Nonlinear Pricing: Evidence of Price Discrimination in the Fluid Milk Market.

Ben Wallace

October 11, 2017

Abstract

Firms may use nonlinear price schedules as a method to second degree price discriminate at the detriment of consumers to increase profits. However, it is possible that these nonlinear price schedules may also increase welfare from the consumer's perspective if they are able to serve portions of a market that firms may otherwise ignore. The more extensive theoretical literature on price discrimination shows how firms may use nonlinear price schedules to price discriminate by distorting product characteristics from their efficient levels given heterogeneous preferences among consumers. Additionally, the theoretical literature finds that whether or not this practice is welfare decreasing or increasing from the consumer's perspective depends on a variety of things,[1] thus implying that the welfare implications of price discrimination from the consumer's perspective is an empirical question. Subsequently, some recent empirical papers have found this form of second-degree price discrimination to be welfare decreasing from the consumer's perspective, where as papers have found it to be welfare increasing depending on which market the paper studies. First, I model how consumers choose between types and sizes of fluid milk and some outside good. Variation in the ratio between marginal price per ounce and marginal cost per ounce over the menu of options is evidence that firms are using nonlinear pricing schedules to price discriminate. Once this evidence is shown, I use counterfactual analysis to determine the welfare implications of this practice within the market for fluid milk from the consumer's perspective. I find that firms within this industry are using nonlinear pricing to price discriminate, showing that approximately 70.4% of the markup differential between sizes can be explained by price discrimination. Through my counterfactual analysis, I find this pricing practice to be welfare decreasing from the consumer's perspective.

1 Introduction

Nonlinear pricing schedules are a pricing strategy in which firms vary the marginal price per unit of their goods over several sizing options. Generally firms use nonlinear pricing to provide quantity discounts to high demand type consumers, but it is also possible that firms may utilize pricing menus that provide a quantity premia.[2] Dolan (1987) states that firms use nonlinear pricing to either price discriminate, incentivize consumers to purchase the options which have a lower marginal cost, to remain competitive or some combination of the previous three. First, by varying marginal price per unit over size, firms are able to charge consumers with lower preferences for the good in question higher marginal prices per unit than those who prefer the good more and extract additional surplus

[1] Such as the distribution of consumer heterogeneity.

[2] Since firms within the fluid milk industry exclusively utilize nonlinear pricing schedules that provide quantity discounts, this paper will focus entirely on nonlinear pricing in terms of quantity discounts.

from the market. Firms may also use nonlinear pricing schedules to incentivize consumers to choose the larger option which is cheaper for the firm to produce per unit. Additionally, if a firm's competitors are offering cheaper prices per unit for larger sizing options, then they may have to follow suit or lose business. This paper will focus only on firms using nonlinear pricing as a way to price discrimination.

With price discrimination as the reason for nonlinear pricing in mind, we may see firms manipulate either the price[3] or the size of their products as to increase the marginal price/marginal quantity per unit ratio of a small version above the efficient ratio to extract additional surplus from the lower demand consumers who purchase the smaller sizes. The theoretical literature on price discrimination suggests that this is not possible for the largest sizes of a product offered, as high demand-type consumers have relatively more elastic demand functions than low demand-type consumers.[4] This finding from the theoretical literature suggests that the largest sizes of a product must be priced and sized at the efficient marginal price/marginal quantity per unit ratio. Since I will observe product menus that have nonlinear pricing through sizing options, there is concern that the smallest sizes are inefficient in terms of their marginal price/marginal cost per unit ratios. Since the practice of second-degree price discrimination may be at the detriment to consumers from a welfare perspective, this concern leads me to analyze the welfare implications of these nonlinear pricing schedules, for which the previous empirical literature has found differing answers.

The main empirical issue for estimating whether or not the theory holds in the case of nonlinear pricing schedules in the past had been due to data availability. Many other papers have estimated demand models using only product characteristics, prices and aggregate sales data. One of the strengths of this paper is that with the Nielsen Consumer Panel, I am able to observe household shopping decisions, as well as the demographic characteristics of the households, for every shopping trip they make for every year they participate in the panel. The Nielsen data allow me to use not only product characteristics to estimate the preferences of the households, but characteristics of these households to obtain a better fit to my discrete choice model by allowing for additional heterogeneity in tastes.

The market in question for this paper is fluid milk.[5] I choose to focus on fluid milk market because it has several convenient properties for the estimation of consumer preferences on the consumer side of the model, as well as marginal costs on the firm side. First, the USDA regulates and publishes raw fluid grade milk prices in the United States which provides a reasonable starting point for estimating the marginal cost per ounce of fluid milk. Estimating costs has been difficult for several other empirical studies and these regulations allow me to relax some of the assumptions that other papers have needed when modeling the firm's side of the model. Additionally, fluid milk sizing options in their current state are exogenous in that they follow the Imperial measurement system in the United States. This market norm allows me to focus this paper only on pricing and not have to worry about firms changing their menus through size. This is a potential weakness of the previous literature which chooses products for analysis where no such market norms exist. The main reason why the endogeneity of size may be a concern is that firms likely have the ability to choose both prices and quantities in the majority of other markets, a fact which has been mostly ignored by the majority of the previous literature and may lead to biased parameter estimates. Due the exogeneity of sizes in the fluid milk market, this potential source of bias is no longer a concern. Another strength of fluid milk for this type of analysis is that the outside good in question, or

[3]From here on forward, when I refer to price, I am referring to price per unit.

[4]Assuming the Spence-Mirrelees condition holds.

[5]Which I define to be fluid milk only from cows. This definition excludes non-fluid milks, i.e. evaporated or condensed milk, milks that come from non-cow animals and milks from plants. This is done for simplicity.

substitute good, can be well defined. I will assume that the outside good for fluid milk is any other consumable liquid product.[6]

There are a couple necessary conditions required for firms to price discriminate. These necessary conditions include: firms must face a downward sloping demand curve, firms must be able to prevent resale of their product, and consumers must have heterogeneous tastes for the product. One may question that these necessary conditions are met in this market, since fluid milk appears to be a fairly homogeneous good. However, several simple observations of this market make it clear that these conditions are met.

There are many different types of fluid milk that are offered by many different firms. Some examples of the ways that firms vary their fluid milk's product characteristics are through a variety of sizes[7], butterfat contents[8], flavors[9] and whether or not the product has the USDA organic seal. I even observe some producers of fluid milk produce several different brands targeted at different segments of the market. Since all of these different product characteristics in all types of combinations exist for fluid milk on stores shelves, consumers must have heterogeneous tastes for fluid milk, otherwise there would be no reason to have this much variety within the market.

Additionally, it seems reasonable to assume that resale is not an issue in this market for a couple of reasons. Fluid milk is a relatively small portion of the average consumer's budget in that generally one to two gallons per week are plenty for households with even strongest of preferences for milk and the average price of a gallon of fluid milk is a couple of dollars. Also, if resale were a concern in this market, we would observe consumers purchasing the largest size of one type of fluid milk at the efficient marginal price/marginal quantity per ounce bundle and then pour the portion that they do not plan to consume out into another container and attempt to resell it on some secondary market. I am not aware of such a market. Another reason this is a reasonable assumption is that resale would require a large amount of coordination between households, which would seem to have a much higher marginal cost in terms of effort than the marginal gain from saving a couple of pennies.

One may think that firms within the fluid milk market may not face a downward sloping demand curve and have no market power thus would be unable to price discriminate due to fluid milk being a fairly homogeneous good. For this, I provide a couple of observations as anecdotal evidence of price discrimination within the market. On one shopping trip at a Kroger store in October 2016, I observed that the price for private label[10] gallon containers of 2% fluid milk was listed at \$1.88 whereas the price of private label 2% quart containers was listed at \$1.89. The packaging of both containers were very similar. This observation is hard to explain through a story of channel efficiency alone, as it is not likely that the absolute cost of producing a gallon container of fluid milk is cheaper than producing a quart of milk. With this pricing menu, the large grocery chain is actually paying consumers a penny to take away an additional three quarts of fluid milk.

One of the largest sources of market power that firms likely have in this market are search costs. Since fluid milk is such a small portion of the budget constraint, household on their weekly shopping trip are highly unlikely to make a trip to a different store to purchase fluid milk, even if they know that they can purchase the product for \$0.50 less per gallon at a different location. Firms must have a decent idea as to what these search costs are for consumers and set their prices accordingly.

[6] The outside good includes all substitutes for fluid milk such as: pop, juice, water, etc.

[7] One and a Half Gallon, Gallon, Half Gallon, Quart, Pint, Cup, and a variety of multipacks.

[8] Skim, .5%, 1%, 1.5%, 2% and Whole fluid milk.

[9] Regular, chocolate, strawberry, etc.

[10] Private label meaning store brand product.

Even more evidence of firms within this market having market power was the difference in price between the non-organic regional brand of milk and the private label brand. A gallon of regional brand 2% fluid milk at this grocery store was listed at \$4.99 a gallon, more than twice the price of the private label of \$1.88. Unless this regional brand producer uses a wildly different production process, this more than two and a half times price differential is hard to explain through differential costs alone. There are two different explanations for this price differential. Regional firms could be exerting market power to charge a higher price for their "different, higher quality product." One concern with this explanation is that, all in all, the retailers have the final say in what price is seen by consumers, though many of these regional firms are known to provide incentives to retailers to set the price they prefer. An alternative explanation for this observation, based upon the fact that retailers have final say in prices, could be some profit maximizing behavior by retailers attempting to induce consumers to purchase their private label brand by having the shelf price of the regional brand good be significantly higher.

There are two main contributions of this paper. First, I show that firms use nonlinear pricing schedules to price discriminate among heterogeneous consumers with consumer choice micro data rather than aggregate sales and simulated consumer behavior. After estimating household preferences for fluid milk, I then use these parameter estimates, which allow for additional heterogeneity in tastes among households through the micro data, to compute the welfare implications of this form of price discrimination from the consumer's perspective. To estimate the welfare implications of price discrimination within this marketplace, I compare the total welfare in a hypothetical world where price discrimination via nonlinear pricing schedules is not allowed to the total welfare of the market place as it exists currently. This has been done in the past[11] but as stated previously, has yet to have been agreed upon within the empirical literature. Since this paper uses actual consumer choices during shopping trips, I find my result more compelling as the previous papers use only aggregate sales data and simulated consumer decisions.

This paper is organized as follows. Section 2 will provide a layout of the composition of the market for fluid milk. Section 3 will present a literature review of both the existing theoretical and empirical literature on price discrimination as it relates to this paper. Section 4 will discuss the datasets used for this paper. Section 5 will present a simple model for nonlinear pricing in the case of quantity discounts and discuss its implications. Section 6 will provide preliminary evidence of nonlinear pricing being used to price discriminate in this market. Section 7 will present the model that I will estimate for consumer demand as well as a strategy for estimating the marginal cost of milk. Section 8 will present the results of both the consumer and firm sides of the model, as well as where the consumer welfare counterfactual analysis is presented and section 9 will conclude.

2 Market Description: Fluid Milk in USA

With many steps of the production process, the market structure, and regulation of bulk milk prices, fluid milk retail pricing can be somewhat complex. This section of the paper will discuss all of these in three subsections. The first subsection will discuss the production process and market structure for fluid milk, whereas the second will discuss the regulations that are in place for the fluid milk market, and lastly the third will discuss the implications of the market structure and regulations.

[11] See Leslie 2004 and others.

