Price and Variety in Supermarkets: Can Store Competition Hurt Consumers?

Andre Trindade*

November 22, 2010

Abstract

Looking at a large number of markets, I find that prices and variety are higher when there are two competing supermarkets than in those with a single store. This pattern persists after controlling for differences across markets in a variety of ways and when looking into markets where the number of competitors changes over time.

I present a model that explains these patterns. Stores choose prices and the number of products to carry and consumers decide which store to go to and what to buy. In the model: 1) Incentives to increase variety are higher for duopoly supermarkets because of the business stealing effect; 2) As more products become available, the potential surplus for each consumer increases, which allows stores to raise prices and still induce a purchase. These two forces combined result in equilibrium predictions consistent with the patterns in the data.

In order to answer whether consumers are better-off in duopoly, when prices and variety are higher, I estimate consumer preferences. I find that consumer welfare is higher under competition. However, that is a result from the wider choice of products rather than lower prices.

*Department of Economics, Northwestern University. mailto: andretrindade@northwestern.edu
I would like to thank my committee Aviv Nevo, Igal Hendel and Eric Anderson for support and guidance during this project. I would also like to thank Guy Arie, Agustin Casas, Rita Costa, Pierre Dubois, Rachel Griffith, Kei Kawai, David Matsa, Ketan Patel, Patrick Rey, Mariano Tappata and Mauricio Varela for helpful comments and suggestions. This research was funded by a cooperative agreement between the USDA/ERS and Northwestern University, but the views expressed herein are those of the author and do not necessarily reflect the views of the U.S. Department of Agriculture. Financial support from Fundação para a Ciência e Tecnologia is also gratefully acknowledged.
For the latest version of this paper go to http://www.depot.northwestern.edu/~agt129/paper.pdf

I Introduction

In many industries, reducing prices is just one of many ways in which a firm can attract consumers. Increasing the quality of the product or improving the shopping experience are examples of decisions that can also affect consumers' choices. In the retail sector, in particular, the number of products offered in a store (henceforth "variety") is one of the most important factors that consumers consider when deciding where to shop (Bureau of Labor Statistics (2010)). Despite this fact, most of the empirical literature on competition has looked at prices as the single strategic variable that firms choose.

In this paper, I study the impact of competition on a firm's decision of products and prices. I focus on grocery stores, an industry where variety is particularly relevant. The reason for this is that consumers shop frequently which makes travel costs an important part of the grocery decision and induces consumers to concentrate their purchases in only one store (Information Resources, Inc. (2002)). As such, the introduction of a new product by one supermarket has the potential to capture the total grocery expenditure of the households that value that new product highly.

I study the behavior of supermarkets in over 700 markets across 43 different product categories and over four years. I center my analysis on small markets where there is either a single store or two competing ones.

After controlling for city differences in a variety of ways, I find that supermarkets in duopoly sell more products at higher prices than monopolist stores. I first control for demographic differences between markets in a set of linear regressions. I include the dependent variables in levels and in logs and the results are similar. Furthermore, I look within store groups by introducing chain fixed effects and still conclude that supermarkets with a nearby competitor offer more products at higher prices. The conclusion is unchanged if I include store size as a covariate.

I then control in a more flexible way for observable market differences through propensity score matching. First, I use market characteristics to measure the propensity of each of them to be a monopolist or a duopoly. Then, I compare stores with a similar propensity score and look at systematic differences between the two groups. I allow for polynomial functions of the observables to enter the propensity function, which permits those variables to affect the market structure in a variety of non-linear ways.

Finally, I control for unobserved differences across markets. I do this in two ways. First, I use a sample of markets where there is a change in structure during the period covered in the data. Second, I separate the competitive effect from the impact of other market characteristics by looking at duopoly markets where both stores belong to the same chain. The results with both strategies are consistent with competition increasing product variety and prices.

I provide a theoretical explanation for these facts, by setting up a model where stores compete for consumers by offering more variety and lower prices. I assume that consumers are heterogeneous in their brand preferences and are also imperfectly informed about prices before entering a store, which is likely in supermarkets where prices change every week. Competition has two effects. First, the traditional direct effect where stores fight for consumers by offering more variety and lower prices. In doing that, each supermarket takes into account the additional fixed costs from carrying more products. Then, a second, indirect effect: as a wider range of products becomes available, each consumer will find their preferred product available with higher probability and the expected utility from grocery shopping increases. Stores are then able to increase prices. This upward effect on prices can dominate the traditional direct effect of competition.

In order to answer whether consumers are better or worse off, I construct and estimate a consumer choice model. The utility from each store is a function of fixed storage attributes such as distance, parking and so on, and the utility from the goods that the consumer will likely buy. At the time of choice, the consumer does not know the prices that he will face at the store and so, uses the expected value of each product category to make the decision. The expected category inclusive value can be computed exactly from my choice model.

I estimate the model by combining the store level data with information on the purchases made by a sample of households. For now, the assumptions in the model allow me to estimate recursively the category and store choice decisions. I use maximum-likelihood methods to obtain the parameters of interest in both levels of the consumer decision.

With the results from the model, I look at what would be the choices of duopoly city consumers if they were buying at the price and variety level of monopolist stores. I restrict consumers to visit the same number of stores with the same frequency in order not to increase the welfare in a duopoly artificially.

The average consumer is found to benefit from increased competition. However, the increase in consumer welfare does not come from the expected mechanism - lower prices - but rather from a wider selection of products.

One important caveat of the paper is that I am only looking into the impact of moving from a one-store city to two stores. It is not clear that the effect would be similar if a new store is introduced in a market which is already multi-store. Furthermore, I am only looking into a particular segment (beverages) of a specific industry (conventional grocery store). I expect results in other product categories to be similar but I cannot measure it with my current dataset.

This paper presents evidence that supermarkets compete mainly by increasing the set of products that they sell. That is an important fact for a better understanding of retail competition. Moreover, I present evidence and a theory to support it that show that competition can have the opposite effect on prices from the one expected. As far as I know, this is the first paper to describe empirically this effect. All the results have important implications for antitrust policy.

The remaining of the paper is organized as follows. The next section reviews related literature. Section 3 describes the data. Section 4 presents the main empirical results. Section 5 sets up a simple model of store competition that explains the patterns in the data. Finally, section 6 estimates the structural store choice model and measures the impact of competition on consumer.

II Related Literature

This paper looks at the effect of market structure on supermarket assortment and pricing decisions. A recent set of empirical papers that focus on the relation between competition and product variety or other quality variables relates closely to mine. For example, Watson (2009) studies the impact of a nearby retailer on variety decisions in the eyewear industry. He finds a non-monotonic effect of the number of competitors on a store's assortment. A change from zero to one competitor (which is the setting used in my paper) has a positive effect on variety. However, an increase in the number of retailers in markets with more players has the opposite effect. Olivares & Cachon (2009) find that General Motors' dealerships have a larger inventory if a rival is located nearby. Like me, Matsa (2010) looks at the supermarket industry. However, he focus on the effect in stockout rates resulting from changes in the competitive environment. He finds that the presence of a Walmart induces stores to improve their service to consumers by restocking faster the products that disappear from the shelves.

The main force that drives the variety increase in my paper is the desire to capture consumers from competing stores. In that way, this paper is closely related with the theory of free entry being excessive (Mankiw & Whinston (1986)). In an aplication of that framework to radio broadcasters, Berry & Waldfogel (2001) find that a higher concentration of firms decreases the number of radio stations. My paper studies the same idea applied to individual products in supermarkets rather than to the entry of new firms. The forces are similar: when a store introduces a new product in the shelves, it does not take into account the externality on the competitor (i.e., the business stealing effect).

My paper is also related to the literature on endogenous sunk costs (ESC) from Shaked & Sutton (1983) and Sutton (1991), which was applied to supermarkets in Ellickson (2007) and to restaurants and newspapers by Berry & Waldfogel (2010). Like that branch of the IO literature, I also have predictions for the relation between market size and competition. However, an increase in the size of the market in my paper increases variety because the competitive forces will be stronger. In the ESC literature, the effect comes through a larger consumer base that makes any investment costs worthwhile. To understand the differences better, say that a store enters a market where there was already an incumbent monopolist supermarket. My paper predicts that the incumbent store will increase variety because it will have to fight harder to attract consumers. The ESC theory would predict the opposite effect because the consumer base for the incumbent store is reduced and the incentives to invest are now smaller. Another application of the ESC theory, now to Mutual Funds, can be found in Gavazza (2010). That paper looks not only at investment in variety, but also at its effect in equilibrium prices. In that sense, it is closer to my paper.

My paper claims that the indirect effect of competition on prices, through more variety, can dominate the traditional competitive forces. As far as I know, it is the only empirical paper to document this fact. On the theory side, Chen & Riordan (2008) set up a model similar to mine where more competition drives prices upward. In their model, like here, heterogeneous consumer preferences cause that effect. The main difference is that their model does not allow for firms to compete in product variety. Anderson & de Palma 1992 derive the equilibrium in a model with a multiproduct firm. They capture consumer preference for variety in a reduced form through two parameters.

Finally, I estimate consumer preferences for supermarkets. From within the literature on estimating store choice models, the papers that look at supermarket choice are the closest to mine. Like Smith (2004) and Dubois & Jodar (2010), I use household level data and estimate the parameters of interest through maximum likelihood methods. I build on those models to allow preferences for product variety to enter the consumer utility in an explicit way. Hausman & Leibtag (2005) also construct a store choice model where consumers can decide for a traditional supermarket or a supercenter. However, that paper has a different focus: finding ways to correct biases in the CPI. With a different approach, Katz (2007) also looks at supermarket decisions by consumers. He uses moment inequalities to be able to ignore consumers' choice of bundle and be able to focus on the remaining variables that drive the store decision. That strategy, however, only allows him to set identify the parameters of interest.

III Data

The data used in this paper has four components: Store level data from Nielsen Scantrack; Household data from Nielsen Homescan; a list of all the supermarkets in the country in 2004 from Market Scope; and Demographic information from the 2000 Census.

The main data set comes from Nielsen (Scantrack). It includes weekly information from June 2002 to March 2006 on prices and quantities for all the products in the beverage categories, particularly water, fruit juices, soft drinks, teas and milk. The complete data has information for roughly 10000 stores within the U.S. in 3507 different cities. However, I restrict the analysis to those cities in which all the supermarkets available to consumers are in the Scantrack dataset. A residual number of cities that are left have more than two stores. After dropping those, I am left with 707 cities, each with one or two supermarkets. I use the information from Market Scope to see which cities have all their stores surveyed by Nielsen. This stores are located all around the U.S. (figure 10), with a special predominance in the East coast where most of the population lives.

I aggregate week level data to quarterly observations. Then, I construct a price index for each store from the product, UPC level, data. With this, I create three variables for each supermarket/quarter: regular non-sale price, average price and variety. The difference between the two price series is that the first one excludes temporary price reductions while the second does not. The algorithm to remove sales from the original price series and more details about the construction of all the variables are described in the Appendix. I end up with a panel of 983 stores, each with up to 15 quarters, for a total of 10789 observations.

The reason to use a price index based on a fixed basket of goods rather than prices of individual products is that it gives a better indication of what consumers will pay in each store. By giving higher weights to those products in which the average consumer spends more money, I can measure more accurately the impact that going to a different store would have in that consumer's monthly expenditure. For example, if store A has a price index 5% higher than store B, it means that consumers will spend 5% more if they buy *the same bundle* in the first store. The analysis at the UPC level [1] could bias this measurement in

[1] UPC level regressions are somewhat equivalent to using a price index where the weight of each UPC in

both directions. If chains keep most of the prices homogeneous across stores and concentrate the changes in a few products with larger sales, UPC regressions would understate the expenditure difference. On the other hand, if competing firms use loss-leaders more often in a few important products, the impact in a consumer's wallet would be overstated. Using a price index has its drawbacks. In particular, it ignores possible substitution effects to cheaper products when prices go up.

This dataset also includes the address of each store. This information is geocoded using ArcGIS, providing the exact coordinates of each location. I then combine the location information with 2000 Census demographic information at the block group level.

The second set of Nielsen data has detailed information on purchases made by 5345 households. This set is then merged with the store level data used to estimate consumer preferences and compute welfare. The structural model in the last section could be estimated using only the store level data. However, that would put too much pressure on the parametric assumption regarding the distribution of consumer preferences. Furthermore, individual level data helps in the process of controlling for endogeneity in the estimation.

Descriptive Statistics

Table 1 describes the markets used in this paper. Out of the 707 markets, 586 have only one store and the remaining 121 have two. Demographic characteristics are similar for the two groups of stores, with the exception of the median income which is slightly higher for duopolistic cities.

Table 2 presents the average characteristics for stores in the two types of markets. The average store in a duopoly market offers 12% more products and charges higher prices. A consumer living in a city with two stores will face regular (average) prices which are 2.7% (1.7%) higher than a person in a one store market. Finally, the typical monopolist store has quarterly revenues of roughly $2.6 million dollars. That is lower than the duopolist's revenue which is $3 million.

Table 3 decomposes the variety differential. Most of the increase in the number of products of duopolist stores is explained by the 9.1% higher number of brands available[2]. The fact that those stores have 3% more flavors (for each combination of brand and size) also

the Index is 1/J (J being the total number of products).

[2] I am using the Nielsen definition of brand which may be different from the common definition for some products. For example, Regular Pepsi, Diet Pepsi and Pepsi Cherry are considered different brands. However, Tropicana orange juice and Tropicana lemon juice are different flavors of the same brand (Tropicana).

has an impact, but smaller. The number of different formats (sizes or packages) per brand is actually smaller for duopolist stores. This occurs because the brands that are available in every store are the ones that are more established and have larger product lines.

The information that Nielsen provides for private label products is less detailed. For this reason, this table was constructed without those products (roughly 10% of all UPCs).

The last line of table 3 shows the percentage of products, on average, that stay in the shelves into the following quarter. The difference between the two types of market structure is negligible.

The composition of the assortment is very similar for cities with one or two supermarkets (4). That shows both that monopolist stores are similar to duopolists and that the increase in variety for the latter group of store is uniform among all product categories.

