta^f \in \mathcal{E}$. Let $\mathcal{J} = |\mathcal{A}|$ and $\mathcal{E} = \mathbb{R}^{\mathcal{J}}$, so that η^f is a vector with elements $\eta^f(a^f)$. In the second stage, firms observe the offered products and choose prices. I assume firms play a Nash Equilibrium in prices in the second stage, and that they anticipate this in the first stage. Thus a firm knows that if it offers a^f and the opposing firm offers a^{-f} in demand state D it will receive variable profits $\Pi^f(a^f, a^{-f}, D)$. This separates the game into a stage where fixed costs are determined, and a stage where variable profits are determined. The firm does not completely observe the state of demand when the choice of which products to offer is made. Divide D into two parts $D = (d, e)$, where d is the observed or known state of demand in the first stage and e is the unobserved state of demand (random from the perspective of the firm) in the first stage. Define the expectation with respect to the demand shock e of second stage profits as $\pi^f(a^f, a^{-f}, d) \equiv E_e \Pi^f(a^f, a^{-f}, D)$.

Definition 1. Product Choice Game

- Set of players $\{f, -f\}$
  - Set of types for each player $\mathcal{E}$
  - Set of actions for each player $\mathcal{A}$
  - Payoff function: $\pi(a^f, a^{-f}, d) - C(a^f, \eta^f)$

In order to characterize product choice in the first stage, it is necessary to characterize the solution to the price setting game in the second stage. After firms choose what products to offer, they observe shocks to marginal cost and demand and then set prices. Let p denote the full vector of all prices, with p^f being the prices of firm f's products and p^{-f} the prices the other firm's products. Similarly, I will denote the market shares as $s = \left(s^f, s^{-f}\right)$, and note that the market shares are all functions of the prices and other parameters so that the share of good k is $s_k = s_k(p, \vartheta, D)$. Denote the per unit costs $c = \left(c^f, c^{-f}\right)$, and the markups as $b = p - c$. Define the firms profit function as $\mathbb{P}^f(p^f, p^{-f}, \vartheta, D) = \left(p^f - c^f\right) s^f(p^f, p^{-f}, \vartheta, D)$. The firm solves:

$$
\begin{aligned}
&\underset{p^f}{max}\ \mathbb{P}^f(p^f, p^{-f}, \vartheta, D) \\
&\quad s.t.\ p^{-f} = \underset{p^{-f}}{argmax}\, \mathbb{P}^{-f}(p^f, p^{-f}, \vartheta, D).
\end{aligned} \tag{1.1}
$$

For each firm this leads to a first order condition:

$$
s^f + (p^f - c^f) D_{p^f} s^f = 0. \tag{1.2}
$$

The matrix $D_{p^f} s^f$ is the Jacobian of the market shares with respect to prices and has elements $\left[D_{p^f} s^f\right]_{kh} = \frac{\partial s_k(p,\vartheta,D)}{\partial p_h}$, where k and h index products owned by firm f. The Nash equilibrium prices can be characterized by stacking each firm's first-order conditions on top of the other. This is because in equilibrium both firms first order conditions must be satisfied by definition of the pricing problem. Morrow and Skerlos (2010) refer to these conditions together as the simultaneous stationarity conditions and prove that solutions to this system must be Nash equilibria. Let $D_p s$ be defined with elements $\left[D_{p^f} s\right]_{kh} = \frac{\partial s_k(p,\vartheta,D)}{\partial p_h} 1\left\{a_k^f a_h^f = 1 \ or \ a_k^{-f} a_h^{-f} = 1\right\}$. I follow the terminology of Morrow and Skerlos (2010)again and call $D_p s$ the inter-firm Jacobian of market shares and its elements are 0 whenever the product k is not owned by the firm setting the price h. Using this new Jacobian matrix the simultaneous stationary conditions can be written as:

$$s + (p - c)\, D_p s = 0. \tag{1.3}$$

Computing Nash Equilibrium prices and the equilibrium profits reduces to solving this non-linear system of equations. (1.3) implicitly defines the equilibrium price function, and (1.2) implicitly defines firm f's reaction function (or best response function) which characterizes how it responds to changes in other firm's products (or an exogenous change in the price of one of its own products). In the next section, I will use the equations (1.2) and (1.3) to characterize how prices and market shares change when a new product is introduced.

The following assumptions describe the information structure of the game and are the key to the identification and estimation procedures. I will briefly describe each assumption here, and then describe their role in full when they are used in the following sections.

Assumption 1. *Complete Information:*

- Assume that the vector η^{-f} is known to firm f when it makes its product choice decision
- $g(\eta^f|D)$ is the same for both firms, and η^f does not depend on the actions of the other firm
- $\eta^f \perp \eta^{-f}$

Assumption 1 states that the product choice game is a game of complete information. This assumption motivates the use of the Nash equilibrium concept. In Nash equilibrium firm's choices or products are best responses. This generates inequalities of the form: given a^{-f} and d, $\pi(a^{f*}, a^{-f}, d) - C(a^f, \eta^f) \geq \pi(a^f, a^{-f}, d) - C(a^f, \eta^f)$ for any $a^f \neq a^{f*}$. These inequalities for the basis for estimating the fixed cost parameters. $\pi(a^f, a^{-f}, d)$, the expected variable profits, are computed using estimates of demand and variable cost. Then given a parametrization of the cost function I make pairwise comparisons between the chosen set of products a^{f*} and alternatives a^f, and maximize the number of times the observed choice is optimal (which is Manski's score function).

Assumption 2. *Exchangeability:*

- Let $g(\eta^f|D)$ denote the distribution of the firm's type vector conditional on the state variable D, and let g be absolutely continuous with full support
- Assume that $g(\eta^f|D)$ is exchangeable in the elements of the vector η^f

Assumption 2 makes the joint distribution of elements of η^f (conditional on the state) exchangeable. This is one of the sufficient conditions for consistency of the pairwise maximum score estimator (Fox, 2007). The exchangeability property is key to making the choice set rank ordered by the observable component of payoffs. Also note that the distribution does not depend on a^{-f}. This means that regardless of what the other firm chooses the fixed costs disturbances are exactly the same for any action. Estimation of fixed costs based on this type of exchangeability assumption and the N.E. necessary conditions derived from Assumption 1 can also be found in Ellickson et al. (2013).

Assumption 3. *Conditional Independence:*

- $E(e|\eta^f) = 0$
- $E(e|d) = 0$

Assumption 3 makes the firm's fixed cost draw (type) uninformative about the marginal cost and demand shocks that are captured by the unobserved state variable e. This assumption is one of the sufficient conditions for the identification strategy proposed by Eizenberg (2014) for the variable profit parameters. Assumption 3 means that the choice of a^f is made only based on the observable shifters of demand and variable costs–d, and the fixed costs unobservable η^f which is uninformative about demand and variable cost. In turn, this means that the actions a^f should be uncorrelated with the $e$the second stage unobservable demand and variable cost shocks. This exclusion restriction is the basis for instrumenting for prices and ruling out selection bias when estimating demand. I further assume that there is one marginal cost shock and one demand shock per product-brand pair in each market quarter and that these shocks are independent and identically distributed.

So based on Assumption 3 I am able to consistently estimate demand and variable costs using standard methods. This is the first step in estimating the fixed costs of offering products and can be thought of as estimating the parameters of the function $\Pi(a^f, a^{-f}, D)$. I do this estimation in Chapter 2. The second step is to simulate $\pi(a^f, a^{-f}, d) \equiv E_e\Pi(a^f, a^{-f}, D)$ in order to estimate fixed costs. For this I use the residuals from the first stage as estimates of e. I simulate $\pi(\cdot)$ for many different alternative a^f that were not chosen, and the third step uses these alternatives to form the pairwise maximum score estimator for the parameters of the fixed cost function $C(\cdot)$. I describe the simulation of expected variable profits and the estimation of the fixed costs in Chapter 3.

1.3 Characterization of Incremental Profits

A firm will choose to add a product when the incremental profits from doing so are positive. Specifically, if a^{f+} denotes a^f with one product added, then firms will offer the set of products a^{f+} if:

$$\pi^f(a^{f+}, a^{-f}, d) - \pi^f(a^f, a^{-f}, d) \geq 0$$

If the demand and marginal cost shocks are i.i.d. (in particular their distribution does not depend on the number or type or products offered|), then the difference in expected variable profits can be characterized by the incremental variable profits, letting the new product be indexed j and the existing products be indexed by 1 up to J:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\Pi^f(a^{f+},a^{-f},d,e)-\Pi^f(a^f,a^{-f},d,e)\,dF(e_1)...dF(e_J)dF(e_j)$$

The term $\Pi^f(a^{f+},a^{-f},d,e)-\Pi^f(a^f,a^{-f},d,e)$ is the incremental variable profits and will determine the profitability of entry.

In order to characterize the incremental profits, I will re-write them in terms of the equilibrium prices and market shares. First, define p^* to be the vector of equilibrium prices if the product is introduced. The equilibrium prices when the new product, product j, is not in the market can be usefully thought of in terms of the virtual price of product j (Hausman and Leonard (2002)). There are some technical issues with virtual prices[3], here I use them simply as a notational convention to denote prices without product j. Call this price p_j^v, and denote the equilibrium prices when product j is priced at its virtual price: p_{-j}^v. Then the incremental profits from offering product j are:

$$\left(p_j^{*f}-c_j\right)s_j^f+\left(p_{-j}^{*f}-c_{-j}\right)^T s_{-f}^f(p_{-j}^*,p_j^{*f})-\left(p_{-j}^{vf}-c_{-j}\right)^T s_{-j}^f(p_{-j}^v,p_j^{vf})$$

This equation divides the incremental profits into a benefit: the markup and revenue from

[3]The idea is that if the price is large enough, it is possible to make the consumer's demand for product j arbitrarily close to 0, and with the market share essentially 0 treat the product as unavailable. Further, it is possible to choose the virtual price such that the firms revenue from product j is arbitrarily close to 0, and the prices that solve the simultaneous stationarity conditions (equation 1.3) when the price of product j is exogenously set to the virtual price are arbitrarily close to the equilibrium prices when the product is removed from the market. It is necessary that the price sensitivity is bounded in some sense (Morrow and Skerlos (2010) give some conditions for this and discussion of this issue). In particular, if there is a random coefficient on price in the consumer's utility function which has unbounded support then it is possible that there are consumers for whom prices increase utility. For these consumers, no price would set their demand to 0, and the existence of such consumers would imply that the limit of the aggregate demand as price increases would not go to 0 as price rises (which is what is required to apply the virtual price idea). In my empirical model of demand I do not place bounds on random coefficient on price so technically virtual prices are not applicable to my model, but in practice, the measure of consumers with positive price coefficients is too small to matter.

the new good, $\left(p_j^{*f} - c_j\right) s_j^f$, and the change in profits of existing products is the remaining two terms. Given the the new product is a substitute for existing products the second two terms would seem to represent a cost. I will refer to the change in profits of existing products as the cost of adding a product in the discussion that follows. However, these are differences in equilibrium profits, therefore, it is not just consumers that respond the introduction of a new product by substituting toward it, but also firms who may respond by altering prices. This means there would be additional work required beyond noting that the products are substitutes to fully sign the cost term. For example, if the firm raises the prices of its own goods, it is not clear, without additional work, that profits will fall.

It is important to consider both the price changes and substitution patterns, and to illustrate these I will subtract and add the term $\left(p_{-j}^{*f} - c_{-j}\right)^T s_{-j}^f(p_{-j}^v, p_j^{vf})$ to the cost of adding a product to decompose it into two effects:

$$\left(p_{-j}^{*f} - c_{-j}\right)^T \left(s_{-j}^f(p_{-j}^*, p_j^{*f}) - s_{-j}^f(p_{-j}^v, p_j^{vf})\right) + \left(p_{-j}^{*f} - p_{-j}^{vf}\right)^T s_{-j}^f(p_{-j}^v, p_j^{vf})$$

This divides the cost of the new product into two parts. The first part might be termed the substitution effect, and the second termed the price effect. There are several prices that might change in the substitution effect: the firm's own prices, the other firms' prices (if there is another firm), and the price of the new good falling from its virtual price. Adding and subtracting suitable terms decomposes the substitution effect into three parts:

- Own price substitution: $s_{-j}^f(p_{-j}^{*f}, p_{-j}^{*-f}, p_j^{*f}) - s_{-j}^f(p_{-j}^{vf}, p_{-j}^{*-f}, p_j^{*f})$
- Price competition effect: $s_{-j}^f(p_{-j}^{vf}, p_{-j}^{*-f}, p_j^{*f}) - s_{-j}^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{*f})$
- Cannibalization: $s_{-j}^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{*f}) - s_{-j}^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{vf})$

The cannibalization effect will be negative in the random coefficients demand model that I estimate, and that this effect is negative can be easily verified for the simple logit specification[4]. Let u_k for $k \in \{1, ..J\}$ be the indirect utilities that consumers get from consumer the existing products when the are priced according to $\left(p_{-j}^{vf}, p_{-j}^{v-f}\right)$, and take the new product $j \notin \{1, ...J\}$ with utility u_j when its price is p_j^*. Then the cannibalization effect for product k owned by firm f can be written:

$$
\begin{aligned}
s_k^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{*f}) - s_k^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{vf}) &= \\
\frac{e^{u_k}}{1 + e^{u_j} + \sum_{h=1}^{J} e^{u_h} 1\{a_h^f = 1 \, or \; a_h^{-f} = 1\}} - \frac{e^{u_k}}{1 + \sum_{h=1}^{J} e^{u_h} 1\{a_h^f = 1 \, or \; a_h^{-f} = 1\}} &= \\
\frac{e^{u_k}(-e^{u_j})}{\left(1 + e^{u_j} + \sum_{h=1}^{J} e^{u_h} 1\{a_h^f = 1 \, or \; a_h^{-f} = 1\}\right)\left(1 + \sum_{h=1}^{J} e^{u_h} 1\{a_h^f = 1 \, or \; a_h^{-f} = 1\}\right)} &= \\
-s_k^f(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{vf}) s_j(p_{-j}^{vf}, p_{-j}^{v-f}, p_j^{*f}) &=
\end{aligned}
$$

The cannibalization effect is therefore negative as expected. In general, cannibalization will affect the oligopoly less than the monopolist. Consider the case where firm f offers the same products if it is in competition or if its a monopolist, so that the only difference between the two cases is whether there are any products where $a_h^{-f} = 1$. These added products enter the denominator of both $s_k^f(\cdot)$and $s_j^f(\cdot)$ in the same way (all products that are not product j are held at their p_{-j}^v values), and therefore the multiplication of $s_k^f(\cdot)s_j^f(\cdot)$ will be smaller for the oligopoly case. The existence of competing products softens cannibalization. This effect might be referred to as business stealing and has been pointed to as driving the incentives for competitive firms to enter products in other settings (see Tirole (1994)).

On the other hand, the price competition effect does not enter into the monopolist's 'cost' term lowering his costs of adding products. I expect the price competition effect to be negative, since by an application of the mean value theorem it can be written for $\tilde{p}_{-j}^{-f}$ in the convex hull of p_{-j}^{*-f} and p_{-j}^{v-f} as:

[4]The proof for the logit model given here generalizes to the random coefficients model as long as the distribution of the random coefficients does not change with the introduction of a new product.

$$D_{p_{-j}^{-f}} s_{-j}^{f}(p_{-j}^{vf}, \tilde{p}_{-j}^{-f}, p_{j}^{*f}) \left(p_{-j}^{*-f} - p_{-j}^{v-f}\right)$$

The derivatives of the firm's market shares with respect to the other firm's prices are positive, and in empirical work, the average change in price for competitors after the introduction of a new product has been negative (Giacomo (2008)). It should be noted that in theory (and even in practice) prices may not all fall in response to the introduction, so the overall magnitude and possibly sign of this term is uncertain. However, the countervailing effects of decreased price competition and increased cannibalization for the monopolist in comparison to the oligopoly make predicting the effect of market structure on product entry and exit difficult.

