Market Size Matters

Jeffrey R. Campbell and Hugo A. Hopenhayn

Market Size Matters[*]

Jeffrey R. Campbell[†]   Hugo A. Hopenhayn[‡]

August, 2003

Abstract

This paper characterizes the effects of market size on the size distribution of establishments for thirteen retail trade industries across 225 U.S. cities. In most industries we examine, establishments are larger in larger cities. Models of large-group competition in which markups fall after adding competitors can reproduce this observation.

JEL Classification: L11, L16, L81

This paper empirically examines the effects of market size on producers' sizes in industries with many producers. A robust prediction of oligopoly theory is that larger markets are more competitive and have lower price-cost markups. Because producers in more competitive markets must recover their fixed costs by selling more at a lower markup, our estimated market size effects indicate whether or not this prediction of oligopoly theory carries over to large-group competition. Our analysis uses observations from thirteen narrowly-defined retail trade industries in 225 metropolitan statistical areas (MSAs), each of which we identify with a separate market. In all of the industries we consider, almost all of these MSAs contain a large number of establishments. Our primary data source is the 1992 Census of Retail Trade (CRT), from which we calculate establishments' average sales and employment in each market. We supplement these measures with observations of the empirical *c.d.f.* of establishments' sizes from the 1992 County Business Patterns (CBP). We regress these statistics from the size distribution against the MSA's market size and a set of control variables. The control variables account for differences in MSAs' factor prices and demographics, which covary with market size and can by themselves affect retailers' sizes.

Figures 1 and 2 are representative of our results. For Women's Clothing and Specialty Stores (SIC 562,3), Figure 1 plots observations from 224 MSAs of the logarithm of establishments' average sales versus the logarithm of the MSA's 1992 population. In Figure 2, establishments' average employment replaces their average sales. In these figures, both variables are defined as residuals from regressions against our control variables. The data indicate a clear positive relationship between MSA population (our baseline measure of market size) and establishments' average size. The slope of the regression line in Figure 1 equals 0.10, and the slope of Figure 2's regression line is 0.06. Both estimates are statistically significant at the 1% level. They are also economically significant: Doubling market size increases average sales by 7.1% and increases average employment by 4.4%. Ten of the other twelve industries we consider also display a positive relationship between market size and average establishment size.

Our investigation of market size's effects in industries with large numbers of producers builds on a previous literature that examines similar effects in oligopolies. This literature measures the toughness of competition, defined to be the rate at which the post-entry equilibrium markup falls with the addition of competitors, and the extent to which producers can lessen competition through product differentiation. Bresnahan and Reiss (1991) infer the toughness of price competition in local oligopoly markets by measuring the relationship between market size and the number of producers. We show below that the observed elasticity of producers' average sales with respect to market size is a lower bound for the toughness of price competition in a model of competition between many producers. For example, this result and the regression estimates for Women's Clothing and Specialty Stores imply that doubling the number of competitors decreases that industry's percentage pricecost markup (the Lerner index) by *at least* 7.1%. Davis (2001) and Mazzeo (2002) use direct measures of oligopolists' product characteristics and prices to measure the effects of product differentiation on competition and markups in local cinema (Davis) and motel (Mazzeo) markets. Product differentiation substantially lessens competition in these industries. Berry and Waldfogel (2001) provide evidence that incumbent radio broadcasters crowd their products (stations) together to preempt entry and so lessen competition. Our finding of pervasive market size effects suggests that retailers facing large numbers of competitors cannot use their product placement decisions to protect their markups indefinately.

The remainder of the paper is organized as follows. The next section motivates our empirical analysis by examining a model of competition among large numbers of producers. Section 2 describes our data. Section 3 presents the paper's empirical results, and Section 4 offers some concluding remarks.

1 Market Size, Markups, and Producers' Sizes

In this section, we illustrate the economic content of the observed relationship between market size and firm size with a free-entry model of large-group competition. In the model industry, there is an inexhaustible supply of potential entrants who can choose among several markets to enter. We index markets with i. The characteristics that distinguish the markets from each other are the number of consumers, S_i; factor prices, demographic variables, and other observable market characteristics, X_i; and an error term, U_i. The error term is independent of S_i and X_i and observed by all potential entrants. The variables in X_i and U_i account for differences in cost and demand conditions across markets that are exogenous from the perspective of the industry under consideration. All parameters describing producers' costs and consumers' demand curves are functions of X_i and U_i. Unless the resulting expression is obviously ambiguous or poorly defined, we suppress the dependence of these parameters and the model's endogenous variables.

For simplicity, we assume that each potential entrant can produce in at most one market. To produce in market i, an entrant must incur a fixed cost of entry, ϕ. Thereafter, it produces its own differentiated variety of the industry's product using a technology with constant marginal cost, c.[1] After entry, active producers simultaneously choose prices. If N producers populate market i, the demand of a producer who sets a price of p while its rivals all charge P is $S_i \times q(p, P, N)$. Here, $q(\cdot)$ is the quantity demanded of the producer by a single consumer, which is decreasing in p given P. We assume that

$$(1) \qquad q(P, P, t \times N) = q(P, P, N)/t, \quad t > 0.$$

That is, doubling the number of producers while holding all prices at P cuts each producers' demand by half. This rules out market size effects that are built into the demand system.

---

[1] We consider the robustness of our results to the assumption of a constant marginal cost below in Section 1.3.

1.1   Free Entry Equilibrium

A symmetric free-entry equilibrium consists of a price function $P^\star(N, S_i)$ and a number of producers $N(S_i)$ such that (i) the price $P^\star(N, S_i)$ maximizes the profit of any producer in market i if there are N producers serving that market and each of the others also chooses this price; and (ii) all potential entrants expect to earn exactly zero profits from entering any market.[2] Consider first the determination of $P^\star(N, S_i)$. The condition that choosing the price P maximizes each producer's profits conditional on all others' making the same choice can be written as the familiar inverse demand elasticity-markup rule.

$$(2) \qquad\qquad \frac{P - c}{P} = \eta^{-1}(P, P, N)$$

On the right-hand side of (2), $\eta(p, P, N)$ is the elasticity of a single producer's residual demand curve. To guarantee that there is a unique solution to (2), we assume that it is continuous and increasing in its first two arguments.

The solution to (2) clearly does not depend on market size, so we henceforth drop S_i from its list of arguments and write it as $P^\star(N)$.[3] So that this price is weakly decreasing in N, we also assume that if $N' > N$, then

$$(3) \qquad\qquad \eta(p, P, N') \geq \eta(p, P, N).$$

This monotonicity assumption captures the idea that increasing the number of producers weakly increases the substitutability of any one producer's product with those of its rivals, and so increases that producer's residual demand elasticity. If (3) is a strict inequality, then increasing the number of competitors erodes erodes each producer's market power.

