Does Sutton apply to supermarkets?

Paul B. Ellickson*

This paper presents empirical evidence that endogenous fixed costs play a central role in determining the equilibrium structure of the supermarket industry. Using the framework developed in Sutton (1991), I construct a model of supermarket competition where escalating investment in firm level distribution systems is driven by the incentive to produce a greater variety of products in every store. Employing a store level census and 51 distinct geographic markets, I demonstrate that the supermarket industry is a natural oligopoly in which a small number of firms (between 4 and 6) capture the majority of sales, regardless of market size.

1 Introduction

■ In many retail industries, the most successful firms are the ones that offer the widest selection. Firms such as Home Depot, Staples, and Crown Books became household names by offering a staggering array of products at competitive prices. Maintaining this variety requires substantial firm level investments. Successful "category killers" employ state of the art distribution systems in which advanced logistical software, optical scanners, and specialized sorting equipment allow products to flow from warehouse to store in a nearly continuous process. Store level orders are filled automatically through paperless supply chains. The emphasis on variety and the requisite fixed investments yield tightly contested markets among a handful of rival chains. The fact that these markets do not fragment as market size increases, suggests investment in distribution may scale up with the size of the market, as in Sutton's (1991) model of endogenous fixed costs (EFC). This paper proposes an EFC model of retail competition in which escalating investments in variety enhancing distribution systems yields a natural oligopoly of high quality firms. The model is evaluated empirically using a detailed census of the supermarket industry.

[0]* Duke University; paul.ellickson@duke.edu.

This paper is based on Chapter 2 of my Ph.D. dissertation. My advisors, Susan Athey, Scott Stern, and Richard Schmalensee provided invaluable guidance and advice. I would like to express my gratitude to the Editor and two anonymous referees for their extremely thoughtful comments and to the National Science Foundation, grant SES-0517889, for financial support. I also thank Arie Beresteanu, Elizabeth M. Bailey, Bryan Ellickson, and seminar participants at Berkeley, Chicago, Harvard, Stanford, UCLA, and Yale for many useful comments.

The explanation for why we observe a natural oligopoly among supermarkets is based on Shaked and Sutton's (1987) claim that "entry in certain industries is limited to a small number of firms, not because fixed costs are so high relative to the size of the market, but rather because the possibility exists, primarily through incurring additional fixed costs, of shifting the technological frontier constantly forward towards more sophisticated products." Such continuous innovation is not always feasible. In many cases, the level of fixed costs is determined exogenously and the market share of every firm shrinks as the market expands. Using a model of horizontal differentiation, the authors demonstrate that such fragmentation occurs under quite general conditions: when consumers differ in their preferred choice of product, industries are naturally fragmented. However, when consumers agree in their ranking of products, this result can be broken.

Building on the earlier vertical differentiation literature, Shaked and Sutton demonstrate that when the burden of quality improvement falls predominantly on fixed rather than variable costs, the number of entrants may be completely unaffected by the size of the market. This extreme failure to fragment follows from the fact that (with this unique combination of technology and tastes) market expansions lead existing firms to escalate their fixed costs. Instead of having more firms, larger markets have better products.

To establish how this can occur in retail, I adapt Sutton's (1991) model of advertising to account for some specific features of supermarket competition. In this model, supermarkets compete for customers by offering a greater variety of products in every store, requiring a fixed investment in distribution. Serving a larger share of the market requires building more stores, which is naturally constrained by the price of land. Expanding variety requires building larger stores and more advanced distribution systems. Because variety is a purely vertical form of product differentiation, firms that fail to match the quality increases of their rivals cannot survive. Therefore, as markets grow, existing firms must incur higher costs if they are to remain in the industry, and this escalation in costs discourages entry by other firms. Consequently, markets both large and small are served by roughly the same small number of high quality chains.

However, this natural oligopoly result holds only when product differentiation is purely vertical, a particularly strong assumption in retailing, where spatial differentiation plays so prominent a role. Fortunately, Shaked and Sutton are able to demonstrate that their non-fragmentation result extends, in a somewhat weaker form, to settings with both types of differentiation. In the most general setting, the vertical dimension continues to drive market structure. In particular, so long as firms can achieve a minimal increase in quality through primarily fixed investments, at least one firm will retain a strictly positive market share, regardless of market size. Simply placing bounds on the outcomes that can be sustained as equilibria, this lack of precision is a product of the theory's breadth. The non-fragmentation result is robust to alternative assumptions regarding the form of product market competition, the timing of moves, and whether the players are single or multi-product firms. This breadth is particularly comforting in a setting as complex as retail competition.

The empirical goal of this paper is to demonstrate that both the weak and strong forms of non-fragmentation apply to supermarkets. Both exercises require identifying a suitable set of reasonably distinct markets. Since many category killers operate national chains, empirically distinguishing the endogenous fixed cost model from simple scale economies is not always feasible. However, due largely to the constraints of distributing perishable products, supermarkets actually compete in relatively distinct geographic markets. After demonstrating that this is the case, I use these 51 distribution markets to establish the most general non-fragmentation result. Concentration is indeed bounded below. A comparison with an industry that fragments (barber shops and beauty salons) provides a sharp contrast.

The uniformity of structure in the supermarket industry extends beyond the largest chain in each market. Using the full set of firms, I demonstrate that the industrial organization of each of the 51 spatially defined distribution markets is a natural oligopoly in which a small number of firms (between 4 and 6) capture the majority of sales, independent of market size. The number of firms does scale up with the size of the market, but the expansion is limited to a fringe of firms operating stores of much lower quality than those of the oligopolists. These results accord well with the purely vertical model of supermarket competition, which is easily extended to accommodate two tiers of firms. While not strictly a natural oligopoly, the supermarket industry is remarkably close to one.

Although this is the first study to use an EFC model to explain competition between retail firms, subsequent authors have adopted a similar approach in a number of settings. Dick (2007) applies the methodology developed here (and in Ellickson (2000)) to the banking industry. She finds evidence of endogenous investment in branch quality among large regional banks. Berry and Waldfogel (2003) examine the newspaper and restaurant industries and Latcovich and Smith (2001) study online book sellers.

The paper is organized as follows. Section 2 provides a brief history of the supermarket industry, arguing that endogenous investments have played a central role throughout its evolution.

Section 3 presents a vertical model of supermarket competition. Sections 4 and 5 describe the dataset and demonstrate how distribution networks can be used to identify distinct geographic markets. The empirical results on non-fragmentation and natural oligopoly are presented in Section 6. Section 7 concludes.

2 Distribution and the Evolution of Supermarkets

The evolution of the supermarket industry is marked by three major innovations: the rise of chain grocery stores in the early 1900's, the introduction of the supermarket format circa 1950, and the adoption of technology intensive distribution systems in the 1980's and 1990's. Each innovation involved an expansion in fixed investments that favored larger firms. While chain stores capitalized on economies of scale in production and procurement, the supermarket format introduced economies of scope at the level of the store. By placing a premium on coordination and automation, modern distribution systems solidified the link between firm and store.