The remaining of the paper uses the price of a bundle of goods as a proxy for the price differences that a consumer will find in each store. Figures 5 and 6 show that the price differences between the two types of markets happens as well at the individual UPC level and it does not result artificially from the creation of the bundle of products. The graphs show histograms of the price difference (in percentage) between competitive and monopolist stores. Both charts are skewed to the right because most products are more expensive in duopolist stores.

It is important to keep in mind that I am using only a subset of the beverage products in the construction of the bundle (representing 17% of beverage revenues). Furthermore, the beverage product categories represent only 20% of the total revenues of an average store. However, the fact that I am using the top products in the categories which generate more store traffic makes me confident that the price differences in the sample used are, if anything, understating the true differences.

Table 4 describes the household data used in the paper. The 5345 households are recorded making 242524 different visits to the stores included in the monopolist and duopolist markets studied in the paper. That represents an average of over 45 different visits recorded for each houshold. The typical consumer in the data is recorded during almost 2 years and spends $47.1 on each visit. These patterns are similar for households in duopolist and monopolist cities but the later type includes more observations (consistent with the the larger number of monopolist cities in my store level data).

The data on household purchases includes detailed information on each food product bought by each consumer on each store visit. It includes the price paid and whether the product was on sale. In order to be able to merge this set with the beverage store data, I restrict my analysis to the beverage categories. Within this section, I focus on the six most frequently bought categories: Dairy, Orange Juice, Other Fruit Juices, Soft Drinks, Low calorie Soft Drinks and Water. Those categories comprise roughly 75% of the beverage products purchased. Table 5 shows the propensity to buy a good in each of the six categories, by household demographics. Naturally, consumers with higher income and households with more elements tend to buy more often most products. The notable exception is the regular soft drinks category that is purchased more often for lower income households.

Table 6 describes the ten most purchased brands in each of the categories.

IV Empirical Results

Main Analysis

I will now look into the impact of having a nearby competitor on a store's choice of assortment and price level.

Figures 2 and 3 show that the unconditional mean of variety and prices is higher for stores in duopolist markets and that this effect is persistent over time. Looking at either the regular or average price series, the conclusion is the same. The vertical difference is higher for the regular prices, suggesting that there are more sales with competition.

Tables 7 to 9 present regressions of variety and prices on a different set of controls. The first column of each table includes the regressions with the dependent variable in levels. In the remaining columns (2)-(4), the variable being explained is in logs. Regressions (3) and (4) include chain fixed effects.

Columns (1) and (2) of Table 7 suggest that being a monopolist reduces the number of products that a store has in its shelves. Having a rival in the same city is estimated to increase the number of products in the beverage categories by 116 or 8.5%. Those numbers are obtained after controlling for population, income, age and education.

Using the same set of market controls, Tables 8 and 9 show the impact of competition on prices. I find that regular store prices drop by 1.46% and a similar effect, albeit smaller, if I include sales: 0.92%. The same results are found in analogous regressions run in levels (column (1) of those tables).

In order to compare stores *within the same chain,* I introduce chain fixed effects in the main regressions. In that case, the effect of competition does not change (columns (3) of

Tables 7, 8 and 9). A store with a competitor in the same city will have a 3.7% higher number of products than a store of the same chain without rival. Furthermore, regular prices will be 0.5% higher and average prices, 0.3%. The magnitude of the difference is smaller within a group, which is consistent with the idea that some chains tend to have a uniform policy for all or part of their stores. In that case, if most stores of a chain are in monopolist markets, the few that are in a duopoly may not increase variety and prices that much.

Also, by introducing chain fixed effects, I am effectively ignoring in the regressions a large number of firms. Chains that appear in only one type of market structure (either monopolist or duopolist) are not used to estimate the coefficients of interest. Out of the observations remaining, a large number belongs to a small number of chains, which will have a large impact in the coefficients estimated. If the effect of competition for those chains is not representative of the true effect for other stores, the estimated coefficient will be biased. Columns (4) of Tables 7-9 show the results of the same regressions, now without chains that account for more than 500 observations. This means that the four most prevalent chains in my data: Food Lion, A&P, Stop & Shop and Shoprite are excluded. The coefficients of the monopolist dummy go up in absolute value, indicating that, in fact, the effect is smaller in those four chains. The common characteristic of the chains omitted is that they belong to four of the largest groups in the data in terms of the number of stores. This could be an indication that the effect that I am describing in the paper is stronger for small chains than for large groups. Large chains may have stronger incentives to keep prices and assortment consistent across stores to protect their "brand value". I investigate this hypothesis in the data by correlating the size of a chain with the effect of competition in its stores (not reported) . I could not find enough evidence in its favor.

It is natural to think that a firm that expects ex-ante to offer a wider assortment will build a larger supermarket. Retail space is a variable that can be changed in the long-run but not in the short-term. Under the assumption of no unobserved differences between duopolist and monopolist markets, we can read the coefficients from Tables 7-9 as the long-term effect of competition. When a monopolist and a duopolist stores are in markets with exactly the same characteristics but one store is bigger than the other, that difference has to come from competition. In particular it comes from the desire of the store in a duopoly to carry more products to steal consumers from the competitor. If a store is not able to change its retail space after an increase in competitive forces, it will not be able to increase the assortment available to the consumer as much as it would desire. Table 10 includes the size of a store as a covariate. As such, we can read the coefficients in that table as the impact of competition on a store if no change in the retail area takes place. In this case, competition is estimated to increase the number of products by 44 or 3.4%. Adding a second store in a city previously dominated by a monopolist will have an increasing impact on regular (average) prices of 1.17% (0.69%).

The price index used before is composed of 164 UPCs, chosen for being available in every store. They belong to different beverage product categories (following the Nielsen classification), as can be seen in table 11. With the objective of looking into the effect of competition by product category, I create separate price indexes for each family of products. I do this for the six most represented categories (with more than 10 UPCs)[3]. The unconditional mean of the regular prices is higher for markets with competition in all the segments, except for Cranberry Juice in the first six quarters of data (figure 8). The price differential is higher for the juice categories than for the soft drinks. This may be an indication that variety is more important in the former. If we think that the soft drinks market is more concentrated around Pepsi and Coke then, adding more brands to that segment will not attract many consumers. In fact, table 7 shows that competition has a higher impact in the number of products offered in the three juices categories.

When I incorporate sales in the price index, it cancels part of the price effect in the soft drinks but not in the juice categories (figure 9).

Additional analysis

The previous section documents the difference in prices and variety between stores with and without competitors. The goal is to establish causality between increasing competition and higher prices/variety. The analysis so far requires a careful reading of the results. One concern is that the linear model employed is not capturing correctly the impact of the covariates. That would mean that the observed differences between the two types of markets were not being correctly accounted for and that could explain the differences found. A second concern is the possibility of systematic unobserved differences between the two market structures that might have an impact in the strategic choices of stores. Even if I were to control for a much larger number of market characteristics, the problem would persist. Finally, I am using a city as the market definition. In the real world, however, there are no physical barriers that prevent consumers from doing their purchases in stores outside

[3] The Product Categories are: fruit drinks-other container, soft drinks - carbonated, fruit juice - orange - other container, soft drinks - low calorie, soft drinks - powdered, fruit drinks and juices-cranberry.

of the city. If consumers travel more often beyond the city borders in the monopolist case, results can be biased. In particular, my classification of markets as high or low competition could be inaccurate.

I now address all these concerns in more detail.

Observed market differences

I now control in a more flexible way for observed differences in the markets by using propensity score matching. The objective of this strategy is to separate the effect of the market characteristics on the market structure from the effects on the response variable (i.e. prices and variety). The idea is to collapse all the effect of the covariates into one variable and use it to compare stores of similar characteristics (see Rosenbaum & Rubin (1983) for more details). By doing this, I am able to attribute any systematic differences between the two groups to the effect of competition.

The first step to calculate the estimator is to construct the propensity score for each store, i.e., the probability that a specific location would be a monopolist or duopolist. In order to do this, I use a logit specification with a dummy variable indicating if a store is a monopolist regressed on a set of market characteristics introduced in a flexible way. It allows the covariated to impact the market structure in a variety of non-linear ways. The results of the propensity regression can be found in table 14

The second step consists in matching each monopolist store with duopolist stores of identical propensity and looking for a systematic difference between the two groups in terms of prices and variety offered. Since the propensity score is a continuous variable, it is impossible to do a one to one matching. Hence, I match each monopolist with a group of duopolists of similar propensity. I use a Normal Kernel density to give a larger weight to those stores closer in terms of score. The bandwidth used was $1/10$[4]. Figure 11 shows the distribution of the propensity scores by market structure. By construction, the distribution in the case of the monopolist stores is further to the right than that of duopolists. However, there is still a large overlap between the two distributions. This condition is required for this type of analysis.

Table 15 shows the results of matching supermarkets by propensity score. I estimate that having a competitor in the same city will increase assortment by 9.6%, regular prices by 1.4% and average store prices by 0.9%. Those results are in line with the linear model

[4] Alternative bandwidths were tested without changing the results substantially.

presented in the previous section. In table 16, I include store size in the list of covariates that generate the propensity score. Results are still significant but smaller in magnitude. If a chain that decides to open a supermarket will condition the size of that store on the competitive environment of the location chosen, table 15 is the correct one to look at. If, on the other hand, the size of store is a good predictor of whether it will be a monopolist or a duopolist, table 16 will provide the correct impact of competition.

I also look at the effect by propensity score quartile (Table 17). In the lowest region, i.e. markets more likely to be duopolists, there are almost no observations. Looking at the remaining three quartiles, I conclude that the absence of competition has a stronger negative effect on prices and variety for markets with "monopoly characteristics". Moreover, for the second quartile (propensity score between .25 and .5), monopolist markets actually have higher prices and more variety. Within that group of stores, competition decreases regular (average) prices by 0.8% (0.7%) and reduces variety by 5.6%. Stores with a propensity score higher than .5 (roughly 90% of all stores in my data) are estimated to have the opposite effect.

The finding in the previous paragraph is consistent with both an idea of competition in variety and with previous literature on the relation between market size and the number of products offered (for example Berry & Waldfogel (2010)). Monopolist markets tend to be smaller. If two stores operate in a market with "monopolist characteristics", I expect competition to be much stronger than between two similar stores in a market with "duopolist features". An increase from one to two stores in locations with a high propensity score will induce a strong competition to attract consumers. Cities with a lower score will be able to accommodate better the two stores and the competition intensity will be smaller. On the other hand, a supermarket without rival will have a higher residual demand, which will result in more incentives to increase variety. The net effect of introducing a second store depends on the relative strengths of the increased competition and the lower residual demand. In "monopolist type" markets, the competition effect dominates and the net impact on variety will be positive, which ends up increasing prices as well. For "duopolist type" markets, the opposite is true.

Unobserved Differences

If there are further differences between the two types of market structure which are not controlled for, the estimated coefficients can be biased. This might be the case if those factors affect both the likelihood of a specific market structure and the elasticity of the demand curve.

Fortunately, within the time span of my data, there are some cities that observe a change in structure. Some of them move from one store to two while others face the reverse process. Overall, 6% of the markets change the number of stores at some time during the 15 quarters of the data. This allows me to use market fixed effects to capture city characteristics - observed and unobserved - and still identify the effect of competition. Table 19 presents regressions of the variables of interest on a monopolist dummy, size of store and market and time fixed effects. These regressions measure the reaction of stores to changes in the competitive environment. It is natural to think that the full impact of those changes upon prices and variety (especially the latter) takes more than a few quarters to be felt[5]. Still, the coefficients estimated are already significant and in a direction consistent with my previous results. The first row of table 19 shows an average reduction of variety of 1.6% with decreased competition. Also in line with previous numbers, supermarkets increase regular (average) prices by 0.9% (0.6%) after the appearance of a new rival store in the same city.

Market fixed effects capture all the unobserved characteristics of a city which are *constant over time*. If those markets are changing in some unobserved way (population is growing faster than in other cities, for example), results can still be biased. To make sure that that is not the case, I focus now on markets where the change in structure was quickly reversed. Those are mostly markets in which a store closes temporarily and reopens after a few quarters with a different name. In markets like those, there are no changing trends that affect the number of players that co-exist in equilibrium. The reason for the temporary closing of one of the stores is related usually to some reconstruction work so that it can open with a new brand name. Table 20 show that the same results hold when I restrict the sample to markets where the change of structure was reversed.

Further evidence that competitive forces are causing the higher variety and prices comes from duopoly markets where the two stores belong to the same chain. That is the case in 5% of the markets. Those stores operate in markets with duopolist characteristics but without the competitive forces. I run the basic regressions including a dummy for duopolist markets with stores from the same chain. Results from table 18 show that, without the competitive pressure, those markets have prices and variety similar to monopolist stores.

[5]This is the reason why "store size" is included as a regressor. A sudden entry/exit of a rival will only have an impact in the incumbent's store size in the very long run.

Market Definition

I now discuss the market definition used in this paper.

I am treating cities with only one store as "monopolistic" or "low competition" markets. Similarly, cities with two store are called "duopolistic" or "high competition". Cities, of course, do not have physical boundaries that prevent consumers from buying groceries elsewhere. If the competitive pressure from stores outside the city exists but is independent of the market structure, the results from the main regressions hold. Problems may arise, however, if cities with only one store face stronger outside competition. One can think of an hypothetical reason why this may occur. Suppose that the presence of a nearby Walmart reduces the likelihood of a second store entering the city (as in Grieco (2010), for example). In that case, if the Walmart pressure on prices and variety is stronger than that of a second traditional supermarket inside the city, the interpretation of the coefficients estimated will be wrong. It would still be true that cities with two stores would have higher prices but that effect would come from *smaller* rather than stronger competition.

