Market Size Matters

Jeffrey R. Campbell* Hugo A. Hopenhayn†

July, 1999‡

Abstract

This paper characterizes the organization of the U.S. retail trade sector by comparing establishment sizes and numbers across cities with different populations. In most two digit retail trade industries, large cities have larger establishments but fewer establishments per–capita than do small cities. These observations are inconsistent with free entry models where markups are independent of the number of producers. Models in which adding competitors reduces markups can reproduce these results.

JEL Classification: L11, L16, L81

*University of Chicago and NBER

†University of Rochester and Universidad Torcuato DiTella

‡We thank Victor Aguirregegabiria, Eric Hanushek, Sarit Markovich and numerous seminar participants for their insightful comments on previous drafts of this paper. Emin Dinlersoz provided expert research assistance. Much of this research was completed while Campbell visited the Bradley Center for Policy Research at the University of Rochester's William E. Simon Graduate School of Business. We thank that institution for its generous research support. The National Science Foundation supported Campbell's research through SBR–9730442. Please address correspondece to:

Jeffrey Campbell
Department of Economics
University of Chicago
1126 East 59th St.
Chicago, IL 60637

e–mail: jcampbe@midway.uchicago.edu

1 Introduction

This paper characterizes the organization of the U.S. retail trade sector by comparing retail establishment sizes and numbers across cities with different populations. If individuals' demand curves for retail trade services do not depend directly on city size, there are a large number of identical potential entrants, and producers take prices as given, as in Hopenhayn's (1992) competitive industry dynamics model, then large and small cities' retail trade industries should differ only in the numbers of establishments they support, which will be proportional to their populations. Wolinsky's (1986) and Fishman and Rob's (1998) free entry models of consumer search and monopolistic competition between large numbers of producers share these predictions.

The common characteristic driving these results is the invariance of producers' postentry conduct to the number of competitors. If competition becomes "tougher" and markups fall with the addition of new producers, as in Cournot competition, then the number of customers per producer must increase with city population. Otherwise, producers' variable profits would fail to cover their fixed costs. In this case, producers' average output must also rise with city population because they serve more customers and charge lower prices. These differing theoretical predictions of market size's impact on producer numbers and sizes motivates our empirical investigation.

Our primary source of observations of retail trade industries is the 1992 Census of Retail Trade (*CRT*), which reports the number of operating establishments, the value of their sales, and their total employment for each two–digit industry in each metropolitan statistical area (*MSA*), which we identify with a market. We regress the number of establishments per capita and their average employment and sales against the *MSA*'s population and a set of control variables. Our control variables account for factor price differences, productive spillovers from urbanization, recent growth experiences, and demographic differences. Although our set of control variables is extensive, it does have its shortcomings. For example, one potentially relevant missing variable that may vary with city size is the cost of advertising. Our regression results should be interpreted with the caveat that some potentially interesting relevant variables are unavailable to us.

In the regressions for most two-digit industries, *MSA* population is statistically significant. Figures 1 and 2 are representative of our results. For our sample of 297 *MSA*'s, Figure 1 plots establishments per capita in SIC 57, Apparel and Accessory Stores, against *MSA* population. Both variables are expressed in logarithms and are the residuals from regressions against a constant and our control variables. The figure displays a clear tendency for the number of Apparel and Accessory Stores per capita to decline with *MSA* population. The corresponding regression line, also plotted on Figure 1, has a slope of -0.098. The associated t–statistic equals -4.60. In retail trade as a whole and nine of its ten two digit industries, we estimate similar, statistically significant, negative coefficients. To gauge the magnitude of population's estimated impact on Apparel and Accessory Stores per capita, it is helpful to compare the regression's predictions for two cities, one with the population of Rochester, New York $(1,062,470)$ and the other with the population of Detroit, Michigan $(4,266,654)$, which are otherwise identical. The estimated coefficient implies that establishments per capita should be 12.7% lower in the larger city compared to the smaller.[1]

Figure 2 is analogous to Figure 1, except the logarithm of average establishment size, measured with the value of sales per establishment, replaces establishments per capita.[2] This figure displays a positive relationship between establishment size and *MSA* population. The regression line has a slope of 0.079 with a t–statistic of 5.34. Again comparing our hypothetical cities with the populations of Rochester and Detroit, the regression predicts that sales per store should be 11.6% greater in the larger city.[3] Although our specific estimates vary across the remaining two digit industries, those from Apparel and Accessory stores qualitatively resemble most of them. Only one of the estimated coefficients on *MSA* population is negative, and six of those that are positive are statistically significant.

For the purpose of characterizing an entire sector of the economy rather than a particular industry, the Census of Retail Trade's breadth and uniformity across industries and geography are clearly essential. To the best of our knowledge, no other data set with similar characteristics exists. The return to taking such a relatively broad view of local retail competition is the wide applicability of our results. The cost of doing so is that we are unable to refine our observations using industry specific details.

One notable shortcoming of the Census of Retail Trade is that it reports no information on establishments' firm affiliations at the state, county, or *MSA* level. Although many models of retail competition make the simplifying assumption that an establishment's firm affiliation is irrelevant, the converse is clearly true in the data. In Food Stores, for example, the average sales of establishments belonging to multiestablishment firms are 6.3 times higher than the average sales of single establishment firms. However, 66%

of establishments in Food Stores belong to such small single establishment firms.[4]

Food Stores is one of the industries where we find weak effects of *MSA* population on average establishment sales and employment. One potential explanation for this is that only conduct at establishments owned by large, "dominant" firms differs across different sized markets. In this case, the presence of a sizeable "competitive" fringe reduces the response of average establishment size to these changes below what it would be if the fringe were removed from the data set. Our results from SIC 53, General Merchandise Stores, supports this theory. In that industry, single establishment firms account for only 27% of establishments and we find very strong effects of market size on establishments' average sizes and numbers. More thorough examination of this hypothesis must await the development of retail trade data sets that include characteristics of establishments and their parent firms, such as the Bureau of the Census's forthcoming Longitudinal Business Database.

Our use of observations of establishment numbers places us in the footsteps of previous work which attempts to infer the nature of competition between small groups of producers in local markets from their observed entry and exit decisions. Berry (1992) evaluates airlines' costs of entry by observing their previous airport presence and entry decisions. Chevalier (1995) measures the impact of firms' financial decisions on product market competition by documenting supermarket firms' entry and exit responses to changes in a rival's financial leverage. Unlike these papers, we use only aggregate data for each *MSA*, so we can say nothing about the entry and conduct of individual producers.

For three service industries and two retail trade industries in a sample of isolated rural towns, Bresnahan and Reiss (1991) found that the number of consumers in a market necessary to support the entry of N firms grew approximately linearly with N for N greater than two or three. Their interpretation of this result is that competitive conduct changes little with entry. The apparent contradiction between these results and ours for much larger markets can be partially resolved if we look more closely at the particular retail trade industries, druggists and tire dealers, that Bresnahan and Reiss consider. Drug Stores is the only two digit industry for which we fail to find statistically significant effects of market size on establishments per capita, although the relevant t–statistic is very close to the 5% critical value. The two–digit industry to which tire dealers belong, Automobile Dealers, displays significant size effects. When we repeat our analysis for tire dealers' parent three-digit industry, Auto and Home Supply Stores, we find no market size effects on either establishments per capita or average establishment size. Thus, it appears that tire dealers is an exception to the general rule that market size matters in retail trade.

The remainder of the paper is organized as follows. The next section uses a canonical two stage free entry model of competition to illustrate how the comparative statics of establishment numbers and sizes with respect to market size can reflect the nature of postentry competition. Section 3 describes our data sources and econometric procedures. Section 4 presents the paper's linear regression estimates, and Section 5 compares our results with Bresnahan and Reiss's (1991) in more detail. Because some interesting imperfect competition models predict the impact of market size on establishment sizes and numbers to diminish for large markets, the assumption that the regression function is linear may be inappropriate. In Section 6 we address this by considering the robustness of our results to estimating a semilinear regression function, which is linear in the control variables but an arbitrary function of *MSA* population. Section 7 offers concluding remarks and discusses useful directions for future research.

2 A Canonical Model of Competition

Before proceeding with the empirical analysis, it is helpful to use a canonical free entry model of competition to illustrate what can be learned about producers' interactions and conduct from our regressions. Towards that end, consider a symmetric model of competition between N firms serving an *MSA* with S consumers. The *MSA*'s consumers have identical downward sloping demand curves, $D(p)$. The firms have access to a common production technology which displays constant marginal cost c. To map the predictions of this model into our data, we assume that each firm operates a single establishment. To emphasize this assumption, we hereafter refer to these firms as establishments. We defer a discussion of this assumption's accuracy for U.S. retail trade industries until Section 3.