The rise of the chain stores, led by the Atlantic and Pacific Tea Company (A&P), was primarily driven by the efficiencies of integrating backward into both wholesaling and manufacturing. Although the chains operated stores that were physically identical to their rivals, their prices were dramatically lower.[1] By manufacturing many of their own products and distributing them through their own vertical networks, the chain stores eliminated several layers of middlemen, substantially lowering per-unit costs (Adelman, 1958). Quantity discounts from major suppliers were passed on to consumers through lower prices. The escalating levels of investment quickly led to a shakeout, as the smaller "independent" grocers could no longer compete. As the chain store format spread, concentration rose sharply. Between 1919 and 1932, the share of the top 5 firms in the U.S. increased from 4.2% to 28.8% of total grocery sales (Tedlow, 1990), with A&P alone accounting for 16.3%. The independents responded by forming wholesale cooperatives that matched the scale of the chain format and, by the late 1930s, prices had begun to equalize across the two segments. However, before the independent grocers could fully recover, the introduction of the supermarket permanently shifted the balance of power back to the chains.

This rise of nationally branded products and the diffusion of the automobile in the 1940s and 1950s created a natural incentive to build larger, less centrally located stores. Situated in the suburbs to economize on land costs, supermarkets were 5 times larger than a typical grocery store, carried far more products, and required customers to serve themselves (Tedlow, 1990).

[1] Several price studies performed in the late 1920s and early 1930s found that chain store prices were 4.5-14% lower than their independent counterparts (Tedlow, 1990).

Furthermore, these new stores were heavily advertised, adding a second dimension of firm level investment that wholesale cooperatives and third-party distributors were slow to replicate. Supermarkets now competed on the basis of both price and *variety*, linking stores to firms through both advertising and distribution. Although advertising certainly helped drive the diffusion of the supermarket format in the 1950s and 1960s, it was the new focus on product variety that distinguished the dominant supermarket firms from the competitive fringe of independent grocers. Supermarkets were now vertically differentiated.

The introduction of computerized logistical and inventory management systems in the 1980s allowed chain supermarkets to stock an ever expanding array of products. The explosion in both product variety and store size in the supermarket segment is striking. The number of products offered per store increased from 14,145 in 1980 to 21,949 in 1994 (Messinger and Narasimhan, 1995) to over 30,000 by 2004 (Food Marketing Institute). To accommodate the greater brandwidth, store size has increased an average of 1,000 square feet per year for the past three decades.[2] Every major supermarket firm invests in proprietary information technology and logistical systems aimed at increasing variety while minimizing storage and transportation costs. Pioneered by mass-merchandisers such as Wal-Mart, these information intensive procurement and distribution system require a high degree of coordination, both in terms of scheduling and technology adoption, that is difficult to achieve through an arm's length contract.[3] The potential savings from integration are significant. According to a 1997 report by the Food Marketing Institute (FMI), operating costs are 25 to 60 percent lower for self-distributing chains. In 1998, 49 of the top 50 supermarket firms were vertically integrated into distribution,[4] operating state of the art, climate controlled warehouses and specialized trucking fleets. Although independent grocers (and third-party wholesalers) continue to capture around 25% of overall food sales, they do not compete directly with the dominant supermarket chains, focusing instead on smaller niches like ethnic foods and rural locations.

The physical requirements of distribution also determine the level at which firms compete, providing a natural geographic market definition. Due to the historical importance of rail transport and the short distances over which perishable goods can be shipped, supermarket distri-

[2] Modern distribution technology may also favor larger stores by allowing products to be replenished at a faster rate. Holmes (2001) argues that faster inventory cycles lead to more frequent but smaller deliveries which, given the fixed costs of loading a truck, increase the number of products per delivery (and hence, efficient store size).

[3] For example, even getting independent grocers to implement a standardized scanner system can be a major hurdle for third party wholesalers (Schiano, 1996).

[4] The only top 50 firm that utilized a private wholesaler was K-Mart, which was in the process of converting its dry goods superstores to supercenters that included groceries. The firm has since filed for bankruptcy protection.

bution areas are relatively distinct and well defined. Most firms cluster their warehouses in 50 major cities (near a rail head) and serve surrounding markets via overlapping hub and spoke networks. For example, all of the major Southern California supermarkets chains operate warehouses in East Los Angeles, serving markets as far away as San Diego and Las Vegas from the same cluster of facilities. While the radius of operation of a typical distribution center varies across regions[5], the patterns are remarkably consistent across firms within regions. Indeed, the leading marketing directory for the supermarket industry (Trade Dimension's *Marketing Guidebook*) is organized into 52 distribution areas, each roughly the size of a state.[6] Within each distribution area, firms make strategic investments in distribution and store size that allow them to stock the widest array of products at the lowest possible prices. The following model is motivated by the importance of these investments.

3 A Pure Vertical Model of Supermarket Competition

■ This section adapts Sutton's (1991) advertising model to retail chains. Apart from assuming that firms compete in distinct distribution areas, the setup remains purely vertical. The only substantive changes to Sutton's formulation involve the cost function, which has been modified to reflect the distinction between building larger stores and expanding the size of the chain and to include the price of land.

Within any distribution area, supermarket chains are vertically differentiated, differing only in their level of quality z, which represents the brandwidth or variety provided in each of their stores. I assume that $z \geq 1$, with $z = 1$ representing the minimum level of quality and that higher quality, prices held fixed, appeals to all consumers. There are M identical consumers with income Y and utility

$$u(x_1, x_2, z) = (1 - \alpha) \ln(x_1) + \alpha \ln(z x_2) \tag{1}$$

where x_1 is the quantity of the composite commodity and x_2 is the quantity of groceries with quality z. Let x_1 be numeraire and $p(z)$ denote the price of groceries.

A supermarket firm producing quantity q_j of quality z_j has cost function

$$C_j = p_L \sigma + \frac{\lambda p_L}{\gamma}(z_j^{\gamma} - 1) + c q_j \tag{2}$$

[5] Markets have a radius of roughly 100-150 miles in the western U.S. versus 50-100 miles in the east, reflecting the substantial geographic variation in population density and transportation infrastructure across the U.S.

[6] My own analysis, described later, yielded 51 markets.

where p_L is the price of land and c is marginal cost. The parameters σ and λ are strictly positive and $\gamma > 1$. Marginal cost (c) is given by

$$c = \phi_1 w + \phi_2 p_g + \phi_3 p_L \tag{3}$$

where the parameters $\phi_i > 0$, w is the wage, and p_g is the cost of goods sold. Applying Shephard's Lemma to the cost function, the demand for land by a firm producing quantity q_j of quality z_j is

$$h_L = \sigma + \frac{\lambda}{\gamma}(z_j^{\gamma} - 1) + \phi_3 q_j \tag{4}$$

Holding quality z_j fixed, the demand for land increases linearly in q_j since serving additional consumers requires building more stores. At the minimal level of quality ($z_j = 1$), the middle term in both (2) and (4) drops out, so that fixed costs are exogenous. However, as the firm improves quality, fixed costs escalate and the demand for land increases as the size (rather than the number) of stores expands. As shown below in equation (6), equilibrium quality z depends inversely on the price of land, so high land prices can dampen quality improvement if the required expansion of store size is too expensive.