The condition that all entrants earn zero profits following the entry of N producers is

$$(4) \qquad\qquad \phi = S_i \times q(P^\star(N), P^\star(N), N) \times (P^\star(N) - c).$$

---

[2] In this definition, we abstract from integer constraints on the number of producers that naturally arise in oligopoly models but are less likely to be important in models of large-group competition.

[3] Although the price does not depend on S_i, it may depend on the other market characteristics.

4

The right-hand side of (4) is strictly decreasing in N, so there is a unique number of active producers, $N(S_i)$ consistent with symmetric free-entry equilibrium. We denote the price those producers charge and the revenues of an individual producer with $P(S_i) \equiv P^\star(N(S_i))$ and $R(S_i) \equiv P(S_i) \times S_i \times q(P(S_i), P(S_i), N(S_i))$.

We are now in a position to consider the comparison of observable industry characteristics across large and small markets. As noted above, (3) implies that $P^\star(N)$ is weakly decreasing in N. Hence if $t > 1$, then

$$(5) \qquad\qquad N\left(t \times S_i\right) \;\leq\; t \times N\left(S_i\right),$$

$$(6) \qquad\qquad P(t \times S_i) \;\leq\; P(S_i),$$

$$(7) \qquad\qquad R\left(t \times S_i\right) \;\geq\; R\left(S_i\right),$$

The first two results are immediate. That increasing the number of consumers in market i weakly increases the revenues of each producer follows from (6) and rewriting the zero profit condition as

$$(8) \qquad\qquad \frac{P(S_i) - c}{P(S_i)} \times R(S_i) = \phi.$$

If (3) is a strict inequality, then so are the inequalities in (5), (6) and (7). That is, if additional competition increases producers' residual demand elasticities, then an increase in market size increases the number of producers less than proportionally, decreases producers' prices and price-cost markups, and increases the value of each producer's sales. On the other hand, if (3) always holds with equality, then so do (5), (6), and (7). In this case, producers' sizes and the price-cost markup are both invariant to the market's size.

1.2   Empirical Implications for the Price-Cost Markup

For our sample of U.S. cities, which we identify with distinct markets, our empirical work regresses statistics from producers' size distribution on market size and other market characteristics. For example,

$$(9) \qquad\qquad \ln R_i = m\left(S_i, X_i\right) + \epsilon_i,$$

where $R_i = R(S_i, X_i, U_i)$ is producers' average sales revenue in market i;

$$m(S_i, X_i) \equiv \mathbf{E}[\ln R_i | S_i, X_i]$$

is the regression function of $\ln R_i$ on S_i and X_i; and the error term ϵ_i reflects the unobserved market conditions U_i. It has mean zero and is uncorrelated with S_i and X_i by construction. Although the model is silent regarding the effects of market size on producers' employment, we expect that it grows and shrinks with sales and production. For this reason, we also estimate regressions in which the logarithm of producers' average sales and the empirical *c.d.f.* of their employment distribution replace $\ln R_i$ in (9).

When the regression's dependent variable is $\ln R_i$, (8) implies that $\partial m(S, X)/\partial \ln S$ measures the average rate at which equilibrium markups fall as the market expands. This quantity has a close connection with $\partial \ln \left(\frac{P^\star(N) - c}{P^\star(N)} \right) / \partial \ln N$, the rate at which the markup declines with additional entry. Sutton (1991) calls a similar partial derivative the "toughness of price competition", and so we adopt that terminology here. To connect the toughness of price competition with the slope of our regression, differentiate (8) with respect to $\ln S_i$ to get

$$(10) \qquad \frac{d \ln R(S_i)}{d \ln S_i} = -\frac{\partial \ln \left((P^\star(N(S_i)) - c) / P^\star(N(S_i)) \right)}{\partial \ln N} \times \frac{d \ln N(S_i)}{d \ln S_i}.$$

The right-hand side of (10) is the absolute value of the toughness of price competition multiplied by a quantity closely related to those measured by Bresnahan and Reiss (1991), the rate at which expanding market size induces entry. Because $0 < d \ln N(S_i)/d \ln S_i \leq 1$, the rate at which additional producers lower the post-entry equilibrium markup will exceed the rate at which additional consumers lower the free-entry equilibrium markup. Applying this inequality to (10), explicitly recognizing dependence on X_i and U_i, and averaging over U_i yields

$$(11) \qquad \frac{\partial m(S, X)}{\partial \ln S} < E \left[\left| \partial \ln \left(\frac{P^\star(N(S_i, X_i, U_i), X_i, U_i) - c(X_i, U_i)}{P^\star(N(S_i, X_i, U_i), X_i, U_i)} \right) / \partial \ln N \right| \right].$$

That is, the regression function's derivative with respect to market size provides a lower bound to the average absolute value of the toughness of price competition. This interpretation provides one way of judging our estimates' economic significance.

1.3 Robustness

Before proceeding to our empirical analysis, we consider the robustness of the model's predictions with respect to its key simplifying assumptions: symmetry and a constant marginal cost of production. It is straightforward to show that (5) and (6), and (7) still hold good if we assume that marginal cost is weakly increasing in output. Furthermore, (8) continues to hold if we replace the constant marginal cost c with average variable cost in free entry equilibrium, $AVC(S)$. In this case, the geometry of cost curves immediately implies that

$$\left| \partial \left(\frac{P(S) - AVC(S)}{P(S)} \right) / \partial S \right| \leq \left| \partial \left(\frac{P(S) - MC(S)}{P(S)} \right) / \partial S \right|,$$

where $MC(S)$ is each producer's marginal cost in a free-entry equilibrium. This and the result that the elasticity of $N(S)$ is less than one together imply that $\partial m(S, X) / \partial \ln S$ continues to bound the average toughness of price competition from below.

If marginal cost is increasing, then the possibility arises that the relevant regression coefficient could exceed the average toughness of price competition. However, in that case it is still possible to infer from the regression whether or not a class of models Hart (1985) and Wolinsky (1986) call "Chamberlinian monopolistic competition" characterize the industry under consideration. In those models, competition is anonymous, in the sense that no single producer's actions affect any other producer's profit. Campbell (2002) analyzes a general model of Chamberlinian monopolistic competition that allows for post-entry producer heterogeneity; arbitrary cost functions; sunk investment at the time of entry; non-trivial product placement decisions; and competition across an arbitrary number of dimensions including price, advertising, and quality. In that model, doubling the number of consumers simply doubles the number of producers. The distributions of their sizes and actions do

not change. We can infer from a positive estimate of $\partial m(S, X)/\partial \ln S$ that Chamberlinian monopolistic competition does not characterize the industry under consideration.