Rather than explicitly model the interactions between establishments, we simply assume that a symmetric Nash equilibrium exists in which S/N customers visit each one of them. In this case, each establishment's profits are

$$\frac{S}{N} D(p(N))(p(N)-c),$$

where $p(N)$ is the common equilibrium output price. For simplicity, we assume that $D(p)$ is concave, so that variable profits are a decreasing function of the price for all prices below the unique monopoly price. We also assume that the equilibrium price $p(N)$ is less than the monopoly price and is a weakly decreasing function of N. Note that the assumption that a common equilibrium output price exists does not presuppose that establishments compete in prices or sell homogeneous goods. The assumption that the equilibrium price only depends on N and not on S is appropriate given our assumptions because S affects profits multiplicatively and so should have no impact on postentry behavior.

The number of active establishments, N, is determined in a game of free entry. Specifically, there is a large number of identical potential entrants that simultaneously make their entry decisions. Those that do not enter receive a payoff of zero, while those that do enter pay the fixed cost, F, and then collect the postentry variable profits. Equilibrium in the free entry game requires that the total costs and benefits of entry are equal.

$$F = \frac{S}{N} D(p(N))(p(N)-c) \tag{1}$$

This statement of the free entry condition clearly ignores the constraint that N be an integer. This would be problematic when studying small markets with few establishments, but it should provide an acceptable approximation for the large markets we consider. Because variable profits are strictly decreasing in N given S, there is a unique N that satisfies Equation 1. If we define $\pi(N)$ to be each establishment's variable profits per customer, then Equation 1 can be written more compactly as

$$F = \frac{S}{N} \pi(N) \tag{2}$$

Given our assumptions, $\pi(N)$ is a decreasing function of N.

Our regressions provide estimates of the elasticity of establishments per capita with respect to total population. In this simple model we can calculate this elasticity by applying the implicit function theorem to Equation 2. If we define $\eta(N)$ to be the elasticity of profits per–customer with respect to N, then the resulting expression is

$$\frac{d\ln(N/S)}{d\ln S} = \frac{\eta(N)}{1-\eta(N)}. \tag{3}$$

Because $\eta(N)$ is weakly negative, this elasticity must also be so and be greater than -1. In Wolinsky's (1986) model of large group monopolistic competition, equilibrium profits per–customer are invariant to the number of establishments. Therefore, this elasticity equals zero and the number of entrants must be proportional to the number of consumers. In Salop's (1979) model of competition on the circle, the addition of one establishment makes all producers' goods more substitutable with each other, so markups fall. In this case $\eta(N)$ is strictly negative, and the number of establishments per capita falls with population. Within the context of this simple framework, our regressions discriminate between models where $\eta(N)$ equals zero and models where $\eta(N)$ is strictly negative by estimating Equation 3's left hand side.

Our regressions also provide estimates of establishments' average sales elasticity with respect to total population. To demonstrate that this elasticity is always positive in this model, denote the quantity sold by one producer with $q(N)$ and divide and multiply Equation 1 by $p(N)$ to get

$$p(N)\,q(N)\left(\frac{p(N)-c}{p(N)}\right) = F. \tag{4}$$

So each producer's postentry profit equals the product of its sales' value with the percentage price–cost markup. Because $p(N)$ is decreasing in N and N is increasing in S, the price–cost markup must decrease with S. To offset this, sales' value must be increasing in S. That is, if we take logarithms of Equation 4 and differentiate with respect to $\ln S$, we get

$$\frac{d\ln pq}{d\ln S} = -\frac{d\ln((p-c)/p)}{d\ln S}. \tag{5}$$

Again, within the context of this framework, our regressions allow us to infer how equilibrium markups change with population, although they tell us nothing about markups' levels.

3 Data Sources

For our empirical analysis, we wish to characterize the following linear regression function.

$$y = \beta_0 + \beta_s s + \beta_x' x + u \quad (6)$$

The dependent variable, y, is a characteristic of a particular retail trade industry in a particular *MSA*, for example the logarithm of Apparel and Accessory Stores' average sales in Detroit. The independent variables, s and x, are the logarithm of that *MSA*'s population and a vector of control variables. By the definition of a regression function, the expected value of the error term, u equals zero if one conditions upon s and x. Throughout our empirical work we will also maintain the assumption that the disturbances u are independent across *MSA*'s.

To estimate β_s and β_x, we use observations of retail trade industries in 297 *MSA*'s. The goods produced by the retail trade sector are largely locally consumed, so it is reasonable to suppose that a sequence of geographically distinct markets characterizes the retail trade sector. However, whether the *MSA* is the correct definition of a market is questionable. Because both producers and consumers are spatially differentiated, correct market definition requires determining the demand for a product at each geographic location within an *MSA*, as in Davis's (1998) study of movie theaters' spatial differentiation. Nevertheless, using observations at the *MSA* level is appropriate if *MSA*'s are aggregations of perfectly competitive markets, because neither the per capita numbers or sizes of these markets' establishments should exhibit a relationship with their market sizes. Using observations at the *MSA* level is also informative if the model of interest is one of spatial differentiation across an entire city. Nevertheless, the correspondence between *MSA*'s and markets may be far from perfect. Subject to this qualification, we refer to "*MSA*'s" and "markets" interchangeably.

3.1 Observations of Retail Trade Industries

From the 1992 Census of Retail Trade, we obtain observations of the number of operating establishments in each metropolitan statistical area (*MSA*) and primary metropolitan statistical area (*PMSA*), their total employment, and their total sales during 1992. In the empirical work, we group *MSA*'s and *PMSA*'s together and refer to them simply as "*MSA*'s." The observations are available for retail trade as a whole and for its constituent two-digit industries.[7] Dividing total sales and employment by the number of establishments yields the two measures of average establishment size we consider. To construct establishments per capita, we measure each *MSA*'s population by summing the populations of their constituent counties from the 1990 decennial census, as reported in the 1994 County and City Data Book. Because these two data sources use two different definitions of *MSA*'s in the New England States, we exclude those *MSA*'s from consideration. The resulting sample consists of 297 *MSA*'s.

As we noted in the introduction, the *CRT* provides no information regarding establishments' affiliations with firms at the MSA level. To empirically gauge the influence of firm affiliation on retail trade establishments, the first two columns of Table 1 report the average sales of those owned by a firm with two or more establishments, labeled "Multiunit", and those owned by a firm consisting of only that establishment, labeled "Single Unit". The column labeled "Sales Ratio" reports the ratio of these two averages. If firm affiliation were irrelevant, we would expect these ratios to be close to one. To show how establishments with different firm affiliations compose the retail trade data, the table's final column, labeled "Fraction Single," reports the fraction of all establishments that are part of firms with only one establishment. These statistics are reported for the whole retail trade sector and for each of its two digit industries. Table 1's entries are constructed from reports of these two groups' establishment numbers and total sales for the United States as a whole from the *CRT*'s 1992 "Establishment and Firm Size" volume.[7] The Bureau of the Census does not report analogous statistics for any geographic partition of the United States, so we cannot observe variation in firm sizes and numbers across *MSA*'s.

Retail Trade's firm affiliation characteristics are unsurprising. For the retail trade sector as a whole, a majority of 65% of all establishments belong to single unit firms, but establishments at multiunit firms have average sales that are about three times larger than these. The heterogeneity in the statistics across two digit industries is somewhat more interesting. In one industry, Automobile Dealers, firm affiliation appears to be nearly irrelevant, while the average General Merchandise Store in a multiunit firm is 17 times larger than its average counterpart in a single unit firm. The other industries' sales ratios generally fall between 2.5 and 4.0. The exceptions to this are Food Stores, with 6.31 and Gasoline Service Stations, with 1.64. Although a majority of a two digit industry's establishments generally belong to a single unit firm, this is not the case in General Merchandise Stores and Apparel and Accessory Stores, where only 27% and 39% of establishments belong to such small firms.

Overall, Table 1 leaves the impression that both the association between firm affiliation and establishment size and the representation of relatively small establishments in single unit firms vary widely across two digit industries. Many of the models of competition among large groups of producers fail to make a clear distinction between firms and establishments. For example, location models such as Salop's (1979) and search models like Wolinsky's (1986) typically assume that a separate firm operates each establishment. In industries where establishments' firm affiliations appear to matter little, such as Automobile Dealers, and in industries where establishments in single unit firms are very small and few in number, such as General Merchandise Stores, simplifying establishments' firm affiliations in this way may be innocuous. However, in industries like Food Stores with numerous independent stores that are much smaller than chain stores, a model that treats all establishments symmetrically such as that in Section 2 is probably suspect.

For our sample of *MSA*'s, Table 2 reports summary statistics for establishments per capita, average establishment sales, and average establishment employment for the retail trade sector as a whole and its five largest two-digit industries, measured by total sales or employment.[7] There is substantial variation across industries and *MSA*'s in the number of establishments per capita. The average numbers of General Merchandise Stores and Eating and Drinking Places per 1000 people are 0.14 and 1.79. The other industries fall between these two extremes. For all of the two-digit industries, the standard deviation of establishments per capita's logarithm falls between 0.15 and 0.30. Variance across and within industries also characterizes the two measures of average establishment size. Using either sales or employment to measure size, General Merchandise Stores are much larger on average than establishments in any other industry. The average General Merchandise Store in the average *MSA* employs approximately 75 people. The smallest and largest average employment for these stores are approximately 29 and 142. The standard deviations for the two average size measures' logarithms are between 0.20 and 0.35 for most of the two-digit industries.