Competition is modeled as a three-stage game. In the first stage, firm j chooses whether or not to enter the market, incurring entry cost $p_L\sigma$. In the second stage, each firm chooses a level of quality z_j, incurring the additional fixed cost $\frac{\lambda p_L}{\gamma}(z_j^{\gamma} - 1)$. In the third and final stage, firms compete in the product market, which is modeled as Cournot. Following Sutton, I assume that the parameter γ is large enough to guarantee a symmetric equilibrium in both quantity and quality.[7] Solving the game by backwards induction yields equilibrium quantity and price

$$q = \left(\frac{N-1}{N^2}\right)\frac{S}{c} \quad \& \quad p(z) = \left(\frac{N}{N-1}\right)c \tag{5}$$

and equilibrium quality

$$z = \left(\frac{2S(N-1)^2}{N^3 \lambda p_L}\right)^{\frac{1}{\gamma}} \tag{6}$$

where $S \equiv \alpha Y M$ is the size of the market (total revenue) and N is the number of firms who have entered at stage one. While fixed costs clearly expand with the size of the market, the level of quality depends inversely on the price of land, which can constrain its growth.

All that remains is to determine the number of entrants. Ignoring integer constraints, free entry will drive profits to zero, yielding the zero profit condition

[7] In this version of his model, Sutton's condition for a symmetric equilibrium becomes $\gamma > \max\left\{1, \frac{2}{3}\frac{\lambda p_L}{\sigma}\right\}$.

$$\left(\frac{\lambda p_L - \gamma p_L \sigma}{S}\right) N^3 = 2N^2 - (4+\gamma)N + 2 \tag{7}$$

The right hand side (RHS) equation (7) is a convex quadratic function with root[8]

$$N^+ = 1 + \frac{\gamma}{4} + \frac{1}{4}\sqrt{8\gamma + \gamma^2}$$

The left hand side (LHS) of equation (7) is a cubic that passes through the origin. The shape of this cubic determines the asymptotic behavior of N, yielding two possible outcomes[9]:

1. If $\lambda - \gamma\sigma < 0$, then the LHS has a concave graph and lies below the horizontal axis. As shown on the left side of Figure 1, the equilibrium number of firms N^* lies in the interval $(0, N^+)$. Because the slope of the LHS decreases (in absolute value) as S increases, the equilibrium number of firms increases as market size increases. This effect can be offset to a greater or lesser extent by an increase in the price of land as market size expands.

2. If $\lambda - \gamma\sigma > 0$, then the LHS has a convex graph and lies above the horizontal axis. Because the slope of the LHS decreases as S increases, this case has the somewhat counter-intuitive implication that the equilibrium number of firms will decrease as market sizes expands, an effect that will be reinforced if land prices also increase. This case is illustrated on the right side of Figure 1.

In either case, it is possible to be at a corner solution where $z = 1$ for small values of S, corresponding to a situation in which the LHS and RHS intersect for $N^* \leq 1$ or not at all.[10] As we will see in the data, the empirically relevant case appears to be the first one, where the number of firms increases and then flattens out, tempered by an increase in the price of land.

4 Data

■ The data for the supermarket industry are drawn from Trade Dimension's Retail Tenant Database for 1998. Trade Dimensions collects store level data from every supermarket operating in the U.S. for use in their *Marketing Guidebook* and *Market Scope* publications, as well as selected issues of *Progressive Grocer* magazine. The data are also sold to marketing firms and food manufacturers for marketing purposes. Their establishment level definition for a supermarket is

[8] The second root of this quadratic is always less than 1.

[9] When $\lambda - \gamma\sigma = 0$, the number of entrants is identically N^+.

[10] See Sutton (1991) for a more complete discussion of these comparative statics.

the government and industry standard: a store selling a full line of food products and generating at least $2 million in yearly revenues. Foodstores with less than $2 million in revenues are not included in the dataset.[11]

Each store is assigned a unique identifying code. For each store in the database, its place in the organizational hierarchy is documented, indicating the operating name, ultimate parent company, the number of stores in the chain, and the principal supplier. When the firm is vertically integrated into distribution, the warehouse locations and, in some cases, their sizes are provided. Warehouse locations are also provided for third-party distributors. At the store level, Trade Dimensions collects information on average weekly volume, store size, number of checkouts, number of full and part time employees, whether scanners are in operation, and the presence or absence of various service counters (e.g. deli, seafood) as well as other measures of quality (e.g. ATM, check cashing). This information is gathered through quarterly surveys sent to store managers which are then compared with similar surveys given to the principal food broker assigned to each store. Market demographics are drawn from the 2000 Census of the United States.

5 Market Definition

■ The central implication of the EFC framework is that when fixed costs are determined endogenously, market structure is relatively invariant to changes in market size. It is fixed costs, rather than the number of firms, that increase with the scale of the market. Testing this prediction empirically requires identifying a set of reasonably independent markets that vary in size. Retail industries, which are clearly spatially differentiated, provide a unique opportunity to test the theory using the geographic submarkets of a single industry. However, they also present a challenge: since retail industries are frequently dominated by national chains, the appropriate market may also be national. If this is the case, the EFC framework cannot be distinguished from exogenous scale economies using non-fragmentation alone.[12] A central claim of this paper is that the fixed costs most relevant to supermarket competition, namely investments in distribution and store size, are primarily incurred at low enough levels, and are sufficiently localized, to yield several distinct geographic markets. Specifically, by exploiting the physical constraints of

[11] Food outlets this small are classified as convenience stores. Offering a very limited range of products, they compete only with the smallest grocery stores.

[12] For example, demonstrating that McDonald's has the same share of restaurant sales in Wichita as it does in Chicago does not constitute a valid test of the EFC framework. Since the primary mechanism of differentiation is national advertising, the same investments that limit entry in one market also constrain it in others.

distribution systems, we can construct a reasonable number of relatively independent submarkets using the observed networks of stores. The purpose of this section is to verify this claim.

The appropriate market definition must be both large enough to contain the relevant fixed investments and distinct enough that firms cannot achieve the requisite scale by operating in more than one market. If distribution areas are smaller than the minimum scale of a chain, this should be reflected in the size distribution of firms. The upper left panel of Figure 2 contains a histogram of the total number of stores operated by each of the top 50 supermarket firms in the U.S. They range in size from 43 to 1,399 stores. While there are a number of very large chains, half of the top 50 firms operate less than 150 stores. These smaller chains are not fringe firms: within individual distribution areas, the average share of each of the 5 smallest chains in the (national) top 50 is over 10%. The set of firms (and distribution of sizes) is nearly identical if, instead of choosing the top 50 national chains, we select the top 2 from each distribution area. There are too many successful chains operating 50 to 100 stores for the minimum efficient chain size to be 500 or even 200 stores.[13]

So why is a distribution area the appropriate market? It is useful to start with a market definition that is clearly too small: the Metropolitan Statistical Area (MSA). The bottom left panel of Figure 2 shows the total number of stores for every MSA in the U.S.[14] The empirical distribution shows that the typical MSA is smaller than almost every dominant chain. Indeed, two-thirds of the MSAs hold less than 50 total stores. Although quite large in population, even the largest MSAs are not independent on the cost side: supermarkets firms routinely serve stores in several MSAs from the same distribution center. For example, the top four firms in Southern California serve stores in the San Diego, Riverside, Ventura, and Orange County MSAs from distribution centers in east Los Angeles. Every MSA is too small to exhaust the economies of distribution.