To address this concern, I look into the distance of each city to the closest supermarket outside of its boundaries. Table 12 shows that monopolist markets are, on average, more distant from external stores. This is an indication that, if anything, the coefficients estimated are understating the true effect of competition. To confirm this idea, I run the standard regressions restricting the sample to markets where the nearest outside supermarket is further than X miles (for $X = \{2, 10\}$). Most consumers do not travel more than 2 miles when doing grocery shopping in traditional supermarkets (Orhun (2005)). However, they may travel longer distances in the case of non-traditional formats like Walmart supercenters. By using either the 2 or the 10 mile cutoff, I find that the presence of a competitor has a positive impact on variety and prices (Table 13). Notably, the effect increases dramatically when I use the 10 mile cutoff. In that case, increasing competition will result in a store with 22% more variety (regression (2) of Table 13). Similarly, it will increase regular and average prices by 6.1% (column (3)) and 4.2% (column (5)), respectively.

One possible explanation for the much larger effect estimated using the 10 mile cutoff is that cities, in general, face some competitive pressure from outside its borders. That exterior pressure will smooth the effect of having an additional player inside the city. When I use the 10 mile cutoff, I look into those consumers that have almost no possibility of buying outside the city. This may be an indication that I was actually understating the true effect of competition upon variety and prices.

V Model

This section describes a simple model that generates equilibrium predictions consistent with the data patterns reported in the previous section. In the model, consumers choose optimally which store to visit and which products to buy and supermarkets compete to attract those consumers.

To keep the intuition of the model easy to understand, I limit the number of products that each store can carry to two. Increasing this number would only add complexity to the set-up without additional benefits in terms of economic forces.

Setting

Consumers

There are N consumers in a market deciding which supermarket to visit and which product to buy. Consumers can only visit one store (shopping time is limited) and will have to choose according to the information that they have before the visit. Without loss of generality, I assume that they will always visit one store.

When a consumer enters a supermarket, he may find either product A, product B or both available. Consumers have heterogeneous preferences in the sense that $1/2$ of them prefer product A over B and the remaining prefer good B over A. The preferred good has a consumption value of v_H while for the the less-preferred that number is only v_L. The 2 goods are close substitutes in the sense that $v_H < 2v_L$. The net utility of a good is the consumption value minus the price. Once inside a store, a consumer buys the good that gives him higher net utility. If all the goods have consumption value below its price, the consumer buys nothing. Let Ω_s denote the set of products available at store s. Then, the value for consumer i of visiting store s is:

$$V_{is} = \max_{j \in \Omega_s} (v_{ij} - p_{js}, 0)$$

If two options provide the same utility, the consumer decides between them with equal probability.

Consumers know the assortment of each store before the visit but not prices. This assumption is realistic for the supermarket industry and reflects the fact that assortment is usually a quarter decision while prices change every week. When a consumer enters a store he knows that, with a high probability, prices will be different from the last visit whereas the set of products available will be the same. The version of the model below assumes that all the consumers are uninformed about prices. In equilibrium, they will correctly infer the true price distribution.

In the appendix, I allow for a positive share of consumers to be informed about the exact prices of each product before visiting the store. I show that the results hold if that share is not too high. With a positive number of informed consumers, the equilibrium will be in mixed strategies (similar to Varian (1980)). In that case, competition increases both the mean and the variance of the price distribution, consistent with what I find in the data.

Firms

Firms maximize profits. In the first stage, each one choose the set of products to sell ($\Omega \in \{A, B, (A, B)\}$). Then, firms compete in prices (knowing the assortment of all the other stores).

A larger assortment comes at a cost. Having more products available to the consumer increases the costs of the firm in terms of storage space, stock-outs, time to change prices and so on. That increase in fix costs is captured by the function $\mathcal{F}(\Omega)$. Without loss of generality, I assume that function F takes the following form:

$$\mathcal{F}(\Omega) = \begin{array}{l} 0 \text{ if } \Omega = A \\ 0 \text{ if } \Omega = B \\ F \text{ if } \Omega = (A, B) \end{array}$$

Furthermore, firms are not able to credibly commit to a specific vector of prices.

Equilibrium

With the previous assumptions, the set of possible equilibrium prices is restricted to $\{v_L, v_H\}$. Since consumers are not informed ex-ante, decreasing prices does not increase store traffic. The only advantage for a firm is that it may induce more purchases from the consumers that are already inside the store. If a store carries both products, profit is maximized by having $p_A = p_B = v_H$. This is because any consumer that enters the store will always purchase one of the goods and will do so at maximum price. If a store only sells one product, it will choose either a price of v_H, if it decides to sell to only one type of consumer, or v_L, if it sells to both, depending on the relative number of consumers of each type that enter the store.

I now derive the equilibrium for a monopolist and a duopolist firm. For each type of firm, I first solve for the optimal prices for each possible choice of products. Then, I go back to the first stage of the firm's problem and choose the optimal assortment.

Monopolist

In a market with only one store, the only decision that the consumer has to make is which good to buy, if any.

Let $\Pi_{\Omega}^{M}(p)$ denote the profit of a monopolist firm that chooses assortment Ω and vector of prices p. If the monopolist firm decides to have only product A available to the consumer (analogous to product B), the following will be true:

$$\begin{aligned} \Pi_A^M(v_L) &= v_L N \\ \Pi_A^M(v_H) &= v_H \frac{N}{2} \end{aligned}$$

With $p = v_H$, only one type of consumer (of size $\frac{N}{2}$) will purchase the good. On the other hand, a price of v_L will induce both types of consumers to buy. Since $v_H < 2v_L$, a low price will be optimal.

Now imagine that the monopolist store chooses instead to have a larger assortment of products, i.e., $\Omega = (A, B)$. In that case, every consumer that enters the store will find a product for which it is willing to pay v_H. Therefore, it becomes optimal to have high prices for both products.

$$\Pi_{AB}^M(v_H, v_H) = v_H N - F$$

The optimal choice of variety in the first stage of the game depends on the magnitude of the fixed cost F. In particular, the solution to the firm problem is:

$$\begin{aligned} \text{for } F &< (v_H - v_L)\, N: \; \Omega^* = (A, B), p^* = (v_H, v_H) \\ \text{for } F &> (v_H - v_L)\, N: \; \Omega^* = A, p_A^* = v_L \text{ or } \Omega^* = B, p_B^* = v_L \end{aligned}$$

Duopolist

The difference from a duopolist store to a monopolist is that the former has to take into acount the impact of its decisions upon store traffic.

I start by solving for the optimal prices for each combination of products that firms choose. Let $\Pi^D_{\Omega_s,\Omega_{-s}}(p_s;p_{-s})$ denote the profit of a duopolist firm that chooses (Ω_s,p_s) as assortment and prices when its competitor has (Ω_{-s},p_{-s}). If both stores decide to have high variety, i.e., $\Omega_s=\Omega_{-s}=(A,B)$, any consumer that enters the store will buy one product if both prices are v_H. Therefore, no store has incentives to lower the price below that and the profit will be:

$$\Pi^D_{AB,AB}\left[(v_H,v_H);(v_H,v_H)\right]=v_H\frac{N}{2}-F$$

Assume now that store s has only one product (without loss of generality $\Omega_s=A$) while store $-s$ carries both ($\Omega_{-s}=(A,B)$): Similar to the previous argument, $p_{-s}=(v_H,v_H)$. Consumers know that stores will only charge prices of either v_H or v_L in equilibrium. So, those consumers that value product B highly know that, by going to store s, they will get at most the same utility than from store $-s$, but possibly less. If they believe that store s will quote a price of v_H with positive probability, all the consumers of this type will go to store $-s$. So, store s will only receive visits from consumers that prefer product A over B and will, in fact, choose $p_{sA}=v_H$ [6]. Consumers of that type will be indifferent and split evenly between stores ($\frac{N}{4}$ each). So, the profit for each firm is:

$$\begin{aligned}\Pi^D_{A,AB}\left[v_H;(v_H,v_H)\right] &= v_H\frac{N}{4}\\ \Pi^D_{AB,A}\left[(v_H,v_H);v_H\right] &= v_H\frac{3N}{4}-F\end{aligned}$$

The final possibility is for both firms to choose one product each. If that is the case, in equilibrium, firms will choose different products in order to differentiate from the competitor and be able to charge higher prices. That means that each type of consumer will visit a different store and the following will be the profit of each firm:

$$\Pi^D_{A,B}\left[v_H;v_H\right]=v_H\frac{N}{2}$$

[6] In fact there is another possible equilibrium here. If consumers believe with probability 1 that store s will have a price of v_L, they will be indifferent between going to either store. In that case, they will split and store s will in fact quote a price of v_L. This equilibrium, however, will be unstable. It will collapse with a very small perturbation in the beliefs of consumers.

Knowing the payoffs for each combination of variety, firms play the game represented in the following matrix (V_i is the number of products of firm i):

$V_s \backslash V_{-s}$	1	2
1	$(v_H \frac{N}{2}, v_H \frac{N}{2})$	$(v_H \frac{N}{4}, v_H \frac{3N}{4} - F)$
2	$(v_H \frac{3N}{4} - F, v_H \frac{N}{4})$	$(v_H \frac{N}{2} - F, v_H \frac{N}{2} - F)$

For $F < v_H \frac{N}{4}$, an assortment with two products is a dominant strategy for each firm.

Comparison across market structures

The thresholds in the fixed cost F for which stores find it optimal to move from high to low variety are different for a monopolist and a duopolist stores. If the difference between v_H and v_L is not too big $\left(\frac{(v_H - v_L)}{v_H} < \frac{1}{4}\right)$, monopolist stores will switch to low variety at lower values of F. In that case, there will be a range of possible values for F where competition drives both variety and prices up, as shown in figure 1.

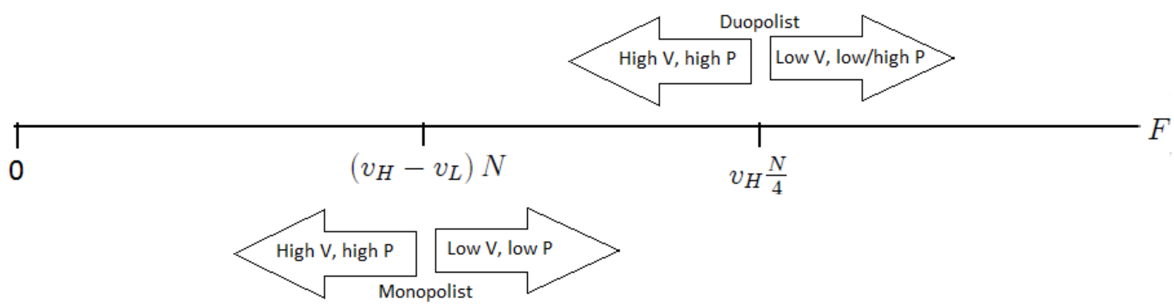

Figure 1: Equilibrium strategy for a monopolist and duopolist store for different values of the fixed cost F

The intuition for the results is simple: each store faces a trade-off when deciding to increase variety. It increases the fixed costs of the firm but it also raises the revenues. The cost increase is the same independently of the number of competitors that each store faces. However, the change in revenues will be different for a monopolist and a duopolist supermarket. An investment in variety increases the revenues on each existing consumer but it also drives more traffic into the store. The first effect is stronger for a monopolist since its consumer base is bigger. However, a duopolist store has more to gain from an increase in store traffic since it may capture consumers from the rival. This business stealing effect can be enough to make the incentives to increase variety higher in stores with competition. When that occurs, competitive prices will be higher as well. The reason for that to happen is that the monopolist firm will lower its prices to sell to both types of consumers.

VI Consumer Welfare

The previous two sections show 1) empirical evidence that competition has an increasing effect on variety and prices, and 2) a theoretical explanation for those patterns. When consumers have a wider range of products to choose from but pay higher prices for them, it is not clear if they are better-off. In order to address whether consumers' welfare increases, I estimate a model of consumer choice and use it to perform welfare comparisons between the two types of market structure. I examine only the impact of changes in variety and prices on consumers and abstract away from other potential welfare implications that increased competition may have (for example, lower distance to the closest store or less stockouts).

A Consumer Choice Model

Consumers choose one store to visit each period. Each visit is an independent event. That has the advantage of allowing for different shopping frequencies for different consumers without explicitly modelling the frequency of visits decision. On the negative side, as with most of the supermarket choice literature (for example Smith (2004) and Dubois & Jodar (2010)) it does not allow consumers to split the purchases over multiple store visits. This model looks at the choice of store *conditional* on the decision to visit one.

Each consumer decides which store to visit based on his preferences for prices, variety and a number of other store characteristics. The relative importance of prices and variety depends on what consumers buy in each product category. For this reason, I model the choices that consumers make at the category level as well.

Let Ω_{jct} and p_{jct} denote respectively the set of products and vector of prices at store j in period t and product category c. Then, the utility for consumer i of going to store j is:

$$U_{ijt} = \sum_{c} \left(\theta_{ic} * E\left(V_{ijct}\right)\right) + \gamma_{ij} + u_{ijt}$$

It includes a household-store specific component (γ_{ij}), which represents the individual characteristics that are constant over time and affect the choice of supermarket. That term captures things like the size of each household, the number of cars, proximity of workplace to each store, consumers' tastes for store music and so on. The error term u_{ijt} captures random shocks to the value of going to a store. It is independent over stores, time and individuals and follows an extreme value (type I) distribution.

The main component of the utility of going to a store comes from the bundle of goods that is purchased, which is assumed to be a linear function of the utility derived from the separate categories. The term $E\left(V_{ijct}\right)$ is short for $E\left(V_{ijct}(\Omega_{jct}, p_{jct})\right)$ and represent the utility that consumer i expects to get from category c prior to visiting the store. When making that decision, consumer i knows which products are available and where but not the prices or the vector of idiosyncratic shocks e_{it} [7] . That means that it will decide which store to visit based on the *expectation* on those two components. In particular, for category c:

$$E\left(V_{ijct}\right) = \int_{e,p} V_{ijct} dF(e,p)$$

The vector of parameters θ in the store utility has two functions in the model. First, it converts category utils into store utils (in terms of the error term u_{ijt}). Second, it varies by category which allows the impact of each in terms of attracting consumers to a store to be different. Within each category, the utility from not purchasing is normalized to zero. If, for example, not buying milk in the supermarket means that the consumer will not consume that product at all while not buying a soda means that he will buy it instead from a small convenience store, the outside utility will mean different things. In that case, the parameters θ_{ic} will play an important role.