3.2 The Regressors

Our regression estimates are meant to provide empirical analogues to theoretical comparative statics exercises. This identification requires the determinants of establishment numbers and sizes omitted from the regression function to be independent of those that are included. Therefore, the validity of this identification depends on the exhaustiveness of the control variables included in x. Because of their obvious importance for establishments' production possibilities, the regressions include the prices of land and labor in the *MSA*. They also include measures of productive spillovers associated with urbanization, which Jacobs (1969) argues are the economic cause of city formation. In models of monopolistic competition, characteristics of the population unrelated to the market's scale can impact establishment sizes by changing the elasticity of consumers' demand functions and consequentially firms' optimal markups. Therefore, x also includes demographic characteristics of the *MSA*'s population; median income, the percentage of the population that is Black, and the percentage of the population that holds a bachelor's degree or higher. Finally, we include in x the growth rate of the *MSA*'s population between the 1980 and 1990 decennial censuses to control for the effects of short run dynamics on industry composition.

Table 3 lists the regressors we include in our estimation. Our measure of s_i is the logarithm of the *MSA*'s population estimate from the 1990 decennial census. We describe the remaining independent variables and the importance of their inclusion in x below.

3.2.1 Factor Prices

We include two measures of factor prices in x, the retail wage rate and the cost of real estate, both of these in logarithms. Dividing the retail trade sector's annual payroll by its employment, both reported in the Census of Retail Trade, yields our measure of the retail wage. An ideal measure of the price of real estate would be the price per square foot of retail space. Unfortunately, nothing of that sort is available to us. An imperfect but available substitute is the median rent of a renter occupied residential unit, which is reported in the County and City Data Book for each county. We aggregate these measures into an *MSA* wide rental rate with a population weighted average of each county's rate. Because land and other construction resources can be used to construct either residential or commercial real estate, this measure should be a good indicator of the cost of commercial real estate if the cost of new construction pins down all real estate prices. On the other hand, prices for residential real estate will fail to track those of commercial real estate if zoning policies create a significant difference between commercial and residential construction costs and the size of this difference varies widely across *MSA*'s.

3.2.2 Urbanization Spillovers

Cities are not exogenous agglomerations of people. Jacobs (1969) hypothesized that spillovers associated with spatial concentration of diverse activities are the primary return to city formation. In support of this view, Glaeser, Kallal, Scheinkman, and Shleifer (1991) showed that wages in *MSA*'s with a diverse set of industries grew faster than in those with more uniform employment opportunities. The theoretical impact of such spillovers on retail trade establishment sizes and numbers is inherently ambiguous. The direct effect of a productive spillover on an industry's establishments is to lower their costs, but the spillover can either raise minimum efficient scale, reduce it, or leave it unchanged. In a competitive free entry model, the ambiguous impact of spillovers on minimum efficient scale directly produces corresponding ambiguities in establishment sizes and per capita numbers.

There is no reason to suppose that such productive spillovers do not impact retail trade technology, so we include two measures of them in x. The first is a measure of the *MSA*'s economic diversity, the share of employment in the *MSA* accounted for by its five largest two-digit industries. This measure of industry concentration uses the employment data from the 1992 County Business Patterns, and is similar to that used by Glaeser, Kallal, Scheinkman, and Shleifer (1991). When industry concentration is high, the *MSA*'s economy lacks diversity. Our second measure of productive spillovers is the geographic density of economic activity, which Ciccone and Hall (1996) found to have a significant positive impact on U.S. states' measured total factor productivity. To construct the density measure, we divided a county's total employment reported in the 1992 County Business Patterns database by the county's total land area, in square miles, for each of the *MSA*'s constituent counties. The *MSA*'s density measure is then the employment weighted average of the measures for its constituent counties. This density measure is similar to the theoretically ideal measure considered by Ciccone and Hall (1996), with the exception that we follow Ciccone (1997) by using employment rather than output to measure economic activity. Density enters the regression as a logarithm.

3.2.3 Demographic Characteristics

In standard formulations of competitive industrial organization models, the demographic characteristics of the market's customers affect only the demand specification, so they impact only establishment numbers and not their sizes. For our purposes, this conclusion is hasty because the industries we consider are rather broadly defined. For example, SIC 58, Eating and Drinking Places, encompasses restaurants and bars of all kinds. If relatively wealthy consumers prefer French food to pizza while the opposite is true for their poorer counterparts, then we should expect that the average establishment size in a high income community would differ from that in a low income community simply because the technologies for French food and pizza differ. If the service of food preparation is a normal good, then wealth can influence the scale of an individual consumer's demand for restaurant meals. Cultural differences across communities can also influence the nature and quantity of different retail services demanded at given prices, which in turn affect retail establishment sizes and numbers. As we noted above, these demographic characteristics can also impact the elasticity of demand, and hence the sizes and numbers of establishments in a monopolistically competitive equilibrium. For these reasons, we include three demographic variables in x that measure the income, educational attainment, and racial composition of an *MSA*'s residents.

All of our demographic observations are reported at the county level in the 1994 County and City Data Book. We used population weighted averages of their values in the *MSA*'s constituent counties to construct analogous measures for each *MSA*. To measure income, we use median family income, and we enter it into the regression as a logarithm. Our measure of racial composition is the percentage of the *MSA*'s residents who are Black. Finally, the percentage of an *MSA*'s residents over twenty-five years old who have a bachelor's degree measures the *MSA*'s educational attainment. These two variables enter the regression in levels.

3.2.4 Population Growth

At any point in time, the establishment size distribution corresponds to an aggregation of different age cohorts. For a panel of Wisconsin retail trade firms, Pakes and Ericson (1998) documented that the firm employment distribution tends to increase stochastically with age. Gort and Klepperer (1982) also showed that markets for new products experience an initial period of growth in the number of operating establishments, followed by a shake-out in which large numbers of them exit. Together, these results suggest that a rapidly growing *MSA* will have more but smaller establishments than an otherwise similar, but stable, market. If large *MSA*'s correspond to stable, mature cities and smaller *MSA*'s correspond to growing ones, then these growth effects could induce spurious correlation between *MSA* population and establishment sizes and numbers. To control for this possibility in our regressions, the vector x includes the *MSA*'s population growth rate between the 1980 and 1990 decennial censuses. The growth rate is measured as the difference in the *MSA*'s population divided by its average population in the two censuses. It appears in the regression in levels.

3.2.5 Descriptive Statistics

Table 4 reports summary statistics for population and the regressions' other independent variables. In the first, second, and third columns are the average, minimum and maximum values of the variable. The fourth column reports the standard deviation of the variable as it is entered in the regression. For example, the standard deviations for population and employment density are those of the logarithm while that for industry concentration is of the level. The table's right most column, labeled "Corr. w/ Pop.", reports the correlation of the given variable with population's logarithm. Table 4's first row reports the summary statistics for population. Unsurprisingly, population varies substantially across our sample of *MSA*'s. The least populated *MSA*, Enid, Oklahoma, has $56,795$ residents. The *MSA* with the largest population is Los Angeles, California. Density also varies considerably across *MSA*'s. The *MSA* with the lowest employment density, with just under four employees per square mile, is Casper, Wyoming. The *MSA* with the largest density is New York, New York. Although population and employment density are clearly related, they are not identical to each other. The correlation of their logarithms equals 0.74, which seems sufficiently below one to include them simultaneously in the regressions without fear of colinearity. Industry concentration also varies considerably across the sample. The lowest five-industry concentration measure is Decatur, Alabama's, 27.32%, and the highest ratio is Kokomo, Indiana's, 61.92%. The industry concentration ratio is negatively correlated with population, so larger markets tend to have more industry diversity. All of the remaining regressors are positively correlated with population.

4 Linear Regression Results

The analysis of establishment counts and sizes for the retail trade sector as a whole and its two digit industries produces 33 regressions. To conserve space, we report the complete regression results for one industry as an example. For the remaining industries, we report only the estimates of β_s and summarize the estimates of the control variables' regression coefficients. For SIC 57, Apparel and Accessory Stores, Table 5 reports the estimated regression coefficients, their t–statistics, and the R^2's and sample sizes for the regressions of establishments per capita, and establishments' average sales and employment on s and x. All of the t–statistics reported in this paper are robust to general forms of heteroskedasticity in the regression disturbances. The R^2 from the establishment regression is 0.16, which is the lowest R^2 from any of our estimated regressions. The R^2's from the sales and employment regressions, 0.44 and 0.30, are more typical of the regressions for the other industries.