The bottom right panel of Figure 2 shows the number of stores that fit into 51 distribution areas defined using the observed networks of stores and warehouses.[15] The average distribution area is large enough to absorb three median sized chains. Five of the top 50 firms operate in only one distribution area and half operate in 3 or less. Of course, the biggest chains span

[13]The firm that Consumer Reports ranked second in the nation in 2003 (Wegman's) operates 68 stores in upstate New York.

[14]The total number of stores that fit into a market as a function of population follows a tight linear relationship (one store for every 11,000 people), regardless of how markets are defined.

[15]This is the method employed by Trade Dimensions in constructing the 52 distribution areas reported in their *Marketing Guidebook*. My own analysis produced only four changes (two of which accord with their most recent (2003) redefinitions), resulting in a total of 51 markets.

several distribution areas. Kroger (the largest firm) operates stores in 27 of these 51 markets. However, like most multi-region chains, Kroger operates separate divisions in each distribution market.[16] In many cases, the divisions operate under different names. An extreme example is Royal Ahold (the sixth largest chain), which is simply a holding company for six different chains (two of which are in the process of being sold off). While there is certainly some cost sharing that occurs across markets, the existence of so many successful chains in the 50 to 150 store range (and the recent spate of downsizings) makes it highly unlikely that the returns to scale extend much beyond a single distribution area.[17] If the efficient scale for a chain was larger than this, *most* chains would operate 500 or more stores. This is not the case.

Finally, at least in terms of physical shipments, these markets are relatively self-contained. Since I observe the distribution center that serves every store, it is possible to calculate the percent of sales that originate "within market". On average, stores supplied by a distribution center in the same distribution market account for 82% of total sales. Since these markets are not geographically isolated, it is natural to expect some spillover at the boundaries.[18] In testing the EFC framework, we should be concerned only if the spillover is caused by small markets "piggybacking" on the distribution facilities of larger markets (as is clearly the case with MSAs). If this were the case with distribution areas, we would expect the percent of outside sales to be negatively and significantly correlated with the size of the market. However, a regression of the percent of outside sales on the log of population yields a coefficient of -.01 (robust standard error .017). Since the coefficient is both small in magnitude and insignificantly different from zero, there does not appear to be any small market bias. Having identified an appropriate set of markets, I now turn to the central empirical exercises of the paper.

6 Does Sutton Apply to Supermarkets?

■ This paper suggests that the structure of the supermarket industry is best explained by a model of endogenous fixed investment. The following sections test this claim by verifying

[16] The physical structure of the distribution networks is also consistent with the EFC framework: vertically integrated supermarkets operate much bigger distribution centers (DCs) in larger markets. For example, Safeway operates a 300,000 square foot DC in Spokane, WA and a 900,000 square foot DC in neighboring Seattle.

[17] The primary benefit of national scale (buying power) does not require integration: the two largest third party distributors move as much volume as the biggest chains. Firms mainly integrate to facilitate *coordination* which, at least in the supermarket industry, occurs primarily at the level of the distribution center.

[18] Some of this spillover is also due to recent entry by neighboring firms. Supermarket firms typically enter a new market by opening a few trial stores and supplying them from the closest facility. Once they have a foothold, they will build a dedicated distribution center.

two predictions of the EFC framework: non-fragmentation and natural oligopoly. Starting with the more robust implication, I demonstrate that the share of the largest supermarket firm remains bounded as markets expand. This non-fragmentation result is robust to alternative measures of both size and structure. I then compare this outcome to that of a second industry in which investments are exogenous. Consistent with the theory, the share of the largest firm is monotonically decreasing. Finally, I evaluate the more restrictive natural oligopoly hypothesis. Using the full set of firms, I find that 3 to 6 firms dominate each market, regardless of size. While larger markets do have more firms, the expansion is limited to a fringe of low quality stores.

□ **The lower bound to concentration.** Figure 3 contains scatter plots showing the relationship between concentration and market size. In the upper panels, concentration is measured by C_1, while the lower panels use the share of the six largest firms (C_6). Results for C_4 and the Herfindahl index are provided in a separate (online) appendix. The left and right panels correspond to two measures of market size. In the left panels, size is measured using $\ln(\alpha Y M)$, the log of total revenue from the theoretical model. Population (M) and income (Y) for each market were drawn from the 2000 Census, while α was assumed to be .05 (FMI's estimate of the percent of income spent on groceries in 1998). In the right panels, total revenue is taken directly from the sales data[19].

In all four panels, we observe a range of values for both C_1 and C_6. While the average value of C_1 is 28%, it ranges from 12.8% in Milwaukee to 59.2% in San Antonio. C_6 is 70% on average, but ranges from 44.3% to 95.4%. The scatter of points is consistent with the multiplicity of equilibria we expect in retail industries, where spatially differentiated, multi-store chains compete in markets that often developed decades apart. For example, an early entrant (H.E. Butt) managed to capture half the market in San Antonio, but the top seven firms in Milwaukee have been in place for decades and have roughly equal shares. Such difficult to measure features as the toughness of price competition and degree of horizontal differentiation are likely to vary by market, implying that we should not *expect* to uncover a tight functional relationship between concentration and size. This is Sutton's motivation for focusing on bounds.

Indeed, there is a clear *lower bound* to concentration in all cases. Focusing first on C_1, the minimum level of concentration is highest in the smallest markets, decreases for a range, and then hits a lower limit beyond which it does not fall and might even increase. The same pattern occurs for C_6 and is robust to alternative measures of both concentration (C_4, Herfindahl) and

[19] The correlation between the two measures of size is .973.

size (population, total stores). In each case, the non-fragmentation results obtains: concentration is strictly bounded below.

Sutton (1991) develops a formal test of the non-fragmentation result using bounds estimation. His proposed estimator requires a parametric assumption on the distribution of maximum shares. Sutton follows the standard statistical literature (Smith, 1994) in selecting the Weibull distribution, based on its inclusion in the extreme value family.[20] A logit transformation of the relevant concentration ratio ($\widetilde{C}_1 = \ln(\frac{C_1}{1-C_1})$ and $\widetilde{C}_6 = \ln(\frac{C_6}{1-C_6})$ in this case) ensures that the predicted concentration measures are between 0 and 1. To parameterize the lower bound, Sutton suggests the following functional form for $\widetilde{C}_1$ (or $\widetilde{C}_6$):

$$\widetilde{C}_{1i} = \beta_0 + \frac{\beta_1}{\ln(S/p_L\sigma)} + \varepsilon_i \quad (\varepsilon_i > 0) \tag{8}$$

where the residuals ε_i are distributed as a two parameter[21] Weibull

$$F(\varepsilon) = 1 - e^{\left(-\frac{\varepsilon}{b}\right)^a}, a > 0, b > 0 \tag{9}$$

The functional form for the bound is loosely based on Sutton's (1991) advertising model and includes the term $p_L\sigma$ (σ in Sutton's formulation) to make the regressor unit free and help with scaling. Sutton equates σ (here $p_L\sigma$) to a proxy for set-up costs that is intended to capture "the minimal level of sunk cost that must be incurred by each entrant to the industry prior to commencing production". In the context of supermarket competition, this corresponds to building a small chain of medium sized stores and a basic distribution center. Since the smallest vertically integrated firms consist of about 28 stores, I assume that $p_L\sigma$ is \$250 million.[22] In the current setting, the bound is so flat that this choice of scaling parameter has no impact on the results.