Further complicating the comparison between oligopoly and monopoly, is that own price substitution and direct price effects do not have obvious signs and magnitudes. In the simple logit and nested logit models reviewed in Section 1.1.1 the sign of these terms were negative for the indirect and positive for the direct effect. New products, as long as they were of higher quality than existing products, allowed the firm to raise markups on all of its products. In the random coefficients logit model markups are not necessarily equal, however, the intuition from the simple logit case may be a good prediction. Comparing the monopoly to the oligopoly is not straightforward, since even though the monopoly may charge higher prices the change in prices after the product introduction is what matters for the incremental profits not the absolute size of the prices. Still if the monopolist enjoys greater pricing power, perhaps it faces more inelastic demand, then the indirect price effect may be smaller for the monopolist increasing the returns from adding a product.

Overall, there are two forces at work. The monopolist's pricing power, or lack or price competition, relative to the oligopoly reduces his cost when adding a product. It may also allow him to charge a higher price on the new product than the oligopolist other things being equal. But the oligopolist benefits from the reduced cannibalization, which reduces his cost of adding a product. Because of the countervailing forces at work the difference between the incremental profits for the two market structures is unclear without estimates of the demand and prices before and after the introduction of a product.

1.4 Conclusion and Preview of Chapters 2 and 3

Product choice plays a key role in market conduct. The goal of the remaining chapters is to examine the question of product choice's role in consumer welfare in Chapter 2, and in Chapter 3 to look at the effect of market structure on the incentives of firms to add or drop product offerings. Both of these chapters will estimate elements of the product choice game using data from the U.S. Yogurt Industry. In Chapter 2, I estimate the demand and variable cost curves in the U.S. Yogurt Industry from 2001-2011. This allows me to compute the incremental profits that arise when products are added or dropped by firms, which will be an important input to work done in Chapter 3. I also compute the changes in consumer welfare that have occurred after the introductions of 15 new products during the sample period. In 3, I estimate the fixed costs of offering products using the information about the incremental profits obtained from the demand and variable cost estimates. I then counterfactually change the market structure in order to predict the effect of a merger on the incentives of firms to add or drop products.

2 Product Variety in the U.S. Yogurt Industry: Demand and Welfare

2.1 Introduction

When firms adopt a new product, previous work has estimated large increases in consumer welfare[1]. These studies take a new product introduction or adoption as an exogenous event, uses estimated demand and marginal cost curves to compute welfare changes for firms and consumers, and are quick to point out that they do not say anything about the costs of firms to offer these new products. In all studies, welfare changes are in the millions of dollars for both industry profits and consumer welfare. Closely related to my work, Giacomo (2008) studies the introduction of a new low-fat and regular fat brands in the Italian yogurt industry. She finds that there is only a small 6 million euro gain by consumers directly from variety. Rather than creating new product categories firms in her paper are adopting existing product characteristics, either low or regular fat, so it is unsurprising that the direct effect of variety is small. Consumers gain 389.6 million euros due to the decreases in prices that occur after the addition of these products. The firm introducing the brand increased its average mark-up, while its competitors lowered their prices in response. Giacomo also finds that industry profits fell by 365 million euros. These studies clearly indicate that product variety plays a key role in determining

[1] For example, by Petrin (2002) in the consumer automobile industry, Hausman (1996) in the ready to eat cereal industry, Hausman and Leonard (2002) in the bath tissue industry, and Giacomo (2008) in the Italian yogurt industry.

the strength of price competition and the resulting consumer welfare, but studies in this literature take the adoption of a new product as given when it is actually a key strategic variable for firms. Additional work has found that introductions of new product categories in the U.S. may have important implications for consumer diets (Taylor et al. (2017)), and that consumers value positively many of the health related product attributes that new yogurt products have been marketed with in the last 10 years (Bonanno (2016)).

In this chapter I will estimate demand for yogurt using a random coefficients logit model, and use the results to document changes in consumer welfare that result from the introduction of 15 new yogurt product lines during my sample. I replicate the qualitative result of Giacomo (2008) that price effects account for an order of magnitude more of the change in consumer welfare than direct effects. The demand and variable cost curves estimated in this chapter will also form the basis of Chapter 3, where I estimate of fixed costs of offering products, and simulate the effect of potential mergers on the products offered in the industry. In that chapter, I hope to address the limitation of previous work that takes these introductions as exogenous and is therefore limited in what it can say about the effect of market structure on incentives to offer products.

I model firms as playing a two stage game (see 1 for more details). In the first stage firms choose which products to offer, and then, conditional on the set of products offered by all firms, firms set prices in the second stage. Eizenberg (2014) shows that this timing assumption allows demand and variable costs to be estimated using standard methods e.g. Berry et al. (1995) and Berry (1994). This estimation strategy has been standard across several recent works on product entry and exit, for example: Wollmann (2014) and Fan and Yang (2016). Given estimates of demand and variable cost, it is possible to simulate the variable profits firms would receive for offering different sets of products. Differences in variable profits across alternative choices with different numbers of products provides the variation that I use in Chapter 3 to identify the fixed costs of offering product lines. In this chapter I will use the estimates of demand to compute the changes in consumer welfare changes that occurred after the introduction of each of the new categories of yogurt in my data. I find that introductions of new products in the yogurt industry resulted in a total of $205.4 million dollars of increased consumer welfare from 2001-2011.

2.2 Data and Industry

I am going to study the relationship between market structure and product variety in the U.S. Yogurt Industry, using data from the IRI marketing database and auxiliary sources. The IRI marketing database contains scanner data on products typically offered in super markets. The data allow me to observe sales, prices, and product characteristics for yogurt at the store level weekly from 2001-2011 in a cross section of 48 geographic markets defined by IRI. The IRI geographic markets vary considerably in population from about 0.5 million to 19 million and the number of constituent counties in an IRI market ranges from 1 county to a maximum of 78 counties. I aggregate the data across weeks to quarters, and across stores to the market level. I also aggregate up the sales and price data to the product line level using my own definitions of product lines based on the IRI product characteristics, see section 2.2.2 for a detailed description of how products are defined, section 2.2.1 for a discussion of the yogurt industry and top brands, and appendix 2.1 for more detail on data aggregation.

The IRI data lists the counties contained in each IRI market, and I am able to use this information to supplement the IRI marketing database with demographic and cost data. I use data from the 2000 U.S. Census to explain difference in demand for yogurt across markets. I include data on the median income, percent white, and percentage of household with children. In order to model variable costs: I pull data on county level average weekly wages from the Quarterly Census of Employment and Wages, and annual state level commercial electricity retail prices from the U.S. Energy Information Administration. In addition I was able to determine the locations of the manufacturing facilities where all of the main brands of yogurt are produced from their websites, customer service representatives, and local news articles accessed online. I use this information to calculate the average distance from each plant to the counties in each geographic market. The distances to manufacturing facilities turn out to be an important cost shifter and a strong valid instrument for prices. The last auxiliary data comes from the annual intercensal population estimates from 2000-2011 provided by the U.S. Census bureau, which I use to determine the number of possible yogurt sales in a geographic market and time period. I calibrate the market size to be one yogurt per person per week–so thirteen times the population. This strategy matches that used by Villas-Boas (2007) and Giacomo (2008)for estimating demand for yogurt. I update the market size annually to allow for changing population over the decades.

Tables 2.1, 2.2, 2.3 give descriptive statistics for demographics, prices and competitive conditions, and cost data respectively. Table 2.1 reports some standard demographics from the 2000 U.S. census; median income is reported in thousands of U.S. dollars and population in thousands. 'Dist. to Closest Plant' is the distance in kilometers to the nearest manufacturing facility for a products brand. ' The large maximum of 3356 km is closer than the distance from New York, NY to Los Angeles, CA and is consistent with the location of several smaller brand's manufacturing facilities in upstate NY and New England (e.g. Chobani in New Berlin, NY, and Stonyfield in Londonberry, NH). The two largest brands Dannon and Yoplait have several manufacturing facilities spread out across the midwest, south, and west, which sets them apart from the smaller producers. Table 2.2 reports descriptive statistics for variables that vary across quarters and overtime . 'Electricity Price' is the average price paid for electricity for commercial use in cents per kilowatt. The average weekly wage is the average wages in U.S. dollars per week for workers in a market-quarter. Each product-quarter-market observation is an aggregation of several UPCs across weeks and stores, and the variable 'Avg. Frac. Featured' records the fraction of UPC-week-store observations where a product was recorded by IRI as having been featured. 'Avg. Frac. of Stores with Product' looks at each product and calculates the fraction of stores during that market-quarter observation that carried that product line. 'Frac. of Stores with Private label' is simply the fraction of stores in a given market-quarter that carried one or more private label yogurt brands. I also report the average fraction featured and fraction of stores with product for each brand in a market-quarter. 'Num Stores' is the number of retail stores (in the IRI sample) in a market quarter observation. 'Retail HHI' is the Herfindahl Index computed for the retailers using IRI's estimated total sales per store. Finally, Table 2.3 shows descriptive statistics for the price of a standard 6oz unit of yogurt and other competitive variables. It also reports 'Frac. Featured' which is the number of store-weeks a given product is featured–this is aggregated to become 'Avg. Frac. Featured', and it reports fraction of stores that offered a given product ('Frac. Stores with Product')–which is aggregated to become 'Avg. Frac. of Stores with Product'. The distributions of the disaggregated versions of these variables are similar to the aggregated versions, implying that there is little within brand-market-quarter variation across products in these variables and supporting the use of the aggregate versions when estimating demand.

Table 2.1: Descriptive Statistics: Demographics

Variable	n	max.	mean	min.	std. dev.
Median Age 2000	47430.0000	38.3487	35.0380	27.4391	1.8502
Median Income 2000	47430.0000	64.9065	44.5703	29.9851	6.7108
Perc. White 2000	47430.0000	96.5593	77.4488	54.4219	11.3994
Population 2000	47430.0000	18893.1120	3629.6957	526.6240	3475.3177

Descriptive statistics for data that is only at the market level. For demographics: variables reported are weighted averages across counties within IRI geographic markets (using census weights) of variables from the 2000 U.S. census. "Median Age 2000" is the average median age in years. "Median Income 2000" the average median income in 1000s of U.S. dollars. "Perc. White 2000" is the average percentage of the population that is white. "Population 2000" is the total population in 1000s of persons of each the IRI market in 2000. 'Dist. to Closest Plant' is the distance to the closest plant for each manufacturer and is measured in kilometers.

2.2.1 Industry and Firms

The yogurt industry in the United States is dominated by two major brands: Dannon and Yoplait. Combined these two firms control about eighty percent of sales over the sample period, and each controls around forty percent of sales. Dannon and Yoplait each

Table 2.2: Descriptive Statistics: Market level variables

Variable	n	max.	mean	min.	std. dev.
Median Age	48	38.3487	34.9358	27.4391	1.9266
Median Income	48	64.9065	44.4122	29.9851	6.8852
Dist. to Closest Plant	48	3356.7326	974.5851	42.3331	714.4255
Percent White	48	96.5593	77.2394	54.4219	11.5610
Population (2000)	48	18893.1120	3532.9965	526.6240	3390.4963

Descriptive statistics for price and promotion data. "Brand Avg. Frac. Featured" is the average fraction of store-weeks in a market-quarter pair that a brands products are featured. "Brand Avg. Stores with Product" is the average fraction of stores carrying a brands products in a market-quarter observation. "Frac. Featured" is the fraction of store-weeks in a market-quarter pair that a product is featured. "Frac. Stores with Private Label" is the fraction of stores in market-quarter observation that offer a private label yogurt product. "Frac. Stores with Product" is the fraction of stores carrying a brands products in a market-quarter observation. "Price" is the average price of a product per standard 8oz unit of yogurt during a market-quarter observation. "Stores per Market" is the number of stores per IRI geographic market during a quarter.

offer several product lines in most of the IRI markets. While the industry is dominated nationally by the leading manufacturers, competition from regional and niche producers plays an important role in determining the degree of price competition and profitability of yogurt products.

There are several smaller brands that appear in the data. Some smaller brands are also associated with premium products, quality or local ingredients, and particular styles of yogurts. Others simply represent regional dairies that have traditionally produced yogurt. Finally some brands are new entrants into the Yogurt market. For example, greek yogurt made Chobani one of the top yogurt brands in the country with about 25% of total yogurt sales in 2015Giammona (2015). In order to account for competition from secondary brands, I will model demand for their products and allow them to adjust prices in response to changes in product variety, but for simplicity I leave them out of the product level entry and exit model by assuming they would offer the same products regardless of what Dannon and Yoplait choose to offer. In this way I include the next ten largest brands, after Dannon and Yoplait, by total sales over the sample. I do not consider the demand for any brands outside of this set of 12. Figure 2.1 gives the complete

Table 2.3: Descriptive Statistics: Product level variables

Variable	n	max.	mean	min.	std. dev.
Frac. Featured	47430	1.0000	0.0820	0.0000	0.0830
Frac. Stores with Product	47430	7.8177	2.0956	0.6326	0.7022
Price per 6oz	47430	0.9474	0.5281	0.0056	0.2233

Descriptive statistics for cost shifters. "Avg. Weekly Wage" is computed using the county average weekly wages reported in the Quarterly Census of Employment and Wages averaged over the counties in the IRI geographic market. "Electricity Price" is the average retail price of electicity for commericial use average over the counties in the IRI geographic market. "Min. Dist. Firm." is the average distance in 1000s of km from the geographic centers of the counties in the IRI geographic market to the nearest yogurt manufacturer. "Retail HHI" is an HHI index computed for retail chains in each IRI geographic market using total sales reported by IRI.

list of 12 brands that I consider, and shows their share of industry sales by quarter over the period[2].

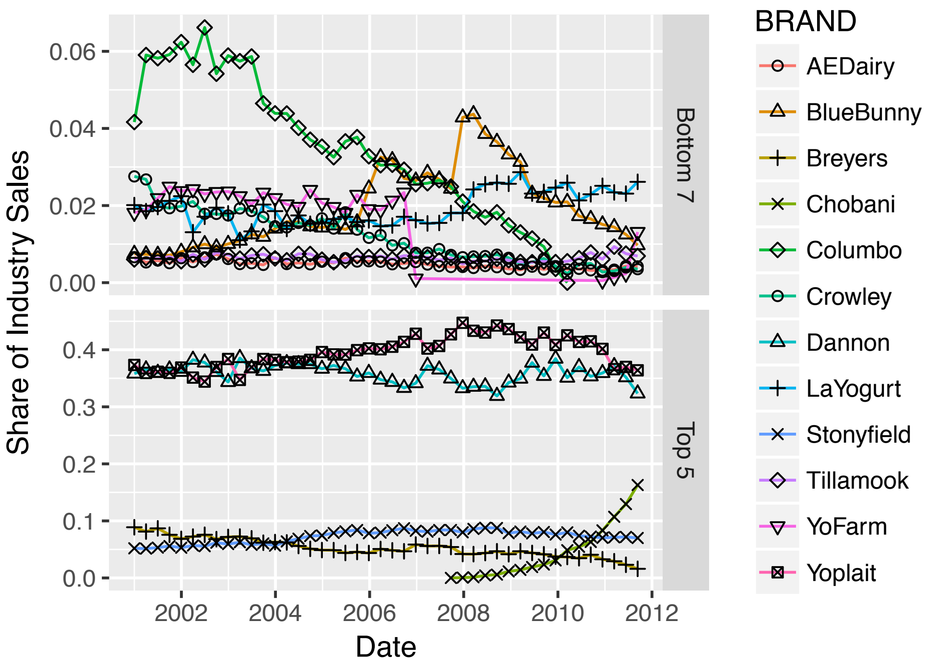
Figure 2.1 shows that Yoplait and Dannon are consistently getting between 30 and 45 percent of industry sales. The next three producers Chobani, Breyers, and Stonyfield each at some point acheive around 10% of industry sales. Chobani of course is a successful new entrant with rapidly increasing sales. Stonyfield entered the yogurt market producing organic and premium yogurts prior to the sample and have been successful throughout. They were acquired by the same holding company that owns Dannon in 2003, but I model their brand separately since they appear to have maintained their own brand identity throughout the sample. Finally, Breyers yogurt has been declining throughout the sample. Ownership of the brand changed hands twice during the sample as its value decreased. Further complicating matters is that the YoFarm brand was marketed by Breyers from 2005-2010, which explains YoFarm's lack of sales over this period. Besides the ownership change for YoFarm the rest of the bottom five yogurt producers in my sample have relatively consistent or declining and small shares of the industry.