2.1 Production Process and Market Structure

The first part of the production process starts at a farm where farmers raise cows that produce bulk milk. There are also seasonal supply fluctuations for production, as is true with almost all agricultural products. In the spring and early summer, cows produce more milk, whereas production of milk by a cows lowers in the fall and winter. Once the cows produce the bulk milk it is next sent to a processor that separates, produces and packages the various dairy products that come from the bulk milk. Once the production process is complete, the finished fluid milk product is sent to a retailer where, according to Manchester and Blayney (USDA 2001) [MB01], the retailer then sets the price the final good and consumers make their consumption choices.

Dairy farmers may also have an incentive to form cooperatives to help gain market power when dealing with processors. These cooperatives control the supply of milk sent to processors and help eliminate some of the risk that a dairy farmer may face due to daily, season and other types of market fluctuations.

The production process for fluid milk has lead to an interesting market outcome in this industry. Many of the large grocery chains are fully integrated and have their own processing plants. Unfortunately, this is not the case for all large grocery chains within the industry. If it were, a possible way to identify which UPCs come from a vertically integrated production process would be via private label packaging vs regional brand. However, Nielsen masks the identity of each store's parent company by just assigning a common identifier among stores who are owned by the same parent company due to confidentiality reasons. As such, information on which large chains are vertically integrated is not useful, as it is not possible to line up the parent companies with the stores within the data.

Another difficulty with identifying which goods are produced via a vertically integrated production process is that some firms are vertically integrated in some parts of the country but not in others. For example, consider the following quote from Dean's, the largest regional brand fluid milk firm found in my data, website:

> The Company (Dean's) is one of the nation's largest processors and direct-to-store distributors of fluid milk marketed under more than 50 local and regional dairy brands and private labels.

This quote illustrates another issue with identifying which products come from a vertically integrated production process or not, as some of the "private label" brand goods, which are usually what is thought of when discussing a product that is produced within a vertically integrated firm are not necessarily from a vertically integrated production process.

Additionally, which processing plant that grocery stores are receiving their fluid milk from is not readily available due to confidentiality as well, though where potential processing plants that supply the stores are known.

All smaller convenience stores deal with independent milk processors. Many of these stores only carry one brand of fluid milk in only a few sizes. Due to this, as well as the generally higher prices, it is likely more appropriate to treat these convenience stores as a completely different market.

According to MB01, in the past many large grocery chains had several[12] regional brands on their shelves that were fully serviced by a representative from each brand respectively. In this regard, fully serviced means that each of the brands would have a representative bring the fluid milk to the store and have that representative stock, rotate and care for their brands space. In recent years,[13] there has been a shift from grocery chains to only have one or no regional brands

[12]They say two to six.

[13]MB01 states the 1990s onward.

in their floor plan. In addition to this, these regional brands no longer fully service these stores, instead they ship their product to the store and it is handled by the retailer. Once the product hits the loading dock of a retailer it becomes their property, although many of these regional brands will refund the retailer for damaged and outdated goods.

2.2 Regulations

Throughout this production process, there are many regulations set in place by the government. These pricing regulations are all placed on the processors and aim to benefit domestic dairy farmers. According to MB01, these regulations include: Federal milk price support, Federal milk marketing orders, import restrictions, export subsidies, domestic and international food aid programs, state level milk marketing programs and multi-state milk pricing organizations. Each of these regulatory practices has its own goal, some of which will be discussed below.

Federal milk price support was first put into place in 1949. These price supports provide a price floor for the price of bulk milk, though the price floor often does not bind. Federal milk marketing orders (1937) also provide a similar function in that they set pricing minimums for bulk milk, regulate the quality of fluid milk and limit the monopsony power of dairy processors.

The federal government also attempts to protect domestic dairy farmers through import restrictions and export subsidies. These two processes in conjunction give domestic dairy farmers the upper hand against foreign competition.

In addition to Federal pricing support, many states have their own policies which vary from state to state.

Overall, the goal of these regulations are to protect domestic dairy farmers both from international competition and dairy processors. These regulatory policies are stated to aim to help domestic dairy farmers by ensuring a minimum price in the market for bulk milk as well as ensure the quality of the fluid milk that comes from the processor. It also seems that the regulations that exist aim to lower risk for dairy farmers, though this is not explicitly stated by the USDA.

These data on bulk milk prices are available from the USDA's National Agricultural Statistics Service. The butterfat, bulk and skim milk prices are reported monthly at the state level.

2.3 Implications

According to MB01, even with the complex market structure and regulations in place in the fluid milk industry, ultimately the retailer is allowed to set any price they see fit in regards to the price that consumers face.[14] Even though each may set whichever price they see fit in terms of the prices that consumers face, many of the regional brands provides have a strong incentive to nudge retailers to set prices that the regional brands would prefer.[15] There is no federal regulation that requires a retailer to set a certain price for each good,[16] thus retail prices are determined by some combination of consumer demand, competition between retailers, and wholesale price.[17]

[14] Though the range of prices that stores may set depends on which state they are located in. It is my understanding that in states such as California, grocery stores are allowed very little freedom to set their own prices due to the state level price regulations implemented by the California legislature, but in other states, no such pricing regulations exist at the state level at all. I am currently working on finding this information out for the states in my sample, but this information has been difficult to find.

[15] One such way of doing so, which I have heard is the case in the soda industry, is to provide large cost reductions to stores if they set the price at $X during a specific period of time.

[16] Though there may be state level regulations that impact this decision.

[17] Which is a function of the Federal regulation on butterfat, skim and bulk milk prices.

3 Literature Review

In this section, I summarize the current literature that exists on nonlinear pricing and price discrimination, starting with the theoretical literature then the empirical literature. The theoretical literature on price discrimination is quite extensive, as is the case with the majority of topics within the industrial organization field, whereas the current empirical literature is not nearly as extensive, mainly due to data limitations.

The story for price discrimination starts with Dupuit (1849). Dupuit laid out the necessary conditions for price discrimination by a profit maximizing monopolist as well as provided an example. In his paper, Dupuit discusses the several classes of tickets offered by trains. These different classes were used to charge higher demand consumers higher prices as well as used to induce the marginal consumer to purchase a higher class ticket then they might otherwise purchase. Pigou (1920) and Robinson (1933) both more formally define the degrees to which a monopolist can price discriminate as well as named them.

The first couple of modern papers to discuss second-degree price discrimination were Mussa and Rosen (1978) and Maskin and Riley (1984). These papers consider a monopolist's design of a price and sizing menu when consumers' preferences towards the product are unknown to the firm, but have some distribution that is known to the firm. The equilibrium that is found by both papers is that the highest demand consumer is offered a product that is efficiently sized and all other consumers self select into a smaller option that is tailored to their preferences though inefficiently sized.

Many papers have used this result and attempted to explain why firms may use nonlinear pricing in such a way. Dolan (1987) provides three possible stories as to why firms may use nonlinear pricing as a mechanism for price discrimination: heterogeneous consumers, channel efficiency and competitive bidding.

A quote from Buchanen (1953) sums up the heterogeneous consumers reason best"the demand schedule of small buyers is more inelastic over the relevant price range than that of large buyers". This variation in consumer type allows firms to exploit small demand consumers while having to somewhat cater to those with high demand in lower pricing per unit.

The channel efficiency argument for price discrimination is that firms may have cost savings if consumers buy specific quantities of their goods. In an anecdotal example of this Dolan shows why Sealed Air Corporation (SAC) provides lower marginal prices to those who purchase larger quantities. If SAC's product is sold in quantities that are smaller than a truckload, they must send one truck to multiple locations to serve many consumers, but if one consumer purchases an entire truckload of their product, then they only have to send the truck to one location. Having a truck serve a single location only results in considerable shipping cost savings for the firm. To induce consumers to purchase entire truckloads of their products, SAC lowers marginal price as the quantity that consumers purchase increases.

The last reason for price discrimination discussed by Dolan is competition over market share. Dolan states that if only a portion of firms within a market place utilize nonlinear price schedules to provide quantity discounts, then the firms who do not may not be able to compete with their competitors for consumers whose demand is relatively more elastic. The example he uses is electronic components suppliers. When interviewed as to why they offer lower marginal prices to large buyers of their products, they stated that it was because the other firms do it too.

Another paper that looks at the welfare implications of price discrimination is Sharkey and Sibley (1993) [SS93]. In this paper, SS93 model two types of consumers and test the standard finding within the literature that marginal price per unit is equal to marginal cost per unit for the highest demand users. In their model, SS93 find that the distribution of consumer types determines if marginal price per unit is equal to marginal cost per unit for the highest demand users of the goods. Depending on how types are distributed, equilibriums of marginal price per unit higher than marginal cost per unit and marginal cost per unit higher than marginal price per unit can be supported. SS93 also test to see if price discrimination can exist under competition. With their model, they find that price discrimination can exist under duopoly, but only if there is a regulator that favors high demand users of the good over low demand users.

The final theoretical paper that I will discuss that takes the welfare implications of price discrimination into account is Reiss and White (2006) [RW06]. In their paper they provide a method for welfare analysis that estimates consumer surplus in the face of menu pricing. They too find that the distribution of demand determines the welfare implications of nonlinear pricing.

Due to recent advances in empirical techniques as well as increases in data availability, the empirical literature on price discrimination has been growing in recent years. Some of the earlier work uses reduced form estimation techniques, as opposed to the newer empirical papers on nonlinear pricing which use structural models to estimate demand for the goods in question. One downfall of the current literature is that almost all of these papers use aggregate sales data rather than actual consumer choice. The first of these types of papers I present Friebel et. al (2015) [FOG15]. FOG15's finds evidence for price discrimination in the market for wheat in Russia, a fairly homogeneous good similar to fluid milk. Their paper deviates from mine in that they focus on price discrimination in the export market rather than in the consumer goods market. This allows them to use a reduced form approach that is inappropriate for the questions that this paper attempts to answer. FOG15 finds evidence for price discrimination by Russian firms in 25 out of 61 destination countries within their data set over the 2002-2011 period.

The more closely related literature to my paper are papers that use structural models of demand. These papers include: Leslie (2004), Cohen (2008), Liu and Shen (2012) [LS12], Miller and Osborne (2014) [MO14], and McManus (2007).

Leslie (2004) uses data from the Broadway show *Seven Guitars* to estimate the welfare and profit implications of price discrimination. In this paper, Leslie uses a structural model that allows him to use counterfactual analysis that allows Leslie to estimate the welfare implications for price discrimination. Leslie finds that the price discrimination observed by the firm increases profits by 5% but has little to no effect on consumer welfare, thus price discrimination in this case seems to increase total surplus.

Cohen (2008) estimates how packaging size is used in the paper towel market as a vehicle for price discrimination. In this paper, Cohen uses a discrete choice model with product characteristics and aggregate sales data to estimate model parameters for demand. This allows Cohen allows to perform counterfactual analyses similar to Leslie 2004. Cohen finds quantity discounts that are consistent with second degree price discrimination rather than what could be attributed to cost differences across the sizes of the goods. Cohen also finds that consumers are better off with price discrimination as there is more competition amongst firms in the multi-roll package size segment of the market due to the pricing strategy.