Under ideal conditions, the entire set of parameters can be estimated using maximum likelihood (MLE). However, Smith (1994) has shown that the MLE may not exist if the parameter $a < 1$ and is not normally distributed if $1 < a < 2$.[23] He proposes an alternative two-step

[20] Since the share of the largest firm (C_1) is the largest value drawn from a sample of market shares (generated by some underlying share distribution), C_1 (and therefore any monotonic transformation of C_1) is an extreme value. A similar argument applies (somewhat loosely) to C_n.

[21] The three parameter Weibull was rejected in all cases by a standard likelihood ratio test.

[22] According to the Food Marketing Institute, the cost of building a new 40,000 square foot store is roughly \$7 million, so a 28 store chain would cost \$196 million to build and equip. The remaining \$54 million is a rough estimate of the cost of building a small distribution center, based on SEC filings and various media announcements.

[23] Specifically, Smith shows that for $a = 2$, MLE is asymptotically efficient and normally distributed (with a different rate of convergence than for $a > 2$), for $1 < a < 2$, the MLE exists, but is not normally distributed and may not be efficient, while for $a \leq 1$, no local maximum of the likelihood function exists in general.

minimum distance (MD) estimator that is consistent for all a and asymptotically efficient for $a < 2$. Since my estimates of a are all less than 1, I will use the MD estimator throughout the analysis. However, I will also present graphical results from the MLE procedure. For statistical inference, I use the asymptotic formulae derived by Smith (1994).

Estimates of the lower bound using both procedures are illustrated in Figure 4. The solid lines are the MD (Smith) estimates and the dashed lines correspond to the MLE. The parameter estimates from the two-step MD procedure are presented in Table 3 in the appendix. In all four panels of Figure 4 (which correspond to the same measures of size and concentration used in Figure 3), the lower bounds are remarkably flat and clearly asymptote to positive levels. Focusing on C_1, the limiting level of concentration (C_1^{∞}) is 12.4% for the first measure of size ($\ln(\alpha YM)$) and 12.2% for the second ($\ln(Sales)$). The corresponding 95% confidence intervals for C_1^{∞} are $(11, 13.9)$ and $(10.8, 13.8)$ respectively. Both intervals are bounded well above 0. Similarly, the limiting levels of C_6 are 35.2% and 31.2% respectively, with 95% confidence intervals $(28.6, 42.3)$ and $(25.1, 37.9)$. Concentration clearly does not fragment in the supermarket industry. To emphasize the strength of this result, the following section demonstrates how quickly markets fragment when investments are exogenous.

□ **Fragmentation in a horizontal market.** Non-fragmentation requires a sharp focus on quality. In order for escalation to drive out weaker products, consumers must agree on their ranking. Furthermore, the burden of improving quality must fall primarily on fixed costs. If either condition fails, the market will fragment, as the following example illustrates. To highlight the contrast between exogenous and endogenous fixed costs, I gathered data from an industry where pure vertical differentiation is largely absent: barber shops and hair salons.

While hair cuts clearly differ in price, it seems unlikely that consumers would choose to frequent the same type of shop if prices were equalized. These businesses differentiate themselves through location, the talent of their stylists, the quality of their products, and (to some extent) the gender of their clientele. Most shops do not advertise at all and, while there are a few national chains, the vast majority are sole proprietorships. There are no distribution networks and no role for R&D. In short, this is a classic exogenous cost industry.

Store level data on barber shops and beauty salons were drawn from the 1998 (second) edition of the American Business Disk (ABD). The ABD contains information on the identity and locations of retail firms based on listings in the yellow and white pages of local phone books. The continuously updated entries are cross-checked by direct phone calls to local businesses and comparisons with other directories. The ABD provides location and estimated sales data for each store in their database. I extracted every establishment that listed either barber shop (7241) or beauty salon (7231) as one of their three primary Standard Industrial Classification (SIC) codes, dropping only clearly irrelevant entries like beauty schools and barber colleges. Since the ABD does not include firm codes, I grouped stores by firm using store names only. Hoover's online database was used to identify chains such as Regis International that operate stores under several different names (e.g. Supercuts and Hair Cuttery). The final dataset included 57,588 stores operated by 38,321 distinct firms.

Since people are unlikely to travel very far for a haircut, we can continue to exploit the existence of independent geographic markets in estimating empirical bounds. However, the distribution based markets used for the supermarket industry clearly do not apply here. Instead, I present two alternative definitions of local markets: metropolitan statistical areas (MSAs) and counties. I calculated the sales share of the largest firm in each geographic market and constructed the measure of market size ($\frac{\alpha YM}{p_L\sigma}$) used in Sutton's model[24] . Population (Y) and income (M) were again taken from the 2000 census, while α was set at .002 (the ratio of total industry sales to total U.S. income). I chose the value \$20,000 for the scaling parameter $p_L\sigma$. This corresponds to the lower end of the asking prices at a website where small businesses are sold[25]. As in the previous analysis of supermarkets, I also calculated the value of total sales directly from the data.[26] Scatter plots of C_1 versus market size are presented in Figure 5. The two upper panels show C_1 versus market size for MSAs. Concentration is low and appears to be decreasing. However, the scatter of points suggests that the MSA may be too large a market definition. Switching to counties (in the two lower panels) the pattern is much clearer: the lower bound to concentration decreases monotonically to zero. Again, there is a scatter of points above the bound, but the minimum value of concentration clearly asymptotes to zero.

Figure 6 contains lower bounds estimated using both the two-step and MLE procedures. For MSAs, the bounds are fairly linear, reflecting the low levels of concentration that hold over the entire range of markets. However, the bounds are monotonically decreasing and asymptote to less than 1% in both cases. For counties, the bounds show greater curvature, although the sensitivity of the MD procedure to outlying observations is evident in the lower right panel.

[24] When quality is fixed at its minimum level ($z \equiv 1$), fixed costs are exogenous in Sutton's model. It is easy to show that the number of entrants increases monotonically with the size of the market $\left(N = \sqrt{\frac{\alpha YM}{p_L\sigma}}\right)$, consistent with the fragmentation result.

[25] http://www.businessesforsale.com/

[26] Since the ABD reports estimated sales by range, I used the midpoint in constructing shares. The ranges they use are somewhat coarse, so the majority of stores fell into only three categories. Using store counts or number of employees instead of sales yielded similar results.

Nonetheless, in both cases concentration decreases monotonically to less than 1%.

The parameter estimates are presented in Table 4 in the appendix. Focusing first on MSAs, the limiting level of concentration (C_1^∞) is .23% for the first measure of size ($\ln(\alpha YM)$) and .06% for the second ($\ln(Sales)$). The corresponding 95% confidence intervals are $(.08, .67)$ and $(.02, .14)$ respectively. Both intervals are strictly below 1%. Similarly, the limiting levels of C_1 in counties are .42% and .15% respectively, with 95% confidence intervals $(.05, 3.62)$ and $(.02, 1.22)$ that are bounded well below the corresponding intervals for the supermarket industry.