The competition from regional and niche producers means that in many markets there will be several firms offering any given product line. In a counterfactual that removes

[2]Share of industry sales is not the same as market share. Market share is the share of potential sales, while share of industry sales is the brands share of total realized sales by the industry

this competition, there will be three potential effects on the variety of products that are offered. First, removing niche producers will decrease competition in these product categories increasing the incentive to enter. On the other hand, there will be a reduction in the business stealing incentive. This is because when a firm introduces a product, mechanically more of the added product's sales will be cannibalized from the firm's own existing products rather than competitors' products. Finally, prices for traditional yogurt products that where Dannon and Yoplait face the most competition from regional producers may rise when that competition is removed. Increases in prices for traditional yogurts may reduce consumer to substitution away from niche products toward cheaper traditional offerings.

Figure 2.1: Share of Industry sales by quarter



Plot of each brand's share of total industry sales using standardized 8oz units. The top graph shows the top 2 brands and three other producers with national level distribution and brand recognition. The bottom graph shows smaller producers who are largely regional producers of yogurt with limited brand recognition.

2.2.2 Product Categories

The yogurt industry in the U.S. has been driven by innovation. Originally, plain yogurt was a staple food for some european immigrant groups, but in the 1950s Dannon added fruit to their yogurt creating a sweeter product that appealed to a broader group of American consumers. From the 1980s onward yogurt expanded to new categories. Dannon launched Danimals(TM) target to kids in 1994. Yoplait responded with its successful Go-Gurt(TM) brand aimed at kids in 1999. The pattern of product innovation continued in the 2000s. Yogurt companies responded to diet fads with high-fiber and low carb yogurts, and added yogurt beverages with Dannon beginning production of its drinkable yogurts in 2000[3]. The history of new product introductions makes the yogurt industry an ideal setting to learn about product variety in consumer goods industries.

I divide the yogurt market into seven product categories: Greek, Lite, Active, Fiber, Carb, Kids, and Drinks[4]. Products can belong to several categories. I define a product line to be a unique combination of category membership. A yogurt might be both targeted towards Kids and Health. In order to construct the product categories and assign UPC codes to product lines I combined the existing product characteristics coded in the IRI data with information that could be gleaned from reading the product description leaf[5].

[3]Information on these product introductions comes from the manufacturer webpages. Introductions in the period 2001-2011 are covered by my data.

[4]Lite is yogurt that appears targeted towards health conscious consumers sometimes with restricted calories, but I do not reliably observe calories and instead rely on the branding containing the term 'lite' or 'light'. Active yogurt contains pro-biotic or active bacterial cultures. Fiber yogurt has added dietary fiber, and Carb has lower carbohydrates which would usually require reduced sugar content. Kids yogurt is marketed toward kids, and the branding usually makes this clear–I include in this category tie-in products with Disney characters or other media properties under the assumption this is directed at children. Greek yogurt is a thicker yogurt, which is based on a traditional method of straining moisture out of the yogurt, but many firms rely on thickeners and other techniques to replicate the traditional method. Drink yogurts are thinner yogurts designed to be consumed as a beverage.

[5]The product description leaf L5 provides the "brand name" of a yogurt e.g. 'Dannon Light'. These descriptions provide some information in and of themselves: 'Dannon Light' is a light yogurt, further, the product description can be google searched and the results of the search used to further confirm the product characteristics. Often searching for the product description returns images of the packaging, a manufacturer's page with a more detailed description, nutritional information, advertisements, and even customer reviews. I made heavy use of images of the packaging to gather information about product characteristics, since the packaging tends to contain marketing information about a products intend ended category.

Product lines abstract away the many different versions of a product line that might be offered. Yogurt producers usually offer several flavors in each of their product lines. In addition I observe several other characteristics that firms could use to create different versions of their product lines like fat content, and textures. I capture differentiation within products by including brand-product interactions in demand estimation. This allows each firm to occupy a different position along a single axis of differentiation within each product market. I am forced to limit the number of categories because while I am able to handle a high dimensional choice set when firm's choose products using maximum score, demand estimation is still based on the logit model and additional products requires additional product fixed effects increasing the number of parameters. The type of product differentiation helps to predict the equilibrium product variety. Since the industry is characterized by a set of discrete product characteristics and a corresponding discrete set of products it is not as easy to create close substitutes to a competitors product that differ in more ways than brand identity. If there is more scope for differentiation along observable characteristics firms would be able to reduce the competition for their products. In the way I model differentiation firms have a limited ability to spread out within product markets leading to increased competition within markets, and removing competition from other firms within each market should therefore make entry more profitable.

Product lines are the key unit of observation for this study. I aggregate the UPC level price observations to the product line level, by calculating an average price per ounce for each product line. Using these prices and the observed quantities I estimate demand for product lines and the decision of by firms of which product lines to offer in each market and time period. Table 2.4 shows all of the potential product lines. These are combinations of product categories that I observe in the sample. Firms are not limited to choose from among this set since when I estimate the model. The method of maximum score is still consistent even when only part of the choice set is observed. Restricting attention to only product lines that I observe prevents me from considering products that were actually outside of the choice set. I also avoid ad hoc restrictions on the choice set.

Table 2.5 reports descriptive statistics for the number of product offered by each brand. Dannon and Yoplait stand out with an average number of products over the sample of 9.5 and 5.61 respectively, and both of their maximums are a bit more than twice their minimum number of products, which is driven by the net entry of products that has occured in the industry. BlueBunny, Breyers, and Stonyfield offer several products and have national distribution for most of the sample period. Stonyfield focuses on niche products, while Breyers and BlueBunny focus on more traditional yogurt products. The remaining firms are all regional diaries during the sample, except Chobani which does have national level distribution by the end of the sample even though it focuses on only one product.

2.3 Demand Model and Estimation

In order to evaluate the welfare impacts of the introduction of new products, I will need to predict prices in counterfactuals where those products are not offered. This requires estimating both the demand and marginal cost curves to use as inputs for those counterfactuals. In Chapter 3 I will also use both curves together to assess the profitability of offering alternative sets of products, and use differences in profitability between the observed set of products and alternatives that were not offered to estimate the fixed costs of offering products. Estimating these two components can be done using the familiar two step procedure of estimating demand (recovering the revenue function) and then using the N.E. pricing first order conditions to recover marginal cost. First in section 2.3.1, I will lay out the demand specification, then explain the how I take the model to the data, and identify the parameters of interest. Second in section 2.3.2 I show how the demand estimates can be used to recover variable costs and estimate a model for those costs using my data on cost shifters.

Table 2.4: Potential Product Lines

Product Name	Greek	Kids	Drink	Lite	Active	Carb	Fiber
Regular							
Lite				X			
Carb						X	
Kids		X					
Kids Drink		X	X				
Active					X		
Lite Active				X	X		
Lite Carb				X		X	
Drink Lite Carb			X	X		X	
Lite Fiber				X			X
Drink			X				
Drink Lite			X	X			
Drink Active			X		X		
Drink Lite Active			X	X	X		
Drink Carb			X			X	
Greek	X						
Fiber							X
Active Fiber					X		X
Greek Active	X				X		
Greek Kids	X	X					
Greek Drink	X		X				

"X" indicates that a product line is in the product category indicated by the collumn. These 21 product lines are the combinations of product categories observed in the data.

2.3.1 Demand Estimation

Estimating demand when the set of offered products is endogenous presents a possible selection problem. Shares and prices are only observed when firms choose to offer a product. Consistent estimation of the demand parameters will therefore have to rely on the two stage timing of the pricing and product choice decisions, and on Assumption 3 which makes the firm's information at the time it chooses products uninformative about later shocks to demand and marginal cost. At the time that the firm chooses a^f_{tm} the firm has observed: η^f_{tm}, η^{-f}_{tm}, and d_{tm} (which contains the demographics d_{0m} as well as marginal cost shifters). Since $\Delta\xi^f_{mjt}$ is one of the unobservable e_{tm} state variables, $\Delta\xi^f_{mjt}$ has mean zero expectation conditional on each variable in the information set by Assumption 3. Because a^f_{tm} is a function of only the variables in the information set, $E(\Delta\xi^f_{mjt}|a^f_{tm}) = 0$.

Table 2.5: Number of Products by Brand

Brand	Mean	Std. Dev.	Min.	Max.
AEDairy	2.00	0.00	2.00	2.00
BlueBunny	3.82	1.39	2.00	7.00
Breyers	2.66	0.78	2.00	4.00
Chobani	1.00	0.00	1.00	1.00
Columbo	1.97	0.16	1.00	2.00
Crowley	2.41	0.76	1.00	3.00
Dannon	9.50	2.95	4.00	13.00
LaYogurt	2.00	0.00	2.00	2.00
Stonyfield	3.27	0.73	2.00	5.00
Tillamook	1.00	0.00	1.00	1.00
YoFarm	1.07	0.26	1.00	2.00
Yoplait	5.61	1.48	3.00	8.00

Descriptive statistics, computed across all market and quarter observations, for the number of product lines offered by each of the 12 brands.

This is essentially a selection on observables argument, which may not be appropriate if firms have access to special knowledge about demand shocks (like market research) that I do not observe[6]. However, there are only 6 instances where a brand offers a product for a year of less. There are also no cases where a product persists in being offered in only a single market. This suggests that when firms offer products they are responding to longer term changes in consumer taste rather than market quarter specific shocks.

The model of demand I use is based on the following specification for the indirect utility for consumer i purchasing product j offered by firm f at time t in market m:

$$u^f_{ijtm} = p^f_{jtm}\alpha_{ip} + r^f_{tm}\alpha_r + \mathrm{f}^f_{tm}\alpha_{\mathrm{f}} + \xi^f_{jt} + \Delta\xi^f_{mjt} + \epsilon^f_{ijtm}.$$

The price p^f_{jtm} has an individual specific coefficient α_{ip}. The parameter α_{ip} determines the

[6]For a complete discussion of this assumption see: Eizenberg (2014) who first proposed this modeling paradigm. Recent work by Cilberto et al. (2015) allows for correlation between the fixed cost shocks (firm types) η^f_{tm} and the unobserved state variable e_{tm}. The authors of that paper find evidence for selection bias in the airline industry. I would argue that entry and exit decisions by airlines are more likely to respond to transitory market specific shocks to demand than are product choice decisions, but also concede that a model without the selection on observables argument is probably preferably. Their model is already computational challenging with binary entry and exit choices by firms, and it is not clear that it would be feasible to translate to the high dimensional product choice setting without restrictive assumptions to reduce the size of the choice set.

markups and substitution patterns. Higher average α_{ip} leads to more inelastic demand and higher markups. Cross price elasticities also depend on the distribution of α_{ip}, and when its variance increases the demand system will deviate more from the independence of irrelevant alternatives (IIA). I specify α_{ip} as a random coefficient with distribution $N(\alpha, \sigma_\alpha^2)$. I will estimate the model using the nested fixed point estimator of Berry et al. (1995).

Prices p_{jtm}^f are still endogenous even if selection is ruled out because they are chosen in the second stage after firms observe the demand and marginal cost shocks. Therefore a set of instruments for prices and these variables will be required. Eizenberg (2014) proposes to use counts of the number of products of each category offered by each firm, the standard Berry et al. (1995) (BLP) instruments. Since $E(\Delta\xi_{mjt}^f | a_{tm}^f) = 0$ implies that functions of either firm's chosen set of products will be valid instruments. I also know the average distance of each market to the firms' manufacturing facilities. I use this as an instrument for price, and refer it as the Cost instrument. The Cost instrument is valid under the assumption that the locations of manufacturing facilities are independent of the $\Delta\xi_{mjt}^f$ demand shocks. During the sample period I am not aware of any major changes in the locations of manufacturing facilities in the yogurt industry. Given that the locations are plausibly determined prior to the demand shocks their independence can be justified by the same type of timing argument as the BLP instruments. However, the locations and other cost shifters do not vary by product making these unsatisfactory instruments for prices. Berry and Haile (2014) point out that identification of the random coeficients requires one additional instrument per coefficient[7]. For product j offered by firm f, in a market-time observation, I count the number of products in a_{-j}^f and a^{-f}in each of the seven product categories. This gives me a total of 14 instruments, and in estimation I use a third degree polynomial of each instrument for a total of 42 instruments. This gives

[7] They consider the problem from the perspective of estimating inverse demand functions. Price enters the inverse demand function both linearly through the mean utility with parameter $\alpha$and through the BLP fixed point (see equation 33 in Berry and Haile (2014)). The BLP fixed point in the specification I consider is parameterized by the single random coefficient on price. To identify this parameter seperately and additional exclusion restriction is needed.

me more than enough instruments to identify the coefficient on price and the random coefficient parameter.

The variable ξ^f_{jt} represents the average taste for product brand pair at time t. I handle these parameters in two steps as in Nevo (2001). When I estimate demand the ξ^f_{jt} are treated as brand-product-time (BPT) fixed effects[8]. Later I decompose the BPT fixed effects as a function of the brand, product line, and time period. I use a non-parametric kernel regression for discrete variables from Li and Racine (2007) (Nevo (2001)uses OLS to do this decomposition). Since the regressors in the decomposition are discrete using OLS is equivalent to splitting the sample and taking sample averages. The kernel regression avoids splitting the sample and therefore can perform better in small samples. It also provides some flexibility to handle interactions between brands, products, and time. The estimating equation for this kernel regression can be written:

$$\xi^f_{jt} = h(f, x_j, t) + \dot{\xi}$$

where $h(\cdot)$ is the unknown regression function, and x_j is a vector of dummies indicating the category membership of product j, and $\dot{\xi}$ denotes the deviation of the BPT fixed effect from the mean effect of brand, product characteristic and time period. This means that if I were to predict demand for a product in period t that was not actually offered by firm f the model would combine information from similar products offered by firm f, the same and other similar products offered by the rival firm, and the same and other similar products offered by both firms in the previous and subsequent time periods.

The variable r^f_{tm} is the log of a brand's average fraction of stores offering its products

[8] I also estimate the model with other fixed effect specifications. Estimates without BPT fixed effects lead to inplausably low estimates of αeven in specifications with instruments for prices–the elasticity of demand would imply negative marginal costs. This is because the BPT fixed effects capture important ommitted variables. For example it appears that both brand-product interactions are important as well as brand-time interactions and ommitting them biases estimates. Therefore despite a risk of overfitting the only way to fit a reasonable model is to include the BPT fixed effects. See Appendix 2.2 for estimates using different fixed effect specifications.

within market m at time t. The parameter α_r is intended to model the increase in demand that should occur when products are offered at more stores in a market. Increased adoption of a product by stores lowers the effective price, increasing the indirect utility from the product. This effect is particularly important in the yogurt industry because yogurt is sold in refrigerated isles shelf space for yogurts is more scarce than for other grocery items. Competition by firms to get their products adopted by stores is therefore important in the industry as documented by Hristakeva (2016). Potentially offering more products in an industry may crowd out competing products or the firm's own products. Also an additional product may not be accessible to consumers since it will only be on the shelves at a few stores. By allowing the demand model to capture the response of demand to the number of stores offering a product, I capture the firm's marginal revenue that is generated by store level adoption decisions. I abstract away from the process that determines the equilibrium adoption, and simply include a brand's average fraction of stores offering its products. Modeling the process of store level adoption as a function of the products To understand this simplifying assumption let r^f_{jtm} denote the log of the fraction of stores offering product j, and consider a simple model for r^f_{jtm}:

$$r^f_{jtm} = r^f_{mt} + r^f_{jt} + \Delta r^f_{jtm}.$$

This model decomposes the product specific value into two parts: the average adoption of a brand in a market at time t, the average adoption of a brand-product pair at time t across markets, and a market specific shock to store level adoption of brand f's product j. If the true model of indirect utility included r^f_{jtm} multiplied by a coefficient α_r then substituting and re-organizing terms would yield:

$$
\begin{aligned}
u^f_{ijtm} &= p^{f\prime}_{jtm}\alpha_{ip} + \alpha_r r^f_{jtm} + \mathrm{f}^f_{tm}\alpha_{\mathrm{f}} + \xi^f_{jt} + \Delta\xi^f_{mjt} + \epsilon^f_{ijtm} \\
&= p^{f\prime}_{jtm}\alpha_{ip} + \alpha_r r^f_{mt} + \mathrm{f}^f_{tm}\alpha_{\mathrm{f}} + \left(\xi^f_{jt} + \alpha_r r^f_{jt}\right) + \left(\Delta\xi^f_{mjt} + \alpha_r \Delta r^f_{jtm}\right) + \epsilon^f_{ijtm}.
\end{aligned}
$$

This can be re-interpreted as the original model proposed for indirect utility if the term $\alpha_r r^f_{jt}$ becomes part of the BPT fixed effects, and the term $\alpha_r \Delta r^f_{jtm}$ becomes part of the market specific demand shock. Implicit in this reduced form approach is that store level adoption of product j depends only on the total number of products offered through the coefficients β_{r1} and β_{r2}, that the deviations Δr^f_{jtm} are treated as random mean zero shocks when firms make their choices of products in the first stage. Estimating the simple linear model for r^f_{jmt} gives an adjusted R^2 of 80%, which implies that this simple specification even if unrealistic does explain a significant portion of the observed variation in store level adoption of products. This reduced form approach also makes predicting demand after counterfactual changes in products simple, avoiding the need for an additional structural model to make these predictions. I also include, f^f_{tm}, the average percentage of store-weeks in a market quarter that a brands products are featured (advertised or promoted in some way). This variable can also be treated in the same reduced form way as the store level adoption, and the coefficients in the indirect utility function re-interpreted again.