2   Data Sources

To examine empirically the relationship between market size and the producers' sizes, we use observations from thirteen retail trade industries in 225 *MSA*s.[4] Our definition of a producer is an establishment. Models of competition among large numbers of producers typically assume that each firm operates one establishment. Two examples of relevance for retail trade industries are Wolinsky's (1985) Chamberlinian monopolistic competition model with consumer search and Fisher and Harrington's (1996) model of retail store location and competition with consumer search and central places. Allowing firms to operate multiple establishments does *not* change the implication of Chamberlinian monopolistic competition that market size affects the size distributions of neither firms nor establishments. If one interprets our work as subjecting these models to empirical scrutiny, then our use of establishment-based size distribution data is not problematic. We further discuss the possible effects of multiple-establishment firms on our results in Section 3.3.

2.1   Observations of Retail Trade Industries

We selected our industries from all those below the two-digit level for which the *CRT* reports data for all *MSA*s. Because we focus on competition between large numbers of producers, we required that at least 95% of *MSA*s have ten or more establishments serving the industry. Table 1 lists the thirteen industries satisfying this criterion and their constituent SIC codes. For most of these industries, the smallest number of establishments serving *any MSA* exceeds ten. Seven of the industries are conventionally defined three-digit SIC industries, and two of them (Building Materials and Supplies and Women's Clothing and Specialty Stores) are

---

[4]Appendix A describes the data used in this paper and its sources in much more detail.

aggregates of two three-digit industries within the same two-digit industry. One industry (Furniture Stores) is a four-digit industry, and one (Homefurnishings Stores) is an aggregate of three four-digit industries. The remaining two industries, Restaurants and Refreshment Places, are each part of SIC 5812, Eating Places. The Census primarily distinguishes a restaurant from a refreshment place by the provision of table service.

From the *CRT*, we construct the dollar value of establishments' average sales and their average employment for each industry in each *MSA* of our sample.[5] The Census uses two different definitions of *MSA*s in the New England States, so we exclude those *MSA*s from our sample. We also exclude Consolidated Metropolitan Statistical Areas (*CMSA*'s), which are large urban areas such as Los Angeles and Chicago, because we doubt that our set of control variables adequately captures the differences between a typical *CMSA* and a much smaller *MSA*. Finally, we exclude any *MSA* for which the measure of commercial real estate rent that we describe below is unavailable. For the resulting sample of 225 *MSA*s, Table 1 reports sample quartiles (across *MSA*s) for establishments' average sales and employment. The sample quartiles reveal substantial variation in establishments' average sizes across *MSA*s. For both average size measures, the ratio of the interquartile range to the median is between 1/4 and 1/3 for most of the industries.

For each SIC industry and each county in the United States, the *CBP* reports the number of establishments in several employment size categories. We use these observations to construct the empirical *c.d.f.* of employment for each of these industries in each *MSA* of our sample, evaluated at three of the categories' upper boundaries: 9, 19, and 49 employees. We henceforth denote these with $F(9)$, $F(19)$, and $F(49)$. The *CBP* does not report separate

---

[5] The Census sometimes withholds these observations for a particular industry-*MSA* pair when their publication would reveal private information from a particular producer. Because we examine industries in metropolitan areas with relatively large numbers of producers, these instances of data suppression are rare in our data set. However, they do occur in eight of our thirteen industries. In our empirical analysis, we simply drop these observations. Appendix A reports the extent of this problem for each industry. We note here that it is particularly severe for Homefurnishings Stores, where it eliminates 18 *MSA*s from the analysis.

observations for Restaurants and Refreshment Places, so we examine instead the four-digit industry to which they belong, Eating Places. These observations allow us to detect effects of market size on establishments' sizes that have little or no effect on their *average* size, such as an increase or decrease in dispersion. For each industry, Table 1's final three columns report the average values across MSAs of $F(9)$, $F(19)$, and $F(49)$. In the average MSA, the majority of establishments have fewer than ten employees in all industries but New and Used Car Dealers and Eating Places. At least 90% of the establishments in the average MSA have fewer than 50 employees in all of the industries except Grocery Stores and New and Used Car Dealers. It appears at least one of the three points at which we evaluate each industry's empirical *c.d.f.* lies in the relevant support of its size distribution.

2.2   Measures of Market Size

In our regressions, we use three distinct measures of market size. Our baseline measure is the simplest, MSA population in 1992. This comes from the 1994 County and City Data Book ($CCDB$). If producers primarily differentiate their products across geographic space, then the relevant market size measure is geographic population density, our second measure of market size. We measure an MSA's population density with the population-weighted average of population density in each of its constituent counties. We consider our third measure of market size, the value of industry sales, because it may reflect heterogeneity of consumers across MSAs that our control variables do not adequately measure. If industry demand is unit elastic, the value of industry sales is invariant to producer conduct and accurately measures market size.

2.3   Control Variables

If producers face higher fixed costs in larger markets, then they will contain larger stores even if adding competitors does not lower markups. The same observation can arise if larger markets contain consumers with more elastic demand curves. Thus, our ability to infer how

10

additional competitors change markups from our regressions depends on first controlling for observable factor prices and demographic variables; as in the model of Section 1.

Table 2 lists the independent variables we include in our baseline regression specification and their sources and reports their basic summary statistics. To control for differences in retailers' costs of production across MSAs, we include the prices of labor, advertising, and commercial real estate. All of these inputs are locally traded, so their prices should vary across cities. To measure the price of labor, we divided the first-quarter payroll of the MSA's retail trade sector by its Mid-March employment count, both as reported in the CRT. The price of advertising is the price per 1000 exposures of a standard column inch in a Sunday newspaper, and the price of commercial real estate is the median rent per square foot of retail space in the MSA's strip malls.[6] These prices all appear in logarithms in our regressions. We also include demographic characteristics to control for differences across MSAs in consumers' preferences that can have direct effects on producers' sizes by changing the composition of goods the industry produces. The demographic characteristics we include are the MSA's average personal income, the percentage of the MSA's residents who are Black, the MSA's adult college attainment rate, and the number of vehicles per household. Average personal income enters our regressions as a logarithm, while the other demographic characteristics appear in levels. Table 2's final column reports these variables' correlations with MSA population's logarithm. None of these correlations are above 0.4 in absolute value. With the exception of average personal income, those corresponding to the demographic characteristics are particularly small. Wages and rents tend to be somewhat higher and advertising costs somewhat lower in larger MSAs.

---

[6]We collected the data on these last two prices ourselves. Appendix A provides much more detail regarding their original sources and construction.