4.1 *MSA* Population Coefficients

For all three regressions and all of the industries we consider, Table 6 reports the OLS estimates of β_s and their associated t–statistics. As we noted in the introduction, the regression's estimated slope for Apparel and Accessory Stores equals -0.0983, and it is statistically significant. The estimated coefficients for the other industries indicate that Apparel and Accessory Stores is not an anomaly in this data. All of the estimated coefficients for the two digit industries are negative, and nine of them are statistically significant at conventional levels. The t–statistic for Drug Stores equals -1.94, which is just below the relevant 5% critical value. Although the estimated coefficients generally indicate that there are fewer establishments per capita in larger markets, the magnitude of this difference varies greatly across industries. The coefficient for General Merchandise Stores has the largest magnitude of any of those estimated, -0.146, while that for Eating and Drinking Places has the smallest, -0.039. Most of the remaining coefficients have magnitudes similar to that of Apparel and Accessory Stores. To better understand the quantitative significance of these estimates for competition in retail markets, it is helpful to interpret them within the framework of Section 2's simple model. If we plug our estimates of β_s into Equation 3's left hand side and solve for the elasticity of variable profits per–customer with respect to the number of establishments, this elasticity equals -0.17 for General Merchandise Stores, -0.11 for Apparel and Accessory Stores, and -0.04 for Eating and Drinking Places. Thus, the addition of new competitors appears to have an economically significant impact on profits per–customer, within the context of that simple model.

Apparel and Accessory Stores' positive and significant estimate of β_s in its average sales regression is also typical of the other two-digit industries. For nine of them, the estimates of β_s in the average sales regressions are positive. The exception to this is Building Materials and Garden Supplies' coefficient of -0.005, which is statistically and economically indistinguishable from zero. Of the nine positive coefficients, five of them are statistically significant. As with the establishment regressions, General Merchandise Stores has the largest estimated coefficient, 0.116. All of the other statistically significant coefficients equal or exceed Apparel and Accessory Stores' estimate of 0.079. These estimates are substantially greater than those for the four industries with positive but insignificant coefficients, which range from 0.015 for Drug Stores to 0.030 for Miscellaneous Retail. The estimate for All Retail Trade, 0.042, reflects populations' strong effect on average sales in some industries and its weak or non–existent effect in the others. The former group of industries is sufficiently important to lend this coefficient statistical significance. The simple interpretation of these coefficients from Equation 5 in Section 2 is that their negatives are the equilibrium elasticities of establishments' percentage markups with respect to market size. Therefore, within the context of that model, we observe that establishments' markups fall with market size for most two–digit industries, which is consistent with the estimates and interpretation of the establishment regressions.

For Apparel and Accessory Stores, *MSA* population tends to increase establishments' average employment as well as their average sales. The estimate of β_s from that industry's employment regression, 0.0495, is smaller than the estimate for its sales regression, but it is still statistically significant. The estimates of β_s for all of the other industries' employment regressions are also positive and seven of them are statistically significant. The five industries with statistically significant coefficients from the sales regressions also have statistically significant coefficients in the employment regressions. The remaining two industries with significant coefficients in the employment regression are Eating and Drinking Places and Miscellaneous Retail. A quick test of the null hypothesis that MSA population influences neither measure of average establishment size can be conducted using the conservative Bonferroni procedure based on the largest of the two coefficients' t–statistics. For a test with a 5% size, the relevant critical value is 2.24. The largest t–statistic for Eating and Drinking Places is 2.40, while that for Miscellaneous Retail is 2.07. For Building Materials and Garden Supplies, Food Stores, and Drug Stores, neither of the relevant t–statistics is above 1.96, while the largest t–statistics for the remaining industries are all well above the critical value. To summarize, the Bonferroni procedure rejects the null hypothesis for six of the ten two digit industries. According to the 1992 Census of Retail Trade, the total employment and payroll of these six industries in 1992 was $12,415,802$ persons, and $\$140,355,873,000$. Together, they account for 63% of the retail trade sector's payroll and 67% of its employment, so they represent a significant volume of economic activity.

As we noted in the introduction, Food Stores is among those industries for which the estimated coefficients in the two average size regressions are positive but statistically insignificant. Recall from Table 1 that Food Stores displays a large number of small, single unit firms and large differences in average establishment sales between establishments in multiunit and single unit firms. These features also characterize the other industries for which the Bonferroni tests fail to detect market size effects on average establishment size. Indeed, if we order the two digit industries by the average sales ratio in the third column of Table 1, Building Materials and Garden Supplies, Food Stores, Drug Stores, and Miscellaneous Retail are all among the five industries with the largest multiunit to single unit average sales ratio. What distinguishes these four industries from General Merchandise Stores, where we find large market size effects and the largest average sales ratio, is the presence of a large fringe of small, single unit firms. If markups at the "dominant" multiunit firms fall with market size but those at the "competitive" single unit fringe do not, then average sales and employment, which reflect the conduct of both types of producers, may only weakly reflect the "dominant" firms' behavior. On the other hand, any change in "dominant" firms' conduct may also change the equilibrium size of the "competitive" fringe, and so be reflected in the number of establishments per capita. More firmly establishing the role of different firms in producing our observations requires more information on establishments' firm affiliations than is available in the *CRT*, and so is beyond the scope of this paper.

4.2 Control Variable Coefficients

Table 5 shows that the estimated coefficients for many of the control variables in Apparel and Accessory Stores are statistically significant, justifying their inclusion. For example, the percentage of adults over 25 years old with bachelor's degrees significantly increases both establishments per capita and their average sales and employment for that industry. Table 7 provides a more complete overview of the control variables' estimated coefficients in the regressions. Each of the table's rows corresponds to one of the control variables included in x, and each of its columns corresponds to one of the three dependent variables we consider. The first element in each cell is the number of regressions using two digit industry data in which the corresponding t–statistic is greater than 1.96, and the second element is the number of such regressions for which the t–statistic is less than -1.96. All of the control variables appear significantly in at least five of the 30 regressions, but their importance for the exercise as a whole clearly varies. The three demographic characteristics, median income, the percentage of the population that is Black, and the percentage of the adult population with a bachelor's degree, are usually statistically significant. Furthermore, the signs of their estimated coefficients are fairly consistent across the ten industries. For example, *MSA*'s with a higher percentage of Black residents tend to have more, but smaller, establishments.

The average retail wage, median rent, and population growth are somewhat less frequently significant. When they are significant, their signs are consistent across the average sales and employment regressions. Even in a simple free entry model of price–taking competitive behavior with homogeneous establishments, the comparative statics of wages and rent on the dependent variables cannot be unambiguously signed. Therefore, it is impossible to state whether or not these coefficients' signs are surprising. In five of the ten regressions, population growth is associated with either higher average sales or higher average employment, which weakly challenges the intuition offered in Section 3.2.4 that the small new producers that enter following population growth reduce average establishment sizes. Our measures of urbanization spillovers appear to add little to the regression specifications. The unimportance of density in our regressions contrasts with Dinlersoz's (1998) results for manufacturing's two–digit industries. In seven of those industries, population density increases establishment sizes. Population itself increases establishment size in only three industries and decreases it in five industries. Thus, it appears that our finding of larger establishments in larger *MSA*'s does not simply reflect the effects of "urbanization" on establishments in all sectors.

5 Comparison with Bresnahan and Reiss

In a sample of small rural markets, Bresnahan and Reiss (1991) showed that once a market is large enough to support two or three producers, the number of customers necessary to support each additional producer is roughly constant. To attempt to resolve the apparent discrepancy between our results and Bresnahan and Reiss's, we can examine observations from *MSA*'s of the industries they study in detail. Three of their industries, doctors, dentists, and plumbers, are outside of the retail trade sector, so we are unable to address their behavior with our data. The *CRT* does contain data for the remaining two industries, druggists and tire dealers. Recall from Table 6 that Drug Stores, SIC 591, was the only industry that failed to display statistically significant market size effects in its establishment regression, although the t–statistic was very close to the appropriate 5% critical value. Although the formal statistical test reveals no necessary contradiction between our results and Bresnahan and Reiss's for this industry, the large estimated coefficient is still puzzling and warrants further examination.

The tire dealers in Bresnahan and Reiss's sample belong to SIC 55, Automobile Dealers. Their parent three-digit industry, Auto and Home Supply Stores, accounts for 43% of the two-digit industry's establishments but only 7% of its sales.[8] Therefore, we should hardly expect the results for that two-digit industry to reflect only tire dealers. Fortunately, the *CRT*'s observations for SIC 55's constituent three-digit industries are relatively complete and free of disclosure problems, so we can repeat our regression analysis for the relatively large New and Used Car Dealers and relatively small Used Car Dealers, Auto and Home Supply Stores, and Miscellaneous Auto Dealers separately.[9]

Table 8 reports the estimated coefficients on *MSA* population for these establishment, sales, and employment regressions.[10] In Auto and Home Supply Stores, *MSA* population fails to significantly impact either establishments per capita or total employment. Indeed, the estimated coefficient in the establishment regression is *positive.* The remaining three digit industries all display negative and significant coefficients in the establishment regressions and positive and significant coefficients in the two average size regressions. Indeed, the estimated effects of *MSA* population on these statistics tend to be much larger than the corresponding effects estimated using the more aggregated two-digit data. These results indicate that Bresnahan and Reiss's analysis of tire dealers is consistent with observations of that industry in much larger markets, so the discrepancy between our results and theirs does not appear to reflect counterintuitive differences in producer conduct across large and small conduct. Rather, Auto and Home Supply Stores seem to be not representative of the retail trade sector.