To complete the model of demand, I let $u_0 = -d_{0mt}\gamma_d + \epsilon_0$ be the indirect utility from consuming the outside option, where d_{mt} is a vector of market level demographics. I take three demographsic from the the 2000 U.S. Census and they are: normalized median household income, percentage of the population that is white, and the median age. I also include the number of stores that offered private label yogurt in a market during that quarter. Any of these demographics could also be added to the equation for store level product adoption and their coefficients re-interpreted as reduced form estimates.

Table 2.6 reports the coefficients on price, the elements of P^f_{tm}, and demographics. I normalize all of the demographic variables, so the coefficients are marginal effects of increasing a variable by one standard deviation. The cost instruments and BLP instruments provide comparable but different estimates of the price sensitivity parameter. For more details about the instruments see Appendix 2.2. Estimating the model with a random coefficient on price provides similar estimates of mean price sensitivity. The own price elasticity evaluated at the mean price, mean market share, and mean price sensitivity α_p is approximately . The mean marginal effect of price on market share is .

Table 2.6: GMM Estimates of Demand

	No IV	BLP (poly.)	Rand. Coef.
Price	-1.0480^{***}		
	(0.0178)		
Brand Avg. Frac. Featured	2.2404^{***}	0.4631^{**}	0.8706^{***}
	(0.0856)	(0.1445)	(0.0005)
Brand Avg. Stores with Product	5.0461^{***}	3.9842^{***}	4.2277^{***}
	(0.0358)	(0.0657)	(0.0026)
Frac. Stores with Private Label	-1.7766^{***}	-1.0842^{***}	-1.3883^{***}
	(0.0376)	(0.0622)	(0.0031)
Med. Income	0.2651^{***}	0.3999^{***}	0.4009^{***}
	(0.0045)	(0.0082)	(0.0046)
Perc. White	0.0729^{***}	-0.0657^{***}	-0.0286^{***}
	(0.0049)	(0.0089)	(0.0046)
Med. Age	0.2563^{***}	0.3311^{***}	0.3241^{***}
	(0.0048)	(0.0077)	(0.0046)
Price (IV)		-5.4944^{***}	-4.3392^{***}
		(0.1527)	(0.0101)
Num. obs.	47430	47430	47430
σ_α			0.4633
s.e. σ_α			0.0106

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Estimates of demand for yogurt; all of the models have brand-product-time fixed effects. The first four variables are defined in Table 2, and the next three in Table 1. "No IV" reports estimates without instrumenting for price. "BLP (poly.)" reports estimates instrumenting for price with polynomials of the BLP instruments. "Rand. Coef." reports that model with a random coefficient on price. The standard deviation of the random coefficient is listed as σ_α.

2.3.2 Variable Costs

Given estimates of the price sensitivity parameter and the observed market shares and prices it is possible to infer the variable costs using the first order condition for N.E.

pricing. To see this organize the shares and prices in a market at a given time into vectors: $s_{tm} = \left(...s^f_{jtm}...\right)$ and $p_{tm} = \left(...p^f_{jtm}...\right)$. Let $c_{tm} = \left(...c^f_{jtm}...\right)$ be the per unit cost of yogurt. Then the firm's first order conditions can be written:

$$s_{tm} + (D_p s_{tm})(p_{tm} - c_{tm}) = 0$$

The matrix $(D_p s_{tm})$ is the 'intra-firm' Jacobian of the market shares: $(D_p s_{tm}) \equiv \frac{\partial}{\partial p_k} s_{jtm} 1[a^f_{jtm} = a^f_{ktm} = 1 \; or \; a^{-f}_{jtm} = a^{-f}_{ktm} = 1]$. The first order condition can be solved for the variable costs:

$$c_{tm} = (D_p s_{tm})^{-1}(s_{tm}) + p_{tm}$$

Under the assumption that firms choose prices according to the NE first order condition, the observed shares and prices can be plugged in to yield estimates of the variable costs. Table 2.7 gives descriptive statistics on the recovered mark-ups and variable costs. Having recovered the variable costs I regress them on cost shifters denoted d_{ctm}, BPT fixed effects denoted ζ^f_{jt}, which leaves a residual $\Delta\zeta^f_{jtm}$:

$$ln \, c_{jtm} = \nu d_{ctm} + \zeta^f_{jt} + \Delta\zeta^f_{jtm}$$

I again decompose the BPT fixed effects using a non-parametric regression: $\zeta^f_{jt} = h_c(f, x_j, t) + \dot{\zeta}$. Table 2.8 reports the estimates of the marginal cost model, and I normalize electricity prices and wages for regression, so coefficients should be interpreted as marginal effects of one standard deviation changes in a given variable.

Table 2.7: Markups and Variable Cost Descriptive Data

Variable	n	max.	mean	min.	std. dev.
Marg. Cost	47430.0000	7.5927	1.8629	0.3997	0.7010
Markup	47430.0000	0.3192	0.2327	0.2149	0.0034
Price	47430.0000	7.8177	2.0956	0.6326	0.7022

"Marg. Cost" is the marginal or equivalently variable cost recovered using the N.E. pricing conditions. "Markup" is the difference between the price and variable cost per unit. "Price" is the price, as defined in Table 2, for comparison.

Table 2.8: OLS estimates of Marginal Cost model

	Log Marginal Cost
Intercept	-0.484449^{***}
	(0.067835)
Min. Dist. Firm	0.000069^{***}
	(0.000000)
Commercial	0.009184^{***}
	(0.000235)
Avg. Weekly Wage	0.000406^{***}
	(0.000007)
R^2	0.881318
Adj. R^2	0.877322
Num. obs.	47430
RMSE	0.117374

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Regression of log marginal cost on cost shifters (see Table 3 for definitions) and brand-product-time fixed effects. Coefficients reported as average marginal effects. 'Min. Dist. Firm' is in kilometers here.

2.4 Product Variety and Welfare

When new products are introduced they cause changes in prices and give consumers new options to purchase. Hausman and Leonard (2002) provide a strategy for decomposing the total change into the two effects. Their decomposition takes advantage of the idea of virtual prices (see Chapter 1 Section 1.3 for more on virtual prices). The virtual price is a price that sets the demand for a product equal to 0, in the case of demand based on the logit model with no income effects such a price does not exist but the demand for a good can be made arbitrarily close to 0[9]. The compensating and equivalent variation for a change in price from p^b to p^a for an individual in logit models without income effects are both equal to the difference in the individual's inclusive values divided by their price sensitivity parameter (McFadden (1981)). The aggregate consumer welfare is the expectation of the individual equivalent variations with respect to the distribution of the random coefficients:

$$EV_i(p^a, p^b) = \int_{-\infty}^{\infty} \frac{IV_i(p^a) - IV_i(p^b)}{\alpha_i} dF(\alpha_i)$$

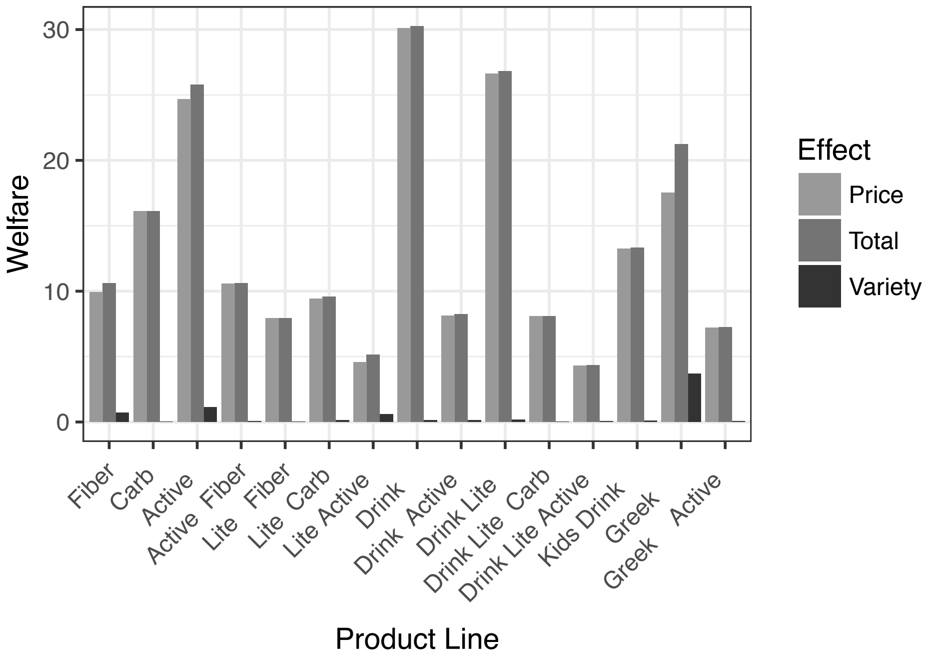
Let p^a be the prices after a new product is introduced, and p^b be the prices before hand. The decomposition of the total welfare change of a product introduction is based on calculating $EV_i(p^v, p^b)$ and $EV_i(p^a, p^v)$, where p^vsets the price of a new product to the virtual price. $EV_i(p^a, p^v)$ measures the direct effect of the new product, while $EV_i(p^v, p^b)$ isolates how welfare changes after the other products prices adjust for the product introduction.

I implement these measures for 15 new products introduced during my sample period. These products are each by different brands, and are not introduced in every geographic market in very period. Therefore I calculate the welfare changes at the for each brand-product pair's introduction to each geographic market. I then aggregate across brands and geographic markets to get the total effects for each product. For each brand-product-market pair's introduction date, I determine a post-introduction period. I define the post introduction period as the quarters after and including the first quarter the product is introduced where all products and firm identities in the market remained the same. In this way I do not conflate the welfare effects of later product introductions (and eventual exits) with the effect of the new products. Figure 2.2 shows the results for

[9] In practice computational software will often have a special constant the exponential of which is defined to be exactly 0 and I use this to get 'exact' virtual prices. Also note that the virtual prices here only need to make it as if the product is not available to consumers–prices are simulated based on the product actually being in the market or out of the market.

each of the 15 product lines in millions of U.S. dollars. The price effects account for most of the total welfare changes, about 95% on average. This is similar to the findings in Giacomo (2008) where the direct effect of the introduction of two product lines by a single brand was 6 million euros compared to a total change of 395.6 million euros. In both cases many of these product's characteristics do not differ much from existing products. Greek yogurt is one of the main exceptions with a direct effect of around \$3.6 million dollars. This introduction is the first product line in the Greek category and of course has been successful since its introduction (the total effect accumulating till the present is likely larger than my sample indicates). Other categories like Drink with \$1.1 million and Fiber with \$0.71 million also have large direct effects when their first product line is introduced. Clearly, the price effects are important, and consumers may gain significantly with the introduction of products that are not innovative like Greek yogurt has been. This means that concerns about product choice and variety are relevant to mergers even outside of industries traditionally associated with these issues due to their innovativeness or emphasis on vertical quality. In Chapter 3 I will consider the incentives of firms before and after a merger to introduce products (or adjust the set of products offered) in the U.S. Yogurt Industry.

Figure 2.2: Welfare Effects of New Products



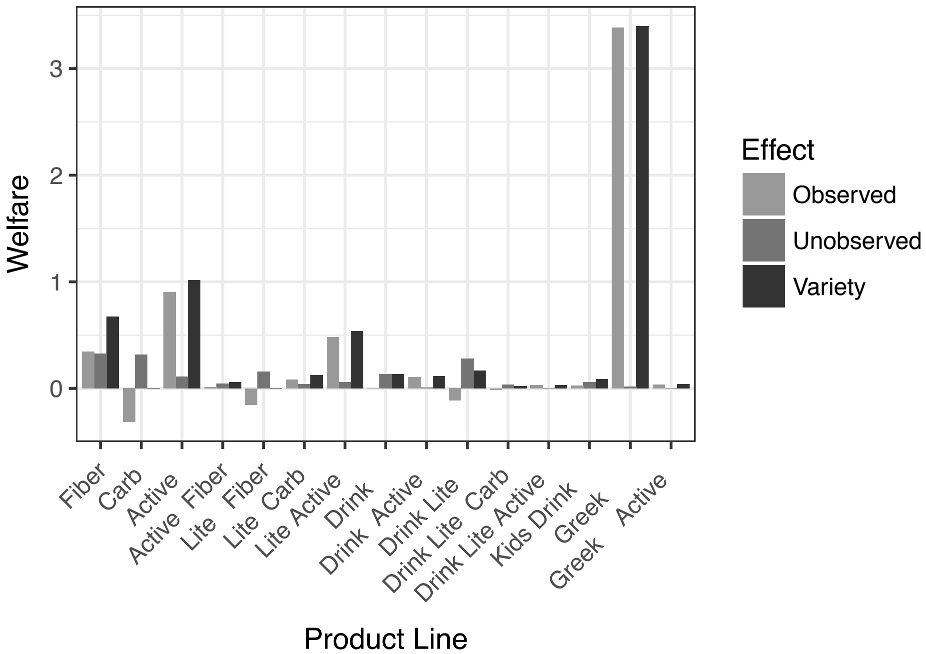
Effects on consumer welfare, in millions of U.S. dollars, of introductions of 15 products appearing for the first time from 2001-2011. Calculated based on average post-introduction costs and taste.

A concern when computing welfare with demand curves based on the logit model, is that the value of a new product may be partially due to the addition of a new draw from the distribution of the type-1 extreme value errors (T1EV). Both Petrin (2002) and Ackerberg (2005) raise concerns about the role of the unobserved errors in measures of consumer welfare. Petrin (2002) finds that without random coefficients and detailed demographic data unobserved product quality and the T1EV errors account for most of the value of a new product and overstate the total value of the new product (in fact in the simple logit specification he reports that consumers dislike the observable product attributes and gained welfare purely from the unobserved attributes). Ackerberg (2005) are similarly concerned that new products in the standard logit type models do not crowd each other out. This is because each product has its own independent T1EV, essentially allowing the product space to grow with the number of products. I have not included this in the demand specification, but it may be an important extension for future work to consider.