3 Estimation Results

In this section, we report the results of estimating our multivariate regression specification using our three measures of market size and a variety of estimation techniques. Because their interpretation is intuitive and straightforward, we begin by considering the estimation of simple linear regression equations for establishments' average sales and employment. That is we assume that

$$(12) \qquad\qquad m\left(S_i, X_i\right) = \beta_S \ln S_i + \beta_X X_i$$

Economic theory imposes no particular functional form on our regression equations, and a linear regression for the measures of the empirical *c.d.f.* seems particularly inappropriate because this variable must be between zero and one. For these reasons, we also follow an alternative estimation strategy which imposes no functional form assumptions on $m\left(S, X\right)$. The regression function's density-weighted average derivatives are defined as

$$(13) \qquad\qquad \begin{aligned} \delta_S &\equiv \mathbf{E}\left[\frac{\partial m\left(S, X\right)}{\partial \ln S} f\left(\ln S, X\right)\right] / \mathbf{E}\left[f\left(\ln S, X\right)\right] \\ \delta_X &\equiv \mathbf{E}\left[\frac{\partial m\left(S, X\right)}{\partial X} f\left(\ln S, X\right)\right] / \mathbf{E}\left[f\left(\ln S, X\right)\right], \end{aligned}$$

where $f\left(\ln S, X\right)$ is the joint density function of $\ln S$ and X across markets and expectations are taken with respect to the same joint density function. If (12) describes the true regression function, then $\delta_S = \beta_S$ and $\delta_X = \beta_X$. Powell, Stock, and Stoker (1989) provide a simple instrumental variables estimator of δ_S and δ_X which converges to the true parameter values at the parametric rate of $\sqrt{N}$. We apply this estimation method to both of the average size measures and the three evaluations of the size distribution's empirical *c.d.f.*

3.1 Linear Regression Results

We begin by considering the results of simple linear regressions on our data. The analysis of the two average size measures and three market size measures across our thirteen industries

produces 65 regressions. To conserve space, we report complete results for only one industry using our baseline measure of market size. For the remaining industries and market size specifications, we only report the estimates of the regression coefficients on market size and summarize the estimates of the control variables' coefficients.

Table 3's first two columns report the estimated regression coefficients for Women's Clothing and Specialty Stores, estimates of their standard errors, and the regressions' R^2 measures.[7] The estimated elasticities of establishments' average sales and employment with respect to MSA population are 0.10 and 0.06. As noted in the introduction, both of these are statistically significant at the 1% level. The R^2 measures from these regressions are 0.32 and 0.26. In the average sales regression, two of the control variables enter significantly, the college attainment rate and vehicle ownership. These variables, the retail wage, and average personal income have significant coefficient estimates in the average employment regression. Table's 3's last two columns provide an overview of the remaining regressions' estimated coefficients. Each cell reports the number of industries for which the corresponding t-statistic is greater than 1.96 and the number for which it is less than -1.96. The coefficient on MSA population in the average sales regression is positive and statistically significant in seven of the thirteen industries. The analogous coefficient in the average employment regression is positive and statistically significant in six of the industries.

The control variables' importance for the exercise as a whole clearly varies. The only control variable that appears significantly more frequently than not in both regressions is the college attainment rate, which always enters positively. The percentage of MSA residents who are Black also enters significantly in nine industries' average sales regressions and in six industries' average employment regressions. This variable's sign varies across industries. Average personal income appears positively and significantly in three industries' average sales regressions and in five industries' average employment regressions. Vehicle ownership appears to be relatively unimportant. Of the three factor prices in our regressions, the retail

---

[7]We report White heteroskedasticity-consistent standard errors for all linear regression coefficients.

wage is clearly the most important. It appears positively and significantly in five industries'
average sales regressions. In the model of Section 1, these positive coefficients could reflect
these industries' overhead labor requirements. In two of these industries the retail wage
also appears positively in the average employment regressions. In four *other* industries it
appears negatively, as if these industries producers can substitute other inputs for labor as
the wage rises. Our measure of commercial real estate costs is never statistically significant,
a point to which we return below. The advertising cost is only significant for one industry,
Homefurnishings Stores.

For all of the industries we consider, Table 4's first column reports the coefficients on
MSA population from the average sales regression, and Table 5's first column reports those
coefficients from the average employment regressions. The two regressions yield mutually
consistent inferences regarding population's effect on average establishment size for ten of
the thirteen industries. The three industries for which the estimates' statistical signifi-
cance at the 5% level is not the same across the two equations are Building Materials and
Supplies, Gasoline Service Stations, and Furniture Stores. The estimates from the aver-
age employment regressions tend to be smaller than those from the average sales regres-
sions. For both sets of regressions, the industry with the largest estimated coefficient is
Radio/TV/Computer/Music stores. To assess the importance of the control variables' for
these results, the second columns of Tables 4 and 5 report the analogous coefficients from
bivariate regressions. For most of the industries we consider, adding the control variables
changes the estimates and their statistical significance very little. Most of the thirteen in-
dustries display positive and statistically significant market size effects on establishments'
average size.

3.1.1   Instrumental Variables Estimates

Because the U.S. Census provides no comprehensive measure of the cost of commercial real
estate for all *MSA*s, we constructed our own measure based on quoted rents per square foot

of strip mall space in the 1993 *Shopping Center Directory*. For some of the *MSA*s in our sample, this measure is based on a small number of quoted rents, so it is possible that the failure of commercial rent to appear significantly in our regressions reflects measurement error. To account for this possibility, we have also estimated our equations using both the median rent of a renter-occupied housing unit and the median value of an owner-occupied housing unit as instruments for our potentially error-ridden measure of commercial rent.

The third columns of Tables 4 and 5 report the estimated coefficients on *MSA* population from this instrumental variables procedure. Accounting for possible measurement error in commercial rent substantially changes three industries' estimated market size effects. For New and Used Car Dealers and Furniture Stores, the estimated coefficients are much larger, nearly doubling in the former industry. The statistical significance of *MSA* population in Furniture Stores' average employment regression rises to the 5% level. In both of these industries, commercial rent appears negatively and is significant at the 10% level (5% in the case of New and Used Car Dealers' average sales). In Gasoline Service Stations, the estimated coefficient from the average sales regression drops from 0.055 to 0.022 and loses its statistical significance, while the coefficient in the average employment equation switches signs and drops in magnitude. For this industry, commercial rent enters positively and significantly— at the 10% level in the average sales regression and at the 5% level in the average employment regression. The only remaining industry where commercial rent appears to be important is Drug and Proprietary Stores, where it enters positively and significantly at the 5% level in both regressions. However, this does not change the inference that there are no effects of market size on average establishment sizes in that industry. Overall, accounting for possible measurement error in commercial rent makes this variable appear to be significant for a few industries, and for two of these the pattern of statistical inference regarding market size effects is unchanged.

15

3.1.2 Alternative Market Size Measures

The fifth columns of Tables 4 and 5 present OLS coefficient estimates and their standard errors from regressions that use population density instead of population to measure market size. If the primary dimension of product differentiation is geography, then we expect this to more accurately measure market size. The coefficient estimates and the patterns of inference are largely unchanged for nine of the industries. However, four of the industries where measured market size effects were weak or non-existent using population display positive and significant effects of population density on establishments' average sales and employment. In Building Materials and Supplies, Shoe Stores, and Refreshment Places, both estimates are positive and statistically significant at the 1% level. For these three industries, the hypothesis of geographic product differentiation seems particularly worth further pursuing. The estimated coefficients in Drug and Proprietary Stores' regressions are both positive and statistically significant at the 5% level, but this significance is not robust to instrumenting commercial rent with residential real estate prices.