In a working paper, Liu and Shen (2012) [LS12] estimate the degree to which firms in the carbonated soda market are able to price discriminate based upon firm type; i.e. private label brand or national brand products. LS12 estimates a discrete choice model using a subset of the data that this paper uses, the Nielsen Scanner dataset, but do not use the Nielsen Panel data and thus, must simulate individual consumer preferences. LS12 also uses the assumption that firms in this market compete in a Bertrand-Nash equilibrium setting that allows them to estimate the marginal cost of the sodas in their model, though this assumption is hard to believe as these firms clearly do not compete in such a way. This assumption allows LS12 to identify to what extent both private labels and national labels are able to price discriminate. Their main findings are that private labels are able to price discriminate just like national brands within this market, which they claim at least somewhat explains the growth of market shares of private label sodas.

Miller and Osborne 2014 [MO14] use data of the Portland cement industry in California, Nevada and Arizona to estimate a structural model of spatial differentiation and price discrimination. MO14 finds that if price discrimination were not allowed, consumer surplus in the cement industry would increase by approximately $30 million dollars per year, thus implying that price discrimination is welfare decreasing from the consumer's perspective,[18] a result not commonly found within the previous literature.

McManus (2007) uses a structural discrete choice model to estimate whether or not specialty coffee shops are able to price discriminate using sizing for their good. McManus also looks at how nonlinear pricing may not only lead to different marginal prices per ounce along sizing options, but also how it may distort product characteristics away from their efficient levels. He finds that nonlinear pricing allows firms to make size of the smallest cups of specialty coffee "inefficiently small" for the price that is set, which McManus concludes that firms do so in part to incentivize consumers to either buy the next size up or have to pay additional rents to the firm. He also finds that these distortions do not exist for the highest demand consumers, a finding that is consistent with the theoretical literature.

The model that will be estimated in this paper is based on Cohen 2008 model, but deviates in two regards. First, I observe consumer demographics for each purchase, thus I do not have to use Cohen's noisy measure of the population demographics within physical market to deal with potential sources of consumer heterogeneity. As such, my second deviation from Cohen 2008 is that I do not utilize the random coefficients framework for estimating my model, which Cohen uses to address issues of substitution patterns that exist within multinomial logit modeling, the interdependence of irrelevant alternatives (IIA). My reason for not addressing this within this paper is that by utilizing household characteristics, the IIA issue is less of a concern. For example, suppose a regional brand firm introduces a new skim milk product to the marketplace. Household types that that are more likely to consume products similar this type of product, such as households that are more likely to a different regional brand skim milk, are far more likely to switch than households who are more likely to consume a far different product, such as private label whole milk. Since I have household characteristics, I identify which type of household prefers each product type, making the IIA issue far less of a concern.

In summary, the theoretical literature on price discrimination defines price discrimination and has attempted to rationalize this pricing strategy commonly used by firms, whereas the newer empirical literature on price discrimination uses structural models to estimate demand for goods, then discusses the price discriminatory implications of their parameter estimates via counterfactual analyses. The main theme from the theoretical literature is that since the demand for the good is relatively inelastic for low demand users as compared to high demand users, firms are able to set higher marginal prices per unit for low demand users and it largely agrees that the highest demand user will always have an option where marginal price is equal to marginal cost.

The empirical literature largely answers two main questions: are nonlinear pricing schedules used as a way for firms to price discriminate and if so, what are the welfare implications within the market place. The main deviation within the empirical literature are how the supply side/marginal

[18]The change in total welfare is unknown to Miller and Osborne, as they do not model firm profits, but can infer that price discrimination must be profit increasing as otherwise firms would not have an incentive to utilize the practice.

costs are dealt with. Some papers (Liu and Shen (2012) and Cohen (2008)) model and estimate marginal costs of a representative firm, whereas other papers either do not have to worry about marginal costs (the Miravete papers, Ayral (2014) and Leslie (2004)) as the marginal costs are sufficiently small in their industry, or they have discussions with firms to find what the costs of the products are (McManus (2007)). Estimates for marginal cost are important in this literature as many of the findings could also be explained as costs savings by the firm without them.

4 Data

Data for this paper come from the Nielsen Panel and Scanner Datasets for the years 2008 and 2009.[19] In the subsections below, I discuss the characteristics of these datasets, as well as any other supplementary datasets that have been used.

4.1 The Nielsen Panel Dataset

The Nielsen Panel dataset includes 40,000 households from 2004-2006 and was expanded to 60,000 households starting in 2007 and thereafter. These data include each and every item purchased by a household within the panel year at the household-shopping trip level. Since the observation level of these data is the household-shopping trip, the Nielsen Panel dataset includes household demographics for each participating household that participates, as well as the Universal Product Codes (UPC) of all items purchased by the household purchases during the particular shopping trip to the store in question. Due to potential differences in preferences that may vary within regions,[20] I use households from the sample that live in FIPS Division 3: Midwest East North Central.[21] This FIPS division has some convenient qualities that will be discussed later on as well.

The strength of using the Nielsen Panel data as opposed to datasets used in previous studies is the observation of actual purchasing choices by households rather than aggregate sales by firms. With these data I am able to use real household characteristics[22] coupled with actual choices rather than simulate consumer heterogeneity as done previously by Liu and Shen (2012), Cohen (2008), Leslie (2004) and McManus (2007). One potential drawback of this panel is the high attrition rate of the panelists. This may be due to the somewhat weak incentives for participation that will be discussed later. The Nielsen states that it generally retains 80% of their panel from year to year[23]. This is not particularly important to this paper as I am more interested in firm behavior, so I will be treating this panel as a repeated cross section from year to year, though I use the panel aspect of the data to construct a psuedo-purchasing history variable, the time (days) since the household

[19] The Nielsen Panel itself goes from 2004-2016, but due to a variety of reasons including: potential changes in preferences that may occur over the long-run, macroeconomic factors, and most importantly, the quality of these data are much better over this two year period than any other two year period in the panel. After removing inactive panelists and trimming down to an individual region, 13,352 households with 22,154 household years are left in my sample. As such, I omit all other years from the analysis.

[20] And also in part due to the size of these data.

[21] The states included in this region are: Illinois, Indiana, Michigan, Ohio and Wisconsin.

[22] Which include: household income, household size, type of residence, household composition, age and presence of children, male and female head employment status, male and female head education, male and female head occupation, marital status of the head of household, race, location of the panelist, a dummy variable for WIC program participation, and the age of the male and female heads of household. See appendix D for tables with summary statistics for the panelists.

[23] In my subsample of these data, I find that 2,306 households only participate in the year 2008, 2,244 household only participate in 2009 and 8,802 households participate in both years.

last purchased fluid milk, to include how the stock of fluid milk the household currently has may affect future or current purchasing decisions.

Nielsen uses a stratified, proportional sampling technique to create a representative sample of the continental United States. Households are randomly selected and invited to join the panel either through a mail or email invitation. Nielsen does not directly pay the households selected to participate in the panel, but it does provide some other incentives. These incentives include: monthly prize drawings, gift points awarded for weekly transmission of data and a sweepstakes. Nielsen states that they also try to encourage participation and create enthusiasm through ongoing communication. They use the following methods to do so: a monthly newsletter, telecommunications, Q & A section with helpful tips and reminders, personalized computer tips and reminders after transmitting, notice of monthly sweepstakes winner, personalized letters for reporting problems and questions, letters from the president, gift point statement, help desk, an 800 number for panelists to call and exit interviews.

In addition to providing rich household characteristics, this dataset also has a variety of product, brand, and retailer demographics as well as the overall purchases by households.[24] Sales of UPCs sold that are in sizes that do not utilize the Imperial measurement system are discarded from this sample for estimation purposes, but kept for the construction of other variables.[25] With that being said, the discarded portion of the data are less than 1% of the overall sample. Additionally, the removal of these sizes allows me to clearly define the top of the menu, as previously it was somewhat muddled. Summary statistics for this data cleaning procedure are included in appendix E.

4.2 The Nielsen Scanner Dataset

Ideally, I would observe not only what households choose to purchase during a shopping trip, but all of the choices the household did not choose as well. The Nielsen Panel dataset only provides information on the choices that households make on a particular trip, rather than the full menu of options that household faced. Since this is the case, I utilize the Nielsen Scanner dataset to simulate the remaining portion of the choice set the households did not choose. The Nielsen Scanner dataset contains the weekly sales of each UPC sold at a store as well as product characteristics similar to what are provided within the Panel Dataset. See appendix F for notes on how the remaining portion of the menu that households face is simulated using the Nielsen scanner data.

4.3 Costs of Raw Fluid Grade Milk

The USDA regulates the price that farmers sell butterfat, bulk and skim milk to dairy processors. The prices of these goods are all regulated by the USDA are published in the National Agricultural Statistics Service's (NASS) monthly report. I use these sales prices as a way to control for the input cost of fluid milk.

[24]Observable with this dataset are: the date of each purchase, a store code and location which provide information on the type of retailer, the amount spent by the consumer on each shopping trip, a product UPC for each item purchased, a retailer code which provides information on the type of retailer (grocery store, mini mart, etc.), the price paid for each item, the packaging size for each item, an organic claim dummy variable for each product, various product demographics, and the brand of each product.

[25]UPCs with containers of the following sizes remain: gallon, half gallon, and quarts.

5 A Simple Model of Nonlinear Pricing: The Case of Quantity Discounts

The predictions from a simple theoretical model can lend us useful insights in this case of utilizing nonlinear pricing as a method of second-degree price discrimination. Consider a representative monopolist offering nonlinearly priced menu options of some good Q to consumers with heterogeneous preferences. In this model, I assume that the nonlinear pricing schedule the representative monopolist constructs offers a quantity discount for their good.

Model The representative firm offers three price quantity bundles for a good: (p_l, q_l), (p_m, q_m) and (p_h, q_h). Consumers in this market may purchase only one price-quantity bundle and there is only one period in the model. I assume that consumers preferences for the good are heterogeneous and that an optimal solution of menu with three choices exists. I also assume that $p_l < p_m < p_h$ and $q_l < q_m < q_h$, which is required for the menu that is offered to provide a quantity discount. Additionally, assume that the representative firm is unwilling to sell any quantity above q_h to any consumer, such that the price of all quantities higher than q_h is set at $+\infty$. In the case of a quantity discount, at least one of the following inequalities must be true:

$$\frac{p_l}{q_l} < \frac{p_m}{q_m} \leq \frac{p_h}{q_h} \tag{1a}$$

$$\frac{p_l}{q_l} \leq \frac{p_m}{q_m} < \frac{p_h}{q_h} \tag{1b}$$

$$\frac{p_l}{q_l} < \frac{p_m}{q_m} < \frac{p_h}{q_h} \tag{1c}$$

A graphical depiction of both the menu of options and the budget constraint that consumers face in case of (1c) are provided as figure 1 and figure 2 in appendix A. For simplicity, I linearize the budget constraint that consumers face though it is important to note that consumers do not have the option to choose any quantity other than q_l, q_m or q_h to purchase in this model.

This situation yields the following equation for the linearized budget constraint:

$$Y = \begin{cases} I - p_l q, & 0 \leq q \leq q_l \\ I - (p_l - p_m)q_l + p_m q & q_l < q \leq q_m \\ I - (p_l - p_m)q_l - (p_m - p_h)q_m + p_h q & q_m < q \leq q_h \end{cases} \tag{2}$$

where I is one's income, q is the quantity of the good that is consumed and Y is the numeraire good. The price of the numeraire good, Y, is normalized to 1. The way that figure 2 is drawn, if a consumer chooses to consume q_h they spend their entire income which is implicitly assuming that $I = p_h q_h$. If $I > p_h q_h$, then the entire graph shifts upwards by $I - p_h q_h$ which would be the minimum amount of Y that is consumed regardless of their choice of Q.