Instead, I can decompose the direct welfare effects reported in Figure 2.2 into the parts that come from the observed product attributes versus the unobserved T1EV errors. I do this by counterfactually offering consumers the new product at the equilibrium price with the consumer taste (δ^f_{jtm}) set to 0, so that utility comes only from unobservables: from the unobserved demand shock and the T1EV error ($\Delta\xi^f_{mjt} + \epsilon^f_{ijtm}$). Figure 2.3 shows these effects. Greek yogurt, a successful and relatively novel product, shows the largest effect from the observed product attributes. Some products like low Carb yogurt have a negative effect from the observed product attributes, and a positive effect on welfare from the unobservable. For most products the observed product attributes account for a significant proportion of the direct welfare gains from that product's introduction[10]. This suggests that there are real welfare effects from new products. Comparing the magnitude of my results with those in Giacomo (2008) no combination of two product introductions in my data can come close to the 360 million euro increase in welfare she measures from two new product introductions. Her demand is based on a nested logit, and consistent with Petrin (2002), when I approach the problem with a random coefficients logit I get significantly smaller welfare increases from a new product.

[10]A similar decomposition for the price effects is less informative, since the equilibrium price for a product with no observable product quality is low, and the other firm's do not need to respond by adjusting prices when the firm adds this low quality product. The largest welfare change due to prices responding the introduction of a product with only unobserved product attributes is abut -\$2000.

Figure 2.3: Direct Welfare Effects of New Products



Direct effects on consumer welfare, in millions of U.S. dollars, of introductions of 15 products appearing for the first time from 2001-2011. Calculated based on average post-introduction costs and taste. Decomposes introduction into 'Observed' the part stemming from consumer taste for the observed product attributes, and 'Unobserved' the part stemming from the unobserved quality of the products.

2.5 Conclusion

In this chapter, I have estimated a random coefficients logit model of demand in the U.S. Yogurt Industry from 2001-2011. I estimate demand at the IRI geographic market and quarterly level for 21 different product lines offered during the sample period, and I define product lines as unique combinations of 7 underlying product characteristics recorded in the IRI data and supplemented by my own research on the products. The model allows for a random coefficient on price, and I use brand product time fixed effects to allow consumer taste for products to change overtime in a flexible way. The model estimates imply demand is inelastic and are consistent with estimates of demand using the same data found in Hristakeva (2016). I also estimate the variable costs of firms as a function of market level variables like distance to the nearest manufacturing facility for each brand, as well as, on brand product time fixed effects. These estimates form the basis of work in Chapter 3, where I consider the problem of estimating a fixed cost of offering products, and counterfactual simulations of mergers. I also use the estimates to compute the total gains in consumer welfare that have occurred when 15 new product introductions occurred during the sample period. I found that consumers gain considerable welfare from reductions in the prices after products are introduced, these gains represent about 95% of the total welfare gains from new products, and these price effects occur even when the new product line is simply a new combination of product characteristics that were already available to consumers. There are smaller increases in consumer welfare when new product characteristics become available, for example, Greek yogurt generates about $3.6 million dollars. These gains in consumer welfare, especially the large gains from changes in prices after new products are introduced, underline the importance of considering the potential effects of market structure on the number and types of products available in an industry.

3 Market Structure's Impact on Product Choice:

3.1 Introduction

Prices and consumer welfare in an industry depends on the number and types of products available, and therefore product entry and exit is of prime importance when assessing market conduct. Further, the relationship between the market concentration in differentiated consumer goods industries, and the entry and exit of products in those industries is key to predicting the effect of mergers on both consumers and competitors. In order to understand product entry and exit, in this chapter I will adapt existing econometric methods used to study firm entry and exit to the problem of product entry and exit, estimate a fixed cost of offering products in the yogurt industry, and then simulate counterfactual mergers in the U.S. Yogurt industry to learn about the relationship between the number of competitors and the set of products that is offered in the industry.

Entry and exit is usually studied by estimating a discrete game, which is itself and extension of the larger literature on modeling discrete choices. Unlike entry and exit, product choice is not a single discrete choice. Instead it consists of multiple simultaneous discrete choices. Firms choose a subset of a set of potential products. In my data I observe 22 different products lines, and therefore firms could be offer 2^{22} different sets of products. This large choice set poses problems for both estimation and simulating counterfactuals.

I deal with the large choice set by applying Manski's method of maximum score. Fox (2007) shows that the method of maximum score can still provide consistent estimates of parameters even when only a subset of the choice set is actually observed. Because maximum score uses information on the relative ranks of alternatives, and ranks are preserved in any subset of the full choice set, estimation based on the observed set of choices will still be consistent. This property makes the method attractive for estimating fixed costs of offering product lines. The method of maximum score is also attractive for high dimensional choice sets because it involves sums of indicator functions. In comparison the likelihood would be a function of probabilities that are mechanically small due to the large number of potential outcomes, and this leads to numerical instability. Much of the recent existing literature on product entry and exit, e.g. Fan and Yang (2016), Wollmann (2014), and Eizenberg (2014), has applied partial identification methods. These methods also avoid the numerical instability of maximum likelihood, and do not require observing the entire choice set. In order to simulate counterfactuals I will need to make distributional assumptions that I think are reasonable and that are consistent with the identifying assumptions of maximum score. This means I would not be able to take advantage of the primary benefit of partial identification, which is its lack of strong assumptions.

The method of maximum score is based on the ranks of alternatives, and therefore it can only 'ordinally' identify counterfactuals. After a counterfactual merger I can compute the preference ordering for the merged firm over sets of products. For example, whether the merged firm ranks offering 1 additional product higher than leaving the set of products the same as that offered before the merger. However, to compare across decision making regimes–to compare the merged firm with the un-merged firm, I will need more than ordinal information (I need to make an interpersonal comparison of utility). I make this comparison by assuming that the unobservables follow a type-1 extreme value distribution, so that the probability of a set of products being chosen has the logit form. This simplifies the analysis of the counterfactuals because the odds of a firm offering a given set of product compared to another depends only on the difference in their expected profitability.

In the next section (Section 3.2), I explain how I use the estimated demand and variable cost curves to simulate the expected variable profits that firms will receive when the offer a set of products. Under the assumption that the observed choices of firms were best responses–more profitable choices–than unobserved choices, these changes in expected variable profits across alternative sets of products provide the variation that identifies the fixed costs of offering a products. In Section 3.3, I explain the details of the maximum score estimator, and report the estimated fixed cost parameter. The fixed costs of offering an additional product for an entire year nationwide is estimated to be around a quarter of a million dollars. Then in Section 3.4, I explain the counterfactual simulations and report the results. I find evidence that the incentives to add or drop products do not change much when the market becomes more concentrated, but that consumer welfare would be lower at any given number of products. I conclude, that increases in product variety that may occur after a merge, could be better provided by a competitive market.

3.2 Simulating Variable Profits

The next step to estimate the fixed costs is to compute the variable profits firms in stage one expect to receive in stage two when they offer alternative sets of products. In other words I need to be able to evaluate the function:

$$\pi(a^f, a^{-f}, d) \equiv E_e \Pi(a^f, a^{-f}, D)$$

The function $\pi(\cdot)$ is actually a composition of two operations. The outer operation takes expectation over the distribution of the demand and variable cost shocks $e =$

$\left(\Delta\zeta^f_{jtm}, \Delta\xi^f_{mjt}\right)$[1]. The inner operation, evaluates $\Pi(a^f, a^{-f}, D)$, the value function from the firms N.E. pricing problem. I will explain how I implement each operation in turn.

Both the estimation of the marginal cost model and the demand model did not make assumptions about the distributions of any of the residuals beyond conditional mean zero requirements. Therefore, I evaluate the expectation by drawing directly from the residuals, rather than introducing additional modeling assumptions. In both cases the regressions contain BPT fixed effects which makes the mean of the residuals 0 across brands, products, and time periods, however, the mean is free to vary across markets. So when taking draws for simulating expected variable profits in a given market, I draw only from residuals that occurred in that market. Given that there are 40 time periods and 7 or more products in each market, I have at least 210 residuals per market. In order to compute the variable profits of a firm who offers J products, I need to take J draws from that market's demand and variable cost residuals. Then to approximate the expectation I repeat that process five-hundred times, and take an average.

For each potential product line given a time period and firm I can use the decompositions of the BPT fixed effects to predict ξ^f_{jt} and ζ^f_{jt}. Define $\hat{\xi}^f_{jt} = h(f, x_j, t)$ and $\hat{\zeta}^f_{jt} = h_c(f, x_j, t)$. I then add to these the values of νd_{cmt} and γd_{0mt} this gives me, and a draw from the residuals to get:

$$\hat{c}^f_{jtm} \equiv \nu d_{cmt} + \hat{\zeta}^f_{jt} + \Delta\zeta^f_{jtm}$$
$$\delta^f_{jtm} \equiv \hat{\xi}^f_{jt} + d_{0m}\gamma + \Delta\xi^f_{mjt}$$

$\hat{c}^f_{jtm}$ is the predicted variable cost for a product, and δ^f_{jtm} is the predicted mean utility

[1] I ignore the errors $\dot{\zeta}$ and $\dot{\xi}$ because these residuals are extremely small (on the order of 10^{-3} or less) due to the non-parametric regression's ability to achieve a close fit within the sample. Keeping track of these errors does not make sense in the since firm profits and estimators are all computed numerically. Further, previous work on product introduction usually fixes the product attributes in counterfactuals to some post introduction average level. The non-parametric estimator used to predict product attributes is just a particular weighted average of post introduction product attributes.

of consumers for a product. Now for any vector of prices I can compute market shares, variable costs for both firms, and from the market shares the matrix $(D_p s_{tm})$ can also be computed. This allows me to evaluate the first order condition and search numerically for a vector of prices that satisfies the first order condition simultaneously for both firms. The first order condition can be re-written as follows to highlight the fact that it defines a fixed point for N.E. prices:

$$p_{tm} = c_{tm} - \left(D_p s_{tm}(p_{tm})\right)^{-1} s_{tm}(p_{tm})$$

There are several approaches in the literature for doing this computation. Morrow and Skerlos (2010) show that solving the first order conditions as a system of non-linear equations or iterating directly on the fixed point above are not reliable ways to compute prices when starting the search from an arbitrary vector of prices. Instead they propose a fast and reliable alternative fixed point that they call the 'zeta-fixed-point'. This fixed point mapping always moves in the direction of steepest ascent. It therefore has some of the nice properties of more general Newton or quasi-Newton root finding algorithms, but involves less computational burden. They also show that the matrix $D_p s_{tm}(p_{tm})$, which needs to be inverted to solve the first order conditions, can be decomposed into two parts in a way that simplifies the computation of the inverse. I use my own implementation of their method[2]and compute the N.E. prices and then variable profits from each draw of variable cost and demand shocks. Taking the average over many draws completes the evaluation of the variable profit function $\pi(a^f, a^{-f}, d)$.

Finally, I will need to evaluate $\pi(\cdot)$ for the observed choices a^f_{tm} and for a series of alternative choice a^f in order to compare the profitability of the observed choices with choices that were revealed to be unprofitable in equilibrium. As discussed previously the

[2]I have written an R package for computing N.E. prices in logit and random coefficient demand models, which can be found on github at 'joearossetti/SimNashPrice'. See Appendix1 for additional details about the computational approach.

pairwise maximum score estimator is consistent even if estimation is based on a subset of the observed choices. As alternatives I start with a relative wide set of possibilities, and the consider the robustness of the fixed cost estimates to using different assumptions. The main decision is when a product should be available to firms. My preferred specification is to limit firms to offering a product line only beginning in the period it was actually introduced. This prevents the estimation from relying on difference in variable profits between sets of products firms may not have been considering e.g. the addition of Greek yogurt 7 years prior to its first appearance in the U.S.. On the other hand, some product lines may have been known options before they were actually introduced, so I also simulate alternatives where firms have access to all 21 products that are available by the end of the sample. I find the particular alternatives I use by adding and dropping one product at a time, allowing the firm to offer all products, allowing the firm to offer each product alone, I sample 15 ways to add or drop 2 product and 15 ways to add or drop 3 products, and finally, I choose at random 7 different possible numbers of products and sample 15 different sets of products that have each of the 7 numbers of products. I also normalize the variable profits of offering no products to 0, and include this as an option for firms. Let the set of alternative markets be A^f_{tm}. Once the sets of alternative markets have been set up for each market and time period I compute: $\pi(a^f_{tm}, a^{-f}_{tm}, d)$, and $\left\{\pi(a^f, a^{-f}_{tm}, d)\right\}_{a^f_{tm} \in A^f_{tm}}$.

3.3 Fixed Cost Estimation

Estimating the fixed costs is relatively straightforward once the expected variable profits for alternative choices of products have been computed. Recall the firm's total profit from offering a set of products is:

$$\pi(a^f_{tm}, a^{-f}_{tm}, d) - C(a^f_{tm}, \eta^f_{tm})$$

I will parametrize the fixed cost function as a function of the number of products offered $n^f_{tm}(a^f_{tm})$:

$$C(a^f_{tm}, \eta^f_{tm}) = -\theta n^f_{tm}(a^f_{tm}) - \eta^f_{tm}(a^f_{tm})$$

Plugging this into the total profits yields:

$$\pi(a^f_{tm}, a^{-f}_{tm}, d_{tm}) + \theta n^f_{tm}(a^f_{tm}) + \eta^f_{tm}(a^f_{tm})$$

This equation is divided into an observable part which I will refer to as $\tilde{\pi}(a^f_{tm}, a^{-f}_{tm}, d; \theta) \equiv \pi(a^f_{tm}, a^{-f}_{tm}, d_{tm}) + \theta n^f_{tm}(a^f_{tm})$, and the unobservable part $\eta^f_{tm}(a^f_{tm})$. In Nash Equilibrium it must be the case that no firm has a unilateral deviation that is profitable. Therefore if a^f_{tm} is chosen then conditional on the other firms choice it must be the case that:

$$1[a^f_{tm} = a^{*f}] = 1[\tilde{\pi}(a^{*f}, a^{-f}_{tm}, d; \theta) - \tilde{\pi}(a^f, a^{-f}_{tm}, d; \theta) \geq \eta^f_{tm}(a^f) - \eta^f_{tm}(a^{*f}) \; \forall a^f \in A^f_{tm}]$$

A traditional discrete choice estimation procedure could be used to estimate θ if the distribution of $\eta^f_{tm}(a^f) - \eta^f_{tm}(a^{*f})$ was known. For example if the distribution of $\eta^f_{tm}(a^f)$ was type one extreme value, then the familiar multinomial logit model could be used. Such an approach would not be appropriate in this case for two reasons. First, the model is incomplete: there are multiple N.E. for some values of the error terms and therefore no unique outcome predicted by the model. Often this leads researchers to consider the model only partially identified, however, Tamer (2003) suggests that point identification is still possible via a semi-parametric estimation procedure: replace the ill-defined probabilities from the model with probabilities based on the empirical distribution of outcomes. Second, most parametric discrete choice estimators require the researcher to observe or know about all of the possible choices[3].

In order to apply the maximum score estimator the key assumption is that $\tilde{\pi}(a^{*f}, a_{tm}^{-f}, d; \theta)$ rank orders the choices. Rank ordering implies that if $\tilde{\pi}(a^{*f}, a_{tm}^{-f}, d; \theta) > \tilde{\pi}(a^{f}, a_{tm}^{-f}, d; \theta)$ for $a_{tm}^{f*} \neq a_{tm}^{f}$ then $Pr(a_{tm}^{f*}) \geq Pr(a_{tm}^{f})$. According to a lemma in Fox (2007), the assumption 2 implies, by the lemma in Fox (2007), that the choices will be rank ordered. It is important to note that the errors are the same regardless of the choices of the other firm and independently and identically distributed across firms. This means that they are not only exchangeable conditional on the state vector $D = (d, e)$, but also conditional on the other firm's choices. The i.i.d. assumption may be stronger than necessary to guarantee rank ordering, but I have not explored this possibility. This is a strong on the distributions of the unobservables (plural because there is one per choice) conditional on the observables. Fox (2007) points out that heteroskedasticity that does not increase monotonically with the observable component of payoffs will violate the rank-ordering assumption. In single agent discrete choice models this rules out some random coefficient specifications, which rules out important types of unobserved heterogeneity without which it is impossible to generate realistic prediction for consumer behavior. It is not clear what types of behavior are ruled out by the rank ordering assumption in the case of discrete games and whether these would be empirically important.