The estimated coefficients on the value of industry sales from the average employment regressions that use it to measure market size are reported in Table 5's final column. We do not attempt to measure the dependence of establishments' average sales on total sales, because this is equivalent to the exercise of measuring the dependence between the number of producers and market size. If this measure of market size is measured with error, then the coefficients are biased upwards, even if the true market size effect equals zero.[8] All of the estimated coefficients are positive. In all but one industry they are statistically significant at the 1% level. For two of the industries, Gasoline Service Stations and Drug and Proprietary Stores, analogous (unreported) coefficients from the instrumental variables procedure are smaller and not significantly different from zero.

---

[8]This is not the case with our other two measures of market size. In that case, classical measurement error imparts no bias on the OLS estimates if the true market size effect equals zero.

3.1.3   Alternative Specifications

In addition to the regressions described above, we have also estimated versions of (12) using alternative definitions of the control variables and alternative samples. These estimates indicate that the results reported in Tables 4 and 5 are very robust. One control variable we added to the regressions was the *MSA*'s population growth rate between 1980 and 1992. Seminar participants have suggested to us that the market size effects we document may be a reflection of a positive effect of market *growth* on establishment size. This could be so if industry growth primarly reflects the growth of incumbent establishments. If instead entry of new small establishments primarly drives industry growth, then recent market growth should negatively affect establishment size. Adding population growth to our set of control variables changes none of the point estimates substantially and alters no statistical inferences. Furthermore, the estimated coefficient on population growth is not statistically significant in most of the industries' regressions.

Another potential explanation of our results is that they reflect the effects of technological spill-overs that are present to a greater extent in large *MSA*s. To investigate this hypothesis, we have included the measure of spill-overs from urbanization that Glaeser, Kallal, Scheinkman, and Schleifer (1992) found to be most useful in forecasting wage growth, the share of the *MSA*'s employment accounted for by its five largest two-digit industries. Low values of this share indicate that the *MSA* has a diverse industrial base that is a fruitful generator of spill-overs. Adding industry diversity to the regressions' control variables also changes their point estimates and test statistics very little. Just as with population growth, the coefficients multiplying industry diversity tend to be statistically insignificant.

As a final check on the robustness of our results, we have estimated our regressions using only the 75% largest *MSA*s (measured with 1992 population). These estimates are a check against the possibility that our results primarily reflect the transition from an oligopolistic market structure in the smallest *MSA*s to a competitive market structure in the largest *MSA*s. With this sample, the market size effects in Homefurnishings Stores and New and

Used Car Dealers are more sensitive to the definition of market size. The remaining point estimates and statistical inferences reported in Tables 4 and 5 are not substantially changed.

3.2 Nonparametric Average Derivative Estimates

The fourth columns of Tables 4 and 5 report the estimates of density-weighted average derivatives, δ_S, for the average sales and employment regressions using our baseline measure of market size.[9] The estimates for most industries are positive, statistically significant, and exceed the linear regression estimates. The lack of robustness of Gasoline Service Stations' linear regression estimates to controlling for measurement error in commercial rent suggests interpreting this industry's estimates of δ_S with caution. In Grocery Stores, the estimates of δ_S are 0.125 and 0.153 for average sales and employment. The estimate in the average employment regression is similar if we measure market size with the value of industry sales, but these estimates become much smaller and lose their statistical significance when population density replaces population. This leads us to conclude that there *may* be significant market size effects in Grocery Stores.[10] For all of the industries except Grocery Stores, these nonparametric estimates reinforce the conclusions drawn from the linear regression estimates.

Our estimates of δ_S for the regressions with $F(9)$, $F(19)$, and $F(49)$ as the left-hand side variables are reported in the first three columns of Table 6. To illustrate the control variables' effects on our estimation and inference, the last three columns of Table 6 report analogous estimates for bivariate nonparametric regressions. For one industry, New and Used Car Dealers, the estimates of δ_S from all three multivariate regressions are negative

---

[9]Implementing Powell, Stock, and Stoker's (1989) estimator requires choosing a kernel function and a bandwidth parameter. We detail these choices and our results' robustness to them in Appendix B.

[10]For the other twelve industries, the statistical inferences in Table 13 are largely invariant to replacing population with one of our alternative market size measures. However, the point estimates of δ_S vary widely across these specifications.

and statistically significant at the 1% level. That is, the size distribution in a large *MSA* apparently stochastically dominates that from an otherwise identical but smaller *MSA*. In Grocery Stores, Women's Clothing and Specialty Stores, Furniture Stores, and Homefurnishings Stores, at least one of the three estimates of δ_S is negative and statistically significant at the 5% level, while none of the estimated coefficients is positive and statistically significant.

A more subtle pattern of market size effects emerges from this estimation in four industries where effects of market size on establishments' *average* employment are either not strong or nonexistent. In Auto and Home Supply Stores, the estimates of δ_S for the regressions of both $F(9)$ and $F(19)$ are both *positive* and statistically significant at the 1% level, while that from the regression of $F(49)$ is negative and significant at the 5% level. As market size increases, a greater fraction of establishments lie in *either* tail of the size distribution. The pattern of market size effects in Radio/TV/Computer/Music Stores and in Eating Places is similar. Market size also significantly affects the dispersion of establishments' sizes in Drug and Proprietary Stores, but in that industry market size decreases dispersion.[11] For these four industries, the finding of small or unstable effects of market size on establishments' *average* employment seems to reflect more complicated but potentially interesting effects of market size on the dispersion of establishment sizes.

3.3   Summary

If we take as given that large-group competition characterizes retail markets with a large number of *establishments*, then our evidence supports the claim that larger markets are more competitive for most of the industries we examine. If we consider the estimates reported in Table 4, the smallest and largest that are also significant at the 5% level equal 0.04 and 0.24. The corresponding percentage changes to establishments' average sales from doubling

---

[11]Because we found that this industry's establishments' average sales and employment depended on commercial real estate prices, which may be measured with error, we cautiously interpret this result as suggestive of market size effects on the dispersion of establishments' sizes.

market size are 3% and 19%. Because our estimates are lower bounds for the toughness of price competition, they imply that doubling the number of competitors in a market decreases markups by at least this much. Market size tends to affect establishments' average employment less than their average sales.

4   Conclusion

We can think of at least one possible explanation for our findings that does not rely on markups falling with market size. Sutton (1991) has shown that introducing an opportunity to bid for consumers' business by making sunk investments in product quality can result in a market structure with only few firms, even in arbitrarily large markets. Furthermore, as the number of consumers increases, firms' quality bids may increase commensurately. Bagwell and Ramey (1995) emphasize the variety of goods offered for sale at a given establishment as an important dimension of retailers' quality. If stores with larger variety have larger sales and more employees, then competition between a few firms to provide high variety can produce a positive relationship between market size and establishments' sizes. We think that this approach may be particularly relevant in two of our industries well-known for containing large "category killer" firms, Building Materials and Supplies and Radio/TV/Computer/Music Stores. It is also well-known that most local grocery store markets are dominated by a handful of firms, so this explanation may also help explain the moderate market-size effects we find in that industry.[12] Determining the relevance of such competition for the industries we consider is on our agenda for future research.