There are some general conclusions that can be made from this graph. Due to the quantity discounts in this pricing menu, consumers in this market face a nonconvex budget constraint. This causes the kinks that are located at q_l and q_m. Utility maximizing consumers are unable to locate at q_l or q_m unless the numeraire good, Y, and Q are perfect complements. In the case of goods that are not perfect complements, only consumers who consume q_0 and q_h are utility maximizing. This is due to the nonconvexity in the budget constraint.

Consumers with a strong preference for good Q will have indifference curves similar to IC_b and locate closer to or at q_h. Those with weaker preferences for good Q will have indifferences curves similar to IC_a and locate near or at q_0.

Consumers may have indifference curves like IC_a where their tangency to the budget constraint is located between two points. If this is the case, in data I will observe them locating on the discrete point nearest to their tangency to the budget constraint. For this particular consumer, I would likely observe them consuming none of the good Q in data.

Consumers may also have indifference curves that look like IC_b. This consumer would be indifferent between consuming at a point between q_l and q_m or q_m and q_h. Since it is not possible for consumers to consume at either point, I may observe one of two things in data. They may just consume at the point their indifference curve is centered around, in this case q_m, and locate at on an indifference curve that is below their utility maximizing level, $IC_{b'}$. If I introduce dynamics to the model, as seen in the data, they may also consume some combination of the other two quantities over time in an attempt to put them on their utility maximizing indifference curve IC_b. For the case of consumer b, I would observe them choosing q_h some portion of the time and q_l the rest of the time. Which option consumers with indifference curves like IC_b choose depends on where their tangencies are located on the budget constraint. If they are closer to the point they are centered upon, they will likely just choose to consume at that point and be pooled with those whose preferences locate them close to that quantity. If the tangencies are closer to the other points that surround the indifference curve that they are centered upon, then they may exhibit the dynamic behavior described above.

There is one general welfare implications to be discussed with this simple model. Since utility maximizing consumers are unable to locate at points q_l or q_m, consumers who do not locate at one of the corners are worse off due to the nonconvexity of the budget constraint under the assumption that Y and Q are not perfect complements. If the firm were to offer a continuum of options rather than the discrete points, these consumers would be able to locate at their utility maximizing level and would be better off.[26] Thus, the discrete menu of options, along with the quantity discount, create a nonconvex budget constraint reduces consumer surplus for consumers who do not locate at either corner. Unless the representative firm, or the many firms in the case of my data, is gaining profits that are greater than or equal to the reduction of consumer surplus, this simple model would suggest that nonlinear pricing resulting in a quantity discount is welfare decreasing.

[26]Note: Offering a continuum of options would still allow the firm to price discriminate, but consumers would still be better off it would be possible for them to locate at a tangency within the interior of their linearized budget constraint.

6 Preliminary Evidence of Price Discrimination

This section provides simple reduced form evidence that is consistent with the idea that firms within the fluid milk market are using nonlinear pricing as a way to price discriminate. First, I provide several graphs and tables that are consistent with price discrimination, then I will provide regression results from a simple hedonic pricing model. The reduced form regression results that are presented in this section are also consistent with price discrimination.

Table 1 and table 2 present unit-price statistics and figure 3 provides graphs of the mean price, price-cost differential, and price-cost ratio for units sold by the firms in these data. The graphs break down the price statistics into several categories, purchases made with and without coupons[27] and purchases of regional brand vs private label products. Since I assume that fluid milk is only sold in containers that are sized using the Imperial measurement system, all observations fall into one of three categories: gallon, half gallon or quart.

Since the average price-cost differential and price-cost ratio generally decrease as the size of a product increases, figure 3 is consistent with firms using nonlinear pricing to price discriminate. Additionally, for gallon containers, which are assumed to be the largest sized offered by firms, the price-cost differential is very close to zero, suggesting that these products are close to the efficient price-quantity bundle, an observation that is consistent with the current literature. It is interesting to note that in the graphs on the right, it appears that firms are not using nonlinear pricing to price discriminate between cup and pint sized units. These means are very close however and as a whole, there is a downward trend. Additionally, these are just means and do not control for differences other than size.

Another interesting portion of the data is the pricing behavior of private label goods. It is interesting to note that not only are these graphs consistent with a story of firms using nonlinear sizes to price discriminate, but that it holds for the smallest sized containers as well. This suggests that firms are not only able to use nonlinear pricing to price discriminate with their private label brand products, but they are able to do so to a greater degree than regional brands. This observation agrees with the main finding of Liu and Shen (2012) in their study of the soda industry.

After observing these simple trends in price statistics along sizes, I estimate the following price hedonic model:

$$PPO_{ut} = \beta_1 Size_u + \beta_2 CPO_{ut} + X'_{ut}\gamma + \epsilon_{ut}$$

where PPO_{ut} is the average price per ounce of UPC u at time t. $Size_u$ are a vector of size dummies for product u.[28] CPO_{ut} is the price of raw fluid grade milk per ounce of UPC u at time t. X'_{ut} are a vector of various product characteristics, year and market controls that are included in the table.

Additionally, price may vary from brand to brand, thus I also estimate the following model with brand fixed effects:

$$PPO_{ut} = \beta_1 Size_u + \beta_2 CPO_{ut} + X'_{ut}\gamma + \delta_u + \epsilon_{ut}$$

where δ_u is a brand fixed effect for UPC u. The results of these regressions are presented in table 3. The first two columns of table 3 are the simple price hedonic models[29], and third and fourth column

[27]Nielsen does not provide a clean definition as to what a "coupon" is beyond a piece of paper that is given at the register to provides a discount.

[28]In these data, size for each UPC does not vary over time.

[29]The first column without any controls and the second column with controls included.

of table 3 are price hedonic models with brand fixed effects.[30]. All of the coefficient estimates for size in ounces negative and statistically significant at the 1% level of significance, except in the case of models with controls for 1.5% fluid milk[31]. Since all of these regressions include the published raw fluid grade milk prices, these findings are consistent with firms using nonlinear pricing schedules to price discriminate in this market. This finding is not only robust to model specification[32], but also to the different butterfat contents of fluid milk.

These findings do not come without a warning however. First, all of these models implicitly assume that container size is continuous and that the product set is dense in ounces. As such, the interpretation of these coefficient estimates are equivalent to estimating consumers average marginal willingness to pay for an additional ounce of fluid milk. Since these data are measured using the Imperial measurement system, the variation in size that these models are using to identify these coefficients is in large discrete changes in size, rather than small changes in size. Since this is the case, it is clear that these coefficient estimates are not properly identified.

Following an argument originally presented by McManus (2007), this econometric strategy would be appropriate if firms offered a menu of sizing options that were continuous in size, or it were possible to purchase any amount in ounces at the very least. This is the major downfall of reduced-form econometric identification strategies in this scenario, as the model is attempting to identify an effect that is implicitly continuous, with variation that is in large unit changes. Another downfall of this identification strategy is for sales of the largest container size, as I do not observe a size that the household is unwilling to pay. Since I will never see a price they are unwilling to pay for an additional ounce of fluid milk it is impossible to identify their marginal willingness to pay for an additional ounce of milk.

With these two main points in mind, these results should be taken as a signal that firms may be using nonlinear pricing schedules to price discrimination in this market, and a more appropriate approach should be taken to show that this is the case. The structural approach proposed in the next section of this paper does not suffer from these weaknesses, and the marginal willingness to pay for an additional ounce of fluid milk is properly identified even with these large discrete changes in size.

According to Berry (1984), there are several other strengths to estimating structural models instead of reduced form models for demand estimation in addition to previous concerns. He states that it is possible to estimate a reduced form model for demand, but due to the endogeneity of price that will be discussed later, one must find a set of valid instruments to estimate the price elasticities of each good. The problem with just finding instruments and using reduced form techniques in this case is that in a market with N goods there are N^2 elasticities to estimate. Parameterizing the consumer utility function allows one to estimate the N^2 cross-price elasticities from far fewer parameters.

Another strength to structural estimation that Berry points out in his paper is the ability to perform counterfactuals. This strength of structural estimation is particularly crucial to this paper, as I intend to estimate the welfare effects of price discrimination within this market as well as check to see if firms could be better off without following the Imperial measurement system for their menu of options.

The last strength of structural estimation that Berry discusses is the model's ability to allow one to move easily between statements about aggregate demand and statements about consumer

[30] The third column without any controls and the fourth column with controls included.

[31] May be due to sample size since only 2,487 observations of 1.5% fluid milk in my sample of approximately 2.365 billion observations.

[32] Controls or no controls, and OLS versus brand fixed effects

utility. This will be very important for this paper in the welfare analysis section.

With both Berry and McManus' comments in mind, it seems appropriate to estimate a structural model for demand rather than a reduced form model in this context, as I will be able to cleanly identify the marginal willingness to pay for an additional ounce of fluid milk for each consumer type. Additionally, I will also have the ability to perform the interesting counterfactuals that are not possible in a reduced form model.

.

7 Structural Model

In the first subsection of this section of the paper, I estimate the demand for fluid milk using a discrete choice model. In the subsection that follows the demand portion of the model, I introduce a model for estimating firm's marginal costs within the fluid milk market. Both the consumer and firm sides of the model follow closely to Cohen (2008). It is important to estimate the demand side of the market separately from the supply side of the market due to the endogeneity of prices.[33] Following a reduced form approach in estimating this type of model will cause estimates to be biased due to simultaneity of the demand and supply portions of the market.

7.1 Demand for Fluid Milk

I assume that each household, i, chooses to purchase one or zero units of a fluid milk UPC during each shopping trip.[34] For simplicity, my model does not consider the household's decision of whether or not to go on a shopping trip, as well as which store to patron during a trip if they choose to go on one.[35] The decision to purchase zero units of a fluid milk UPC is described as choosing the outside good.[36] Purchasing options are indexed by $j \in J$. I assume that all stores that households choose to patron for a shopping trip offer all possible choices within the choice set J.[37]

Each fluid milk UPC is grouped into a choice, j, within the choice set, J to reduce the vast number of UPCs of fluid milk that the household could potentially choose from. Each choice, j, is characterized by: (1) BFC_j, which is a dummy variable that is equal to one if the UPC has a butterfat content of 2% or greater; (2) size_j, which is a dummy variable that is equal to one if the UPC is sold in a gallon sized container and zero if the UPC is sold in a half gallon or quart sized

[33]In the model that is presented in this section, I will assume that there are some characteristics of each product that are observed by both firms and consumers but unobserved to the econometrician. It is fair to assume that a profit maximizing firm will take these characteristics into account when determining what price to set when constructing each price-quantity bundle. Thus by construction, price is correlated with part of the error term of the model and is endogenous.

[34]I define a shopping trip to be going to a store that offers a menu of products that includes fluid milk

[35]This is important to point out as Thomassen (2017) states that a source of potential bias in cost-side parameter estimates exists from ignoring the cross-category pricing effects of other goods the firm sells. A portion of the motivation for this type of bias is that households consider the whole basket of goods they intend to purchase when choosing which store to patron. By removing the household's choice of when and where to go on a shopping trip, and considering how small of the portion of the average household's budget is in terms of fluid milk expenditures, this potential source of bias is far less of a concern for my model.

[36]Which as previously defined, includes all consumable liquids other than fluid milk from a cow.