6 Semilinear Regression Results

Although linear regression estimation is clearly convenient, it unnecessarily constrains *MSA* population to have a log–linear relationship with the dependent variable. This is potentially problematic because a number of interesting models feature strategic interaction in small markets with few producers and its virtual absence in larger markets with many producers. One example of such a model is Cournot competition with strictly convex cost functions. An approach to characterizing market size effects in our sample without constraining them to be linear is to estimate a semilinear regression model that only constrains the control variables in x to impact the dependent variable in a linear way. For this model, the relevant regression equation is

$$y = f(s) + \beta_x' x + u \qquad (7)$$

In Equation 7, the function $f(s)$ is only assumed to satisfy mild regularity conditions, so nonparametric estimates of $f(s)$ can provide information on the shape of the relationship between *MSA* population and industry characteristics that the linear regression estimates do not capture.

Yatchew (1997) provides a simple estimator of this model based on ordering the observations in ascending order of s and differencing the resulting regression equation. Because $f(s)$ is continuous, $f(s_i)$ is approximately equal to $f(s_{i-1})$ in large samples. Therefore, β_x is the only parameter asymptotically remaining in the differenced regression equation.

$$y_i - y_{i-1} \approx \beta_x (x_i - x_{i-1}) + u_i - u_{i-1} \tag{8}$$

Applying ordinary least squares to Equation 8 yields a consistent estimator of β_x. The function $f(s)$ can then be estimated using a univariate Nadaraya–Watson estimator with s_i and the "corrected" dependent variable,

$$\widehat{y_i} = y_i - \widehat{\beta}_x x_i,$$

as the independent and dependent variables. Again, to conserve space, we report the results in detail for only Apparel and Accessory Stores and we summarize the results for the remaining industries.

Figures 3 and 4 plot the estimates of $f(s)$ from the establishment and sales regressions for Apparel and Accessory Stores. The analogous figure for the employment regression is very similar to Figure 4. The circles in the figures plot the constructed dependent variable $\widehat{y_i}$ versus s_i, and the solid line plots the estimate of $f(s)$. The estimation uses a standard Gaussian kernel function with a bandwidth of 0.2.[11] To facilitate a visual comparison of the estimated curve with a linear function, the dashed lines plot the estimated linear function from regressing $\widehat{y_i}$ against s_i using ordinary least squares.[12] The semi–linear regressions leave the same general impression as the linear regressions. Furthermore, the estimated regression functions both wiggle around the plotted linear function only slightly, leaving the impression that a linear specification for the regression function is probably adequate.

A more formal test of the hypothesis that the true regression function is linear in s_i can be constructed by comparing the sum of squared residuals from OLS estimation of β_s and β_x with the one half of the analogous sum from the application of Equation 8 to estimate β_x. Yatchew (1997) shows that subtracting the latter from the former, dividing by the latter, and scaling the result by the square root of the sample size yields a statistic with a standard normal distribution under the null hypothesis that the regression function is linear. Under the alternative hypothesis that the regression function is non–linear in s, we expect the linear equation to fit poorly relative to the non–linear equation, so we reject the null hypothesis when the test statistic is large and positive. One desirable characteristic of this test statistic is that it does not depend on the bandwidth or kernel selection for the estimate of $f(s)$.

Table 9 reports these tests for each of the thirty-three regressions we estimated using the *CRT* data. All but three of the statistics are below the 5% critical value of 1.645. Two of these rejections occur in the establishment and average employment regressions of Eating and Drinking Places. In both of these cases, the estimated regression curve tends to flatten out in larger *MSA* populations. The third rejection occurs in the average employment regression for Food Stores. In that regression, the estimate of $f(s)$ is flat over most of population's support, but it has two substantial increases, beginning about one standard deviation below and above mean *MSA* population. Thus, our earlier finding of no substantial market size effects on Food Stores' average size *may* be an artifact of our linear regression specification. Surprisingly, the test statistic for Apparel and Accessory Stores establishment regression is close to the critical value, but the remaining test statistics are well below 1.645. With these exceptions, we can conclude that the assumption of a linear regression curve is not grossly at odds with the data.

7 Conclusion

From the perspective of models in which establishments' post–entry profits are invariant to the number of competitors, our results are surprising. In such models, increases in the number of establishments entirely accommodate growth in demand, and the establishment size distribution is invariant to the size of the market. We illustrated these points with a simple two period free entry model of competition in Section 2, but they hold much more generally. Hopenhayn's (1992) model of perfect competition and Fishman and Rob's (1998) model of consumer search and monopolistic competition incorporate sunk costs of entry, producer specific uncertainty, and irreversible exit. Both of these models share the predictions of the simple model. In contrast with such models, we observe in our sample of 297 *MSA*'s that the number of retail establishments per capita declines and the average employment and sales of those establishments increases with *MSA* population. This is true both for the retail trade sector as a whole and most of its constituent two digit industries. Thus, we conclude that market size matters for the organization of retail trade industries.

Versions of Section 2's canonical model in which producers lose their market power with the addition of new competitors, as is typical of oligopoly models, can reproduce our observations, because the decline in profits per–

customer and markups must be matched by corresponding increases in customers and sales so that total profits equal the fixed cost of entry. The success of this simple model at qualitatively reproducing our findings suggests that its further development and application to local competition, as in Davis's (1998) study of cinemas' spatial differentiation, would be fruitful. Versions of these models that incorporate nontrivial producer dynamics are in their infancy. In a framework with three periods, Bagwell, Ramey, and Spulber (1995) model the evolution of a pool of ex ante identical potential entrants into a mature retail industry with a "dominant" superstore and a fringe of small monopolistically competitive firms. Introducing ongoing dynamic decisions of entry, capacity choice, and exit into such a model is likely to be difficult for the same reason that it would be interesting: strategic interaction between producers characterizes oligopoly. Ericson and Pakes (1995) made some progress on this front with a computable model of oligopolistic competition and *R&D* expenditures. The further development of these models and their application to retail trade competition is desirable.

Our empirical results also have implications for future empirical research into the birth, growth, and death of retail trade establishments, which the current development of the U.S. Census Bureau's Longitudinal Business Database is making possible. Some past empirical characterizations of establishment dynamics using manufacturing data, such as Dunne, Roberts, and Samuelson's (1989), have used perfectly competitive industrial organization theory to frame their results and have generally ignored the geographic aspects of their data. Our results suggest that perfectly competitive theory is inappropriate for such studies with retail trade data and they hint at the importance of local interactions for retail establishment dynamics. Establishment level studies can in turn contribute to understanding the ultimate sources of this paper's observations. For example, documenting how birth, growth, and exit patterns vary across *MSA*'s can determine whether establishments are larger in larger *MSA*'s because they are born larger, grow faster, or undergo a more rigorous post–entry selection process. Fischer and Harrington (1996) documented the pervasive, though not universal, geographic clustering of establishments offering similar goods and services to the public, and they modeled a producer's choice to join a cluster of competitors as trading off increased price competition with increased customer visibility. More extensively documenting the importance and development of these competitive clusters in different cities can illuminate their role, if any, in producing our observations.

References

Bagwell, Kyle, Garey Ramey, and Daniel F. Spulber. 1997. Dynamic Retail Price and Investment Competition. *RAND Journal of Economics,* **28**(2), 207–227.

Berry, Steven T. 1992. Estimation of a Model of Entry in the Airline Industry. *Econometrica,* **60**(4), 889–917.

Bresnahan, Timothy F. and Peter C. Reiss. 1991. Entry and Competition in Concentrated Markets. *Journal of Political Economy,* **99**(5), 977–1009.

Chevalier, Judith A. 1995. Capital Structure and Product–Market Competition: Empirical Evidence from the Supermarket Industry. *American Economic Review,* **85**(3), 415–435.

Ciccone, Antonio. 1997. Dynamic Externalities and the Spatial Distribution of Wages in the U.S. Unpublished Manuscript, Universitat Pompeu Fabra.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the Density of Economic Activity. *American Economic Review,* **86**(1), 54–70.

Davis, Peter. 1998. Spatial Competition in Retail Markets: Movie Theatres. Unpublished Manuscript, MIT.

Dinlersoz, Emin M. 1998. Agglomeration and Establishment Size in U.S. Manufacturing. Unpublished Working Paper, University of Rochester.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. Monopolistic Competition and Optimum Product Diversity. *American Economic Review,* **67**(3), 297–308.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989. The Growth and Failure of U.S. Manufacturing Plants. *Quarterly Journal of Economics,* **104**(4), 671–698.

Ericson, Richard and Ariel Pakes. 1995. Markov–Perfect Industry Dynamics: A Framework for Empirical Work. *Review of Economic Studies,* **62**(1), 53–82.

Fischer, Jeffrey H. and Joseph E. Harrington Jr. 1996. Product Variety and Firm Agglomeration. *RAND Journal of Economics,* **27**(2), 281–309.