The maximum score estimator will still be consistent even if the full choice set, $\mathcal{A}$, is not observed. Therefore I am able to use only the choice sets A_{tm}^{f} that I described in Section 3.2. The set $\mathcal{A}$ is large with the full set of potential products there are 2^{21} choices

[3]Another approach to replacing probabilities in the likelihood with empirical analogues is to use the method of maximum score, however, this approach does require more assumptions than either partial identification or the semi-parametric approach of Tamer (2003). These additional assumptions have as a special case the multinomial logit model, and are therefore weaker than many common assumptions. Also in order to simulate counterfactuals parametric assumptions about the distribution of unobservables will turn out to be useful to get clearly interpretable results. Further, the method of maximum score is particularly appealing in this setting due to the high dimensional choice set. The Tamer (2003) would fail because the empirical probability of any given action could not be estimated without observing the entire large choice set. In contrast the maximum score estimator generates point estimates when only part of the choice set is observed.

available. If firms consider all of these when choosing products and they are rank ordered then any pairwise comparison or ranking of a subset will maintain the ordering of the full set. But if firms do not consider every possible choice in $\mathcal{A}$ then the ranking may not hold in every subset. Call the set of products the firm does consider the consideration set and denote it $\tilde{\mathcal{A}}$. From the perspective of the econometrician this set is unobserved, so it makes sense to model it as random. Fox (2007) shows that if the probability of this set being the firms consideration set increases when it contains more choices with higher observable payoffs ($\tilde{\pi}(\cdot)$), then the pairwise maximum score estimator is still consistent. This implies that the observed choices more likely to have been in the firm's consideration set. This offers some robustness to the idea that firms may have been aware of some new products in periods prior to when they are introduced.

The pairwise version of the maximum score estimator is attractive because it is simple to compute. Given the simulated expected variable profits for each alternative choice computed in Section 3.2 and a guess of the parameter θ the pairwise maximum score objective function is:

$$Q(\theta) = \sum_{m=1}^{N} \sum_{t=1}^{T} \sum_{a^f \in A^f_{mt}} 1[\tilde{\pi}(a^{*f}, a^{-f}_{tm}, d; \theta) - \tilde{\pi}(a^f, a^{-f}_{tm}, d; \theta) \geq 0]$$

The term pairwise comes from the pairwise comparison of the chosen action with each alternative. Maximizing this objective function is difficult since it is not smooth, and therefore a global numerical optimization routine is needed. I use the Improved Stochastic Ranking Evolution Strategy algorithm implemented in the NLopt library (Johnson, 2016). Global optimization is not as time consuming in this case compared to other cases because the objective function can be computed quickly: the parameter θ enters linearly, the expected variable profits are pre-computed, and most modern statistical programming languages contain fast vectorized condition checking. The main drawback of maximum score is that it does not have a limiting distribution that can be used to calculate standard errors. The most common solution is to used the smoothed maximum score estimator of Horowitz (1992), however, I have not adopted this method since inference about the fixed cost estimates would also require correcting the standard errors for the presence of the estimated expected variable profits, and this is the primary barrier to inference rather than the maximum score estimator itself.

Maximum score requires that the coefficient on one of the regressors be normalized to 1 (Manksi 1975). In the above exposition, I implicitly normalized the coefficient on variable profits to 1 by omitting its coefficient in the equation for $\tilde{\pi}$, thus θ will be immediately in dollar units. Another thing to note about θ is that it measures only the marginal fixed cost from adding or dropping a product. There may be other fixed costs that a firm faces when it chooses to operate in a market at a given time. These fixed costs may be traditional fixed costs, or they may be fixed costs that result from activities required to maintain brand image or equity. If these fixed costs are additively separable from the fixed costs that result from the number of products θn_{tm}^{f} then these costs would drop out of the observable profit differences, $\tilde{\pi}(a^{*f}, a_{tm}^{-f}, d; \theta) - \tilde{\pi}(a^{f}, a_{tm}^{-f}, d; \theta)$. The pairwise maximum score estimator therefore can be thought of as controlling for brand, market, and time fixed effects, but it does not identify these fixed effects.

Table 3.1 gives my estimates of fixed cost. The first line of the table shows the fixed costs when I limit firms to only offering product lines beginning in the period when they are first introduced. This estimate is much smaller than the estimate in the second line of the table, which allows firms to introduce any product that was available over the whole sample in any period. The second estimate results in fixed costs that are 4.82 times the expected variable profits in the average market. These fixed costs are implausibly large, and therefore I use the estimates from the first row in the counterfactuals. There is a sense in which the first row estimate may be a conservative estimate, but it is not clear what would be the right assumption to generate estimates between the two in the table. The fixed costs are relatively small, about 1/3 of variable profits, but it is important to keep in mind that this is the fixed costs that stem only from the number of products offered–firms may also face fixed costs of being active in the market which are not estimated here.

Small fixed costs of offering additional products do not directly imply that the monopolist may offer additional products compared to the competitive industry, because as discussed in Chapter 1 there are additional costs due to cannibalization of sales that need to be considered. However, the low fixed costs remove a potential brake on offering products. If fixed costs were large, then even if the monopolist faced low cannibalization of sales he might not increase the number of products. Importantly, the fixed costs are not implausibly small. Hristakeva (2016) estimates payments from yogurt manufactures to retailers that are designed to get retailers to offer products (essentially purchasing shelf space). Her estimates are on the same order of magnitude as my estimates (under the assumption that products are only available beginning in the first period they are offered). This suggests that perhaps these payments to retailers are one of the primary contributors to the fixed costs of offering new products in the yogurt industry. An avenue for future research would be to include strategic interactions in the fixed cost estimates to test if firms are able to impact their competitors costs of offering products when they choose to introduce a new product.

Table 3.2 gives estimates of fixed costs under the assumption that firms are only able to offer product lines beginning in the period when they are first introduced. Yoplait offers few product lines than Dannon on average during the sample, and the fixed cost estimates rationalize this difference by making Yoplait's fixed costs significantly higher than Dannons. This difference is fixed cost has important effects on the incentives of the firms after a merger. The low fixed costs of Dannon will generate incentives in all counterfactuals to increase the number of products. However, for Yoplait only in the competitive market will it have an incentive to increase the number of products it offers.

Table 3.1: Estimates of Fixed Cost Parameter

Cost per Prod. per mkt-qrtr ($s)	θ	Avg. Fixed Cost	Avg. Fixed Cost Ratio
Limited Available Lines	1290.9		
All Lines Available	19355		

Estimates based on pairwise maximum score of fixed costs per product in U.S. dollars. 'Avg. Fixed Cost' is the average fixed cost found by multiplying the parameter estimate times by the number of products observed in each market. 'Avg. Fixed Cost Ratio' is the average ratio of fixed costs at each observation to expected variable profits.

Table 3.2: Estimates of Fixed Cost Parameter by Brand

Cost per Prod. per mkt-qrtr ($s)	θ	Avg. Fixed Cost	Avg. Fixed Cost Ratio
Yoplait	4989.3	25938.61	0.87
Dannon	575.8	5009.66	0.19

Estimates based on pairwise maximum score of fixed costs per product in U.S. dollars with a seperate parameter for each brand. 'Avg. Fixed Cost' is the average fixed cost found by multiplying the parameter estimate times by the number of products observed in each market. 'Avg. Fixed Cost Ratio' is the average ratio of fixed costs at each observation to expected variable profits. Standard errors are in progress.

3.4 Counterfactual Product Choice

The goal of estimating the fixed costs of offering products was to compare the product variety offered by the duopoly of Dannon and Yoplait to that offered if either firm took over as a monopolist or they merged. Since I estimate the model using pairwise maximum score I have not imposed a distribution on the unobserved component of fixed costs, and therefore, I cannot yet solve the model directly for counterfactuals. In this section, I impose the assumption that the unobservable $\eta^f(a^f_{tm})$ is i.i.d. type-1 extreme value[4]. This assumption is sufficient to guarantee the exchangeability and therefore rank ordering properties that the maximum score estimator requires. This is an advantage of the point identification approach since the fixed costs estimates are internally consistent with the assumptions needed to evaluate the counterfactuals.

I have in mind two counterfactuals. In the first, Yoplait or Dannon becomes a monopolist

[4]Earlier versions of this paper attempted to use just the rank ordering property to draw qualitative (ordinal) conclusions about the counterfactuals. This is fine for studying the preferences of a single decision maker, e.g. the monopolist, but does not allow comparison across decision makers (as this would be an interpersonal comparison of utility).

(I consider both possibilities). In particular, in this counterfactual I eliminate the competition from fringe and regional yogurt producers. In the second, Yoplait and Dannon merge (and inherit the brand label of one or the other firm–again I consider both possibilities). I drop duplicate products and assume that the products are produced at the monopolist's (or merged firm's) manufacturing facilities. In the event of a real merger no doubt the manufacturing locations would change to take advantage of economies of scale, but without additional information I cannot make this prediction. Also the goal is not to simulate a realistic merger, but instead evaluate the effect of competition on the variety of products offered. I also switch that products to the monopolist's (or merged firm's) brand. This simplifying assumption means that occasionally consumer surplus may rise as a brand that is more preferred to another takes over production in a market. This leads to three different decision making regimes: competition, monopoly, and a merged industry. The question is: which of these decision maker has the strongest incentives to offer products?

I evaluate the counterfactuals using the total profits computed for alternative product assortments like those used in estimating the fixed costs. The type-1 extreme value assumption implies the odds of these alternatives being chosen by each decision making regime are directly related to the change in total profits that decision making regime can expect from changing the set of products it offers. This approach leaves out alternative sets of products that were not used in estimation, and so any predictions from it are conditional on the set of alternative I examine. As in the estimation I try to pick both some sets of products that may be informative like offering each product alone, and some alternatives that are reasonable especially the one product deviations from the observed set of products. The idea of the first approach is that if we consider two sets of products that differ in the number of products offered: a_{tm}^{f+} and a_{tm}^{f} where a_{tm}^{f} has for example one additional product, then the odds of a_{tm}^{f+} being chosen over a_{tm}^{f} can be written:

$$\frac{Pr(a_{tm}^{f+})}{Pr(a_{tm}^{f})} = \frac{e^{\tilde{\pi}(a_{tm}^{f+},a_{tm}^{-f},d)} / \sum_{a^f \in \mathcal{A}} e^{\tilde{\pi}(a^f,a_{tm}^{-f},d)}}{e^{\tilde{\pi}(a_{tm}^{f},a_{tm}^{-f},d)} / \sum_{a^f \in \mathcal{A}} e^{\tilde{\pi}(a^f,a_{tm}^{-f},d)}} = \frac{e^{\tilde{\pi}(a_{tm}^{f+},a_{tm}^{-f},d)}}{e^{\tilde{\pi}(a_{tm}^{f},a_{tm}^{-f},d)}}$$

Taking the log of both sides yields the log odds as a function of the difference in observable profitability between the two options:

$$ln\,Pr(a_{tm}^{f+}) - ln\,Pr(a_{tm}^{f}) = \tilde{\pi}(a_{tm}^{f+}, a_{tm}^{-f}, d) - \tilde{\pi}(a_{tm}^{f}, a_{tm}^{-f}, d)$$

In order to implement this idea I need to pick a reference set of products. In order to address the question of whether the counterfactual monopolist or merged firm would increase the number of products available to consumer, I use the set of products the merged firm or monopolist would inherit. This means that if Dannon and Yoplait merger, I combine the sets of products the offer by dropping duplicate product lines, and assigning all of them to the brand identify of the firm who is leading the merger (I compute the counterfactual for both Dannon and Yoplait as the leader). The base number of products for the counterfactuals is therefore higher than what either Dannon or Yoplait were offering by themselves. In order to get at the effect of this I compute the differences in total profits using the actual sets of products that Dannon (or Yoplait) offered, and a third counterfactual where Dannon is given the set of products it would have if it merged but Yoplait remains a competitor. The average difference in total variable profits for alternatives that add 1, 2 or 3 products and drop 1,2 or 3 products as well as alternatives where the firm offers all available product lines or each product line alone are plotted in Figure 3.1.

All of the decision making regimes for Dannon have an incentive to add products. Dannon has lower fixed costs than Yoplait, which means offering and additional product is more likely to increase total profits. Total profits also decrease when products are dropped.

Previously, a working paper version of these results reported the results for just the 'Monopoly' counterfactual, reporting that the monopolist has a preference to increase the number of products, but did not impose the needed assumptions to make comparisons across counterfactuals nor did it estimate separate fixed costs for Dannon and Yoplait. The incentive to increase the number of products remains for Dannon, although all of the counterfactual decision making regimes appear to prefer to add products with about the same strength. On the other hand, Yoplait once it acquires the products it would have it merged with Dannon no longer has an incentive to add products. Yoplait offers fewer products than Dannon throughout the sample, and faces a higher fixed cost. The 'Combined' counterfactual, where Dannon (or Yoplait) is given the products it would have after the merger is similar to 'Merged' and 'Monopoly' counterfactuals, and these counterfactuals in turn are also similar. This suggests that much of the incentive to add products is driven by the products offered rather than the market structure, so that giving the firm with more products market power does not inhibit the incentive to add products. However, market power also decreases consumer welfare via higher prices. Figure 3.2 plots the minimum consumer welfare for alternative product offerings as a function of the number of products. Conditional on the number of products the consumer is always better off with the more competitive market. If the competitive market will have similar or greater incentives to add products, then from the consumer's perspective this market structure would be preferred.

3.5 Conclusion

I have estimated a model of how firm's choose the products offered in an industry. In particular I model the incentives to adopt a type of product that has already been introduced for the first time. The model had two stages. In the first stage firm's determine which products to offer based on their expected profits in the second stage. After the set of products is determined firms observe demand and variable cost shocks and set prices.