For the six industries where we find large, robust, and positive effects of market size on establishments' average sales and employment, it appears that competition is tougher in larger markets. It is also clearly possible that this is the case for the other five industries where the importance of market size for average sales and employment depends on either

---

[12]See Chevalier (1995) for evidence regarding concentration in U.S. cities' grocery store markets.

the measure of market size or the use of nonparametric estimation. For two industries, Radio/TV/Computer/Music Stores and Eating Places (Refreshment Places and Restaurants), our estimates clearly show that establishments' average size and their dispersion are both larger in larger markets. Simple symmetric models in which competition is tougher in larger markets, such as Salop's (1979), can explain the effects of market size on average establishment sizes, but by assumption they abstract from the possibility that dispersion depends on market size. Bagwell, Ramey, and Spulber (1995) develop a model of a retail trade industry in which the distribution of producers' sizes is the endogenous outcome of competition in cost-reducing investment and the effort to create a reputation for low prices among imperfectly informed consumers, and Ericson and Pakes (1995) provide a general framework for modeling the stochastic evolution of an oligopolistic industry with cost-reducing or demand-enhancing investment. In these models, the size distribution reflects producers' strategic interactions as well as the exogenous shocks emphasized by Jovanovic (1982) and Hopenhayn (1992). Our observations suggest to us that the further development of such models and their application to retail industries is a fruitful area for future research.

We expect the results of our analysis also to help guide future empirical work using establishment-level and firm-level observations from the retail trade sector, such as those examined by Pakes and Ericson (1998) and Foster, Haltiwanger, and Krizan (2001). It is our hypothesis that the positive relationship between market size and dispersion in establishments' sizes that we detect in a few industries is much more pervasive than we can observe using our relatively crude observations. Establishment-level observations from the retail trade sector can immediately determine whether or not this is the case.

Appendix A   Data Appendix

Here we list the original data sources and the methods we use to construct our observations. The three primary original source files we use are the Census of Retail Trade data on the 1992 Economic Census Report Series Disk 1i (*CRT*), the 1992 County Business Patterns file from ICPSR Study # 6488 (*CBP*), and the 1994 County and City Data Book (*CCDB*). We place variable names from the original source files in typewriter font, for example, "`value`".

- **Average Sales and Employment** For each of our industries, the *CRT* reports the total value of industry sales for 1992 (`value`), the number of paid employees for a mid-March pay period (`emp`), and the number of establishments that operated in the *MSA* during that year (`estab`). Our measures of average sales and employment are constructed directly from these observations. As we noted above in Footnote 5, the Census sometimes withholds observations of the value of industry sales and total industry employment for a particular industry-*MSA* pair if their publication would disclose a Census respondent's private information. We drop these industry-*MSA* pairs from our analysis. This results in the loss of one *MSA* each for Women's Clothing and Specialty Stores, Furniture Stores, Radio/TV/Computer/Music Stores, and Restaurants, two *MSA*s for Auto and Home Supply Stores and four *MSA*s for Shoe Stores. This disclosure problem is more severe in Homefurnishings Stores, where it eliminates 18 *MSA*s.

- **Empirical *c.d.f.*** The 1992 *CBP* reports the total number of establishments operating at any time during the year (`testab`) and the number of such establishments with mid-March payrolls falling into the following categories: 1 to 4 employees (`ctyemp1`), 5 to 9 employees (`ctyemp2`), 10 to 19 employees (`ctyemp3`), and 20 to 49 employees (`ctyemp4`). We construct $F(9)$ by adding the `ctyemp1` and `ctyemp2` and dividing by `testab`. The other measures of the empirical *c.d.f.* are constructed analogously.

- ***MSA* Population** This is the *MSA*'s 1992 population, Item 002 of the *CCDB*.

- **Population Density** This is calculated as the population-weighted average across all of the *MSA*'s constituent counties of raw population density, where population and land area for each county are taken from Items 002 and 001 of the *CCDB*.

- **Industry Sales** This is the variable `value` from the *CRT*.

- **Retail Wage** This is calculated as first-quarter payroll for all retail establishments in the *MSA* (`pay1q`)) divided by those establishments' mid-March employment count, (`emp`), from the *CRT*.

- **Commercial Rent** This variable is based on observations from the 1993 *Shopping Center Directory* (National Research Bureau, Chicago). This directory lists shopping malls and their characteristics for each *MSA*. For each *MSA*, we tabulated every report of average rent per square foot given in the directory if the entry was for a strip mall. These are self-reported (by the shopping center's manager) observations, so they are incomplete. Furthermore, the shopping center's manager frequently lists a range, such as $7 − $9 rather than a single average. When a report listed a range, we took the average rent to be the middle of that range. Our resulting data set contains price quotations from nearly 3000 malls. We then eliminated outlying observations by throwing out the smallest and largest 5% of these quotes. From the resulting data set, we measured each *MSA*'s median rent per square foot.

- **Advertising Cost** For each *MSA*, we found the cost of a standard column inch Sunday newspaper advertisement for each newspaper serving it and those newspapers' Sunday circulations from the 1992 *Editor and Publisher International Yearbook* (Editor and Publisher, New York). Our measure of advertising costs is the circulation-weighted average of these advertisements' cost per exposure.

- **Income** This variable is per capita personal income for 1992 taken from the Bureau of Economic Analysis Regional Accounts Local Personal Income data, available on the

world wide web at http://www.bea.doc.gov/bea/regional/data.htm.

- **Percent Black** This is the percentage of the *MSA*'s residents who are Black, calculated from Item 010 (Population by Race, Black, 1990) and Item 005 (Population, 1990) in the *CCDB*.

- **Percent College** This is the weighted average across the *MSA*'s counties of Item 071 (Persons 25 years or older. Percent w/ Bachelor's Degree or Higher, 1990) in the *CCDB*, where the weights are proportional to Item 069 in the *CCDB* (Persons 25 years or older, 1990).

- **Vehicle Ownership** This is the weighted average across the *MSA*'s counties of Item 117 (Vehicles per Household, 1990) in the *CCDB*, where the weights are proportional to Item 035 in the *CCDB* (Households, 1990).

- **Median Rent** This is the weighted average across the *MSA*'s counties of Item 108 (Median Rent of a Renter-Occupied Housing Unit) in the *CCDB*, where the weights are proportional to Item 107 (Renter-Occupied Housing Units) in the *CCDB*.

- **Median Value** This is the weighted average across the *MSA*'s counties of Item 105 (Median Value of an Owner-Occupied Housing Unit) in the *CCDB*, where the weights are proportional to Item 103 (Owner-Occupied Housing Units) in the *CCDB*.