If we knew the utility from each category, V_{ijct}, we would be able to compute the probability that consumer i would visit store j as:

$$\Pr_{it}(j) = \frac{\exp(\sum_c \left(\theta_{ic} * E\left(V_{ijct}\right)\right) + \gamma_{ij})}{\exp(\sum_c \left(\theta_{ic} * E\left(V_{ijct}\right)\right) + \gamma_{ij}) + \exp(\sum_c \left(\theta_{ic} * E\left(V_{i-jct}\right)\right) + \gamma_{i-j})}$$

In order to compute this probability, I can either approximate V_{ijct} as a function of prices and variety or to model each category and compute V_{ijct} exactly. I choose the second option because it allows me to understand exactly where the preference for variety is coming

[7] I am currently working to relax this assumption

from. With that, I am able to change the *set* of products available in a store and measure the impact on consumer choices[8]. Also, I will not rely on parametric assumptions on V_{ijct} for the results. They will, however, be more sensible to misspecification of the category model.

Every store has C product categories and every product category is composed by a number of different products. Let Ω_{cjt} be the set of products available at time t in store j, product category c. Once inside a store, a consumer will go to category c and decide which product to buy, if any (the utility of not buying is normalized to zero). He buys, at most, one different product in each category. The utility from category c is then:

$$V_{ijct} = \max_{b \in \Omega_{cjt}} (\delta_{ib} - \alpha_{ic} p_{jbt} + e_{ibt}, 0)$$

The coefficients δ_{ib} and α_{ic} represent respectively the value of product b and the price elasticity for consumer i. The error term e_{ibt} represents an idiosyncratic shock to the value of product b. It is independent across consumers, time and brands and follows an extreme value (type I) distribution. As stated before, the vector e_{it} is only realized after consumer i has entered the store. This last assumption is not innocuous. It doesn't allow, for example, consumer i to choose a store when the shopping needs are large and a different one when only a few products are needed. The main advantage is computational because it separates the store problem from the brand choices. It is not clear that this assumption will bias the results. If a consumer knows before entering the store that he only needs milk, he restricts his analysis to the dairy category. The store with the highest value for dairy may not always be the one that has the highest value for the sum of the categories but there are reasons to believe that they are strongly correlated (Figure 4).

The probability that consumer i will purchase product b in a given store visit is obtained by integrating over e :

$$\Pr_{ijcbt}(\delta, \alpha) = \frac{e^{\delta_{ib} - \alpha_{ic} p_{jbt}}}{\sum\limits_{b \in \Psi_{cjt}} e^{\delta_{ib} - \alpha_{ic} p_{jbt}} + 1}$$

Note that the addition of a new product will increase the expected utility of that product category even if the new δ is low.

I can now derive the expression for the expected category utility and close the model:

[8] This will be important for the welfare simulation ahead

$$
\begin{aligned}
E\left(V_{ijct}\right) &= \int_{e,p} V_{ijct} dF(e,p) \\
&= \int_{e,p} \ln \sum_{b=0}^{B} \exp(\delta_{ib} - \alpha_{ic} p_{jbt}) dF(p)
\end{aligned}
$$

The independence of p and e together with properties of the logit function (see, for ex., McFadden (1974)) explain the second line.

Estimation

I use the sample of households from Homescan described in table 4 to estimate the model above. I use six product categories that account for roughly 75% of all the purchases in the beverage segment: Dairy Products, Soft Drinks Carbonated, Soft Drinks Low Calorie, Orange Juice, Other Fruit Juices and Water. Ignoring the remaining 25% of the beverage purchases and the non-beverage categories, for which I have no store level data, may affect the results. I am assuming that any differences across stores in the value of the missing categories stay constant over time (therefore, they are captured by γ_{ij}). If the value of those categories changes, it is likely that it will go in the same direction as the beverage categories that were included. In other words, if a store increases the variety and prices of orange juices and sodas, it will probably do the same for other categories like beer and cookies. Under that assumption, ommiting some of the categories from the store choice model bias the estimated magnitude of each category impact. However, it does not change the direction of the welfare analysis, which is the main goal of this section.

For computational reasons, I collapse products into brands . This is a good approximation because only infrequently (8% of the times) do consumers buy more than one brand in a given product category. Furthermore, I allow the ten most purchased brands in each category (see Table 6 for a list) to have different values for consumers but impose $\delta_{ik} = \delta_{ih}$ if k and h are two brands outside of that group. The price index for a brand represents the average price paid, per ten ounces of product, by all the consumers that bought one of the sizes/flavors of that brand in a given week in each store. The implicit assumption is that a consumer in my sample would pay the same price that the average consumer did for a given brand.

The distribution of prices for each brand at time t - $F_t(p)$ - is constructed from the 12 weeks of prices prior to each visit.

Consumers are grouped into six *types* according to their demographic characteristics. In particular, they are divided into three income groups and two household size categories[9]. Within a *type*, consumers are assumed to have the same shopping behavior (i.e., price sensitivity and brand preference). Table 5 shows descriptive statistics by consumer type including the percentage of visits in which each category is purchased.

With the assumptions of the model, the store choice problem, for each type, becomes a sequence of $C+1$ discrete choice problems: one for each product category and another at the store level. For now, I am treating each consumer within a type as having the same preferences. Unobserved heterogeneity can be addressed within this setting by allowing the error term to be known before the store visit and by estimating simultaneously the store and the category decisions. Assuming that the error vector e only occurs after choosing a store, I can estimate the different problems recursively.

Then, the estimation algorithm is:

1) Estimate the product choice model for each consumer type in each of the C categories by maximizing the following log-likelihood function:

$$LL_c(\delta, \alpha) = \sum_t \sum_{i=1}^{N} \sum_{b=0}^{B} y_{ijcbt} \ln \Pr_{ijcbt}(\delta, \alpha)$$

where y_{ijcbt} is an indicator function for a purchase of brand b.

2) With the parameters δ and α known, integrate over e and p to compute each $E\left(V_{ijct}\right)$.

3) Treating $E\left(V_{ijct}\right)$ as data, I go into the store choice model and estimate the parameters θ and γ, again by maximum likelihood

The identification for θ comes from variation over time in the set of products and prices in each store and changes in the share of visits that each consumer does to each store.

Results:

Results for the Orange Juice category are included in Table 21. Similar models (non-reported) were estimated for each of the five remaining categories with similar results.

Finally, the store choice model is estimated (Table 22). For practical reasons, due to the large number of consumers, the individual householdXStore effects were replaced by chain fixed effects. This follows some of the previous literature that estimates store choice from

[9]The income categories are: less than \$30k; between \$30k and \$70k; more than \$70k. The household size categories are: less or equal to two members; more than two members.

individual consumer data (Smith 2004, Dubois et al 2010). Most of the store characteristics, other than prices and variety, are decided at the firm level. So, stores within a chain have similar characteristics in most of the cases.

The results from table 22 show that the Orange Juice and Water categories are less important in attracting consumers to a store.

Welfare Simulation

In order to measure the net effect of competition on consumer welfare, I combine the results from the structural estimation above with the numbers that I obtain from the propensity score analysis in Table 15. In that part, I estimate that a monopolist store has, on average, 9.74% less variety and 0.91% lower average prices. By adjusting the number of products and prices of the duopoly stores, I am able to simulate the behavior of those consumers if they lived in a city with only one supermarket.

An important aspect of my analysis is that I look at what would be the choices of consumers if they were buying at monopoly level variety and prices but keeping constant the number of stores and frequency of purchases. Not doing so would artificially give an advantage to the duopoly markets.

The counterfactual monopolist price for each brand is created by lowering the duopoly prices by 0.91%, i.e.,

$$p_{jbt}^{M} = (1 - 0.0091) p_{jbt}^{D}$$

The adjustment to the set of products is less straightforward. Due to the discreteness of the assortment variable, I cannot remove exactly 9.74% of products from each category in each store. Instead, I drop a number of products necessary for the *average* reduction across stores to be close to 9.74%. The products dropped were chosen randomly across stores and products, excluding the 10 most purchased brands. Also, I do that process separately for each product category in order to keep constant the average structure of the assortment.

With the new vector of prices and set of products, I can recursively compute $E\left(V_{ijct}^{M}\right)$ and U_{ijt}^{M}. Finally, the expected welfare for consumer i (McFadden (1974)) in a monopolist hypothetical world is:

$$W_i^M = \sum_t \ln\left(\sum_j \exp\left(U_{ijt}^M\right)\right)$$

Table 23 compares consumer welfare in a monopolist and a duopoly city. The average consumer is better-off with competition. The same is true for each income group and holds independently of the household size.

VII Conclusion

In this paper I look at the impact of competition on supermarket variety and prices. After controlling for market differences, I conclude that competition not only increases the number of products available in a store but also the price level.

The empirical findings are consistent with a theory of competition where variety is the main variable that stores use to attract consumers. The increase in variety is indirectly responsible for higher prices. In this scenario, consumer welfare will not necessarily be higher in markets with more stores. I find that consumers in different demographic profiles do benefit from increasing store competition. However, that benefit comes from having a wider choice of products.

This paper shows that the effects of competition have to be analyzed carefully. That is specially true in industries where decision variables other than prices play an important role. The notion that there may be some indirect effects from competition that can have consequences in terms of prices is very relevant for competition policy. Accounting for these indirect effects can be determinant, for example, in the decision to allow or not a merger to go through.

I study one particular industry - supermarkets - where variety seems to be an important variable. It would be interesting to look at other industries and see if similar effects are present. It may be important to quantify the indirect effect of competition on prices, even when it does not dominate the traditional effect.

Future research should also study the effect of competition for larger markets, where more than two players are present.

This project, in particular, can be extended by relaxing some of the assumptions in the consumer choice model. The first is to allow consumption shocks to affect both the store and category decisions simultaneously. The second is to allow consumers to visit multiple stores.

References

ANDERSON, S. and A. DE PALMA (1992): "Multiproduct firms: a nested logit approach". *Journal of Industrial Economics* 40, 261–276.

BERRY, S. and J. WALDFOGEL (2001): "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting", *Quarterly Journal of Economics*, 116(3), 1009–1025.

BERRY, S. and J. WALDFOGEL (2010): "Product Quality and Market Size", *Journal of Industrial Economics*, Blackwell Publishing, vol. 58(1), pages 1-31, 03.

Bureau of Labor Statistics (2010), U.S. Department of Labor, Career Guide to Industries, 2010-11 Edition, Grocery Stores

CHEN, Y. & M. H. RIORDAN (2008): "Price-increasing competition," *RAND Journal of Economics*, vol. 39(4), pages 1042-1058.

DRAGANSKA, M., M. MAZZEO and K. SEIM (2009): "Beyond plain vanilla: Modeling joint product assortment and pricing decisions", *Quantitative Marketing and Economics*, Springer, vol. 7(2), pages 105-146, June.

DUBOIS, P. and S. JODAR-ROSELL (2010): "Price and Brand Competition between Differentiated Retailers: A Structural Econometric Model". CEPR Discussion Paper No. DP7847

ELLICKSON, P. (2007): "Does Sutton Apply to Supermarkets?", *RAND Journal of Economics*, 38, pp. 43–59.

GAVAZZA, A (2010): "Demand Spillovers and Market Outcomes in the Mutual Fund Industry", Working Paper.

GRIECO, P. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets,"Working paper. Penn State University.

HAUSMAN, J. and E. LEIBTAG (2005): "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart". NBER Working Paper Series, Vol. w11809, pp. -, 2005.

HENDEL, I and A. NEVO (2006). "Measuring the implications of sales and consumer inventory behavior". *Econometrica* 74 (6), 1637–1673.

Information Resources, Inc. (2002): IRI Insights on Channel Differentiation, Chicago: Information Resources, Inc.

KATZ, M. (2007): "Estimating supermarket choice using moment inequalities". Ph.D. Dissertation, Harvard University.

LEUVEN, E. and B. SIANESI. (2003): "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing".

MANKIW, N. G., and M. WHINSTON (1986): "Free Entry and Social Inefficiency" *RAND Journal of Economics,* 17, PP. 48–58.

MATSA, D. (2010): "Competition and Product Quality in the Supermarket Industry", Kellogg School of Management.

McFADDEN, D. (1974): "Conditional logit analysis of qualitative choice behavior". In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York

OLIVARES, M. and G. CACHON (2009): "Competing retailers and inventory: an empirical investigation of General Motors Dealerships in isolated markets". *Management Science.* 55(9). 1586-1604.

ORHUN, Y (2005): "Spatial differentiation in the supermarket industry". Working Paper, Berkeley.

ROSENBAUM, P.R. and D.B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

SHAKED, A. and J. SUTTON (1983): "Natural Oligopolies." *Econometrica*, Vol. 51, pp.1469-1484.

SMITH, H (2004): "Supermarket choice and supermarket competition in market equilibrium", *Review of Economic Studies*, vol. 71 (1),.235-263

SUTTON, J (1991): "Sunk Cost and Market Structure: Price Competition, Advertising, and the Evolution of Concentration". Cambridge: MIT Press.

VARIAN, H.R. (1980): "A model of sales", *American Economic Review* 70(4) 651-659.

WATSON, R. (2009): Product variety and competition in the retail market for eyeglasses, *The Journal of Industrial Economics*, LVII (2), 217-51.

VIII Appendix A

Data Construction

Demographic Variables

To obtain demographics around each store, we used information from the 2000 U.S. Census.

For each supermarket, we found the census block group (BG)[10] that was closer. Distance here is between the store and the (population weighted) centroid of each BG. The demographic characteristics[11] of the closest BG were taken to be the ones of the store.

All the Variables were taken directly from the 2000 Census except for the Education Index which was constructed in the following way:

$$Education\ Index = \sum_{i=1}^{I} i * w_i$$

where i assigns values to education levels (from 1=No education to 15=Doctorate) and w_i represents the share of population in that BG in the i education category.