These results demonstrate that when the vertical dimension is absent, nearly complete fragmentation can occur quite quickly. This provides a sharp contrast with the supermarket industry, where the share of the top firms remained stable for a wide range of market sizes. The final set of empirical results is aimed at establishing an even stronger result, that supermarkets are a natural oligopoly.

□ **Natural oligopoly with a competitive fringe.** The vertical model of retail demonstrates that, in the absence of horizontal differentiation, the equilibrium number of firms can be completely unaffected by market size. This natural oligopoly result does not strictly hold here; the total number of firms clearly expands with market size. However, as I establish below, this expansion is limited to a fringe low quality stores. The number and realized share of the *dominant* firms is nearly constant in market size. Furthermore, these top firms build much higher quality stores and are far more likely to integrate into distribution than firms in the competitive fringe. Consistent with a simple extension of the pure vertical model, this provides the final and strongest test of the EFC framework.

Table 1 presents several measures of concentration for the 51 distribution areas defined above, broken out by market size (αYM) quartile. Although the size of markets increases by an order of magnitude over these 51 observations, the average levels of concentration are nearly constant. Even the ranges are relatively stable. The majority of sales in both small and large markets are captured by less than 8 firms; the number of firms required to serve 70% of the market (F_{70}) is both stable and bounded above. Only the *total* number of firms increases without bound, consistent with a two-tiered structure.

Figure 7 shows the size distribution of firms. Lorenz curves were constructed for each market by first ranking firms according to market share and then plotting the cumulative share of sales against the cumulative share of firms. The first three panels of Figure 7 contain Lorenz curves for three individual markets: Spokane (WA), Denver (CO), and Baltimore/Washington (MD-DC). Although the markets contain roughly 1.3, 4.7, and 9.8 million people respectively, the size distribution of firms is remarkably similar across the three. In each market, 5 or 6 firms account for the majority of sales. The remainder is split among a large fringe of very small firms. The main difference between these markets is the size of this fringe, which clearly expands with market size. The lower right panel of Figure 7 presents Lorenz curves for the full set of market. The uniformity in outcomes across markets is striking. For the full set of markets, 60 to 70% of sales are controlled by 5 to 7 firms, while the remainder of the market is captured by an expanding fringe of small firms. Although there is clearly some geographic variation, in no region do larger markets tend to have a greater number of *dominant* firms.

This two-tiered structure can be captured by a simple extension of Sutton's vertical model, described on pages 64-66 of his book. When there are two segments of the population, only one of which values quality, two distinct submarkets emerge. In the first submarket, a natural oligopoly of high quality firms serve the consumers who value quality. In the second, a fringe of low quality firms serve those who do not. While the size of the fringe expands with the size of the market, the number of high quality firms remains fixed. In large markets, these two segments are completely distinct. This simple framework fits the structure of supermarkets surprisingly well. I have already shown that there are two tiers of firms. This final set of results demonstrates that they differ in quality.

The ideal quality metric would combine variety with store size, since providing brandwidth requires stocking more products and building larger stores (wide aisles and easily accessible products consistently rate near the top of Progressive Grocer's customer surveys). Because Trade Dimensions does not record the number of products carried by each store, I will primarily use store size (in 1000s of square feet) to capture quality.[27] The quality differential between dominant and fringe firms is documented in Table 2. To assess the differences between the two types of firms, I first split the full set of firms operating in each distribution market on the basis of whether each firm was among the top 6 in that market. Average store level characteristics using store size and three additional measures of quality (see footnote 27 for definitions) were calculated for both types of firm across all of the stores operated by each firm, yielding 7,925 firm-market level observations. The results are presented in the first two columns of Table 2. The top 6 firms offer significantly higher levels of quality along all four dimensions (all differences are statistically significant at the 1% level). In particular, the top firms operate stores that are more than twice the size of the firms in the fringe. They also operate many more stores and capture

[27]As a robustness check, I present three alternatives quality measures constructed from store characteristics: the number of checkouts (cash registers), the number of features present in a store (0-4 among an in-store bakery, restaurant, pharmacy, and deli) and a similar measure for scanning registers and ATM machines (technology).

a much larger fraction of sales. Most importantly, almost three quarters of the top firms are integrated into distribution, while only a negligible portion of the fringe firms supply themselves. This reflects the tight connection between store level quality and firm level fixed costs that the EFC framework requires. To verify this, I divided the sample on basis of whether each firm was vertically integrated (VI). As shown in the last two columns of Table 2, VI firms operate stores that are more than twice the size of their non-VI counterparts. They also operate more stores and serve a much larger portion of each market, echoing the results of the top 6/fringe split.

This observed structure is clearly inconsistent with exogenous fixed costs or simple scale economies: if only 4 or 5 firms fit into the majority of large markets, we should expect to find monopolies in smaller markets, which we clearly do not. On the other hand, if minimum efficient scale can be attained by 3 or 4 firms in smaller markets, then larger markets should have many more entrants. The fact that we observe neither outcome suggests that both scale and the number of entrants are determined endogenously, as the EFC framework predicts.

7 Conclusions

■ This paper presents empirical evidence that endogenous fixed costs play a central role in determining the equilibrium structure of the supermarket industry. Using Sutton's (1991) EFC framework, I propose and evaluate a model of supermarket competition in which the escalating pressure to provide a wider array of products limits the number of firms that can profitably enter a market. Instead of encouraging entry, increases in market size yield a concentrated industry with better products, perhaps explaining the high premiums people will pay to live in larger cities.

References

BERRY, S. AND WALDFOGEL, J. *Product Quality and Market Size.* Working Paper no. 9675, Department of Economics, Yale University, 2003.

CHEVALIER, J.A. "Capital Structure and Product-market Competition: Empirical Evidence from the Supermarket Industry." *American Economic Review*, Vol. 85 (1995), pp. 415-435.

COMPETITION COMMISSION *Supermarkets: A Report on the Supply of Groceries from Multiple Stores in the United Kingdom.* London: The Stationary Office, 2000.

COTTERILL, R.W. AND HALLER, L.E. "Barrier and Queue Effects: A Study of Leading U.S. Supermarket Chain Entry Patterns." *Journal of Industrial Economics*, Vol. 40 (1992), pp. 427-440.

DICK, A. "Market Size, Service Quality and Competition in Banking." Forthcoming in the *Journal of Money, Credit and Banking*, 2007.

ELLICKSON, P.B. *Vertical Product Differentiation and Competition in the Supermarket Industry.* Ph.D. dissertation, Department of Economics, Massachusetts Institute of Technology, 2000.

ELLICKSON, P.B. (2006). "Quality Competition in Retailing: A Structural Analysis." *International Journal of Industrial Organization,* Vol. 24 (2006), pp. 521-540.

GABSZEWICZ, J.J. AND THISSE, J.F. "Entry (and Exit) in a Differentiated Industry." *Journal of Economic Theory*, Vol. 22 (1980), pp. 327-338.

HOLMES, T. "Barcodes Lead to Frequent Deliveries and Superstores." *RAND Journal of Economics*, Vol. 32 (2001), pp. 708-725.

KOCHERSPERGER, R.H. *Food Industry Distribution Center Benchmark Report.* Washington D.C.: The Food Marketing Institute and Food Distributors International, 1997.