[37]As described in Section 4.2, the Nielsen Panel dataset does not allow me to observe the entire menu of goods available for households to choose from, but only those actually chosen by the household. As is the case, I use the Nielsen Scanner dataset to simulate the menu of products the store had available on the date of the shopping trip as outlined in that section.

container;[38] (3) PL_j, which is a dummy variable that is equal to one if the UPC is a private label branded[39] product; (4) ξ_j, product characteristics of the UPC that are observed by economic agents but unobserved to the econometrician;[40] (5) the unit-price, p_j.

Thus, the indirect utility associated with purchasing product j, which I assume to be linear in product characteristics, is:

$$U_{ij} = \beta \text{BFC}_j + \tau \text{size}_j + \alpha p_j + \pi_{ij} + \xi_j + \epsilon_{ij} \tag{3}$$

where

$$\pi_{ij} = (H_i \times \text{size}_j)\rho + (H_i \times \text{BFC}_j)\gamma \tag{4}$$

where H_i is a vector of household characteristics[41] and ϵ_{ij} represents household i's deviation from the mean preference for good j, which I assume to be distributed iid with a type I extreme value distribution (TIEV).

Since I assume ϵ_{ij} is distributed iid TIEV, the household's indirect utility becomes a multinomial logit model (MNL) where the probability in which household i chooses choice j is:

$$S_j = \frac{\exp(\delta_j)}{\sum_{k=1}^{J} \exp(\delta_k)} \tag{5}$$

where $\delta_j = X_j\beta + \tau q_j + \alpha p_j + \xi_j$ is the population mean utility for product j.

A well known issue of the MNL model is known as the "independence from irrelevant alternatives" (IIA)[42] IIA, caused by the iid TIEV assumption of the MNL error term, is the unintuitive substitutional patterns that predicts that households will respond to an increase in the unit-price of product j by substituting to the most popular products rather than substituting to products that have characteristics that are similar to product j. This is far less of a concern for my model's specification however, as the π_{ij} term, which includes H_i, constructs heterogeneous preferences specific to each household based upon their demographic characteristics. In turn, substitutional patterns within my MNL model are not only a function of the mean utility of the products within the choice set, but also a function of each household i's preferences, which are constructed and estimated by including the π_{ij} term. As a result, I find that it is both unnecessary and inefficient to implement an approach that would essentially constructs heterogeneous preferences among households via simulation[43] or via an a priori grouping of choices within the choice set,[44] as I am able to estimate these heterogeneous preferences using the household demographic information instead.[45]

[38]Note: I discard all shopping trips where households purchase fluid milk of sizes that are not gallon, half gallon are quart sized, but I do include them when calculating the measure of how long it has been since they last purchased fluid milk.

[39]Or "store" brand.

[40]I utilize a fixed effects approach to control for the product characteristics that are unobservable to the econometrician but observed by economic agents. In particular, I include a private label dummy as well as the private label dummy interacted with various channel type indicator variables. Using a fixed effects approach with these particular fixed effects is valid if ξ_j only varies at the private label vs. regional brand and channel type levels.

[41]Which includes: race, the number of days since the household last purchased fluid milk, a dummy variable that indicates the presence of children within the household, income, whether or not there is a male household head within the household, education, employment, whether or not the household heads are married, and dummy variable that indicates whether or not the household has ever participated in the WIC program.

[42]Which has lead many of the papers previous in the literature to utilize a either a random coefficients logit model, an ordered logit model or some other model specification of a similar nature.

[43]As would be the case in a random coefficients logit specification.

[44]As would be the case in an ordered logit model specification.

[45]Note: This approach assumes that household preference heterogeneity can be exactly modeled via household

7.2 Firm Behavior

On the firm's side of the model, I assume that firms take the menu of products sold by the firm as given.[46] Additionally, I assume that firms participate in price competition. This allows me to infer marginal costs without firm-level cost data if: (1) unit marginal costs are constant for each product j, (2) unit marginal costs are non-constant, but a function of container size over the product set J. Together these two assumptions imply that the cost of producing the first half gallon container of private label whole milk is the same as the cost of producing the second half gallon container of private label whole milk for a particular firm F. However, the cost of producing one gallon container of private label whole milk is not restricted to be equivalent to the cost of producing two half gallon containers of private label whole milk for the same firm F.

From here, I define marginal costs for product j as a function of the vector X^{Cost}:

$$\text{mc}_{jt} = X^{\text{Cost}}_{jt}\gamma + \omega_{jt} \tag{6}$$

where X^{Cost}_{jt} consists of the butterfat cost for each choice j, the skim cost for each choice j, the total number of firms within the three digit zip code and the Herfindahl-Hirschman Indenx (HHI) within the three digit zip code.

I can then use the demand system/parameter estimates that determine consumer choices to infer marginal costs. For simplicity, I'll refer the market shares by consumers as:

$$S(\alpha, \theta; p, X, H; \xi) \quad \text{or} \quad S(\Theta) \tag{7}$$

where θ are the utility parameters estimated in the demand system (excluding the disutility of price parameter α). p, X and H refer to the prices in the observed prices, observed product characteristics and household demographics respectively.

Before continuing, two key points must be made: (1) due to a lack of data on the menu of products that each firm offers, I use the simulated menu described section 4.2 and assume each firm offers all choices within the set J and (2) due to predicted and observed shares of many of the choices being very small[47] I aggregate the choice-set for costs to the size level. Assuming that unit marginal costs are constant over output, firms' profits are proportional to:

$$\sum_{h \in J} \left[p_h - \text{mc}_h\right] \times S_h(\Theta) \tag{8}$$

where $h \in J | h \neq j$. Then the first order condition for p_j where $j \in J$ is given by:

$$\left[\sum_{h \in J} (p_h - \text{mc}_h) \frac{\partial S_h(\Theta)}{\partial p_j}\right] + S_j(\Theta) = 0 \tag{9}$$

Which expressed in vector form is:

demographic characteristics. As such, if household preference heterogeneity varies within demographic characteristics, then this approach still lead to odd substitutional patterns. For example, perhaps households choose to consume organic products because they are "going green" rather than for some other reason which would be captured via demographic information. However, the within demographic household heterogeneity would have to be the main determinant for households for this approach to be invalid, which I find unlikely.

[46] Cohen (2008) states that if the decision whether or not to offer multiple sizes of products is treated as a choice variable in the model that this choice is likely correlated with the product unobservables ξ and would lead to biased parameter estimates.

[47] Less than 1% in several cases.

$$(-\Delta)\big[p - mc\big] + S = 0 \tag{10}$$

where $\Delta_{jr} = -\frac{\partial S_j}{\partial p_r}$ is the $J-\text{by}-J$ matrix of own-price and cross-price elasticities among the size-aggregated choices.

From here, I obtain marginal costs to be:

$$\text{mc} = p - \big[\Delta^{-1}S\big] \tag{11}$$

where $\big[\Delta^{-1}S\big]$ is the markup term: $M(\alpha, \theta; p, X, H; \xi)$ or, $M(\Theta)$.

The strategy I use for estimating the cost side of the model is to find the parameters that minimize the prediction errors (i.e. the distance between the vector of observed and predicted unit-prices) from the following:

$$p - M(\Sigma, \delta) = MC(\gamma; X^{\text{Cost}}; \omega) \tag{12}$$

8 Results

Demand and cost parameter estimates are provided in table 4 and table 6, respectively.

Note that the parameter estimates presented in table 4 cannot be directly interpreted as marginal effects, but finding the marginal effects are quite simple as doing so requires dividing these estimates by the absolute value of α, the price sensitivity parameter. These marginal effects are shown in table 5. After diving by the absolute value of the price sensitivity parameter, these marginal effects are measured in dollars per relevant unit. For example, the marginal effect of 0.88853 for the gallon sized container indicator variable shows that households value products sold in gallon sized container \$0.88853 more than the smaller sized goods, on average all else constant. This marginal effect is quite reasonable, as it shows that on average households prefer more fluid milk to less.

The marginal effect for the butterfat content $\geq 2\%$ of 1.20723 is consistent with the theory that households choose products based upon a basket of characteristics as it shows that households would prefer to consume more butterfat rather than less.

The marginal effects with the "UFE" label are fixed effects that control for product characteristics that are observed by economic agents within the model, but unobserved to the econometrician.[48] Considering them jointly, we see that households prefer to purchase fluid milk at grocery stores over all of channel types, which seems reasonable.

Marginal effects with the label "HH1 Char" are household demographic characteristics that have been interacted with the gallon sized container indicator variable, whereas "HH2 Char" marginal effects are the same household demographic characteristics interacted with the butterfat content $\geq 2\%$ indicator variable instead. In terms of the majority of these demographic characteristics, there is little to have strong feelings regarding their expected sign ex-ante, thus I will only highlight selected characteristics. One would anticipate that "Days Since Last Milk Purchase" would have a positive marginal effect for both interaction, as such would imply that a household's home stock of fluid milk diminishes, they would have stronger preferences for fluid milk, but I find negative marginal effects in both cases. However, both of these marginal effects are very small, -0.03332 and -0.01769, which may indicate that the home stock for a household is largely unimportant in terms of making their purchasing decisions.

[48]Note: The "UFE: (Channel)" fixed effects are all interacted with a private label indicator variable.

Another interesting finding is for the marginal effects regarding if the household heads are married. I find that married household heads prefer gallon sized containers as opposed to smaller sized containers, the coefficient being 0.21105, but prefer less butterfat in said products rather than more, which has a coefficient of -0.35718.

Table 6, which provides the cost parameter estimates, are all interpreted as a respective one unit change on the change in marginal cost in dollars. For example, the 0.91890 parameter estimate for butterfat cost implies that as the butterfat in the product increases by one pound, the marginal cost for that product increases by $0.91890, on average all else constant. The magnitude may seem large at first, but when considering that gallon sized whole milk products only have 0.2795 lbs of butterfat[49] it seems quite reasonable. My model also finds that both the skim cost and the total number of firms within the 3-digit zip code parameters are not statistically different than zero, which is surprising in regards to both parameters ex-ante. The HHI within the 3-digit zip code parameter has the logical sign, as the more competitive the area of competition, the higher the probability that the firms participate in price wars. The grocery channel has the smallest parameter estimate, but is still positive, which implies that firms within the grocery channel have lower marginal costs than those listed, but marginal costs that are higher than the omitted case, which were the "Specialty Stores" channel.

8.1 Measure of Price Discrimination

Following Cohen (2008), I compute unit-markups and unit-costs for each size for each trip. The unit-price discount is equal to the difference in marginal costs between the two size groupings is ($\text{MC}_{smaller\ size} - \text{MC}_{larger\ size}$) plus the difference in markups ($\text{Markup}_{smaller\ size} - \text{Markup}_{larger\ size}$), all of which are measured in quantities per-unit. By dividing the difference in unit-markups, by the difference in unit-prices (outlined previously), I obtain a measure of the extent to which price discrimination is determinant in unit-price differentials across sizes. In doing so, I find that on average, approximately 70.4% of the markup differential between gallon sized containers and half gallon/quart sized containers is attributed to price discrimination. This differential is much larger than what has been shown in the literature thus far, but seem reasonable given the dynamics of the fluid milk industry in comparison to other that have been studied.[50]

8.2 Welfare Counterfactual

One of the motivating reasons behind structural estimation techniques are the ability to perform counterfactual experiments once the parameters of the model have been estimated, this section of the paper includes these counterfactuals. Since there has yet to be a consensus within either the empirical or theoretical literature on whether or not price discrimination is welfare increasing from the consumer's perspective[51] I test to see the effect of price discrimination from the consumer's perspective via counterfactual analysis. As a brief highlight of the current literature, Leslie (2004) and Cohen (2008) found that price discrimination within their market of focus to be welfare increasing from the consumer's perspective, whereas McManus (2007) and Miller and Osborne (2014) found it to be welfare decreasing.