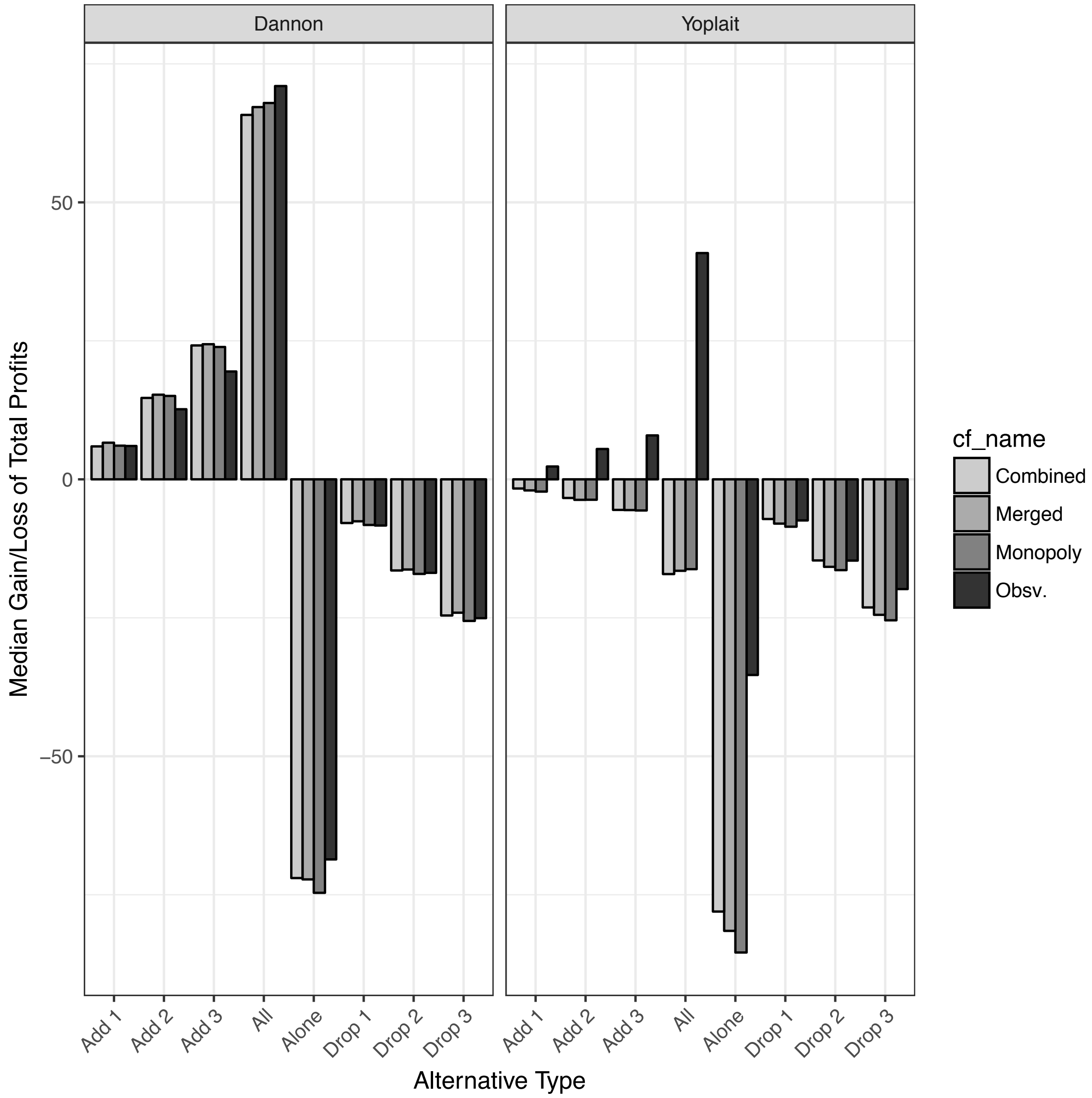
The model was estimated by exploiting the timing of the game to split estimation into two parts. Because the set of products is determined before firms learn about the demand and variable cost shocks their choice of products is independent of the realization of these shocks. This identification strategy based on timing was proposed by Eizenberg (2014) and has been used in other papers studying product variety like Fan and Yang (2016), Wollmann (2014), and Hristakeva (2016). In Chapter 2 I was able to estimate demand using random coefficient logit model using the standard Berry et al. (1995) instruments. In this chapter, given estimations of demand and variable cost, I simulated the variable profits firm's would expect to receive if they offered alternative sets of products. Under the necessary conditions of N.E. the observed set of products should have higher total profits, variable profits minus fixed costs, than any alternative set. Rather than use these inequalities to partially identify the fixed costs as in most previous work on product variety, I use the pairwise maximum score estimator of Fox (2007) to point identify fixed costs under the assumption that the unobservable components of fixed costs follow and exchangeable distribution.

Based on this estimation procedure I find that the fixed costs for offering a product is on average 1/3 of variable profits. Using my estimate of fixed costs, and the added assumption that the fixed cost errors are type-1 extreme value I characterize product variety in two counterfactual settings: one, where the firms merge, and two, where one firm is given a monopoly. The incentives to add products did not appear to change dramatically between the monopoly, merger, and existing competitive market structure. In the case of Dannon, this meant that all three market structure appeared to have incentives to add products rather than drop them. While Yoplait only had an incentive to increase the number of products in the current duopoly market structure. The difference is attributable to the higher fixed cost of Yoplait, and an incentive for Yoplait to add products Dannon is offering but it is not. Even though increased market concentration did not diminish the incentive to add products–which would be good new for consumers, at any given number of products consumer welfare is lower under the more concentrated industry.

The results in this chapter are consistent with the literature. They are consistent with the literature on the welfare effects of new products (Giacomo, 2008) as well as my own estimates in Chapter 2 of the welfare effects of new product in the U.S. Yogurt Industry. In Chapter 2, like the existing literature, I found that much of the consumer's benefit from new product lines was due to changes in price competition. It makes sense then that when the pricing power of firms increases, consumers are unable to benefit from new products. The results in this chapter are also consistent with work in other industries that found decreases in product variety after an increase in market concentration (Fan and Yang, 2016). The policy implication is that even when increases in market concentration provide incentives to increase product variety, those incentives may exist at nearly the same strength under more competitive conditions. If increases in product variety, or at least the same level of variety, can be provided in a competitive setting–as they appear to be in the U.S. Yogurt industry–then that would be preferred.

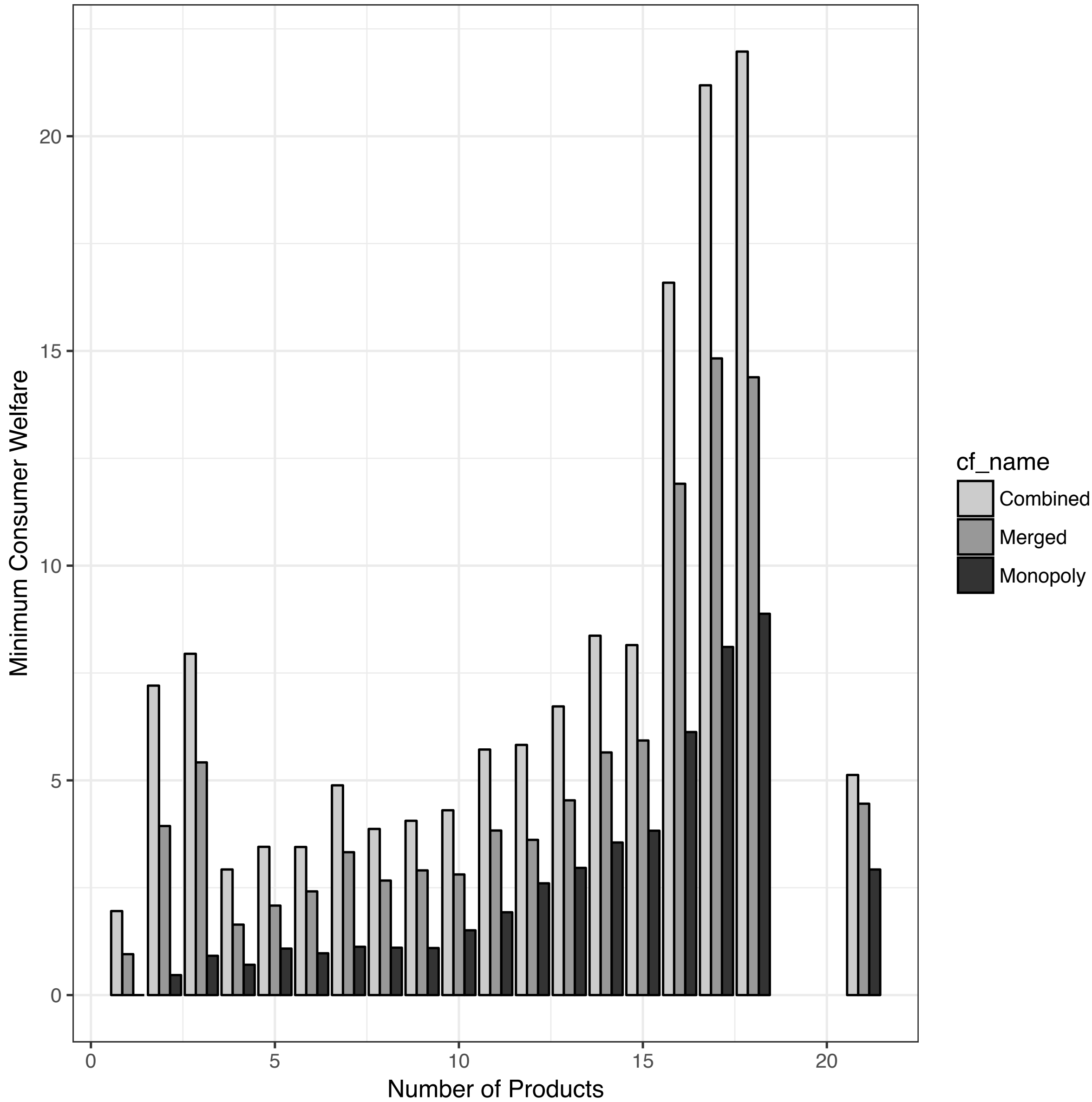
Future, research on this question should focus on the relationship between the demand system and the incentives to add and drop products. The functional form assumptions in the demand estimation play an important role in determining the simulated prices before and after the merge, the strength of cannibalization, and the importance of the price competition term in the cost of adding a product. In more general settings, especially, with more complex consumer hetereogeneity the incentives of the firms to add products and how those incentives change with market structure may be quite different. It also appears possible to simulate the product choices that would occur under different counterfactuals, and even consider the optimal choice of products by a social planner. This is because the conditional probability of a product being offered is tractable under the type-1 extreme value assumption, and therefore Monte Carlo methods might be used to simulate the distribution of product choices.

Figure 3.1: Differences in Total Variable Profits for Categories of Alternatives



Median differences in total profits (1000s of U.S. dollars) for each counterfactual across different types of alternative product choices. 'Combined' gives the firm the products it would have after a merger, but leaves its competitor. 'Merged' merges Dannon and Yoplait but keeps the smaller producers. 'Monopoly' gives the firm a monopoly. 'Obsv.' gives the firm only the products it actaully was observed to offer.

Figure 3.2: Minimum Consumer Welfare and the Number of Products



Minimum consumer welfare (1000s of U.S. dollars) obtained with alternatives that offer a given number of products

Bibliography

Ackerberg, D. (2005): "Unobserved product differentiation in discrete choice models: estimating price elasticities and welfare effects," *RAND Journal of Economics*, 36.

Anderson, S. P. and A. de Palma (1992): "Multiproduct Firms: A Nested Logit Approach," *The Journal of Industrial Economics*, 40, 261–276.

Aydin, G. and J. K. Ryan (2000): "Product line selection and pricing under the multinomial logit choice model," in *Proceedings of the 2000 MSOM Conference*.

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

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25.

Berry, S. T. and P. A. Haile (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

Besbes, O. and D. Saure (2016): "Product Assortment and Price Competition under Multinomial Logit Demand," *Production and Operations Management*, 25, 114–127.

Bonanno, A. (2016): "A Hedonic Valuation of Health and Nonhealth Attributes in the U.S. Yogurt Market," *Agribusiness*, 32, 299–313.

Bresnahan, T. F. and P. C. Reiss (1991): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977–1009.

Cilberto, F., C. Murry, and E. Tamer (2015): "Market Structure and Competition in Airline Markets," Working Paper Accessed from <http://personal.psu.edu/ctm15/research.html>.

Connor, J. M. (1981): "Food Product Proliferation: A Market Structure Analysis," *American Journal of Agricultural Economics*, 63, 607–617.

Draganska, M. and D. C. Jain (2005): "Product-Line Length As A Competitive Tool," *Journal of Economics and Management Strategy*, 14, 1–28.

Draganska, M., M. Mazzeo, and K. Seim (2009): "Beyond plain valnilla: Modeling join product assortment and pricing decisions," *Quantitative Marketing and Economics*, 7, 105–146.

Eizenberg, A. (2014): "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 81, 1003–1045.

Ellickson, P. B., S. Houghton, and C. Timmins (2013): "Estimating network economics in retail chains: a revealed prefernce approach," *The RAND Journal of Economics*, 44, 169–193.

Fan, Y. and C. Yang (2016): "Competition, Product Proliferation and Welfare: A Study of the U.S. Smartphone Market," *CEPR Discussion Paper*.

Fox, J. T. (2007): "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices," *The RAND Journal of Economics*, 38, 1002–1019.

Giacomo, M. D. (2008): "GMM estimation of a structural demand model for yogurt and the effects of the introduction of new brands," *Empirical Economics*, 34, 537–565.

Giammona, C. (2015): "With Chobani Back on Track, Founder is Staying Put," `https://www.bloomberg.com/news/articles/2015-09-10/with-chobani-back-on-track-founder-is-staying-put`.

Hausman, J. (1996): "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan and R. Gordon, University of Chicago Press.

HAUSMAN, J. AND G. K. LEONARD (2002): "The Competitive Effects of a New Product Introduction: A Case Study," *Journal of Industrial Economics*, 50, 237–263.

HOROWITZ, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

HRISTAKEVA, S. (2016): "How Do Vertical Contracts Affect Product Availability? An Empirical Study of the Grocery Industry," *Working Paper*.

JOHNSON, S. G. (2016): "The NLopt nonlinear-optimization package," Tech. rep., MIT, http://ab-initio.mit.edu/nlopt.

KADIYALI, V., N. VILCASSIM, AND P. CHINTAGUNTA (1999): "Product line extensions and competitive market interactions: An empirical analysis," *Journal of Econometrics*, 89, 339–363.

KELL, J. (2016): "General Mills Reveals How it Plans to 'Renovate' Yogurt Products," `http://fortune.com/2016/07/14/general-mills-yogurt/`.

LI, Q. AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

MANKSI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

MAZZEO, M. (2002): "Product Choice and Oligopoly Market Structure," *The RAND Journal of Economics*, 33, 221–242.

MCFADDEN, D. (1981): *Structural Analysis of Discrete Data with Econometric Applications*, MIT, chap. Econometric Models of Probabilistic Choice.

MORROW, R. W. AND S. J. SKERLOS (2010): "Fixed-Point Approaches to Computing Bertran-Nash Equilibrium Prices Under Mixed Logit Demand: A Technical Framework for Analysis and Efficient Computational Methods," *arXiv preprint arXiv:1012.5836*.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Quelch, J. A. and D. Kenny (1994): "Extend Profits, Not Product Lines," *Harvard Bussiness Review*, 72, 153–160.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

Taylor, C., C. Spees, R. Watowicz, S. Martinez, and N. Hooker (2017): "Beware of Greeks bearing gifts: The potential impact of yogurt innovation on dietary intakes," *Journal of Food Composition and Analysis*, 64, 132–137.

Tirole, J. (1994): *The Theory of Industrial Organization*, MIT, 7th ed.

Villas-Boas, S. B. (2007): "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data," *Review of Economic Studies*, 74, 625–652.

Wollmann, T. (2014): "Trucks without bailouts: Equilibrium product characteristics for commerical vehicles," Job Market Paper Accessed from http://faculty.chicagobooth.edu/thomas.wollmann/research/.

Appendix A: Technical Details and Data

1 Simulation of Nash Equilibrium Prices

This appendix deals with the computational details used to compute the Nash equilibrium prices in Section 3.2 of Chapter 3. The methods described here are implemented in an R package, SimNashPrice, available to install from GitHub. Recall, from Chapter 1 that the problem of solving for Nash equilibrium prices reduces to solving:

$$s + (p - c)\, D_p s = 0. \tag{1}$$

Morrow and Skerlos (2010) Proposition 2.3 applies the Leibniz rule to show that the inter-firm Jacobian, $D_p s$, in a random coefficients logit model can always be written as the sum of two matrices: $D_p s = \Lambda s - \Gamma s$, where Λs is diagonal. In order to define these matrices, it is necessary to review the demand system. In the random coefficients demand model the indirect utility of consumer i from consuming product k is denoted:

$$u_{ik} = \alpha p_k + \alpha_i p_k + \delta_k + \mu_{ik} + \epsilon_{ik}$$

where $\alpha$is the average disutility from price and α_i is consumer i's deviation from that average, δ_k is the average utility from consuming product k and μ_{ik} is consumer i's deviation from that average, and ϵ_{ik} is a type-1 extreme value random shock to utility. The probability consumer i purchases good k is $s_{ik} = \frac{exp(\alpha p_k + \alpha_i p_k + \delta_k + \mu_{ik})}{\sum_h exp(\alpha p_h + \alpha_i p_h + \delta_h + \mu_{ih})}$, and the aggregate market share for product k is $s_k = E_{\alpha_i} E_{\mu_{ik}} s_{ik}$. Now define the matrices Λs and Γs:

$$[\Lambda s]_k = E_{\alpha_i} E_{\mu_{ik}} \left(-(\alpha + \alpha_i) s_{ik}\right) \tag{2}$$

$$[\Gamma s]_{kh} = E_{\alpha_i} E_{\mu_{ik}} \left(s_{ik} s_{ih} \left(-\left(\alpha + \alpha_i\right)\right)\right) 1 \left\{a_k^f a_h^f = 1 \ or \ a_k^{-f} a_h^{-f} = 1\right\} \tag{3}$$

Plugging this into the simultaneous stationarity condition gives:

$$\begin{aligned} s + (p-c) D_p s &= 0 \\ s + (p-c)\Lambda s - (p-c)\Gamma s &= 0 \\ p &= c - s\left(\Lambda s\right)^{-1} + (p-c)\left(\Gamma s\right)\left(\Lambda s\right)^{-1} \\ p &= Z(p) \end{aligned} \tag{4}$$

This form of the simultaneous stationarity condition gives a fixed point mapping Z which is called the zeta fixed point mapping by Morrow and Skerlos (2010), and iterating this fixed point mapping proves in my experience and their numerical simulations to be a reliable and efficient method to calculate Nash equilibrium prices. One obvious advantage of this approach is that the matrix that needs to be inverted is diagonal, and inverting a diagonal matrix is much faster than inverting an arbitrary dense matrix (in fact it is as simple as taking reciprocals). Thus most of the computational time is spent computing the expectations $E_{\alpha_i} E_{\mu_{ik}}$ that define the matrices Λs and Γs. By eliminating the need to invert a dense matrix computational speed of computation scales linearly with the number of products rather than polynomially, and I find that the fixed point converges to prices that solve the first order conditions of all firms within machine precision in about 45 fixed point iterations.