Product Variety

The original data has information on each UPC that was sold in each week but not a direct measure of the number of products available to the consumer. I overcome this problem by aggregating the data over a long enough period of time, a quarter in this case (usually supermarket assortments are changed every 3 months). Then, my implicit assumption is the following: if a product is never sold in any of the 13 weeks that compose a quarter, it was not available. This assumption is in line with the assortment literature (...). For each Store/quarter, the variable "Variety" represents the total number of UPCs of which at least one unit was sold.

The second potential concern with the data is the fact that it only includes information about the Beverage Categories. Luckily, a small number of stores in the data have detailed information about *all* the products available to consumers. Using those, one can see that the number of products in the beverage categories ends up being a very good predictor of the total assortment of the store (up to a constant multiple). [add more about this]

[10] This is the closest partition for which the Census has the demographic details that we are interested in.
[11] Median income; education index; median age; and average household Size

Price Index

The idea of this Price Index is to be able to compare prices of different stores in each quarter.

For its construction, only those UPCs that are more widely available are used. Including some products that are sold only in some stores and not in others would complicate the interpretation of the Index. In the end, 164 products were chosen to be included.

The first step was to create the weights that each UPC will have in the final index. The idea is that those products in which consumers spend more money will have a bigger weight. Also, the weights will be constant over time and stores to guarantee that differences in the price index are not a result of weight changes.

So,

$$w_j = \frac{\sum_i \sum_t Revenue_{jit}}{\sum_j \sum_i \sum_t Revenue_{jit}}$$

where $Revenue_{jit}$ is the total revenue of product j at store i in quarter t.

The price series in the original data is in weeks. For each UPC, we created quarterly prices by taking the mean over the 13 weeks of each quarter. The mean is preferred over other possible ways of aggregating weeks over quarters (like mode) because it includes sales prices. If a store has sales more often, it will have lower average quarterly prices. Therefore, it is a better measure of the prices that one could find on a random visit to a store.

Finally, the Price Index is constructed:

$$P_{it} = \sum_j w_j p_{ijt}$$

where P_{it} is the Price Index of store i in quarter t and p_{ijt} is the weekly price.

If a product is not available in a given quarter in a store (which happens infrequently), the weights of the remaining goods are rescaled so that they add up to 1 and keeping the proportions constant.

Sales

According to the definition from Nielsen a sale is a temporary price reduction that is reverted in any of the four subsequent weeks. The algorithm to create the regular non-sale price series from the original average prices is the following:

1) Identify all the weeks where prices dropped

2) Check if there was any price increase in any of the four weeks following that reduction

3) If not, call that price reduction permanent

4) If there is any price increase, call that week a sale and replace the price by the one in the week where it increased.

5) Run the previous four steps four more times to get a smooth enough price series

IX Appendix B: Model Extension

Price informed consumers

I now introduce a positive share of informed consumers in the model described in the main body of the paper. Similar to Varian (1980), the dilemma that firms face when lowering prices is that they lose money on non-informed consumers but attract the informed with a higher probability. Parts of the proof which are similar to Varian (1980) are ommited.

Setting

The only difference with respect to the previous model is the following: out of the N consumers, a share n_I are informed about all the prices at each moment in every store. The remaining $n_U = 1 - n_I$ are aware of the price distribution but not of a specific realization at time t.

Everything else remains the same

Monopolist

Nothing changes for the monopolist solution. Informed and uninformed consumers are similar to the eyes of the monopolist store because all of them will enter the store. When making the actual decision of whether to purchase a good, they already know the prices.

Simillar to the case of $n_I = 0$, the solution to the firm problem is:

$$
\begin{aligned}
\text{for } F &< (v_H - v_L)\, N: \ \Omega^* = (A, B), p^* = (v_H, v_H) \\
\text{for } F &> (v_H - v_L)\, N: \ \Omega^* = A, p_A^* = v_L \text{ or } \Omega^* = B, p_B^* = v_L
\end{aligned}
$$

Duopolist

The existence of a positive number of informed consumers will affect the equilibrium decisions that the firms make. Now, I characterize the symmetric equilibrium.

Proposition 1. Let p_{sH} and p_{sL} denote, respectively, the highest and lowest price available in a given week at store s. Then, $p_{sH} - p_{sL} < v_H - v_L$, $\forall s$.

Proof. The result is immediate for a store with only one product. For a store with $\Omega = (A, B)$, assume the contrary. Then, consumers of both types will purchase the lowest priced good. Now consider reducing p_{sH} to $p'_{sH} = p_{sL} + v_H - v_L - \varepsilon$. Assume, without loss of generality, that A is the good with a high price. Then consumers that value A highly will have the following utility for buying their preferred good: $u_A = v_H - p'_{sH} = v_H - p_{sL} - v_H + v_L + \varepsilon = v_L - p_{sL} + \varepsilon$. That will be higher than the utility of buying product $B : u_B = v_L - p_{sL}$. Since in a symmetric equilibrium with $\Omega_s = \Omega_{-s} = (A, B)$ there will be a positive amount of consumers of both types in each store, this will be a profitable deviation. Increasing p_{sH} increases both the probability of attracting the informed consumers and the revenue made out of each one of them. ■

Corollary 2. Each consumer will buy its preferred product.

Proposition 3. When $n_I > 0$, there is no equilibrium in pure strategies for any product available in both stores

Proof. No price above v_H will be charged since it would attract no consumers. Assume that the prices of a product available in both stores are the same with probability one. Then, if one of the stores lowers the price of all its products by ε, it will capture all the informed consumers of that type and make a profitable deviation. This is true up to the point where both prices equal the marginal cost. But if both prices equal marginal cost, either firm have incentives to raise the prices and sell only to uninformed consumers. Any equilibrium in pure strategies where firms charge different prices for the same product cannot be an equilibrium either. Without loss of generality, assume that firm s charges the lowest price for product A. An increase of $\varepsilon < p_{-sA} - p_{sA}$ in all the prices of firm s will increase the revenues on product A without harming those in product B. ■

Any price equilibrium, in this context, will be a randomization of prices according to some density function $f^*_\Omega(p)$, where $\Omega = (\Omega_s, \Omega_{-s}) = \{A, B, AB\}^2$ and p is the vector of prices

The first step to derive a symmetric equilibrium is to look at the optimal prices for each assortment scenario.

SCENARIO A: $\Omega_s = A$ **;** $\Omega_{-s} = B$

A symmetric equilibrium in this case means that stores will have the same distribution of prices for their products. As such, the uninformed consumers will each go to the store that carries their preferred product.

PROPOSITION 4. If $n_I < \frac{v_H - v_L}{v_L}$, the only equilibrium will have both firms charging prices of v_H

PROOF. First, I am going to show that $p_{sA} = p_{-sB} = v_H$ is an equilibrium and then that it is unique. If both firms charge a price equal to v_H, then each type of consumer (informed or not) will buy its preferred product in the store in which it is available. Clearly, no deviation is profitable for prices above v_H or noone would buy the good. Also, any deviation for prices between v_H and v_L would not be enough to capture new consumers and would lose on the existing ones. The best possible deviation would be the highest price below v_L. That would attract all the informed consumers that were previously buying in the other store and lose the less in the current costumers. With that, total losses would be $(v_H - v_L)\left[\frac{n_U + n_I}{2} * N\right]$ and the gains $v_L\left[\frac{n_I}{2} * N\right]$. A deviation would only be profitable if

$$\begin{aligned} v_L\left[\frac{n_I}{2} * N\right] &> (v_H - v_L)\left[\frac{s_{nU} + n_I}{2} * N\right] \\ v_L\left[\frac{n_I}{2}\right] &> (v_H - v_L)\left[\frac{1 - n_I + n_I}{2}\right] \\ v_L * n_I &> (v_H - v_L) \\ n_I &> \frac{v_H - v_L}{v_L} \end{aligned}$$

Assume that there exists another equilibrium where firm s chooses prices according to some distribution $f_s(p)$. Any equilibrium different from the above has $\int_0^{v_H - \varepsilon} f(p)dp > 0$ for some $\varepsilon > 0$.Let p_s be firm s's minimum price for which $f(p) > 0$. Consider, without loss of generality that $p_s \leq p_{-s}$ (just reverse the argument, otherwise). If $p_{-s} - p_s < v_H - v_L$, informed consumers will split between firms. Then, store s will have no incentives to charge any price below $\min(p_{-s} + v_H - v_L, v_H)$. If $p_{-s} - p_s \geq v_H - v_L$, all the uninformed consumers in the market will buy in store s. Then store $-s$ will want to raise its minimum price to v_H. But at that point, the best possible deviation by firm s is not profitable, as shown before. ■

The equilibrium profit for each firm will then be:

$$\Pi_{A,B}^D\left[v_H, v_H\right] = v_H \frac{1}{2} N$$

SCENARIO B: $\Omega_s = A$ **;** $\Omega_{-s} = (A, B)$

LEMMA 5. If, in equilibrium, $\int_0^{v_L} f(p_{sA}) = 0$, then $p_{-sB} = v_H$

PROOF. In equilibrium, all consumers that prefer B will purchase the good from firm $-s$. So, the profit-maximizing decision is to have the highest price for that product that still induces a purchase. Since, in a symmetric equilibrium, no price for A is lower than v_L, $p_{-sB} = v_H$ is the solution. ■

PROPOSITION 6. If $n_I < \frac{v_H - v_L}{(v_L + v_H)}$ and $p_{-iB} = v_H$, then product A will always be priced above v_L

PROOF. Since there is no equilibrium in pure strategies for product A, the equilibrium will be a price distribution. So, if store s charges a price of v_H for its only product, it will sell only to uninformed consumers and the profit will be $v_H * \frac{n_U}{4} N$. At the minimum price of its equilibrium distribution, it will attract all the informed consumers that prefer product A together with half of the uninformed of that type. Let $\underline{\mathrm{p}}$ denote the minimum price played with positive probability. Then,

$$\Pi^D_{A,AB}\left[\underline{\mathrm{p}}, (p_{-sA}, v_H)\right] = \underline{\mathrm{p}} * \left(\frac{n_U}{4} + \frac{n_I}{2}\right) N$$

for any $p_{-sA} \in f^*(p_{-sA})$

To be part of the equilibrium strategy of store A, it has to be the case that

$$\begin{aligned}
\underline{\mathrm{p}} * \left(\frac{n_U}{4} + \frac{n_I}{2}\right) N &\geq v_H * \frac{n_U}{4} N \\
\underline{\mathrm{p}} * \left(\frac{n_U}{2} + 1 - n_U\right) &\geq v_H * \frac{n_U}{2} \\
\underline{\mathrm{p}} - \underline{\mathrm{p}}\frac{n_U}{2} &\geq v_H * \frac{n_U}{2} \\
\underline{\mathrm{p}} &\geq \left(v_H + \underline{\mathrm{p}}\right) * \frac{n_U}{2} \\
\underline{\mathrm{p}} &\geq \frac{v_H * (1 - n_I)}{(1 + n_I)}
\end{aligned}$$

But since $n_I < \frac{v_H - v_L}{(v_L+v_H)}$, it becomes:

$$
\begin{aligned}
\underline{\mathrm{p}} &\geq \frac{v_H * \left(\frac{v_L+v_H-v_H+v_L}{(v_L+v_H)}\right)}{\left(\frac{v_L+v_H+v_H-v_L}{(v_L+v_H)}\right)} \\
\underline{\mathrm{p}} &\geq \frac{v_H * \left(\frac{2v_L}{(v_L+v_H)}\right)}{\left(\frac{2v_H}{(v_L+v_H)}\right)} \\
\underline{\mathrm{p}} &\geq v_L
\end{aligned}
$$

■

Corollary 7. Under $n_I < \frac{v_H - v_L}{(v_L+v_H)}$, if stores choose $\Omega_s = A$ **;** $\Omega_{-s} = (A, B)$, there is a price equilibrium where $p_{-sB} = v_H$ and product A is always above v_L

Then, prices for prodcut A will follow the cumulative distribution function $F^*(p)$ such that:

$$
\begin{aligned}
v_H * \frac{n_U}{4}N &= p * \frac{n_U}{4}N + (1 - F(p)) * p * \frac{n_I}{2}N \\
v_H * \frac{n_U}{2} &= p * \frac{n_U}{2} + (1 - F(p)) * p * n_I \\
\frac{(v_H - p)}{2p} * \frac{n_U}{n_I} &= (1 - F(p)) \\
F(p) &= 1 - \frac{(v_H - p)}{2p} * \frac{(1 - n_I)}{n_I}
\end{aligned}
$$

That means a density function:

$$
\begin{aligned}
f(p) &= F'(p) \\
f(p) &= -\frac{(1 - n_I)}{n_I}\left[\frac{-2p - 2\,(v_H - p)}{4p^2}\right]
\end{aligned}
$$

and an average price of:

$$
\begin{aligned}
E(p) &= \int_{\underline{p}}^{v_H} p f(p) dp \\
&= v_H - \int_{\underline{p}}^{v_H} F(p) dp \\
&= v_H - \int_{\underline{p}}^{v_H} \left(1 - \frac{(v_H - p)}{2p} * \frac{(1-n_I)}{n_I}\right) dp \\
&= \underline{p} + \frac{(1-n_I)}{n_I} \int_{\underline{p}}^{v_H} \left(\frac{(v_H - p)}{2p}\right) dp \\
& \underline{p} + \frac{(1-n_I)}{n_I} \left(\int_{\underline{p}}^{v_H} \frac{v_H}{2p} dp - \int_{\underline{p}}^{v_H} \frac{1}{2} dp\right) \\
& \underline{p} + \frac{(1-n_I)}{n_I} \left(\frac{v_H}{2} \left[\ln(x)\right]_{\underline{p}}^{v_H} - \left[\frac{1}{2}x\right]_{\underline{p}}^{v_H}\right) \\
& \underline{p} + \frac{(1-n_I)}{2n_I} \left(v_H \ln\left(\frac{v_H}{\underline{p}}\right) - v_H + \underline{p}\right)
\end{aligned}
$$

where the second line comes from integration by parts

Since the minimum price charged, under the previous assumptions is above v_L, then $E(p)$ will also be bigger than v_L.