Fishman, Arthur, and Rafael Rob. 1998. Consumer Inertia, Firm Growth, and Industry Dynamics. Unpublished Working Paper, University of Haifa and University of Pennsylvania.

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Schleifer. 1992. Growth in Cities. *Journal of Political Economy,* **100**(6), 1126–1152.

Gort, Michael and Steven Klepper. 1982. Time Paths in the Diffusion of Product Innovations. *Economic Journal,* **92(**September), 630–653.

Hopenhayn, Hugo A. 1992. Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica,* **60**(5), 1127–1150.

Jacobs, Jane. 1969. *The Economy of Cities.* Random House, New York.

Pakes, Ariel and Richard Ericson. 1998. Empirical Implications of Alternative Models of Firm Dynamics. *Journal of Economic Theory,* **79***(1), 1–45.*

Salop, Stephen C. 1979. Monopolistic Competition with Outside Goods. *Bell Journal of Economics,* **10**, 141–156.

Wolinsky, Asher. 1986. True Monopolistic Competition as a Result of Imperfect Information. *Quarterly Journal of Economics,* **101**(3), 493–512.

Yatchew, A. 1997. An Elementary Estimator of the Partial Linear Model. *Economics Letters,* **57**, 135–143.

Endnotes

1. In fact, Rochester has fewer Apparel and Accessory Stores per capita, 0.49 per 1000 residents, than does Detroit, 0.56 per 1000 residents. This comparison illustrates the importance of both our control variables and the regression error for determining the actual number of establishments serving an *MSA.*

2. The Census of Retail Trade withheld the datum on Apparel and Accessory Stores' sales in Terre Haute Indiana. This *MSA* was simply dropped from this regression, so Figure 2 uses 296 observations.

3. In fact, the average sales of Apparel and Accessory Stores was 11.2% larger in Detroit than in Rochester.

4. See Table 1 for the source of these figures.

5. The Census of Retail Trade reports establishment counts for Gasoline Service Stations (SIC 554) and Drug Stores (SIC 591) separately from those of their parent two-digit industries, Automobile Dealers and Service Stations (SIC 55) and Miscellaneous Retail Trade (SIC 59). This seems like a sensible separation of the data to us, and so we maintain it in our work. With some abuse of conventional terminology, we refer to these industries collectively as "two-digit" industries.

6. These statistics, as well as analogous data for the Retail Trade sector's three-digit industries, are found in Table 3 of that volume.

7. The Census of Retail Trade always reports establishment count data for each industry–*MSA* pair, but the Bureau of the Census sometimes suppresses the reporting of total employment and total sales information for disclosure reasons. These disclosure cases are obviously not a random selection from our sample, but they form a trivial proportion of our observations. In all of the two-digit industries, there are only eight observations which are suppressed for disclosure reasons. Two of these are in SIC 53, General Merchandise Stores, one is in SIC 56, Apparel and Accessory Stores, one is in SIC 591, Drug Stores, and four are in SIC 59, Miscellaneous Retail. Because they form such a small fraction of our sample, we simply dropped these disclosure cases from the estimation of summary statistics and regression coefficients.

8. These figures are derived from the sales and establishment observations in Table 1 of the 1992 Census of Retail Trade's Geographic Area Series volume for the United States.

9. The SIC codes for these industries are 551, New and Used Car Dealers; 552, Used Car Dealers; 553, Auto and Home Supply Stores; and the remainder of SIC 55, Miscellaneous Auto Dealers. This residual catagory includes both boat and motorcycle dealers and is an aggregation of four three-digit SIC codes. With some abuse of notation, we refer to this as a three-digit industry. The ability to conduct our regressions using these observations is the exception, not the rule for this data set. The Bureau of the Census frequently suppresses three-digit sales and employment data for disclosure reasons or simply fails to report *any* data for smaller MSA's, making comparable analyses for most other industries impossible.

10. For all of Table 8's establishment regressions, the full sample of 297 MSA's was used. The sales and employment data were suppressed for Used Car Dealers in 13 MSA's, for Auto and Home Supply Stores in 2 MSA's, and for Miscellaneous Auto Dealers in 15 MSA's. As with the regressions for the two-digit industries, these suppression cases were simply dropped from the sample when estimating the sales and employment regressions.

11. To assess the magnitude of this bandwidth choice, it is helpful to remember that the standard deviation of log population across our sample is approximately one.

12. Note that the estimated slope from this regression does not necessarily equal the ordinary least squares estimate of β_s from the multivariate regression. In practice, however, these estimates are close. The slopes in Figures 3 and 4 are -0.083 and 0.079 respectively.