[49] With whole milk being the product that uses the most butterfat in mind.

[50] As a point of reference, Cohen (2008) finds a range of 34 to 46% of the markup variation can be explained by price discrimination.

[51] Note: It must be profit increasing, as otherwise firms would not utilize the practice.

To test the welfare implications of this pricing strategy from the consumer's perspective, consider the following thought experiment: suppose there is a world where firms who participate within the fluid milk market are no longer allowed to utilize nonlinear pricing schedules, or that firms within this market must charge the same price per ounce for every size container they sell.[52] One approach to finding the overall affect of welfare would be to compare the current world to the fictional world outlined above in terms of total welfare. I attempt to do so by following Ben Akiva (1972), McFadden (1973) and Domencich and McFadden (1975)'s "log sum formula" for finding consumer surplus:

$$\text{Emax } V_i = \frac{1}{\alpha}\left[\log\bigg(\sum_{j \in J} \exp\big(V_j^1\big)\bigg) - \log\bigg(\sum_{j \in J}\big(V_j^0\big)\bigg)\right] \tag{13}$$

where the superscript "1" denotes the fictional world with linear price schedules and the superscript "0" denotes the nonlinear price schedule.

Without solving for the optimal price schedule, I need to make assumptions on which linear price the firm sets in the fictional world. Since it is not immediately clear, table 8 presents welfare calculations using the linear price per ounce of the listed size for all options given the parameter estimates presented in appendix C. The price schedules I assume for firms are using the average price per ounce for gallon and half gallon/quart containers. Additionally, I also present welfare calculations under the average price per ounce of all sizes, which I find to be the most compelling of the three price menus.

The units for the welfare change are in dollars per shopping trip per ounce, which implies that under the Gallon price regime, each household is on average \$0.0785 better off per shopping trip than under current market conditions. Whereas the households are on average \$0.1603 worse off per shopping trip under a half gallon/quart pricing regime. Under the average for all sizes, I find that households are on average \$0.0793 worse off per shopping trip. These estimates imply that households are on average better off with linear pricing in the case of firms choosing the linear price per ounce to be the average gallon, but worse off if the firms chooses to set linear prices equal to the half gallon/quart price per ounce and the average price per ounce of all sizes. Thus, in the case of the gallon regime in comparison to the current world, nonlinear pricing is welfare decreasing, whereas under a half gallon/quart and average of all sizes linear price regime nonlinear pricing is welfare increasing.

This methodology is clearly limited, as firms have not been given a chance to respond and reoptimize in terms of setting their linear price. Additionally, firms within the fluid milk industry may find that it is no longer profitable to sell the smaller sized containers, which is ignored by this methodology. I find the price computations with the average price per ounce of all sizes to be most compelling of the three measures as it would represent a world that firms would seem mostly likely to choose if not given the chance to reoptimize. Thus, I find that nonlinear pricing is welfare decreasing from the consumer's perspective.

With all that being said, I find these calculations to be consistent with a story that nonlinear price schedules in this market are welfare decreasing, however an approach where firms are allowed to reoptimize their prices and the menu of products would provide more compelling evidence.

[52]Perhaps the government disallows nonlinear pricing in this world.

9 Conclusion

This paper attempts to find evidence of second degree price discrimination in the oligopoly market of fluid milk. Using consumer level data provided by the Nielsen Company, I find that the majority of the markup differential in fluid milk across sizes, approximately 70.4%, can be explained by price discrimination.

In addition, through counterfactual analysis, I find that price discrimination is welfare decreasing from the consumer's perspective. Consumers who purchase the smaller sizes, quart and half gallon containers, yield some amount of utility from more convenient sized packages, whereas consumers who purchase gallon sized containers enjoy lower prices as a result of higher competition in this segment of the market. These findings are in line with what the previous literature has found and are possibly more compelling than what previous papers have done given the richness of the data that I utilize.

References

[Aryal (2014)] Aryal, Gaurab. "An empirical analysis of competitive nonlinear pricing." Available at SSRN 2262664 (2014).

[Berry (1994)] Berry, Steven T. "Estimating discrete-choice models of product differentiation." The RAND Journal of Economics (1994): 242-262.

[Buchanen (1953)] Buchanen, J. M. "The Theory of Monopolistic Quantity Discounts." Review of Economic Studies 20.3 (1953): 199-208.

[Chao (2013)] Chao, Yong. "Strategic Effects Of Three Part Tariffs Under Oligopoly." International Economic Review 54.3 (2013): 977-1015.

[Cohen (2008)] Cohen, Andrew. "Package size and price discrimination in the paper towel market." International Journal of Industrial Organization 26.2 (2008): 502-516.

[DeSalvo and Mobinul (2002)] DeSalvo, Joseph S., and Mobinul Huq. "Introducing nonlinear pricing into consumer choice theory." The Journal of Economic Education 33.2 (2002): 166-179.

[Dolan (1987)] Dolan, Robert J. "Quantity discounts: Managerial issues and research opportunities." Marketing Science 6.1 (1987): 1-22.

[Dupuit (1849)] Dupuit, Jules. "On tolls and transport charges." Annales des ponts et chausses. Vol. 11. No. 1962. 1849.

[Escobari and Jindapon (2014)] Escobari, Diego, and Paan Jindapon. "Price discrimination through refund contracts in airlines." International Journal of Industrial Organization 34 (2014): 1-8.

[Friebel et. al (2015)] Friebel, Kerstin, Oleksandr Perekhozhuk, and Thomas Glauben. "Price Discrimination in Russian Wheat Exports: Evidence from Firm-level Data." Journal of Agricultural Economics (2015).

[Hamilton and Thisse (1996)] Hamilton, Jonathan H., and Jacques-Franois Thisse. "Nonlinear pricing in spatial oligopoly." Economic design 2.1 (1996): 379-397.

[Leslie (2004)] Leslie, Phillip. "Price discrimination in Broadway theater." RAND Journal of Economics (2004): 520-541.

[Lewis and Sappington (1994)] Lewis, Tracy R., and David EM Sappington. "Supplying information to facilitate price discrimination." International Economic Review (1994): 309-327.

[Lui and Shen (2012)] Liu, Yizao and Shu Shen. "Price Discrimination with Asymmetric Firms: The Case of the U.S. Carbonated Soft Drinks Market." Working Paper (2012)

[Manchester and Blayney (2001)] Manchester, Alden C., and Donald P. Blayney. Milk pricing in the United States. No. 33612. United States Department of Agriculture, Economic Research Service, 2001.

[Maskin and Riley (1984)] Maskin, Eric, and John Riley. "Monopoly with incomplete information." The RAND Journal of Economics 15.2 (1984): 171-196.

[McManus (2007)] McManus, Brian. "Nonlinear pricing in an oligopoly market: The case of specialty coffee." The RAND Journal of Economics 38.2 (2007): 512-532.

[Miller and Osbourne (2014)] Miller, Nathan H., and Matthew Osborne. "Spatial differentiation and price discrimination in the cement industry: evidence from a structural model." The RAND Journal of Economics 45.2 (2014): 221-247.

[Miravete and Rller (2003)] Miravete, Eugenio J., and Lars-Hendrik Rller. "Competitive non-linear pricing in duopoly equilibrium: the early US cellular telephone industry." (2003).

[Miravete (2007)] Miravete, E. "Are all those Calling Plans Really Necessary." The Limited Gains From Complex Tariffs (2004).

[Miravete (1996)] Miravete, Eugenio J. "Screening consumers through alternative pricing mechanisms." Journal of Regulatory Economics 9.2 (1996): 111-132.

[Moffit (1990)] Moffitt, Robert. "The econometrics of kinked budget constraints." The Journal of Economic Perspectives 4.2 (1990): 119-139.

[Mussa and Rosen (1978)] Mussa, Michael, and Sherwin Rosen. "Monopoly and product quality." Journal of Economic theory 18.2 (1978): 301-317.

[Nevo (2000)] Nevo, Aviv. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." Journal of Economics & Management Strategy 9.4 (2000): 513-548.

[Pigou (1920)] Pigou, Arthur. "C, 1920." The economics of welfare (1920).

[Reiss and White (2006)] Reiss, Peter C., and Matthew W. White. Evaluating welfare with nonlinear prices. No. w12370. National Bureau of Economic Research, 2006.

[Robinson (1933)] Robinson, Joan. "The economics of imperfect competition." (1933): 519-521.

[Sharkey and Sibley (1993)] Sharkey, William W., and David S. Sibley. "Optimal non-linear pricing with regulatory preference over customer type." Journal of Public Economics 50.2 (1993): 197-229.

[Kawaguchi et. al (1997)] Kawaguchi, Tsunemasa, Nobuhiro Suzuki, and Harry M. Kaiser. "A spatial equilibrium model for imperfectly competitive milk markets." American Journal of Agricultural Economics 79.3 (1997): 851-859.

[USDAPR] United States Department of Agriculture, National Agricultural Statistics Service. http://www.nass.usda.gov/

Appendices

A Figures from Section 5

Figure 1: Price Schedule for Good with Discrete Quantity Discount

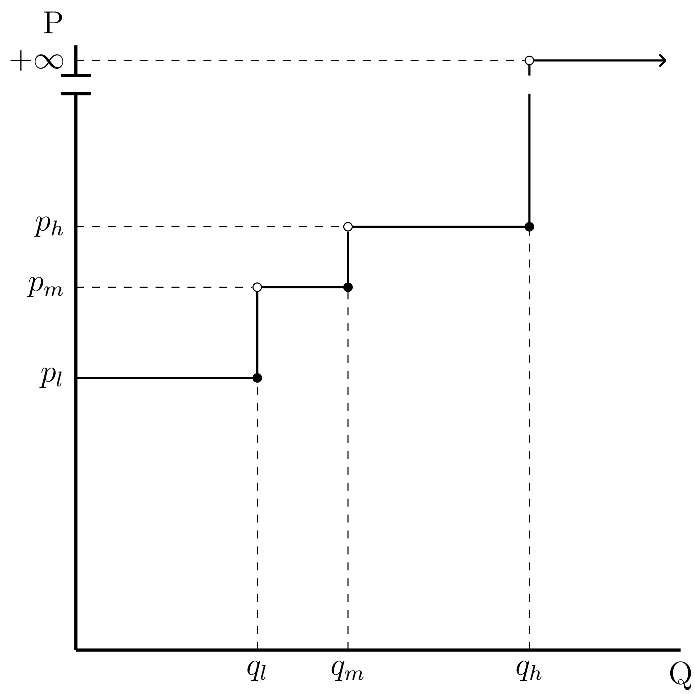
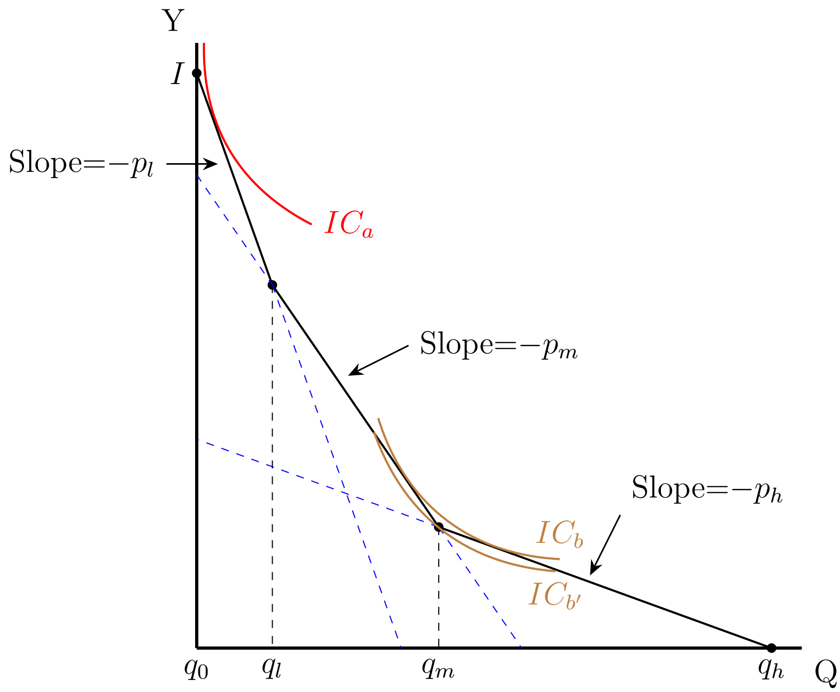