The expected profit, however, will be independent of the prices charged:

$$
\Pi^D_{A,AB}(p^*) = v_H \frac{n_U}{4} N
$$

and

$$
\begin{aligned}
\Pi^D_{AB,A}(p^*) &= v_H \frac{n_U}{4} N + v_H \left(\frac{n_U}{2} + \frac{n_I}{2}\right) N - F \\
\Pi^D_{AB,A}(p^*) &= v_H \frac{n_U + 2}{4} N - F
\end{aligned}
$$

SCENARIO C: $\Omega_s = \Omega_{-s} = (A, B)$

PROPOSITION 8. $F(v_H - \varepsilon, v_H - \varepsilon) < 1$, for any $\varepsilon > 0$.

PROOF. Suppose that there is a maximum $\hat{p} < v_H$ so that $F(v_H, \hat{p}) = 1$. Now think of a pair $(p_A, \hat{p})$[12] that has positive density in equilibrium. With probability 1, no informed consumer that prefer B will purchase at this store. With probability $1 - F_A(p_A)$ no informed consumer

[12] If $\hat{p}$ is not part of the support, replace it by $\hat{p} - \varepsilon$.

of the other type will do that. Now, consider a deviation to $(p'_A, p'_B) = (p_A + v_H - \hat{p}, v_H)$. The utility differential between both options remains the same so, consumers will still buy their preferred products. Then, at those prices, no informed consumers that prefer B will visit this store (by corollary above together with the fact that p_B in the competitor stores is lower with probability 1). But then, an increase in p_B to $p'_B = \hat{p} + \frac{v_H - \hat{p}}{2}$ will increase the store profits. $\blacksquare$

PROPOSITION 9. If $f(p_A, p_B) > 0$, then $\Pi(p_A, p_B) = \frac{n_U}{2} * N * v_H$

PROOF. At prices "very close" to (v_H, v_H), no informed consumer will be captured. So, all the sales will be to uninformed at a price of v_H. That means profits of $\Pi(v_H, v_H) = \frac{n_U}{2} * N * v_H$. Then, at any other set of prices charged with positive density, the expected profit $\Pi(p_A, p_B)$ has to be the same. If $\Pi(p_A, p_B) < \frac{n_U}{2} * N * v_H$, selling to uninformed will be a profitable deviation from (p_A, p_B). If $\Pi(p_A, p_B) > \frac{n_U}{2} * N * v_H$, that would contradict the previous proposition. $\blacksquare$

PROPOSITION 10. If $n_I < \frac{v_H - v_L}{v_H + v_L}$, only prices above v_L will be charged with positive probability.

PROOF. In equilibrium, by corollary X, every consumer will purchase its preferred good. This means that we can write $\Pi(p_A, p_B) = \Pi_a(p_A) + \Pi_b(p_B)$. The remaining of the proof is the same as in scenario B. $\blacksquare$

Then, the expected profit for each firm is:

$$\Pi_{AB,AB} = v_H \frac{n_U}{2} N - F$$

Now, I can solve for the variety decision of each firm (V_i is the number of products of firm i):

$V_s \backslash V_{-s}$	1	2
1	$(v_H \frac{1}{2} N \ , \ v_H \frac{1}{2} N)$	$(v_H \frac{n_U}{4} N \ , \ v_H \frac{n_U+2}{4} N - F)$
2	$(v_H \frac{n_U+2}{4} N - F \ , \ v_H \frac{n_U}{4} N)$	$(v_H \frac{n_U}{2} N - F \ , \ v_H \frac{n_U}{2} N - F)$

If the competitor chooses only one product, a store will follow if

$$\begin{aligned} v_H \frac{n_U + 2}{4} N - v_H \frac{1}{2} N &< F \\ v_H \frac{n_U}{4} N &< F \end{aligned}$$

If the competitor chooses to carry two products, a store will follow if

$$v_H \frac{n_U}{4} N > F$$

That means that high variety is a dominant strategy if $F < v_H \frac{n_U}{4} N$

Comparing different market structures

The thresholds in F for which stores find it optimal to move from high to low variety are different for monopolist and duopolist stores. If the share of informed consumers is small $\left(s_U > \frac{4(v_H - v_L)}{v_H}\right)$, monopolist stores will switch to low variety at lower values of F. In that case, there will be a range of possible values for F where competition drives both variety and prices up.

The following will be true:

$0 < F < (v_H - v_L) N$: Both duopolist and monopolist markets have high variety. Monopolist prices are higher than duopolist

$(v_H - v_L) N < F < v_H \frac{n_U}{4} N$: Duopolist has high price, high variety. Monopolist has low price, low variety

$v_H \frac{n_U}{4} N < F$: Both duopolist and monopolist markets have low variety. Duopolist prices are higher than Monopolist

X Appendix C: Figures

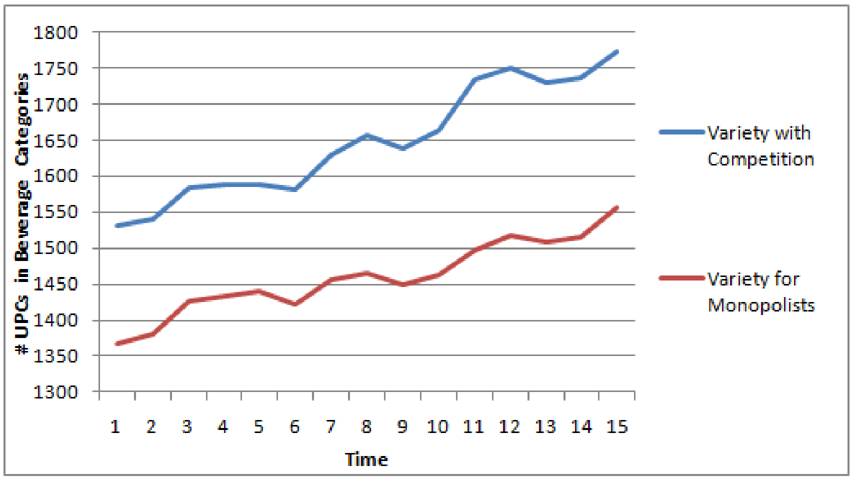

Figure 2: Average Variety per Store type over time

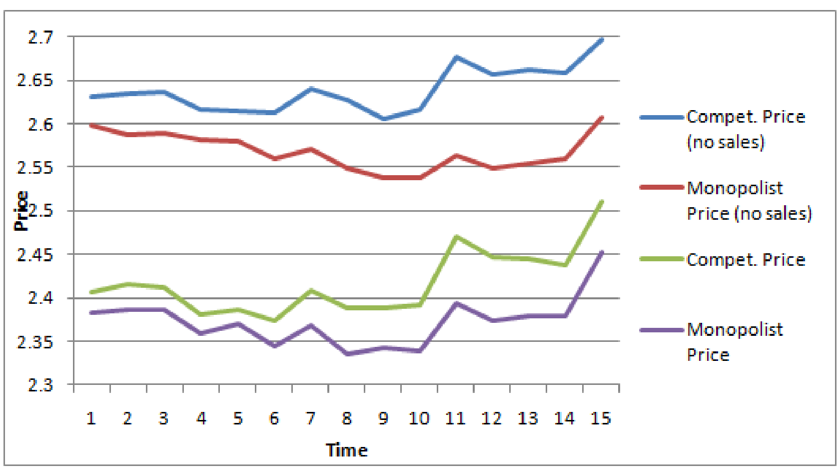

Figure 3: Average Prices per Store type over time

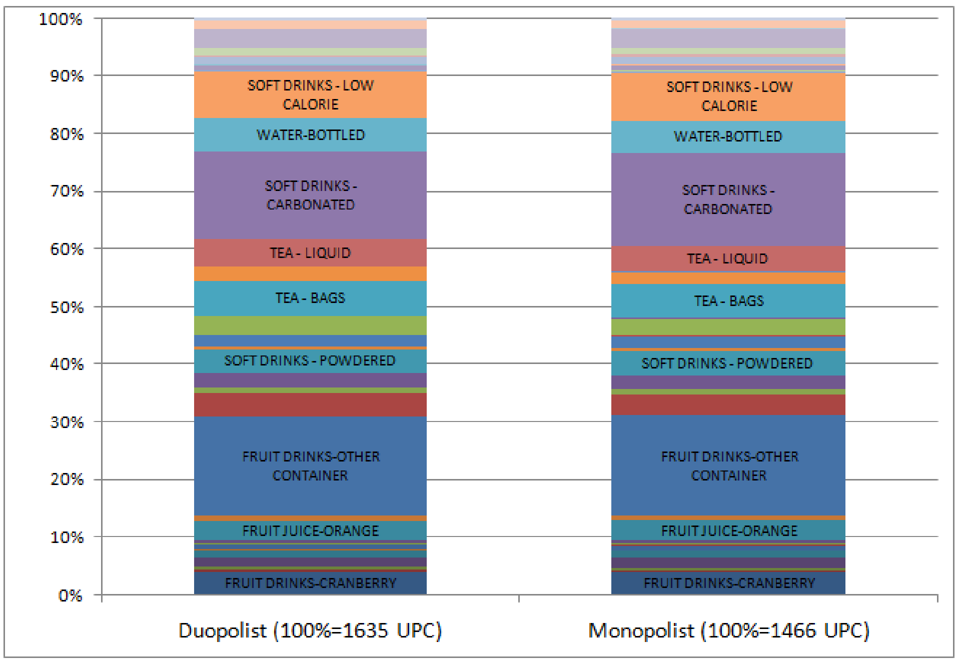

Figure 4: Composition of the assortment of products available in a typical store, by product category

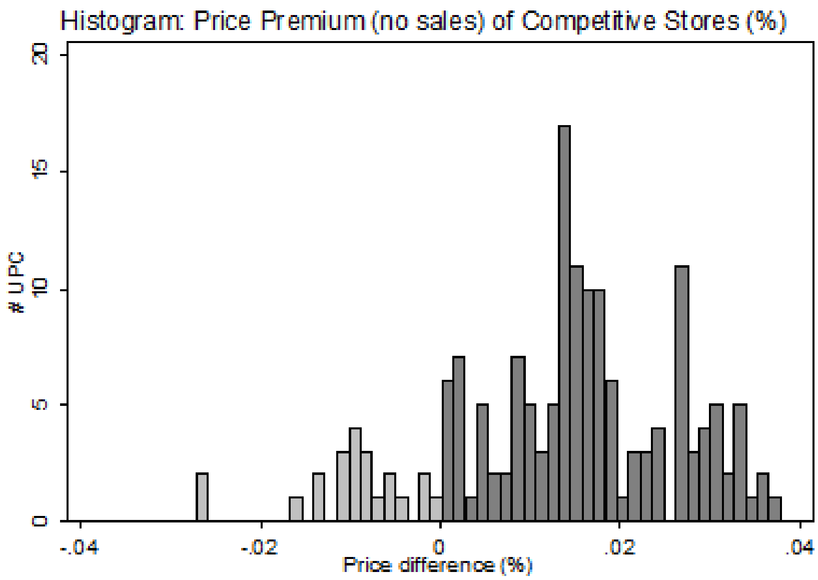

Figure 5: Histogram with the price difference (excluding sales) between monopolist and duopolist stores, for each of the 164 UPCs used. The x-axis is (PriceDuop-PriceMonop)/PriceMonop

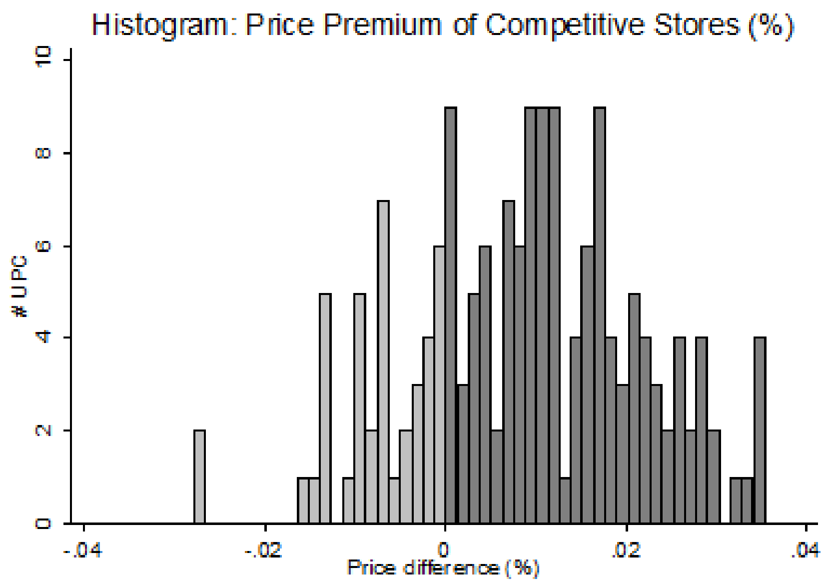

Figure 6: Histogram with the price difference (including sales) between monopolist and duopolist stores, for each of the 164 UPCs used. The x-axis is (PriceDuop-PriceMonop)/PriceMonop