LYONS, B.R., MATRAVES, C., AND MOFFATT, P. "Industrial Concentration and Market Integration in the European Union." *Economica*, Vol. 68 (2001), pp. 1-26.

MESSINGER, P.R. AND NARASIMHAN, C. "Has Power Shifted in the Grocery Channel?" *Marketing Science*, Vol. 14 (1995), pp. 189-223.

PROGRESSIVE GROCER *Marketing Guidebook*, White Plains: MacLean Hunter Media, 1998

ROBINSON, W.T. AND CHIANG, J. "Are Sutton's Predictions Robust?: Empirical Insights into Advertising, R&D, and Concentration." *Journal of Industrial Economics*, Vol. 64 (1996), pp. 389-408.

RONNEN, U. "Minimum Quality Standards, Fixed Costs, and Competition." *RAND Journal of Economics*, Vol. 22 (1991), pp. 490-504.

SCHIANO, W.T. "Spartan Stores Incorporated: Re-engineering for Efficient Consumer Response." Case No. 9-396-263, Cambridge: Harvard Business School, 1996.

SCHMALENSEE, R. "Sunk Costs and Market Structure: A Review Article." *Journal of Industrial Economics*, Vol. 60 (1992), pp. 125-133.

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

______ AND ______. "Product Differentiation and Industrial Structure." *Journal of Industrial Economics*, Vol. 60 (1987), pp. 889-917.
______ AND ______. "Multiproduct Firms and Market Structure." *RAND Journal of Economics*, Vol. 21 (1990), pp. 45-62.
SMITH, H. AND LATCOVICH, S. "Pricing, Sunk Costs, and Market Structure Online: Evidence from Book Retailing." *Oxford Review of Economic Policy.* Vol. 17 (2001), pp. 217-234.
SMITH, R.L. "Nonregular Regression." *Biometrika*, Vol. 81 (1994), pp. 173-183.
SUTTON, J. "Game-theoretic Models of Market Structure." in Kreps, D.M. and Wallis, K.F., eds. *Advances in Economics and Econometrics: Theory and Applications*, Vol. I, New York: Cambridge University Press, 1997.
SUTTON, J. *Sunk Cost and Market Structure: Price Competition, Advertising, and the Evolution of Concentration.* Cambridge: MIT Press, 1991.
TEDLOW, R.S. *New and Improved: The Story of Mass Marketing in America.* New York: Basic Books, 1990.

Table 1: Market Structure by Size Quartile

	Q_1	Q_2	Q_3	Q_4
C_1	28 (11,48)	30 (21,59)	27 (13,52)	27 (15,37)
C_2	41 (18,70)	46 (31,73)	43 (20,80)	44 (29,64)
C_4	58 (31,86)	62 (44,83)	59 (33,88)	62 (46,89)
C_8	72 (49,92)	77 (53,91)	73 (53,92)	76 (58,97)
C_{20}	85 (70,98)	87 (67,96)	84 (70,95)	86 (73,98)
$Herfindahl$	1345 (399,2606)	1506 (796,3769)	1348 (436,3503)	1373 (700,2452)
F_{70}	8.8 (3,20)	7.6 (2,24)	9.1 (2,21)	7.6 (3,16)
$Firms$	91 (28,138)	141 (67,245)	166 (74,249)	224 (65,449)

Ranges in parentheses.

Table 2: Store Characteristics by Firm Type

	Firm Type			
Characteristic	Top 6	Fringe	VI	Non-VI
Size	38.7 (12.3)	16.1 (10.6)	32.5 (14.6)	15.8 (10.4)
Checkouts	12.6 (6.70)	5.61 (3.28)	10.8 (6.81)	5.51 (3.09)
Technology	1.58 (.314)	1.09 (.699)	1.36 (.544)	1.09 (.701)
Features	2.52 (.755)	1.51 (.970)	2.12 (1.08)	1.50 (.961)
Stores	51.6 (53.9)	1.88 (3.18)	30.8 (46.8)	1.80 (3.10)
Percent VI	.778 (.416)	.040 (.196)	1	0
Market Share	.116 (.097)	.002 (.004)	.064 (.092)	.002 (.007)
Observations	306	7689	545	7450

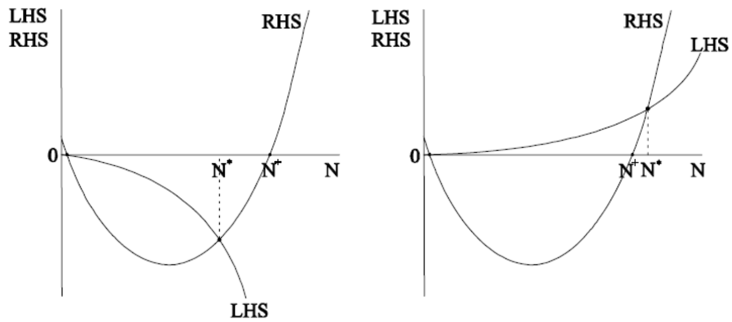

Figure 1: The equilibrium number of entrants

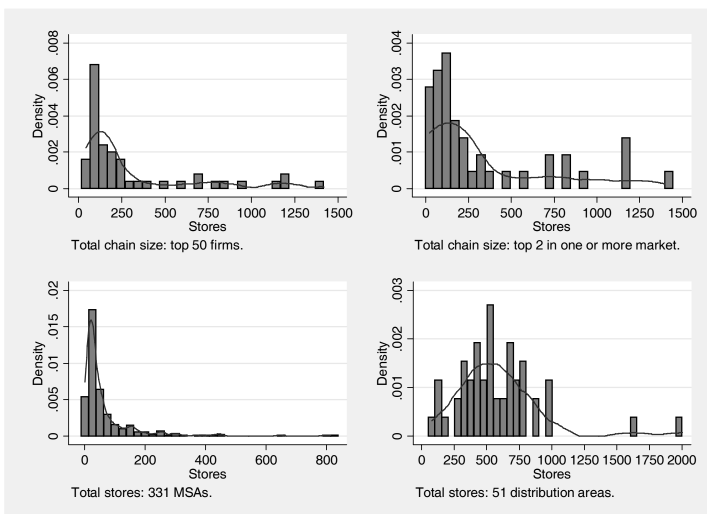

Figure 2: Defining Markets

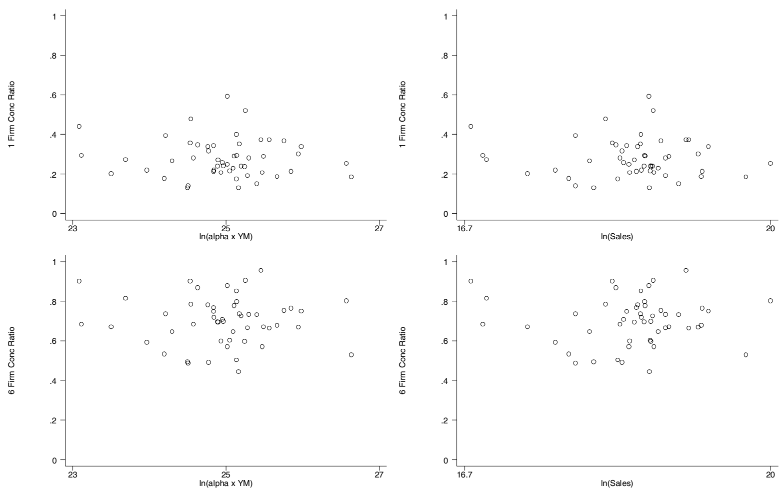

Figure 3: Concentration in the supermarket industry.