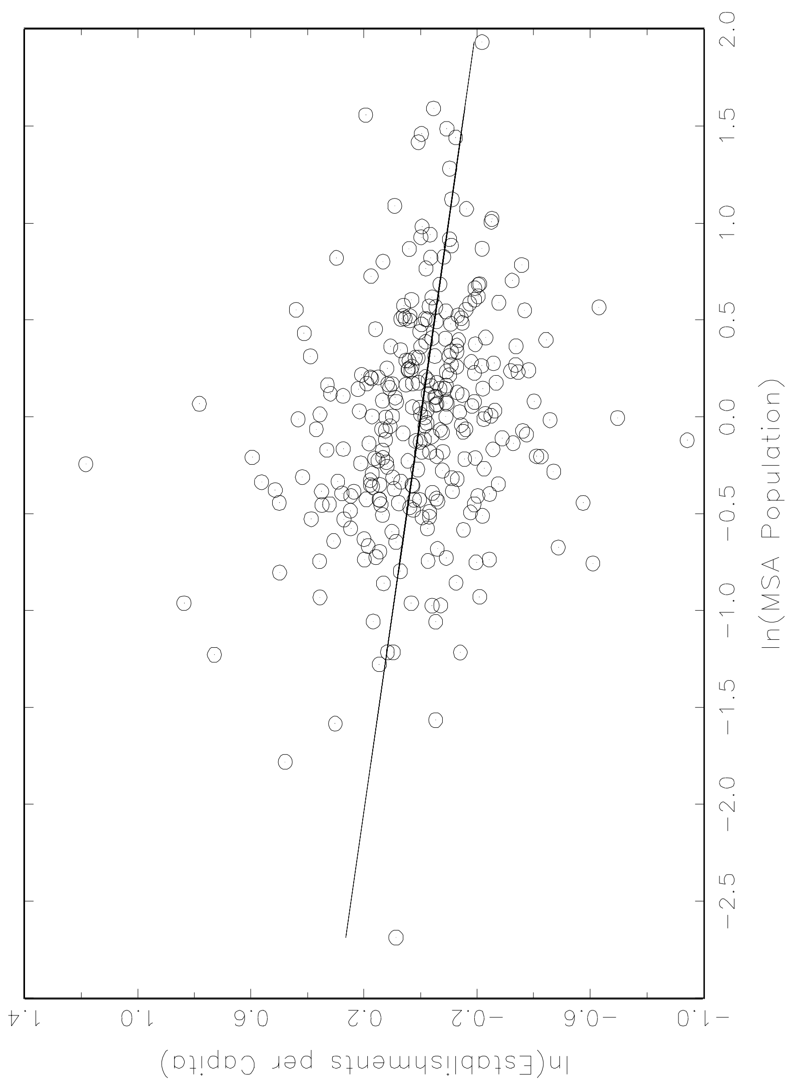

Figure 1: Establishment Regression for Apparel and Accessory Stores

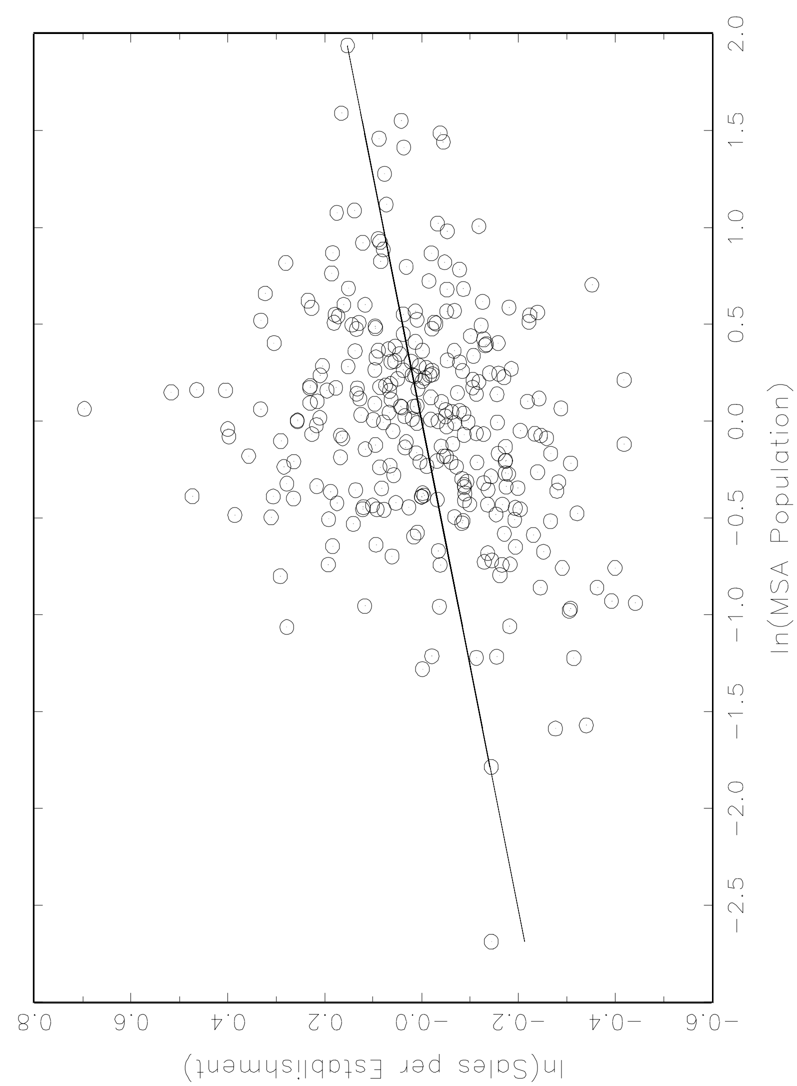

Figure 2: Sales Regression for Apparel and Accessory Stores

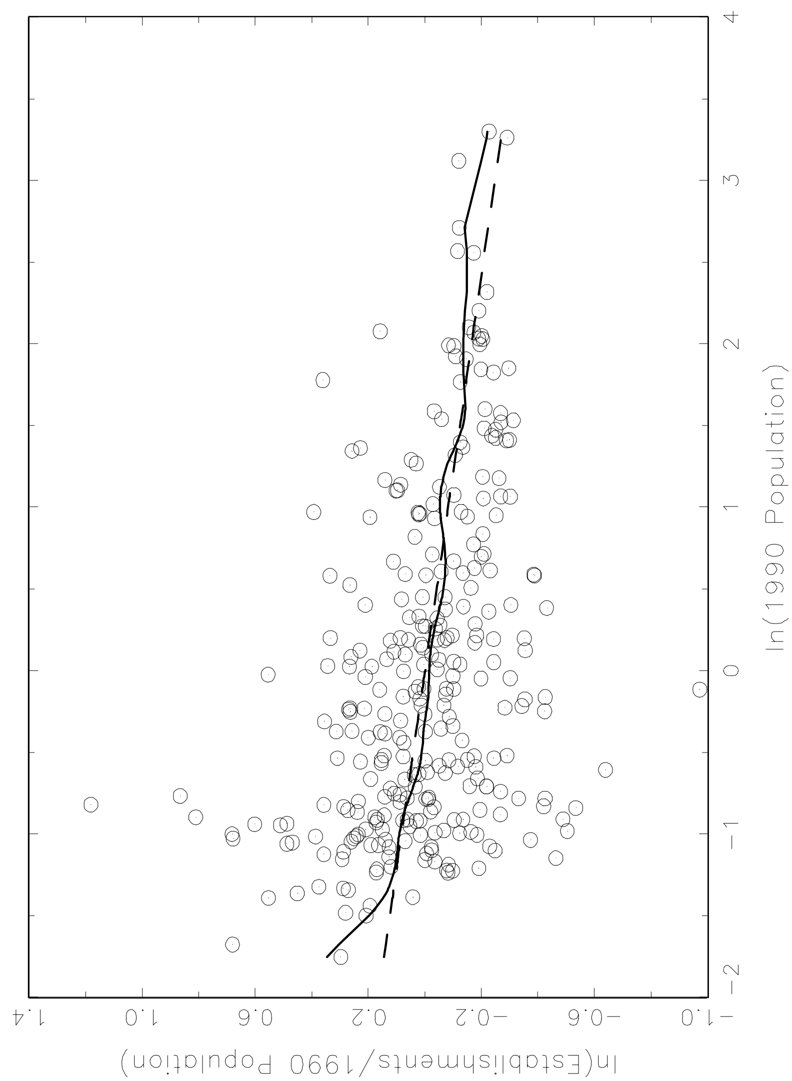

Figure 3: Semilinear Establishment Regression for Apparel and Accessory Stores

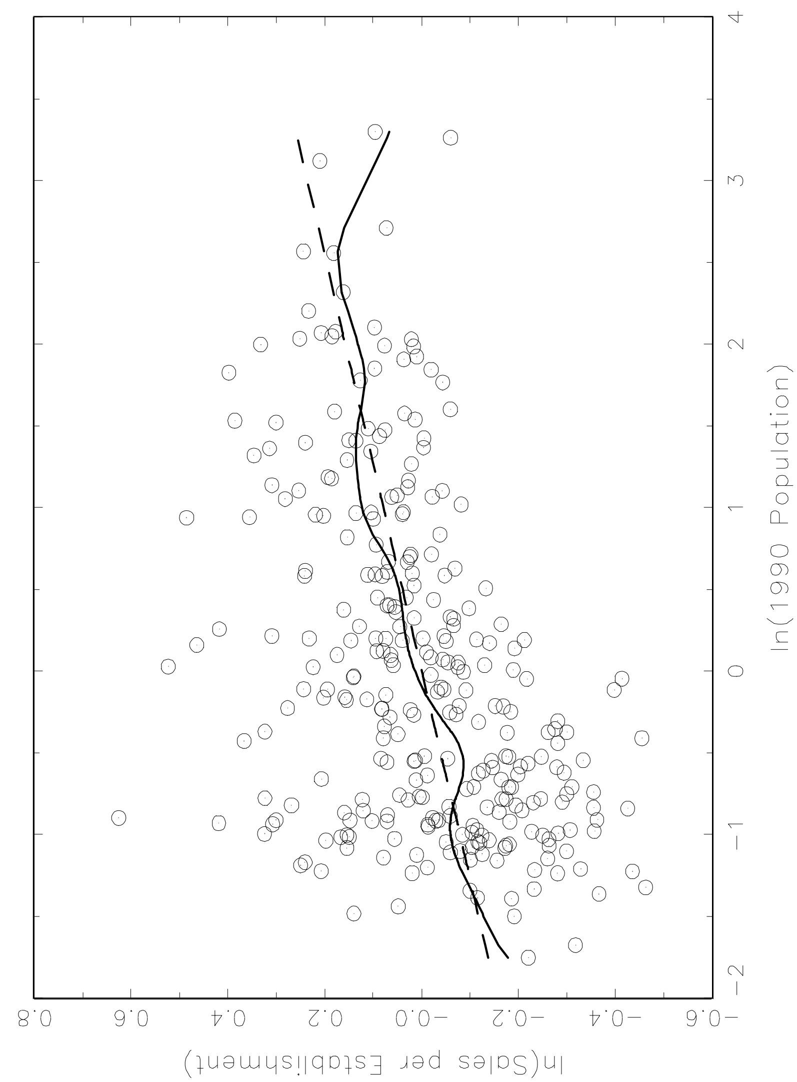

Figure 4: Semilinear Sales Regression for Apparel and Accessory Stores

Table 1: Average Establishment Sales by Firm Affiliation from the 1992 *CRT*

Industry	Firm Affiliation: Multiunit	Firm Affiliation: Single Unit	Multiunit/Single Unit	Fraction Single
All Retail Trade	2151.94	758.76	2.84	0.65
Building Materials and Garden Supplies	3072.50	809.38	3.80	0.73
General Merchandise Stores	9498.06	549.99	17.27	0.27
Food Stores	4615.98	731.08	6.31	0.66
Automobile Dealers	4281.90	4047.30	1.06	0.77
Gasoline Service Stations	1609.09	983.78	1.64	0.53
Apparel and Accessory Stores	918.73	353.64	2.60	0.39
Furniture and Home Furnishings Stores	1436.15	544.93	2.64	0.66
Eating and Drinking Places	828.21	308.28	2.69	0.73
Drug Stores	2609.79	851.40	3.07	0.55
Miscellaneous Retail	1115.22	409.14	2.73	0.72

Notes: "Multiunit" refers to establishments owned by firms with more than one establishment, and "Single Unit" refers to establishments owned by firms with only one establishment. Sales figures reported in thousands of dollars. "Multiunit/Single Unit" is the ratio of the average sales reported in the first two columns. "Fraction Single" is the fraction of establishments that are affiliated with single unit firms. See the text for further details.