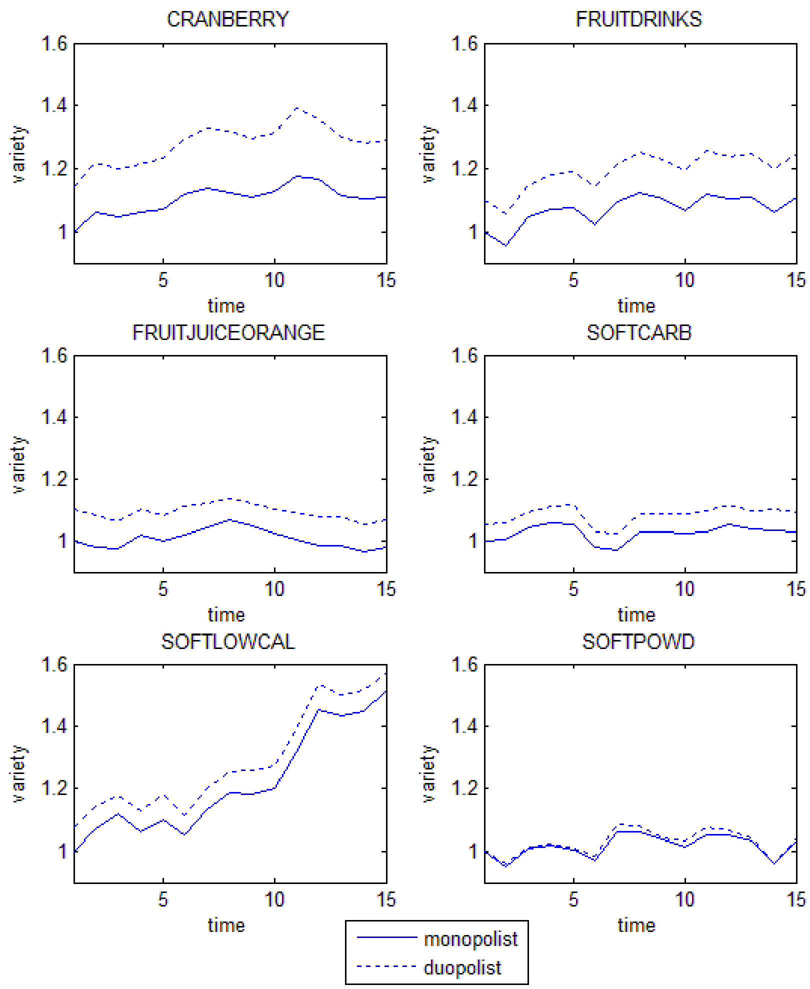

Figure 7: Variety by market structure (for each Product Category)

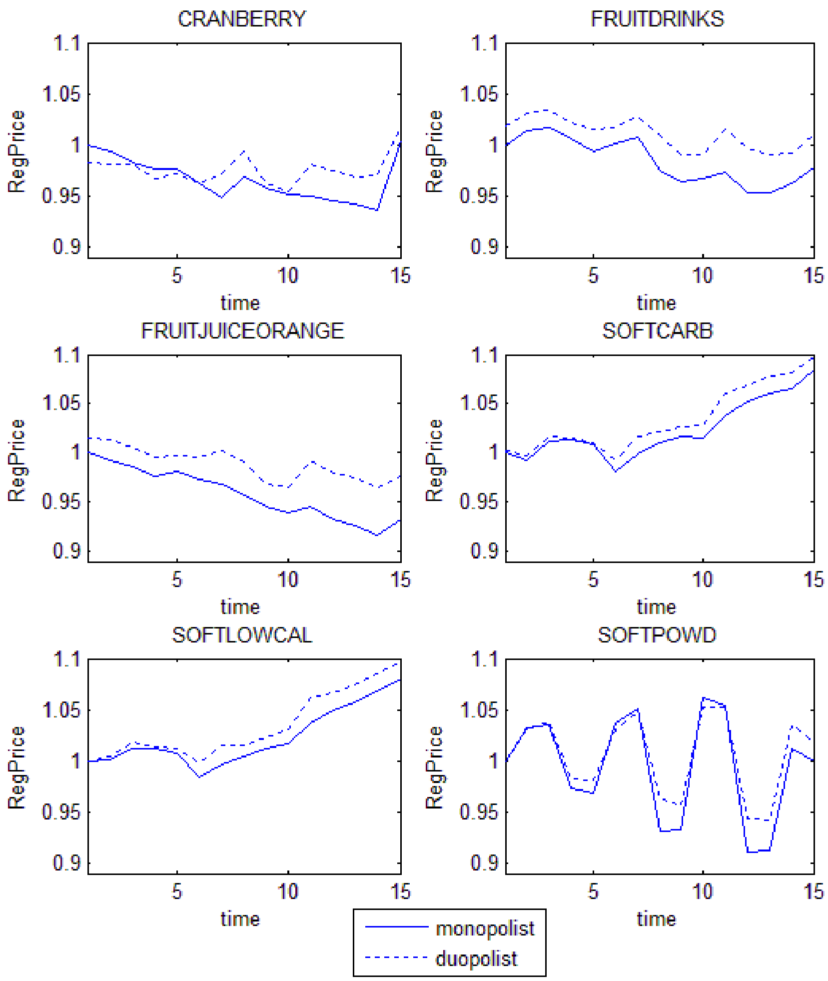

Figure 8: Regular Prices by market structure (for each Product Category)

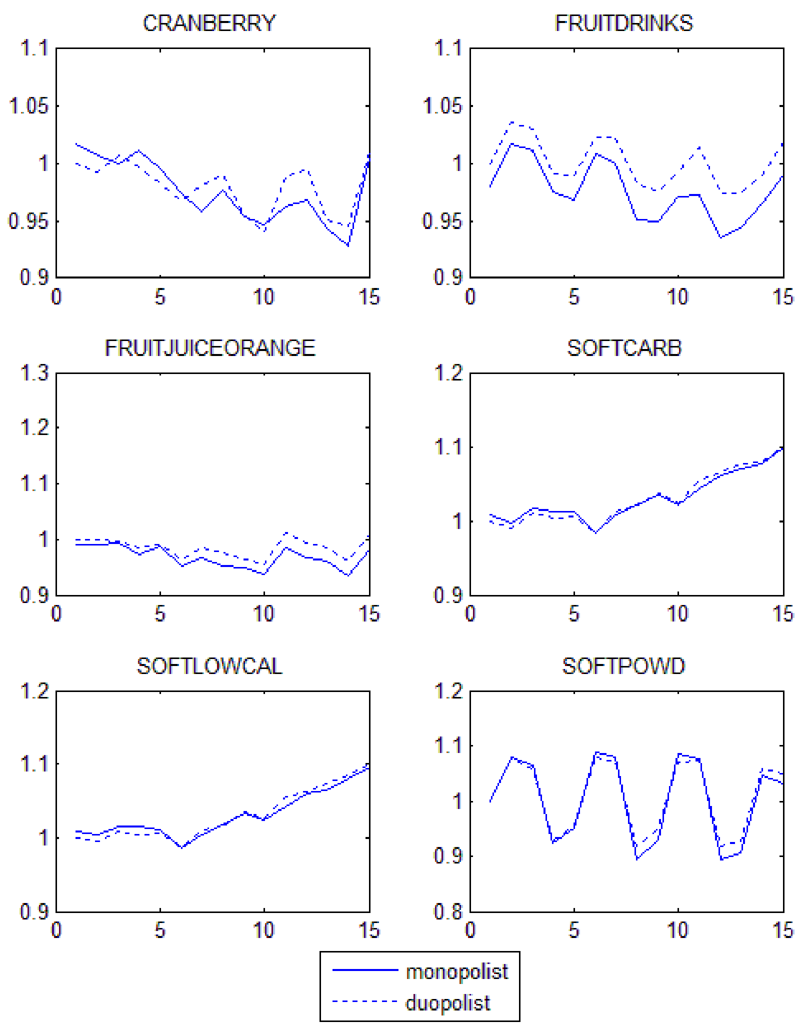

Figure 9: Avg Prices by market structure (for each Product Category)

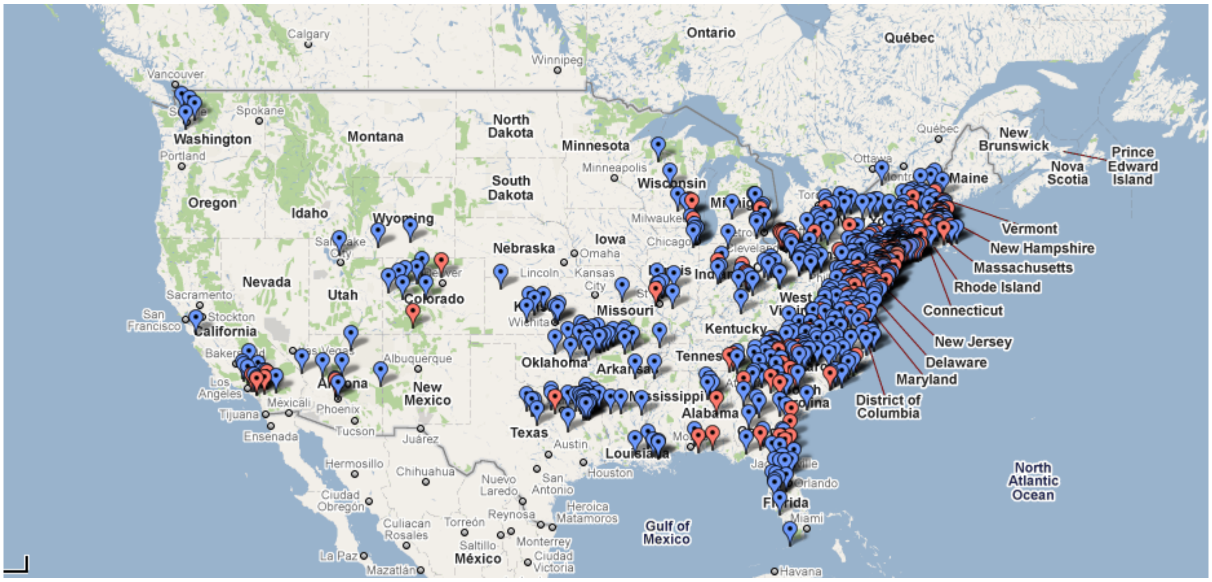

Figure 10: Supermarkets Location

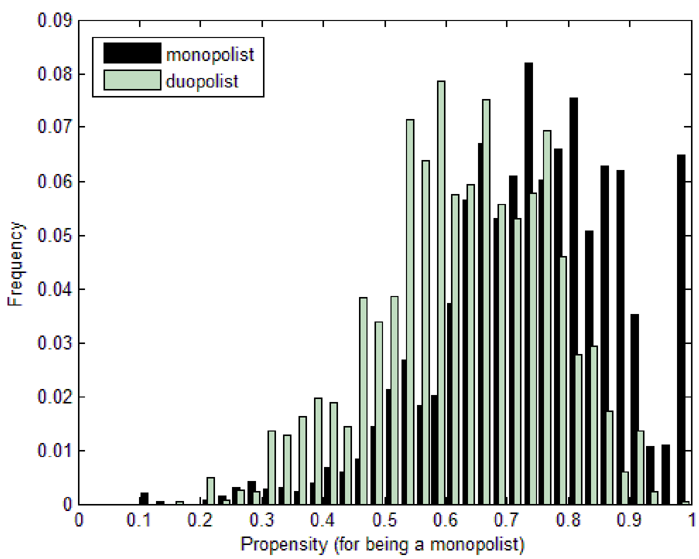

Figure 11: Distribution of Propensity Scores (by market structure)

XI Appendix C: Tables

Mkt type	***POP density***	***Income***	***Education***	***Age***	***Hhsize***	***Count***
Duopolist	1916	44346	8.74	37.54	2.57	121
Monopolist	1208	39467	8.65	38.43	2.50	586

Table 1: *Market level Statistics*
Population density in people per square mile

Monopolist	***variable***	***N***	***mean***	***sd***	***min***	***max***
0	AvgPrice	242	2.42	0.13	1.95	2.69
0	RegPrice	242	2.64	0.18	2.14	2.96
0	NbrProducts*	242	14716	3055	7114	22786
0	Revenues**	242	3012	1794	181	9666
0	SqrFeet	242	33.4	10.3	12	68
1	AvgPrice	592	2.38	0.15	1.94	2.81
1	RegPrice	592	2.57	0.20	2.02	3.02
1	NbrProducts*	592	13111	3420	4760	22440
1	Revenues**	592	2612	1563	418	10116
1	SqrFeet	592	28.9	10.3	6	75

Table 2: *Store level Statistics*
*Total Nbr UPCs (not only Beverage Categories).
**Quarterly, thousand dollars

	Monopolist	***Duopolist***	***(D-M)/M***
Nbr of UPC (Beverages)*	1446	1599	10.58%
Nbr of Brands	257.6	281.2	9.14%
Nbr of Sizes per Brand	2.23	2.21	-0.87%
Nbr of Flavors per (Brand X Size)	2.18	2.24	3.08%
% UPC passing into t+1	0.916	0.914	-0.19%

Table 3: *Breaking variety differences into components*
*Excluding Private Labels

	Monopolist	**Duopolist**	**All Markets**			
Number of Households	3907	1438	5345			
Number of Purchases	171396	71128	242524			
	Mean	*Mean*	*Mean*	*Std. Dev.*	*Min*	*Max*
Number of visits/HH	43.9	49.5	45.4	74.68	1	799
Expenditure($) / visit	45.8	50.6	47.1	62.13	0	1718
Range of days	682.3	706.1	688.8	467.00	1	1826
Visits / week	0.53	0.59	0.55	0.72	0.01	14

Table 4: *Household data descriptive statistics*
The last two lines exclude Households with only one purchase recorded. Range of days is the difference between the last and the first day where the household was recorded

		Annual Household Income		
		Less than \$30k	\$30k - \$70k	More than \$70k
	Households	1011	1583	769
	Total Visits	46435	77797	34977
	Category purchased?			
Household	Dairy	20.4%	22.5%	22.3%
with One or	Orange Juice	9.7%	10.9%	11.4%
Two members	Other Fruit Juices	4.3%	4.9%	4.7%
	Soft Drinks	12.1%	12.1%	12.9%
	Soft Drinks (LowCal)	9.9%	11.6%	14.9%
	Water	4.1%	5.5%	5.6%
	Households	252	887	843
	Total Visits	10533	34982	34977
	Category purchased?			
Household	Dairy	22.8%	28.3%	28.3%
with more than	Orange Juice	9.4%	13.0%	17.1%
Two members	Other Fruit Juices	13.0%	13.5%	13.2%
	Soft Drinks	25.3%	24.5%	20.7%
	Soft Drinks (LowCal)	10.6%	13.6%	15.1%
	Water	5.2%	8.2%	7.3%

Table 5: *Purchase statistics by type of consumer.*
For each Income x Household Size combination, the table presents the number of consumers, number of store visits and share of visits where each product category is purchased