The implementation of these methods relies on R's R6 object-oriented programming system. I represent markets as objects, which carry with them all of the information and functions needed to compute Nash prices (and other quantities of interest like consumer welfare, profits, markups, and shares). There are two advantages to this approach. First, it abstracts away from the product level observations within each market, by treating the entire market as a single object with attributes like prices, shares, and product characteristics. This makes simulating counterfactuals conceptually easy since the object in effect updates its internal state after changes and reports the new equilibrium. Secondly, the R6 system is unique in R in that it allows for pass by reference. In standard R whenever changes were made to an object, for example, a fixed point iteration applied, a copy of that object would usually be created, and this copying would slow down computation. R6 does not create copies, and therefore can offer important speed benefits for code that will be run iteratively or repeatedly.

2 Data and Estimation Appendices

2.1 Data

The data used in this paper largely come from two sources. The first is that provided by IRI. This data is protected by an NDA, and therefore cannot be made publicly available. The second is data that largely comes from the U.S. government, and is freely available for download. Some additional data I construct using publicly available information. In this appendix, I will describe in what detail what processing steps I applied to the IRI data, and how I linked to it the data that is freely available (as well as any processing steps that were done on the freely available data).

The IRI data comes at the weekly, store, and product UPC level. The primary data set records for each UPC and store the total weekly purchases, average weekly prices, and three weekly measures of product promotion. IRI provides this data over 11 years from 2001 to early 2011 in 11 flat files. I pull all of the data on the yogurt industry and use it to construct a Postgresql database. When I do this I divide the data evenly into 13-week artificial quarters. The advantage of doing so is that the data splits evenly into forty-four 13 week quarters, but the disadvantage is that the quarters do not line up precisely with traditional calendar or fiscal year definitions. I also record the flat file that the data is read in order to facilitate matching the purchase data with other IRI data provided on a yearly basis. I convert prices to prices per standard 6 oz unit of yogurt. This size is the standard size given by IRI and used by Giacomo (2008) and Villas-Boas (2007).

I add to the database information provided by IRI about product ownership and characteristics. IRI provides this data for 2001-2006, 2007, and 2008-2011. Because IRI recodes the data three times I will later use the product descriptions and internet research to create a unified list of products and their characteristics. The first step in doing so is to find all of the unique combinations of the L5 product description provided by IRI and 5 of the product characteristics that are provided by IRI: Product type, Fat content, Calorie level, Style, and Type of Yogurt. The L5 product description contains the name of the product and resembles what might be printed on the package in the store or used in advertising. For example 'YOPLAIT EXTRA CREAMY FAT FREE' or 'DANNON DANIMALS'. These product descriptions, as well as the 5 variables mentioned above, provide enough information for me to understand what the product is and what its characteristics are, and to confirm characteristics by googling the L5 description which acts as a product name.

IRI also provides information about the stores at which the products were purchased. The first use of this data is to match the stores with the IRI markets, which are multi-county geographic regions. Then I can match demographic, cost, and population data given at the county level to the appropriate markets and stores. The second use of this data is to provide some additional controls and cost shifters. I keep track of the number of stores and retail chains operating within each region, and the number of stores and retail chains offering each potential product. The availability of the product across the region is an important determinant of demand. I also compute the annual market shares of each chain within the market using IRI's estimates of annual sales. From the annual market shares, I can compute the Herfindahl-Hirschman Index for each market and year. I use the concentration index as a cost shifter since it will reflect the retail chains price-setting power.

Taken together the data on stores, products, and purchases can be matched together and aggregated to the IRI market quarter level. I aggregate by summing the 6oz units sold, averaging prices, and counting the number of store-week observations for which a product was featured, promoted, or in a special display. For the annual measures of stores, chains, and retailer concentration in any 13 week quarters that contain data from multiple years I use the value for the year of the first week of the quarter.

After compiling the quarterly demand data I construct several supplementary data sets that provide demographic controls and cost shifters. First I will discuss the cost shifters. I downloaded the Quarterly Census of Employment and Wages County High-Level NAICS-based datasets for the years 2001-2012 from the BLS. From this, I am able to get county-level average wages that can be matched to IRI-markets using the store location data discussed previously. I got average annual retail electricity prices from 1990 by state and industrial sector from the U.S. Dept. of Energy. I assign each county an annual electricity price and then aggregate up to IRI markets by taking the average (which may involve averaging prices across state lines). The last cost shifter that I construct is based on the distance between Yoplait's and Dannon's manufacturing facilities and each market. I searched for each manufacturer online and found information on their company website, by contacting their customer support, or in local newspapers about where their manufacturing facility or facilities had been located during the sample period. I then use the U.S. Census Bureau's 2010 U.S. Gazetteer File, which provides the latitude and longitude coordinates of each county in the U.S to compute the straight-line distance between each county and each manufacturing facility. I aggregate these distances up by averaging across counties within an IRI market.

I use the U.S. census to construct demographic variables downloaded via the U.S. Census Bureau's FTP server. I used the Profile of General Population and Housing Characteristics for 2010 and 2000, income data from the 2000 Census Summary File 3, and the 2010 American Community Survey 5-year estimates. From these files, I obtained the following demographic variables at the county level: median income, percent of population white, number of households with children under 18, and median age. I aggregate these variables up to the IRI market level using population-weighted averages (using the census population for the year of the data 2000 or 2010). Also from the U.S. Census Bureau, I get the 2000-2010 intercensal population estimates and the 2010-2016 population estimates at the county level. This gives the annual estimated population for each county during my sample period. I sum the population of the counties in each IRI geographic market in each year, and I match this data to my 13 week quarters using the population at the start date of each quarter. I use the changes in population to calibrate the size of the yogurt market.

2.2 Demand Estimation Additional Specifications

In this appendix, I report several alternative demand specifications. In Table A.1 I report several specifications of the instrumental variables. Table A.2 shows the effect of using different fixed effects specifications. Last in Table A.3 I consider instrumenting for additional variables that may be endogenous like promotional intensity, the average number of retailers offering a brand's products, and the number of competing private label brands.

I have two basic sources of exogenous variation. The first is that created by the timing of the model, which make product entry and exit decisions independent of market level demand shocks. This exclusion restriction implies that any functions of the set of products offered by firms will be valid instruments.Berry et al. (1995) suggest for each brand-product pair in a market counting counting the number of products in the same product categories offered by the brand and by competing brands. This forms two types of instruments that I refer to as the BLP instruments. For example, consider Yoplait offering the Lite Active product line. There would be seven instruments of the first type: the number of Lite Yogurts, the number of Active yogurts, the number of non-Drink yogurts, the number of non-Greek yogurts, the number of non-Fiber yogurts, the number of non-Kids yogurts, and the number of non-Carb yogurts offered by Yoplait. The second type generates the same seven instruments except counting products offered by Yoplait's competitors rather than by Yoplait. The second source of exogenous variation is the distance between the IRI geographic markets and the manufacturing facilities of the different yogurt firms. I use these distances to calculate the minimum distance between the market where a brand offers a product and one of that brands manufacturing facilities. I will refer to this as the distance instrument or cost instrument (for perishable goods transportation costs make up a significant part of the variable cost).

Each of these sets of instruments generates a different pattern of variation in prices and a different estimate of the structural parameter. This can be seen in Table A.1, which shows different but relatively similar estimates using several different sets of instruments. BLP 1 is the counts of products of each category offered by the same firm, BLP 2 the counts of products of each category offered by other firms, BLP 12 combines the two, and BLP (poly.) uses third-degree orthogonal polynomials of the BLP 12 instruments. The Cost instrument is simply the distance measure described above, and Cost (poly.) is a sixth degree orthogonal polynomial of the distance measure.

Table A.2 shows how different sets of fixed effects change the estimates of price sensitivity. The model names denote the fixed effects: B stands for brand, P for product, and T for time. When the fixed effects appear additively their abbreviations are separated by a space, and when there is an interaction there is no space in between. In all cases, I have instrumented for prices using both of types of BLP instruments. The main pattern here is that without either brand product or brand time interactions the parameter on price implies an elasticity of demand that is less than unit elastic, and that implies negative marginal costs (when the first order condition for prices is inverted see 2.3.2). The apparent importance of both brand-product effects and brand-time effects leads me to use the brand-product-time (BPT) fixed effects.

It is also possible that other variables are endogenous besides prices, the last robustness check in Table A.3, addresses this by instrumenting for additional variables that might be chosen by firms in equilibrium. These variables are a brand's average promotional intensity, a brand's average fraction of retailers offering its product, and the average number of retailers offering private label competitors to a brand's products. In all cases, I use the BLP 12 polynomials as instruments since instrumenting for five variables will not be possible using just the distance instrument. Adding additional endogenous variables does not change the price coefficient, and this coefficient is the most important for simulating variable profits and computing counterfactuals. Instrumenting for the store adoption and feature do significantly alter the estimates of their effects. For Store, the measure of store adoption the effects are not too large. There is a large change in the coefficient on Feature, the measure of promotional intensity. I plan to check the robustness of the fixed costs estimates and counterfactuals to the possibility that Feature and Stores should be treated as endogenous. Instrumenting for the number of private products changes the sign of the coefficient from negative, the expected direction since more private products should imply reduced demand for name brand products, to positive. I do not report the first stage, but it appears that the available instruments do not predict the number of private products well enough to be valid.

Table A.1: Alternative Instruments

	No IV	BLP 1	BLP 2	BLP 12	BLP IV (poly.)	Cost IV	Cost IV (poly.)	Cost and BLP (poly.)
Price	-1.05^{***}							
	(0.02)							
Feature	2.24^{***}	-0.22	1.20^{***}	0.46^{**}	0.98^{***}	1.32^{***}	1.54^{***}	1.20^{***}
	(0.09)	(0.19)	(0.12)	(0.14)	(0.12)	(0.11)	(0.10)	(0.11)
Stores	5.05^{***}	3.57^{***}	4.43^{***}	3.98^{***}	4.29^{***}	4.50^{***}	4.63^{***}	4.42^{***}
	(0.04)	(0.09)	(0.06)	(0.07)	(0.05)	(0.05)	(0.04)	(0.05)
Private Products	-1.78^{***}	-0.82^{***}	-1.37^{***}	-1.08^{***}	-1.29^{***}	-1.42^{***}	-1.50^{***}	-1.37^{***}
	(0.04)	(0.08)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)
Med. Income	0.27^{***}	0.45^{***}	0.34^{***}	0.40^{***}	0.36^{***}	0.33^{***}	0.32^{***}	0.34^{***}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Perc. White	0.07^{***}	-0.12^{***}	-0.01	-0.07^{***}	-0.03^{***}	0.00	0.02^{**}	-0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Med. Age	0.26^{***}	0.36^{***}	0.30^{***}	0.33^{***}	0.31^{***}	0.30^{***}	0.29^{***}	0.30^{***}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Price (IV)		-7.21^{***}	-3.64^{***}	-5.49^{***}	-4.20^{***}	-3.35^{***}	-2.80^{***}	-3.65^{***}
		(0.26)	(0.17)	(0.15)	(0.11)	(0.10)	(0.09)	(0.08)
Num. obs.	47430	47430	47430	47430	47430	47430	47430	47430

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Estimates of demand for yogurt; all of the models have brand-product-time fixed effects. The first four variables are defined in Table 2, and the next three in Table 1. "No IV" reports estimates without instrumenting for price. "BLP 1" reports estimates instrumenting for price with the first type of BLP instruments, "BLP 2" reports estimates instrumenting for price with the second type of BLP instruments. "BLP 12" reports estimates instrumenting for price with the both types of BLP instruments. "BLP (poly.)" reports estimates instrumenting for price with polynomials of the BLP instruments. 'Cost IV' uses just the distance to the manufacturing center as an instrument for price, 'Cost IV (poly.)' uses a polynomial of the distance, and 'Cost and BLP (poly.)' uses polynomials of both the cost and BLP style instruments.

Table A.2: Alternative Fixed Effects

	P B	P B T	PT B	PT BT	BP T	BPT
Feature	1.67^{***}	1.37^{***}	1.35^{***}	1.86^{***}	1.41^{***}	0.98^{***}
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)
Stores	5.35^{***}	5.41^{***}	5.35^{***}	4.89^{***}	5.17^{***}	4.29^{***}
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)
Private	-1.81^{***}	-1.90^{***}	-1.88^{***}	-1.64^{***}	-1.74^{***}	-1.29^{***}
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Med. Income	0.25^{***}	0.26^{***}	0.26^{***}	0.28^{***}	0.29^{***}	0.36^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Perc. White	0.08^{***}	0.08^{***}	0.06^{***}	0.05^{***}	0.04^{***}	-0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Med. Age	0.25^{***}	0.25^{***}	0.25^{***}	0.26^{***}	0.26^{***}	0.31^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Price (IV)	-1.01^{***}	-1.27^{***}	-1.32^{***}	-1.66^{***}	-2.13^{***}	-4.20^{***}
	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)	(0.11)
Num. obs.	47430	47430	47430	47430	47430	47430

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Estimates of demand for yogurt. The first four variables are defined in Table 2, and the next three in Table 1. "No IV" reports estimates without instrumenting for price. 'P' stands for product fixed effects, 'B' for brand fixed effects, and 'T' for time fixed effects. When two letters are adjacent I use their interaction e.g. 'BPT' is the interaction between products, brands and time.

Table A.3: Instrumenting for Additional Variables

	1	2	3	4	5
Price	-1.05^{***}				
	(0.02)				
Feature	2.24^{***}	0.98^{***}			
	(0.09)	(0.12)			
Stores	5.05^{***}	4.29^{***}	4.12^{***}		
	(0.04)	(0.05)	(0.06)		
Private Products	-1.78^{***}	-1.29^{***}	-1.22^{***}	-1.84^{***}	
	(0.04)	(0.05)	(0.05)	(0.18)	
Med. Income	0.27^{***}	0.36^{***}	0.33^{***}	0.32^{***}	0.36^{***}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Perc. White	0.07^{***}	-0.03^{***}	-0.03^{***}	-0.05^{***}	-0.14^{***}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Med. Age	0.26^{***}	0.31^{***}	0.29^{***}	0.30^{***}	0.38^{***}
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Price (IV)		-4.20^{***}	-4.13^{***}	-4.06^{***}	-4.22^{***}
		(0.11)	(0.11)	(0.11)	(0.12)
Feature (IV)			4.84^{***}	2.49^{**}	5.02^{***}
			(0.59)	(0.88)	(0.99)
Stores (IV)				5.37^{***}	4.39^{***}
				(0.36)	(0.40)
Private Products (IV)					1.29^{**}
					(0.41)
Num. obs.	47430	47430	47430	47430	47430

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Estimates of demand for yogurt; all of the models have brand-product-time fixed effects. The first four variables are defined in Table 2, and the next three in Table 1. Each successive model instruments for one additional short run variable. Starting with Price, then Feature, Stores, and finally Private Products.