Table 2: Summary Statistics for Sample of MSA's

Industry	Variable	Average	Min.	Max.	Std. Dev.
All Retail Trade	Estab.	6.33	4.13	13.14	0.139
	Avg. Sales	1279	849	2158	0.127
	Avg. Emp.	12.83	7.92	18.46	0.119
General Merchandise Stores	Estab.	0.14	0.07	0.38	0.296
	Avg. Sales	9040	2993	30043	0.348
	Avg. Emp.	75.43	28.55	142.31	0.274
Food Stores	Estab.	0.68	0.33	1.38	0.231
	Avg. Sales	2310	1026	4441	0.226
	Avg. Emp.	18.96	7.59	48.06	0.246
Automobile Dealers	Estab.	0.43	0.13	0.77	0.247
	Avg. Sales	4142	1853	8414	0.291
	Avg. Emp.	13.72	7.91	22.95	0.195
Eating and	Estab.	1.79	1.04	4.26	0.180
Drinking Places	Avg. Sales	455	243	752	0.168
	Avg. Emp.	16.26	7.87	23.21	0.175
Miscellaneous Retail	Estab.	1.27	0.70	3.25	0.209
	Avg. Sales	548	255	2042	0.311
	Avg. Emp.	5.94	3.61	12.05	0.190

Notes: Averages, minima, and maxima are always reported for the levels of the respective variables, and standard deviations are reported in logarithms. "Estab." is the number of establishments in that industry per 1000 residents, "Avg. Sales" is average sales per establishment, in thousands of dollars, and "Avg. Emp." is paid employees per establishment. See the text for more details regarding the variables' definitions and construction.

Table 3: Independent Variables Used in the Regressions

Variable	Description	Source
Population	Total MSA Residents	CCDB
Wage	Annual Retail Payroll/Employment	CRT
Rent	Median Rent of a Renter Occupied Unit	CCDB
Employment Density	Emp. per Square Mile	CBP/CCDB
Industry Concentration	Emp. Share of Five Largest Two-Digit Industries	CBP
Income	Median Family Income	CCDB
Pct. Black	% of Population which is Black	CCDB
Pct. Bachelor	% of Population over 25 with a College Degree	CCDB
Population Growth	Growth Rate between 1980 and 1990 Censuses	CCDB

Source Notes: CRT is the 1992 Census of Retail Trade, CBP is the 1992 County Business Patterns, and CCDB is the 1994 County and City Data Book.

Table 4: Independent Variables' Summary Statistics

Variable	Average	Min.	Max.	Std. Dev.	Corr. w/ Pop.
Population	627413	56735	8863164	1.03	1.00
Employment Density	403	3.97	36919	1.32	0.74
Industry Concentration	37.44	27.32	61.93	5.11	-0.28
Rent (in current dollars)	413	277	790	0.21	0.53
Wage (in current dollars)	2699	2140	4054	0.11	0.54
Income (in current dollars)	34078	17619	56884	0.17	0.48
% Black	10.60	0.08	45.77	10.20	0.17
% Bachelor	19.56	9.45	44.00	6.25	0.26
Population Growth	10.54	-15.98	62.00	13.00	0.18

Notes: Averages, minima, and maxima are always reported for the levels of the respective variables. Standard deviations are reported in logarithms if the variable enters the regression in logarithms. Otherwise, it is reported in levels. "Corr. w/ Pop." reports the variables' correlations with population's logarithm. Similarly, the correlation uses a variable's logarithm if that variable enters the regression in logarithms. See the text for more details regarding the variables' definitions and construction.

Table 5: OLS Estimates for Apparel and Accessory Stores

	Dependent Variable		
	Estab.	Avg. Sales	Avg. Emp.
Population	-0.0983	0.0792	0.0495
	(-4.60)	(5.34)	(4.01)
Employment Density	0.0618	0.0178	0.0172
	(3.47)	(1.40)	(1.46)
Industry Concentration	0.0068	-0.0978	-0.1050
	(0.02)	(-0.44)	(-0.55)
Rent	0.3970	-0.1986	-0.2508
	(2.32)	(-1.50)	(-2.40)
Wage	-0.0240	0.8378	0.1301
	(-0.09)	(4.40)	(0.82)
Income	-0.3993	-0.0657	0.1453
	(-2.00)	(-0.40)	(1.14)
% Black	0.0027	0.0010	0.0015
	(2.31)	(0.91)	(1.66)
% Bachelor	0.0086	0.0081	0.0075
	(2.68)	(3.64)	(4.09)
Population Growth	0.1846	-0.0392	-0.0585
	(1.05)	(-0.38)	(-0.63)
R^2	0.16	0.44	0.30
Sample Size	297	296	296

Notes: "Estab." refers to the logarithm of establishments per capita, "Avg. Sales" refers to the logarithm of the value of establishments' average sales, and "Avg. Emp." refers to the logarithm of establishments' average employment. Heteroskedasticity consistent t-statistics appear below each estimated coefficient in parentheses. See the text for further details.

Table 6: *MSA* Population's Regression Coefficients

Industry	Dependent Variable Estab.	Avg. Sales	Avg. Emp.
All Retail Trade	-0.080	0.042	0.042
	(-6.07)	(4.48)	(4.92)
Building Materials and Garden Supplies	-0.112	-0.005	0.026
	(-4.89)	(-0.22)	(1.36)
General Merchandise Stores	-0.146	0.116	0.102
	(-7.51)	(3.89)	(4.49)
Food Stores	-0.065	0.024	0.024
	(-3.23)	(1.42)	(1.18)
Auto Dealers	-0.077	0.096	0.079
	(-4.40)	(5.91)	(6.39)
Gasoline Service Stations	-0.088	0.079	0.044
	(-5.18)	(5.40)	(2.79)
Apparel and Accessory Stores	-0.098	0.079	0.050
	(-4.60)	(5.34)	(4.01)
Furniture and Home Furnishings Stores	-0.100	0.095	0.050
	(-5.18)	(5.66)	(4.23)
Eating and Drinking Places	-0.039	0.020	0.030
	(-2.50)	(1.63)	(2.40)
Drug Stores	-0.050	0.015	0.017
	(-1.94)	(0.62)	(0.87)
Miscellaneous Retail	-0.085	0.030	0.033
	(-5.05)	(1.04)	(2.07)

Notes: "Estab." refers to the logarithm of establishments per capita, "Avg. Sales" refers to the logarithm of the value of establishments' average sales, and "Avg. Emp." refers to the logarithm of establishments' average employment. Heteroskedasticity consistent t-statistics appear below each estimated coefficient in parentheses. See the text for further details.

Table 7: +/- Table for Control Variables' Coefficients

	Dependent Variable		
	Estab.	Avg. Sales	Avg. Emp.
Employment Density	5/0	2/3	0/2
Industry Concentration	0/1	2/1	1/0
Population Growth	3/1	4/0	4/0
Rent	3/2	0/5	0/7
Wages	1/1	5/0	0/3
Income	2/3	6/0	8/0
% Black	7/2	1/6	1/6
% Bachelor	3/5	8/1	8/0

Note: Each cell's first element gives the number of two-digit retail trade industry regressions in which the corresponding t-statistic was greater than 1.96, and each cell's second element gives the number of such regressions for which the t-statistics were less than -1.96. "Estab." refers to the logarithm of establishments per capita, "Avg. Sales" refers to the logarithm of the value of establishments' average sales, and "Avg. Emp." refers to the logarithm of establishments' average employment.

Table 8: Regression Coefficients on *MSA* Population for Auto Dealers' Constituent Industries

	Dependent Variable		
Industry	Estab.	Avg. Sales	Avg. Emp.
Auto Dealers	-0.077	0.096	0.079
	(-4.40)	(5.91)	(6.39)
New and Used Car Dealers	-0.105	0.134	0.117
	(-4.95)	(4.82)	(5.31)
Used Car Dealers	-0.155	0.064	0.091
	(-3.75)	(1.85)	(3.28)
Auto and Home Supply Stores	0.021	0.008	0.006
	(0.91)	(0.39)	(0.36)
Miscellaneous Auto Dealers	-0.112	0.176	0.137
	(-2.43)	(4.70)	(4.03)

Notes: "Estab." refers to the logarithm of establishments per capita, "Avg. Sales" refers to the logarithm of the value of establishments' average sales, and "Avg. Emp." refers to the logarithm of establishments' average employment. Heteroskedasticity consistent t-statistics appear below each estimated coefficient in parentheses. See the text for further details.

Table 9: Tests of Linear Regression Specification Against a Semilinear Alternative

	Dependent Variable		
Industry	Estab.	Avg. Sales	Avg. Emp.
All Retail Trade	1.23	0.61	1.12
Building Materials and Garden Supplies	0.42	-0.65	-1.47
General Merchandise Stores	0.55	0.37	0.03
Food Stores	-0.21	0.32	2.15
Auto Dealers	0.48	-0.40	-0.06
Gasoline Service Stations	-0.24	0.98	0.03
Apparel and Accessory Stores	1.61	1.27	-0.56
Furniture and Home Furnishings Stores	-0.52	-0.86	-0.99
Eating and Drinking Places	1.90	0.30	2.29
Drug Stores	0.14	-1.12	-1.05
Miscellaneous Retail	0.24	1.39	1.09

Notes: Each cell contains Yatchew's (1997) test statistic of the null hypothesis that the regression function is linear in all variables versus the semilinear alternative hypothesis. Under the null hypothesis, the statistics have a standard normal distribution. See Yatchew (1997) and the text for more details regrarding the test's construction. "Estab." refers to the logarithm of establishments per capita, "Avg. Sales" refers to the logarithm of the value of establishments' average sales, and "Avg. Emp." refers to the logarithm of establishments' average employment.