Appendix B   Nonparametric Estimation Choices

Implementation of the density-weighted average derivative estimator requires the choice of a multivariate kernel function and a bandwidth for the preliminary estimation of $f\left(\ln S, X\right)$. We use a higher-order kernel function, as Powell, Stock, and Stoker (1989) recommend for the elimination of their estimator's asymptotic bias. We follow Bierens (1987) by choosing our kernel function, $K\left(u\right)$, to be

$$K\left(u\right) = \sum_{j=1}^{m} c_j \exp\left(u'u/j\right)/\sqrt{2\pi j}^k,$$

where k is the dimensionality of u. In our case, k equals the dimensionality of X plus one. The constants c_j are chosen as in Bierens (1987) so that the first $2m+1$ moments of the vector random variable with "density" $K\left(u\right)$ equal zero.[13] The order of $K\left(u\right)$, indexed by m, is chosen as in Powell, Stock, and Stoker (1989).

The only restrictions which the asymptotic theory of the instrumental variables average derivative estimator places on the bandwidth regard its rate of convergence to zero as the sample size grows to infinity, so theory offers little practical advice regarding the bandwidth's selection given a finite sample. The estimator's asymptotic distribution does not depend on either the choice of a kernel function or the bandwidth sequence, but the possibility that the finite sample distribution does depend on these quantities is clear. To guide our bandwidth choice, we conducted a small Monte Carlo study of the estimator's behavior using the bias-reducing kernel function. Our design mimics Powell, Stock, and Stoker's (1989) study. The true regression function $m\left(S, X\right)$ is linear in $\ln S$ and X, $\ln S$ has a chi–squared distribution with three degrees of freedom, and X and u are scalar random variables with independent standard normal distribution. The experiments used samples of 250 observations generated from this design. We found that the instrumental variables average derivative estimator is nearly unbiased, regardless of the choice of bandwidth. However, the estimator's variance decreases with the bandwidth. Therefore, it appears that "over smoothing" in the first

---

[13]The word "density" is put into quotation marks because $K\left(u\right)$ is not non-negative almost everywhere.

stage estimation of $f\left(\ln S, X\right)$ has no adverse consequences for the estimator's behavior, but "under smoothing" can reduce its informativeness. With this in mind, we chose our bandwidth to equal 2, and we scaled all of the regression's variables to share the standard deviation of MSA population's logarithm, 0.86. The inferences we report from the average derivative estimation are robust to changing the bandwidth to either 1 or 3.

References

Bagwell, Kyle and Garey Ramey. 1995. Coordination Economies, Advertising, and Search Behavior in Retail Markets. *American Economic Review,* **84**(3), 498–517.

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. and Joel Waldfogel. 2001. Do Mergers Increase Product Variety? Evidence from Radio Broadcasting. *Quarterly Journal of Economics,* **116**(3), 1009–1025.

Bierens, Herman J. 1987. Kernel Estimators of Regression Functions, in *Advances in Econometrics: Fifth World Congress, Volume I*, Truman F. Bewley, ed. Cambridge University Press, Cambridge.

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

Campbell, Jeffrey R. 2003. A Robust Prediction of Chamberlinian Monopolistic Competition. Working Paper. (Available at `http://www.nber.org/~ jrc`.)

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

Davis, Peter. 2001. Spatial Competition in Retail Markets: Movie Theaters. Unpublished Working Paper, MIT Sloan School.

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.

Foster, Lucia; John Haltiwanger; and C.J. Krizan. 2000. The Link Between Aggregate and Micro Productivity Growth: Evidence from Retail Trade. Unpublished Working Paper. U.S. Census and University of Maryland.

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

Hart, Oliver. 1985. Monopolistic Competition in the Spirit of Chamberlin: A General Model. *Review of Economic Studies,* **52**(4), 529–546.

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

Jovanovic, Boyan. 1982. Selection and the Evolution of Industry. *Econometrica,* **50**(3), 649–670.

Mazzeo, Michael J. 2002. Product Choice and Oligopoly Market Structure. *RAND Journal of Economics,* **33**(2), 221–242.

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

Powell, James L., James H. Stock, and Thomas M. Stoker. 1998. Semiparametric Estimation of Index Coefficients. *Econometrica,* **57**(6), 1403–1430.

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

Sutton, John. 1991. *Sunk Costs and Market Structure.* MIT Press, Cambridge, MA.

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

Table 1: Summary Statistics from *Census of Retail Trade* and *County Business Patterns* Data[i]

Industry	SIC Code(s)[ii]	Average Sales[iii,iv,v]			Average Employment[iii,iv]			Employment c.d.f[vi]		
		Q(1)	Median	Q(3)	Q(1)	Median	Q(3)	F(9)	F(19)	F(49)
Building Materials & Supplies	521,3	1900	2235	2602	11.4	13.3	15.9	0.66	0.82	0.94
Grocery Stores	541	2354	2820	3516	18.5	22.9	28.0	0.62	0.75	0.85
New & Used Car Dealers	551	11074	13984	18284	32.5	38.8	47.5	0.22	0.35	0.73
Auto & Home Supply Stores	553	612	701	795	6.1	6.9	7.8	0.75	0.94	1.00
Gasoline Service Stations	554	1106	1287	1422	5.9	6.6	7.6	0.81	0.97	0.99
Women's Clothing & Specialty Stores	562,3	431	493	563	6.6	7.5	8.4	0.70	0.94	1.00
Shoe Stores	566	401	444	500	4.6	5.0	5.5	0.90	0.99	1.00
Furniture Stores	5712	746	911	1074	6.3	7.4	8.8	0.74	0.92	0.99
Homefurnishings Stores	5713,4,9	474	559	642	4.6	5.3	6.2	0.85	0.97	1.00
Radio/TV/Computer/Music Stores	573	641	825	990	5.4	6.3	7.6	0.83	0.95	0.99
Restaurants[viii]	5812[viii]	433	494	555	17.0	19.3	21.6	0.42	0.64	0.91
Refreshment Places[vii]	5812[viii]	462	502	547	16.5	18.3	20.0	-	-	-
Drug & Proprietary Stores	591	1275	1500	1880	10.5	12.3	14.7	0.51	0.82	0.97

Notes: (i) For each of these industries, the fifth percentile of the number of establishments across all *MSAs* equals or exceeds 10. (ii) When multiple SIC codes are given, the industry is defined as the union of those industries. (iii) The headings $Q(1)$ and $Q(3)$ refer to the first and third sample quartiles. (iv) The entries in each column are the sample quartiles, across *MSAs*, of establishments' average sales and employment for that industry. (v) Average sales is reported in thousands of 1992 dollars. (vi) The entries in each column are the sample averages across *MSAs* of the empirical *c.d.f.* of employment evaluated at 9, 19, and 49 employees. All entries are reported to two significant digits. The underlying estimates are all strictly less than one. (vii) The reported average *c.d.f.* reported for Restaurants is for its parent industry, Eating Places. (viii) These industries are subsets of SIC 5812, Eating Places. Restaurants are those establishments that provide table service. See the text for further details.