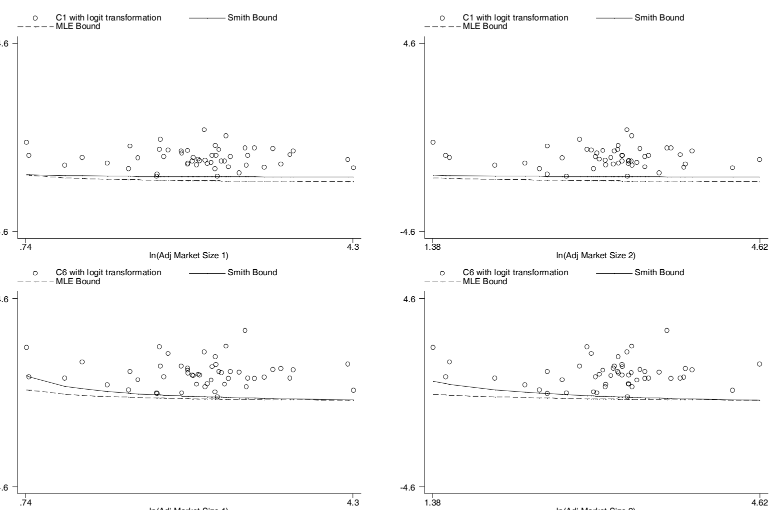

Figure 4: The lower bound to concentration.

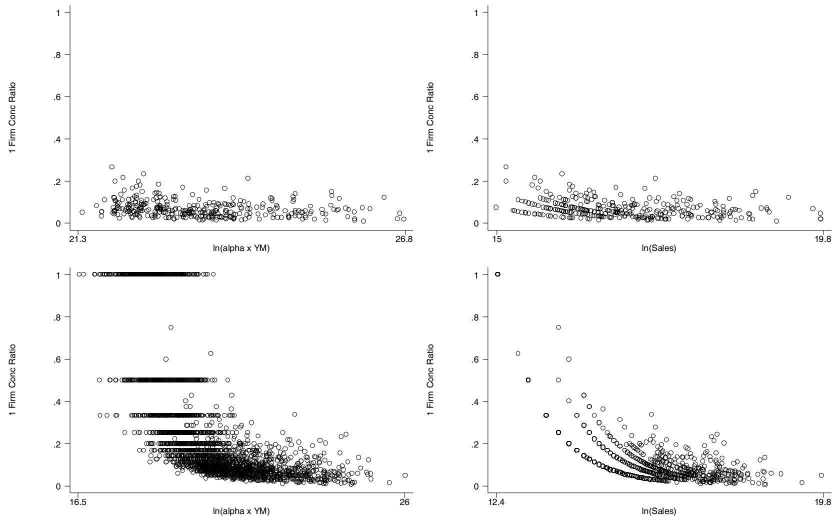

Figure 5: Barber shops and beauty salons.

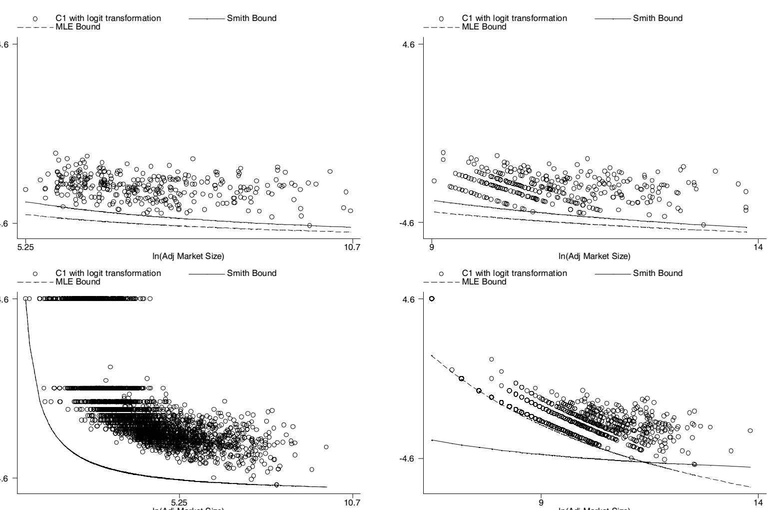

Figure 6: Fragmentation.

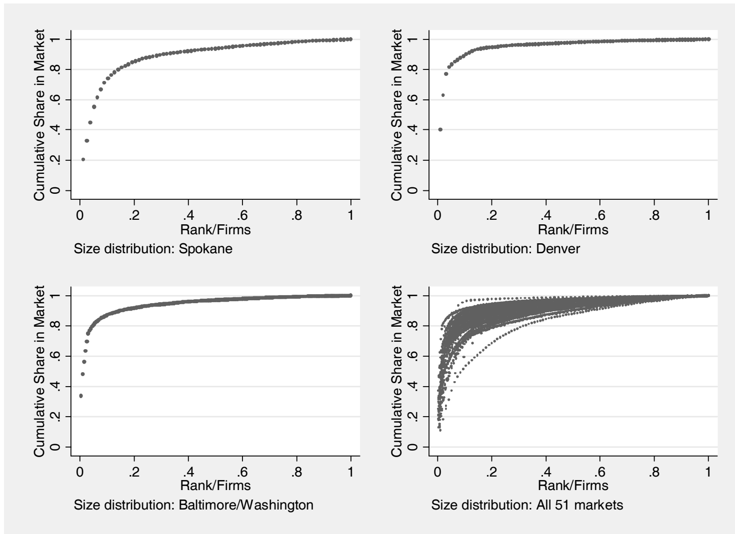

Figure 7: The size distribution of firms.

Appendix

Table 3: Lower Bound Estimates (Supermarkets)

	Measure of Size			
	$\ln(\alpha YM)$		$\ln(Sales)$	
Parameter	C_1	C_6	C_1	C_6
β_0	-1.96 (.057)	$-.612$ (.115)	-1.97 (.097)	$-.792$ (.196)
β_1	.105 (.145)	1.08 (.285)	.171 (.300)	1.86 (.599)
a	.634 (.190)	.893 (.173)	.634 (.190)	.88 (.177)
b	.858 (.085)	1.08 (.114)	.860 (.085)	1.09 (.113)
C_1^∞/C_6^∞	12.4	35.2	12.2	31.2
Observations	51			

Standard errors in parentheses.

Table 4: Lower Bound Estimates (Barbers and Salons)

	Market			
	MSA		County	
Parameter	αYM	Sales	αYM	Sales
β_0	-6.09 (.039)	-7.48 (2.71)	-5.47 (.009)	-6.49 (1.11)
β_1	13.6 (2.59)	36.1 (18.6)	4.17 (.005)	19.1 (3.59)
a	2.53 (.037)	2.08 (.041)	1.85 (.041)	1.83 (.041)
b	1.60 (.112)	1.50 (.096)	3.81 (.025)	3.67 (.024)
C_1^∞	0.22	0.06	0.42	0.15
Observations	326		2789	

Standard errors in parentheses.