Figure 2: Linearized Budget Constraint Facing a Discrete Quantity Discount



B Tables and Figures for Section 6

Table 1: Unit Price Summary Statistics by Size and Coupon

	No Coupon			Coupon		
	Quart	Half Gallon	Gallon	Quart	Half Gallon	Gallon
Mean Price	1.387	1.704	2.375	0.896	1.056	1.502
St Dev	0.354	0.688	0.509	0.444	0.775	0.784
Observations	417,805	161,650	16,439	22,235	8,780	492

Note: An observation is a UPC purchased by a panelist in the years 2008-2009. Prices are in December 2009 dollars. Unit price is defined to be the final price paid by the panelist, minus any coupons, for one unit (which in most cases is a single container) of the UPC in question.

Table 2: Price Statistics by Size and Coupon

	No Coupon		Coupon	
	Mean	St Dev	Mean	St Dev
Quart				
Raw Milk Price/oz	0.009	0.001	0.009	0.001
Price/oz	0.043	0.011	0.028	0.014
Price-Cost Ratio	5.155	1.530	3.240	1.685
Price-Cost Differential	0.03480	0.01123	0.01927	0.01392
Half Gallon				
Raw Milk Price/oz	0.009	0.001	0.008	0.001
Price/oz	0.027	0.011	0.017	0.012
Price-Cost Ratio	3.175	1.378	1.991	1.518
Price-Cost Differential	0.01813	0.01082	0.00808	0.01217
Gallon				
Raw Milk Price/oz	0.008	0.001	0.008	0.001
Price/oz	0.019	0.004	0.012	0.006
Price-Cost Ratio	2.206	0.531	1.401	0.747
Price-Cost Differential	0.01006	0.00403	0.00331	0.00609
Observations				
Quart	16,439		492	
Half Gallon	161,650		8,780	
Gallon	417,805		22,235	

Note: An observation is a UPC purchased by a panelist in the years 2008-2009. Prices are in December 2009 dollars. Raw Milk Price/oz is the raw milk price in ounces at the time of purchase. Price/oz is the unit price per ounce at the time of purchase. Price-Cost Ratio is computed by dividing the unit price per ounce by the raw milk price per ounce. Price-Cost Differential is computed by subtracting the raw milk price per ounce from the unit price per ounce.

Figure 3: Graphical Evidence of Price Discrimination

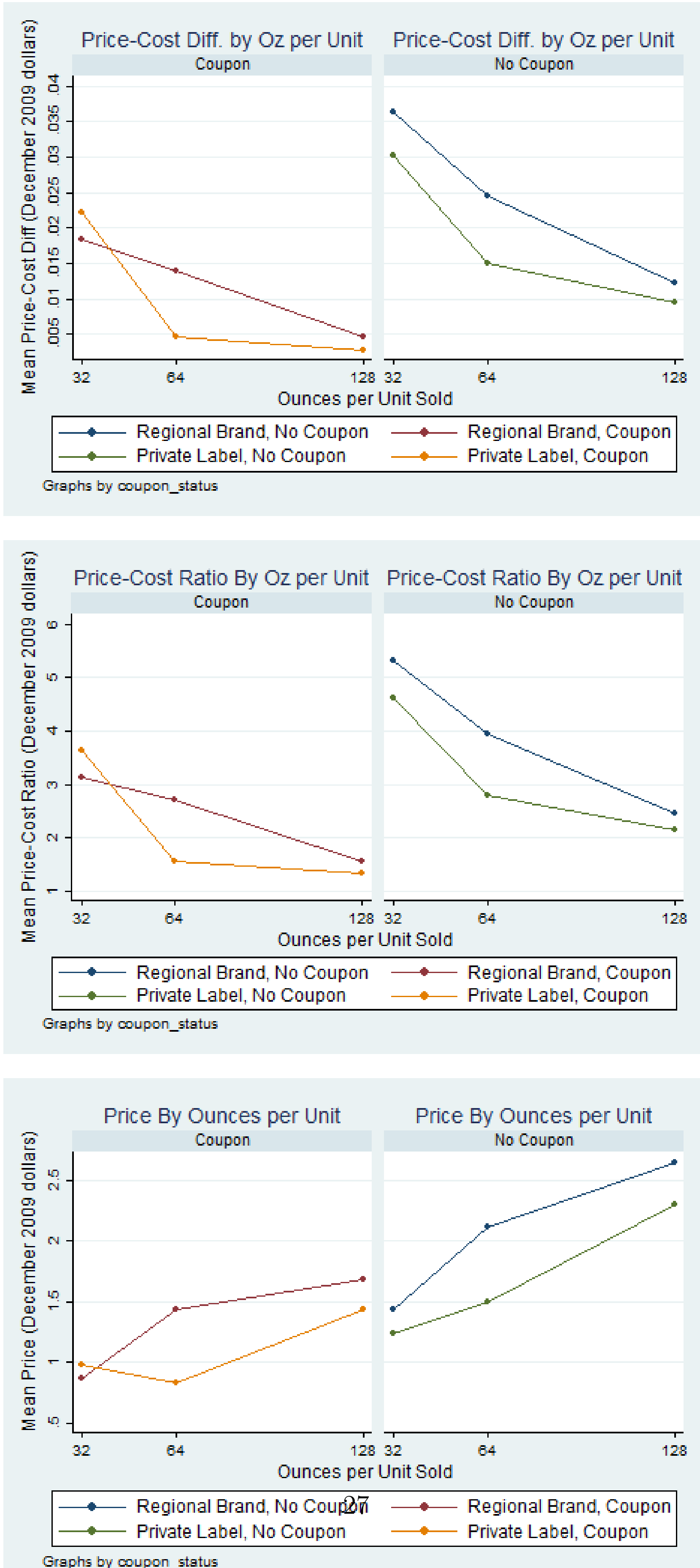


Table 3: Reduced Form Regression Results: Size Dummies

	(1)	(2)	(3)	(4)
Gallon Sized Container	−0.024677***	−0.021815***	−0.021376***	−0.020873***
	(0.00039)	(0.00058)	(0.00019)	(0.00019)
Half Gallon Sized Container	−0.016787***	−0.015495***	−0.015701***	−0.015703***
	(0.00040)	(0.00029)	(0.00026)	(0.00025)
Raw Milk cost/oz	0.057766**	0.020579	0.053478***	0.036165**
	(0.02945)	(0.02467)	(0.01746)	(0.01811)
Private Label Dummy		−0.004687***		−0.002624
		(0.00088)		(0.00178)
Organic Dummy		0.021756***		0.020875***
		(0.00050)		(0.00033)
Flavor Dummy		0.000491		0.000475
		(0.00033)		(0.00037)
Coupon Value		−0.005929***		−0.006116***
		(0.00012)		(0.00013)
Controls	N	Y	N	Y
Brand Fixed Effects	N	N	Y	Y
Obs.	627,401	627,401	627,401	627,401
Brands	.	.	195	195

Note: There are 627,401 sales of fluid milk from a cow sized using this paper's definition of the Imperial measurement system. Controls include channel type indicator variables as well as state fixed effects. Standard errors are clusted at the store level for all regressions. Standard errors are shown in parentheses *** p<0.01, ** p<0.05, * p<0.1.

C Section 8 Tables and Figures

Table 4: Demand Parameter Estimates

	Point Estimate	Std Err
Gallon Sized Container	1.27951	0.04617
Butterfat Content $\geq 2\%$	1.47925	0.05063
α (Price)	-1.44003	0.00246
UFE: (Private Label)	0.02917	0.05894
UFE: (Grocery)	0.43801	0.06051
UFE: (Discount Store)	-0.45266	0.06237
UFE: (Warehouse Club)	-2.02280	0.07320
UFE: (Drug Store)	-2.25680	0.06333
UFE: (GS$C)	-3.24296	0.06048
HH1 Char: White	0.26506	0.03164
HH1 Char: Days Since Last Milk Purchase	-0.04798	0.00033
HH1 Char: At least 1 Child lives in the Household	-0.07930	0.01696
HH1 Char: HH Income above U.S. Median	0.29868	0.02167
HH1 Char: No Male HH Head	0.24836	0.03024
HH1 Char: At least 1 HH Head is Employed at All	0.34182	0.02009
HH1 Char: At least 1 HH Head Ever Attended College	0.06313	0.02268
HH1 Char: The HH Heads are Married	0.30392	0.02476
HH1 Char: HH has Ever on WIC	-0.06662	0.03095
HH2 Char: White	-0.29438	0.03165
HH2 Char: Days Since Last Milk Purchase	-0.02547	0.00033
HH2 Char: At least 1 Child lives in the Household	-0.08426	0.01617
HH2 Char: HH Income above U.S. Median	-0.06030	0.02128
HH2 Char: No Male HH Head	-0.53619	0.03193
HH2 Char: At least 1 HH Head is Employed at All	-0.00626	0.02259
HH2 Char: At least 1 HH Head Ever Attended College	-0.40868	0.02245
HH2 Char: The HH Heads are Married	-0.51435	0.02468
HH2 Char: HH has Ever on WIC	0.35079	0.03272
Observations	134910	

These estimates were obtained from a 5% random sample of the full dataset. Standard errors computed via bootstrap with 100 repetitions. Partly due to how large the sample is, all coefficent estimates are statistically significant at the 99% level of significance. All "HH1 Char:" variables are interacted with a gallon container size dummy variable and "HH2 Char:" variables are interacted with a dummy variable indicating that buttterfat content $\geq 2\%$. UFE represents fixed effects that are included to control for product characteristics that are observed by economic agents within the model but not by the econometrician. All UFE ("Channel") variables are interacted with a private label brand dummy variable. "GS$C" represents the channel that includes gas/service stations and dollar/convenience stores. Stores that could reasonably sell fluid milk, but do not fit into any of the channels listed explicitly are the omitted case for the UFE ("Channel") fixed effects.