JUICE-ORANGE	***FRUIT DRINKS-OTHER***	***SOFT DRINKS***
TROPICANA	CAPRI SUN	COCA-COLA R
FLORIDAS NATURAL	GATORADE	PEPSI R
MINUTE MAID	MINUTE MAID	MOUNTAIN DEW R
SIMPLY ORANGE	SUNNY D	DR PEPPER R
TREE RIPE	TROPICANA	SPRITE R
PRAIRIE FARMS	TROPICANA TWISTER	PEPSI CAFFEINE FREE R
DOLE	HI-C	CANADA DRY R
DONALD DUCK	WELCHS	VINTAGE R
SEALTEST	HAWAIIAN PUNCH	SIERRA MIST R
HILAND	SNAPPLE	COCA-COLA CAFFEINE FREE R
WATER	***SOFT- LOW CAL***	***DAIRY***
POLAND SPRING	COCA-COLA DT	GARELICK FARMS
PROPEL	PEPSI DT	HOOD
DANNON	PEPSI CAFFEINE FREE DT	LACTAID
AQUAFINA	COCA-COLA CAFFEINE FREE DT	LEHIGH VALLEY DAIRY FARMS
DEER PARK	SPRITE ZERO DT	FARMLAND SPECIAL REQUEST
DASANI	DR PEPPER DT	FIELDCREST
FRUIT2O	SEVEN UP DT	FARMLAND
POCONO SPRINGS	A and W DT	SUNNY DALE
ICE MOUNTAIN	MOUNTAIN DEW DT	HORIZON ORGANIC
NESTLE PURE LIFE	FRESCA DT	OAK TREE

Table 6: *Top 10 brands, by Product Category*

	(1) Variety	(2) lnVariety	(3) lnVariety	(4) lnVariety
monopolist	-115.5***	-0.0850***	-0.0371***	-0.0401***
	(7.415)	(0.00480)	(0.00298)	(0.00439)
Quarter F.E.	Yes	Yes	Yes	Yes
Chain F.E.	No	No	Yes	Yes
Exclude 2 largest chains	No	No	No	Yes
N	10789	10789	10789	6743
R-sq	0.225	0.214	0.693	0.612

Table 7: *Monopoly effect on Variety*
The dependent variable in regressions (1) is in levels and in regressions (2)-(4) in logs. In all the regressions, I control for income, age, household size and population (coefficients not reported)
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1) regprice	(2) lnregprice	(3) lnregprice	(4) lnregprice
monopolist	-0.0363***	-0.0146***	-0.00530***	-0.00853***
	(0.00416)	(0.00160)	(0.000967)	(0.00128)
Quarter F.E.	Yes	Yes	Yes	Yes
Chain F.E.	No	No	Yes	Yes
Exclude 2 largest chains	No	No	No	Yes
N	10789	10789	10789	6743
R-sq	0.157	0.155	0.730	0.758

Table 8: *Monopoly effect on Regular Prices*
The dependent variable in regressions (1) is in levels and in regressions (2)-(4) in logs. In all the regressions, I control for income, age, household size and population (coefficients not reported)
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1) price	(2) lnprice	(3) lnprice	(4) lnprice
monopolist	-0.0212***	-0.00921***	-0.00343***	-0.00663***
	(0.00321)	(0.00134)	(0.000810)	(0.00110)
Quarter F.E.	Yes	Yes	Yes	Yes
Chain F.E.	No	No	Yes	Yes
Exclude 2 largest chains	No	No	No	Yes
N	10789	10789	10789	6743
R-sq	0.150	0.149	0.709	0.723

Table 9: *Monopoly effect on Average Prices*
The dependent variable in regressions (1) is in levels and in regressions (2)-(4) in logs. In all the regressions, I control for income, age, household size and population (coefficients not reported)
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1) lnVariety	(2) lnregprice	(3) lnprice
monopolist	-0.0340***	-0.0117***	-0.00692***
	(0.00456)	(0.00168)	(0.00140)
SqrFeet	0.0129***	0.000743***	0.000580***
	(0.000194)	(0.0000716)	(0.0000596)
Quarter F.E.	Yes	Yes	Yes
N	10789	10789	10789
R-sq	0.443	0.164	0.156

Table 10: *Short Term Effect of Competition*
Regressions include size of the store (in square meters) as a regressor. This table represents the effect of competition, when stores are not allowed to change the retail area. In all the regressions, I control for income, age, household size and population (coefficients not reported)
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

Product Category	***Nbr of UPCs***
FRUIT DRINKS-OTHER CONTAINER	40
SOFT DRINKS - CARBONATED	23
FRUIT JUICE - ORANGE - OTHER CONTAINER	16
SOFT DRINKS - LOW CALORIE	12
SOFT DRINKS - POWDERED	11
FRUIT DRINKS and JUICES-CRANBERRY	11
VEGETABLE JUICE AND DRINK REMAINING	8
TEA - BAGS	6
WATER-BOTTLED	5
TEA - HERBAL BAGS	4
FRUIT JUICE - GRAPE	4
REMAINING DRINKS and SHAKES-REFRIGERATED	3
DAIRY-FLAVORED MILK-REFRIGERATED	3
FRUIT JUICE-REMAINING	3
FRUIT DRINKS and MIXES - FROZEN	3
FRUIT JUICE - PINEAPPLE	2
FRUIT JUICE - ORANGE - FROZEN	2
DAIRY-MILK-REFRIGERATED	2
FRUIT JUICE - LEMON/LIME	1
FRUIT JUICE - GRAPE - FROZEN	1
FRUIT JUICE - GRAPEFRUIT - OTHER CONTAINERS	1
TEA - MIXES	1
FRUIT JUICE - UNCONCENTRATED - FROZEN	1
FRUIT JUICE-PRUNE	1

Table 11: *Number of UPCs, by Nielsen Product Category, included in the price Index*

	Distance of market to closest outside store			
	Average distance (miles)	Fraction of observations		
		Dist<2	2<Dist<5	5<Dist
Monopolist	4.79	0.34	0.30	0.37
Duopolist	2.7	0.58	0.28	0.14

Table 12: *Distance of markets to the closest "outside" store*

	(1)	(2)	(3)	(4)	(5)	(6)
	lnVariety	lnVariety	lnregprice	lnregprice	lnprice	lnprice
monopolist	-0.0591***	-0.220***	-0.0113***	-0.0610***	-0.00713***	-0.0421***
	(0.00764)	(0.0197)	(0.00240)	(0.00697)	(0.00201)	(0.00609)
Closest outside store	>2 miles	>10 miles	>2 miles	>10 miles	>2 miles	>10 miles
N	6769	962	6769	962	6769	962
R-sq	0.243	0.243	0.128	0.217	0.120	0.227

Table 13: *Monopoly Effect restricted to completely isolated cities*
Dropped from the sample markets where the closest outside store is further away than X miles (X=2,10) In all the regressions, I control for income, age, household size and population (coefficients not reported)
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1) monopolist	(2) monopolist
Population	0.0000149*	0.00000880
	(0.00000664)	(0.00000674)
Population^2	4.43e-11	7.62e-11*
	(3.80e-11)	(3.89e-11)
log(Population)	-1.224***	-1.110***
	(0.0880)	(0.0888)
Income	-0.0000543***	-0.0000544***
	(0.0000131)	(0.0000131)
Income^2	2.87e-10***	2.79e-10***
	(5.83e-11)	(5.84e-11)
log(Income)	0.207	0.272
	(0.302)	(0.300)
Education	-0.0252	-0.106
	(0.197)	(0.199)
Education^2	0.0107	0.0149
	(0.0116)	(0.0116)
Median Age	0.0133**	0.0117*
	(0.00510)	(0.00516)
Household Size	0.0134	-0.0589
	(0.328)	(0.327)
Household Size^2	-0.0715	-0.0564
	(0.0679)	(0.0678)
SqrFeet		-0.0316***
		(0.00306)
Quarter F.E.	X	X
Chain F.E.	X	X
N	9077	9077
Log Likelihood	-4537.6	-4482.8

Table 14: *First Stage regression to Create Propensity Score*
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnVariety	Unmatched	7.266	7.397	-0.1320	0.0059	-22.34
	ATT	7.266	7.363	**-0.0974**	0.0066	-14.71

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnregprice	Unmatched	0.947	0.969	-0.0223	0.0018	-12.49
	ATT	0.947	0.961	**-0.0142**	0.0022	-6.58

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnprice	Unmatched	0.866	0.881	-0.0149	0.0015	-9.95
	ATT	0.866	0.875	**-0.0091**	0.0018	-5

Table 15: *Propensity Score Matching* (using normal density Kernel for the weights)

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnVariety	Unmatched	7.266	7.397	-0.1320	0.0059	-22.34
	ATT	7.266	7.353	**-0.0874**	0.0067	-13.09

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnregprice	Unmatched	0.947	0.969	-0.0223	0.0018	-12.49
	ATT	0.947	0.959	**-0.0121**	0.0022	-5.54

Variable	*Sample*	*Treated*	*Controls*	*Difference*	*S.E.*	*T-stat*
lnprice	Unmatched	0.866	0.881	-0.0149	0.0015	-9.95
	ATT	0.866	0.873	**-0.0069**	0.0018	-3.75

Table 16: *Propensity Score Matching with store size* (using normal density Kernel for the weights). Includes Store Size in the Covariates used for the Propensity Score

Propensity Score	**Variable**	**Monopolist**	**Duopolist**	**Difference**	**N**
[0 , .25]	lnVariety	7.057	7.278	**-0.221**	44
[.25 , .5]	lnVariety	7.439	7.418	**0.021**	808
[.5 , .75]	lnVariety	7.393	7.418	**-0.025**	4480
[.75 , 1]	lnVariety	7.173	7.313	**-0.140**	3775
[0 , .25]	lnregprice	1.025	0.904	**0.121**	44
[.25 , .5]	lnregprice	0.985	0.977	**0.008**	808
[.5 , .75]	lnregprice	0.967	0.976	**-0.008**	4480
[.75 , 1]	lnregprice	0.924	0.952	**-0.029**	3775
[0 , .25]	lnprice	0.950	0.841	**0.109**	44
[.25 , .5]	lnprice	0.893	0.887	**0.007**	808
[.5 , .75]	lnprice	0.880	0.886	**-0.005**	4480
[.75 , 1]	lnprice	0.851	0.872	**-0.022**	3775

Table 17: *Monopolist effect by propensity score quartil*
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1) lnVariety	(2) lnregprice	(3) lnprice
Monopolist	-0.0939***	-0.0178***	-0.0112***
	(0.00568)	(0.00177)	(0.00148)
Same Chain	-0.0555***	-0.0199***	-0.0121***
	(0.0122)	(0.00380)	(0.00317)
Quarter F.E.	Yes	Yes	Yes
N	10789	10789	10789
R-sq	0.216	0.157	0.150

Table 18: *Monopolist effect for same chain competitors* I including a dummy for stores that share a city with other stores of the same chain
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1)	(2)	(3)
	lnVariety	lnregprice	lnprice
Monopolist	-0.0164*	-0.00924**	-0.00674*
	(0.00637)	(0.00331)	(0.00278)
Store Size	0.00600***	0.000231*	0.000159*
	(0.000176)	(0.0000914)	(0.0000769)
Market F.E.	X	X	X
Quarter F.E.	X	X	X
N	10789	10789	10789
R-sq	0.266	0.031	0.139

Table 19: *Impact within market of a change in structure over time*
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

	(1)	(2)	(3)
	lnVariety	lnregprice	lnprice
Monopolist	-0.0311**	-0.0296***	-0.0204***
	(0.00963)	(0.00791)	(0.00609)
Market F.E.	X	X	X
Quarter F.E.	X	X	X
N	189	189	189
R-sq	0.740	0.490	0.660

Table 20: *Impact for stores where change in structure was reversed* Impact within market of a change in structure over time - only using stores where that change was reversed. Since there is only one store per market, "Market F.E." end up being "store F.E.". As such, store size was dropped from the regression
Standard errors in parentheses (*p<0.05 **p<0.01 ***p<0.001)

HH income	*Less than \$30k*		*\$30k - \$70k*		*More than \$70k*	
HH size	*1 or 2*	*more than 2*	*1 or 2*	*more than 2*	*1 or 2*	*more than 2*
	(1)	(2)	(3)	(4)	(5)	(6)
	Purchase	Purchase	Purchase	Purchase	Purchase	Purchase
Price	-8.281***	-10.65***	-9.032***	-8.104***	-8.070***	-9.860***
	(0.421)	(1.026)	(0.272)	(0.388)	(0.386)	(0.304)
N	83258	15101	172309	55092	82023	90763
Log-Likelihood	-4203.6	-817.3	-9744.3	-4407.4	-4499.6	-6967.9

Table 21: *Parameter Estimates for the Category Choice Model: Orange Juice.* Purchase decision is regressed on a price index, a dummy variable for each of the top 10 brands and another one for all the smaller brands. Only the coefficient on prices is reported. Similar models for the remaining categories were estimated (not reported)

	(1)	(2)
	Store visit	Store visit
$\theta_{OrangeJuice}$	-0.716	-0.406
	(0.371)	(0.554)
$\theta_{OtherFruitJuices}$	16.48***	20.32***
	(1.066)	(1.437)
$\theta_{SoftDrinks}$	3.679***	7.623***
	(0.761)	(0.897)
θ_{Water}	2.746***	-1.769
	(0.719)	(1.156)
$\theta_{SoftLowCal}$	5.589***	2.887
	(1.073)	(1.567)
θ_{Dairy}	0.380**	0.865***
	(0.122)	(0.154)
Chain F.E.	No	Yes
N	59166	59166
Log-Likelihood	-20026.9	-19093.7

Table 22: *Maximum-likelihood estimation of the Store Choice model*
The parameter θ_X is the coefficient on the expected value of category X
Standard errors in parentheses * $p<0.05$ ** $p<0.01$ ***$p<0.001$

	Duopolist (real)	Monopolist (counterfactual)
All Consumers	2.97	2.83
Low Inc, small size	1.75	1.65
Low Inc, big size	5.02	4.83
Med Inc, small size	1.79	1.68
Med Inc, big size	4.72	4.52
High Inc, small size	2.36	2.24
High Inc, big size	5.04	4.84

Table 23: *Consumer Welfare in each type of market*