Table 5: Demand: Marginal Effects

	Marginal Effect
Gallon Sized Container	0.88853
Butterfat Content $\geq 2\%$	1.02723
UFE: (Private Label)	0.02026
UFE: (Grocery)	0.30417
UFE: (Discount Store)	−0.31434
UFE: (Warehouse Club)	−1.40470
UFE: (Drug Store)	−1.56719
UFE: (GS$C)	−2.25201
HH1 Char: White	0.18407
HH1 Char: Days Since Last Milk Purchase	−0.03332
HH1 Char: At least 1 Child lives in the Household	−0.05507
HH1 Char: HH Income above U.S. Median	0.20741
HH1 Char: No Male HH Head	0.17247
HH1 Char: At least 1 HH Head is Employed at All	0.23737
HH1 Char: At least 1 HH Head Ever Attended College	0.04384
HH1 Char: The HH Heads are Married	0.21105
HH1 Char: HH has Ever on WIC	−0.04626
HH2 Char: White	−0.20443
HH2 Char: Days Since Last Milk Purchase	−0.01769
HH2 Char: At least 1 Child lives in the Household	−0.05852
HH2 Char: HH Income above U.S. Median	−0.04187
HH2 Char: No Male HH Head	−0.37234
HH2 Char: At least 1 HH Head is Employed at All	−0.00434
HH2 Char: At least 1 HH Head Ever Attended College	−0.28380
HH2 Char: The HH Heads are Married	−0.35718
HH2 Char: HH has Ever on WIC	0.24360

Note: Marginal effects obtained by dividing each parameter estimate by the absolute value of the price parameter α.

Table 6: Firm Side Model Parameter Estimates

	Parameter Estimate	Std Err
Butterfat Cost	0.91890	0.01536
Skim Cost	−0.00480	0.00864
Total Firms in 3-Digit Zip	−0.00010	0.00017
HHI in 3-Digit Zip	0.11849	0.00156
Grocery	0.45622	0.00586
Warehouse Store	0.87253	0.01129
Discount Club	0.54319	0.00666
Drug Store	0.47896	0.00802
GS$C	0.83623	0.00740
Observations	134910	

Table 7: Welfare Counterfactual

Linear Price Set	Welfare Change
Gallon Price per Ounce	$0.0785
Half Gallon/Quart Price per Ounce	−$0.1603
Average Price per Ounce of All Sizes	−$0.0793

D Household Demographics

This section of the appendix includes summary statistics for households in the Nielsen Panel dataset in years 2008-2009.

Table 8: Panelist Years

	Freq.	Percent
2008	11,108	50.14
2009	11,046	49.86

Table 9: Household Income

	Freq.	Percent
Under $5000	161	0.73
$5000-$7999	214	0.97
$8000-$9999	194	0.88
$10,000-$11,999	284	1.28
$12,000-$14,999	568	2.56
$15,000-$19,999	870	3.93
$20,000-$24,999	1,304	5.89
$25,000-$29,999	1,307	5.90
$30,000-$34,999	1,609	7.26
$35,000-$39,999	1,451	6.55
$40,000-$44,999	1,320	5.96
$45,000-$49,999	1,437	6.49
$50,000-$59,999	2,569	11.60
$60,000-$69,999	2,016	9.10
$70,000-$99,999	4,236	19.12
$100,000-$124,999	1,672	7.55
$125,000-$149,999	399	1.80
$150,000-$199,999	345	1.56
$200,000+	198	0.89

Note: An observation is a panelist year.

Table 10: Household Size

	Freq.	Percent
1 member	5,243	23.67
2 members	9,080	40.99
3 members	3,249	14.67
4 members	2,773	12.52
5 members	1,188	5.36
6 members	415	1.87
7 members	136	0.61
8 members	46	0.21
9 members	24	0.11

Note: An observation is a panelist year.

Table 11: Type of Residence

	Freq.	Percent
One Family House	18,073	81.58
One Family House (Condo/Coop)	285	1.29
Two Family House	647	2.92
Two Family House (Condo/Coop)	76	0.34
Three+ Family House	1,538	6.94
Three+ Family House (Condo/Coop)	848	3.83
Mobile Home or Trailer	687	3.10

Note: An observation is a panelist year.

Table 12: Composition of Household

	Freq.	Percent
Married	14,030	63.33
Female Head OR	1,496	6.75
Male Head OR	636	2.87
Female Alone	3,740	16.88
Female NR	112	0.51
Male Alone	1,503	6.78
Male NR	637	2.88

Note: An observation is a panelist year.

Table 13: Age and Presence of Children

	Freq.	Percent
Under 6 only	780	3.52
6-12 only	1,327	5.99
13-17 only	1,638	7.39
Under 6 & 6-12	727	3.28
Under 6 & 13-17	123	0.56
6-12 & 13-17	899	4.06
Under 6 & 6-12 & 13-17	148	0.67
Under 6 & 6-12 & 13-17	16,512	74.53

Note: An observation is a panelist year.

Table 14: Household Head Education

	Male		Female	
	Freq.	Percent	Freq.	Percent
No Male/Female Head	5,348	24.14	1,863	8.41
Grade School	136	0.61	64	0.29
Some High School	730	3.30	482	2.18
Graduated High School	5,026	22.69	6,011	27.13
Some College	4,692	21.18	6,238	28.16
Graduated College	4,348	19.63	5,562	25.11
Post College Grad	1,874	8.46	1,934	8.73

Note: An observation is a panelist year.

Table 15: Household Head Employment

	Male		Female	
	Freq.	Percent	Freq.	Percent
No Male/Female Head	5,348	24.14	1,863	8.41
Under 30 hours	845	3.81	2,769	12.50
30-34 hours	456	2.06	1,180	5.33
35+ hours	10,810	48.79	8,339	37.64
Not Employed for Pay	4,695	21.19	8,003	36.12

Note: An observation is a panelist year.

Table 16: Household State of Residence

	Freq.	Percent
Illinois	5,392	24.34
Indiana	2,903	13.10
Michigan	4,499	20.31
Ohio	6,470	29.20
Wisconsin	2,890	13.05

Note: An observation is a panelist year.

Table 17: Household Race

	Freq.	Percent
White/Caucasian	19,560	88.29
Black/African American	1,771	7.99
Asian	288	1.30
Other	535	2.41

Note: An observation is a panelist year.

Table 18: Household Hispanic Origin

	Mean	Median	St. Dev	Min	Max
Hispanic Origin	0.0232	0	0.1505	0	1
Observations	22,154	22,154	22,154	22,154	22,154

Note: An observation a panelist year.

Table 19: Marital Status of Household

	Freq.	Percent
Married	14,102	63.65
Widowed	1,697	7.66
Divorced/Separated	3,215	14.51
Single	3,140	14.17

Note: An observation is a panelist year.

Table 20: Age of Household Head

	Male Head Age	Female Head Age
Mean	54.45	53.88
Median	53.97	53.30
St Dev	12.87	13.07
Min	19.16	20.16
Max	109.50	109.50
Observations	16,806	20,291

Note: An observation is a panelist year.

Table 21: WIC Summary Statistics

	Report Currently On	Reported Ever On	Eligible
Mean	0.0093	0.0798	0.0046
Median	0	0	0
St Dev	0.0958	0.2709	0.0680
Min	0	0	0
Max	1	1	1
Observations	22,154	22,154	22,154

Note: An observation is a panelist year.

E Price Summary Statistics

Table 22: Unit Price Summary Statistics by Size and Coupon

	No Coupon			Coupon		
	Gallon	Half Gallon	Quart	Gallon	Half Gallon	Quart
Mean Price	2.369	1.814	1.420	1.531	1.235	1.010
St Dev	0.471	0.762	0.426	0.803	0.873	0.616
Observations	428,077	182,552	19,286	22,570	10,902	558

Note: An observation is a UPC purchased by a panelist in the years 2008-2009. Prices are in December 2009 dollars.

Table 23: Price Statistics by Size and Coupon

	No Coupon		Coupon	
	Mean	St Dev	Mean	St Dev
Gallon				
Price/oz	0.019	0.004	0.012	0.006
Price-Cost Ratio	2.284	0.531	1.490	0.804
Price-Cost Differential	0.010	0.004	0.004	0.006
Raw Milk Price/oz	0.00821	0.00098	0.00809	0.00099
Half Gallon				
Price/oz	0.027	0.011	0.017	0.012
Price-Cost Ratio	3.292	1.446	2.086	1.610
Price-Cost Differential	0.018	0.011	0.008	0.012
Raw Milk Price/oz	0.00824	0.00096	0.00813	0.00097
Quart				
Price/oz	0.043	0.010	0.028	0.014
Price-Cost Ratio	5.306	1.547	3.320	1.761
Price-Cost Differential	0.035	0.011	0.019	0.014
Raw Milk Price/oz	0.00827	0.00096	0.00847	0.00099

Note: An observation is a UPC purchased by a panelist in the years 2008-2009. Prices are in December 2009 dollars.

F Product Menu Simulation Procedure

Here I present how I have simulated the choice set that panelists choose from at stores. Simulation is necessary as the panel and scanner data only match for 29.5% of the 5% random sample dataset on retailer code, channel, 3-digit zip and week. I have done this in two different ways, one using only the scanner data to create the choice set and the other using both the scanner and panel data to create the choice set. I will discuss each separately below, but both are of the same flavor.

F.1 Choice Set Simulation - Scanner Data Only Procedure

1. I start by first taking the weekly sales data provided in the scanner data and creating mean product characteristics for each choice in the choice set on the retailer code-channel-storezip3-week level.
2. I next do the same, but aggregate over a various larger measures of time: month, quarter, half-year, year and the entire sample period.
3. After aggregating the choice set over each time period, drop retailer code and aggregate by channel for each 3-digit zip code over the entire sample period.
4. Next I drop the channel and aggregate over an increased geographical area for each choice, starting with 3-digit zip, the county, state and the entire sample.
5. Once these datasets that include the choice sets at different temporal, geographical and industry levels have been produced, I merge and update the data as described in the following table:

(R=Retailer Code, C=Channel, Z=3-digit Zipcode, W=Week, M=Month, Q=Quarter, H=Half, Y=Year, A=All, Co=County, ST=State)

Dataset	% Unmatched	% Updated	% Total Matched
5% RS Panel Dataset	100%	—	—
R-C-Z-W	70.5%	29.5%	29.5%
R-C-Z-M	70.48%	.02%	29.52%
R-C-Z-Q	70.46%	.02%	29.54%
R-C-Z-H	70.45%	.01%	29.55%
R-C-Z-Y	70.2%	.25%	29.8%
R-C-Z-A	69.16%	1.04%	30.84%
C-Z-A	28.62%	40.54%	71.38%
Z-A	5.07%	23.55%	94.93%
Co-A	1.80%	3.27%	98.2%
ST-A	0%	1.80%	100%
A	0%	0%	100%

F.2 Choice Set Simulation - Both Scanner and Panel Datasets Procedure

1. For this method, I follow the exact same technique, but using both the scanner than observed panel sales. Below is the same table, but using both datasets. Note, the table is somewhat different, as since I am using both datasets, there will always be a match, I will denote how the data is updated.

(R=Retailer Code, C=Channel, Z=3-digit Zipcode, W=Week, M=Month, Q=Quarter, H=Half, Y=Year, A=All, Co=County, ST=State)

Dataset	% Updated
5% RS Panel Dataset	—
R-C-Z-W	100%
R-C-Z-M	15.23%
R-C-Z-Q	5.93%
R-C-Z-H	3.77%
R-C-Z-Y	3.39%
R-C-Z-A	3.42%
C-Z-A	7.41%
Z-A	1.12%
Co-A	0.13%
ST-A	0%
A	0%