Table 2: Independent Variables Used in the Regressions

Variable	Description	Source[i]	$Q(1)$	Median	$Q(3)$	Correlation[ii]
Population	Total MSA Residents	CCDB	136764	254861	471837	1.00
Retail Wage[iii]	First Quarter Retail Payroll/March Employment	CRT	2484	2585	2724	0.39
Commercial Rent[iv]	Median Rent per Square Foot for Strip Malls	Authors' Calculations[v]	7.00	8.00	9.50	0.33
Advertising Cost[iii]	Cost of Standard Ad in Sunday Newspaper	Authors' Calculations[vi]	0.43	0.51	0.58	-0.40
Income	Per Capita Personal Income	BEA	17376	18668	20407	0.37
Percent Black	% of Population that is Black	CCDB	2.72	7.11	16.22	0.09
Percent College	% of Population over 25 with a College Degree	CCDB	14.83	18.04	21.62	0.10
Vehicle Ownership	Vehicles per Household	CCDB	1.66	1.71	1.80	-0.16

Notes: (i) CCDB is the 1994 County and City Data Book, CRT is the 1992 Census of Retail Trade, and BEA is the Bureau of Economic Analysis Regional Accounts File. (ii) These correlations are calculated using the logarithm of population and, depending on how it enters our regressions, either the logarithm or the level of the indicated variable. (iii) In 1992 dollars. (iv) In 1992 dollars per square foot. (v) Our observations of rent per square foot for strip malls comes from the 1993 Shopping Center Directory. (vi) Our observations of Sunday newspaper advertising rates and circulation come from the 1992 Editor and Publisher International Yearbook. See the text and Appendix A for further details regarding the data's construction.

Table 3: OLS Estimation Results

	Estimates for Women's Clothing[i]		+/- Table for all Industries[ii]	
	Average Sales	Average Employment	Average Sales	Average Employment
Population	0.10***	0.06***	7/0	6/0
	(0.01)	(0.02)		
Retail Wage	-0.08	-0.46***	5/0	2/4
	(0.21)	(0.17)		
Commercial Rent	-0.05	-0.03	0/0	0/0
	(0.06)	(0.05)		
Advertising Cost	-0.02	-0.02	0/1	0/1
	(0.05)	(0.05)		
Income	0.19	0.33***	3/1	5/0
	(0.16)	(0.11)		
Percent Black[iii]	0.03	-0.14	4/5	2/4
	(0.12)	(0.09)		
Percent College[iii]	0.55**	0.54**	9/0	10/0
	(0.24)	(0.25)		
Vehicle Ownership	-0.48***	-0.38***	3/3	1/1
	(0.12)	(0.11)		
R^2	0.32	0.26		

Note: (i) Heteroskedasticity consistent White standard errors appear below each estimate in parentheses. The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels. (ii) Each cell's first element gives the number of retail trade industry regressions in which the corresponding t-statistic is greater than or equal to 1.96, and each cell's second element gives the number of such regressions in which the t-statistic is less than or equal to -1.96. (iii) For comparability, the estimated coefficients on these variables and their standard errors are multiplied by 100. See the text for further details.

Table 4: Market Size Effects on Average Sales[i,ii]

	Population				Density
	OLS	No Controls	IV	PSS	OLS
Building Materials & Supplies	0.03	0.07***	0.02	0.19***	0.06***
	(0.03)	(0.02)	(0.03)	(0.04)	(0.02)
Grocery Stores	0.01	0.01	0.01	0.13***	0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
New & Used Car Dealers	0.08***	0.18***	0.15***	0.12***	0.06***
	(0.03)	(0.02)	(0.05)	(0.03)	(0.02)
Auto & Home Supply Stores	-0.01	0.01	-0.01	0.04	0.01
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)
Gasoline Service Stations	0.05***	0.10***	0.02	0.16***	0.02
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Women's Clothing & Specialty Stores	0.10***	0.12***	0.10***	0.22***	0.09***
	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)
Shoe Stores	0.02	0.05***	0.00	0.19***	0.05***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)
Furniture Stores	0.11***	0.12***	0.17***	0.13***	0.12***
	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)
Homefurnishings Stores	0.05**	0.08***	0.08**	0.06*	0.03*
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)
Radio/TV/Computer/Music Stores	0.16***	0.17***	0.19***	0.24***	0.09***
	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
Restaurants	0.05***	0.09***	0.05***	0.17***	0.04***
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Refreshment Places	0.02*	0.01	0.01	0.06***	0.04***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Drug & Proprietary Stores	0.03	0.10***	-0.04	0.00	0.05**
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)

Notes: (i) The table's entries are estimated coefficients on the logarithm of market size from the industy-specific regressions described in the text. Heteroskedasticity-consistent standard errors appear in parentheses. (ii) The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels. See the text for further details.

Table 5: Market Size Effects on Average Employment[i,ii]

	Population				Density	Total Sales
	OLS	No Controls	IV	PSS	OLS	OLS
Building Materials & Supplies	0.06***	0.08***	0.05*	0.14***	0.08***	0.14***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Grocery Stores	0.01	0.01	0.03	0.15***	0.01	0.01
	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)
New & Used Car Dealers	0.06***	0.13***	0.11***	0.14	0.05**	0.08***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
Auto & Home Supply Stores	0.01	0.01	0.03	-0.05**	0.03*	0.05***
	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Gasoline Service Stations	0.02	0.04**	-0.01	0.11***	0.01	0.06***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Women's Clothing & Specialty Stores	0.06***	0.08***	0.09***	0.13***	0.06***	0.09***
	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)
Shoe Stores	0.02	0.01	0.01	0.14***	0.06***	0.06***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)
Furniture Stores	0.05*	0.04**	0.10***	0.09***	0.08***	0.13***
	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Homefurnishings Stores	0.05**	0.08***	0.09***	0.11***	0.05**	0.10***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
Radio/TV/Computer/Music Stores	0.07***	0.08***	0.09***	0.03	0.04	0.13***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Restaurants	0.05***	0.05***	0.07***	0.08***	0.05***	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Refreshment Places	0.02*	-0.01	0.03	0.08***	0.02**	0.04***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Drug & Proprietary Stores	0.03	0.07***	-0.02	0.04	0.04**	0.075***
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)

Notes: (i) The table's entries are estimated coefficients on the logarithm of market size from the industy-specific regressions described in the text. Heteroskedasticity-consistent standard errors appear in parentheses. (ii) The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels. See the text for further details.

Table 6: Market Size Effects on Employment's *c.d.f.*[i,ii,iii]

Industry	Controls Included			No Controls		
	$F(9)$	$F(19)$	$F(49)$	$F(9)$	$F(19)$	$F(49)$
Building Materials & Supplies	2.09**	-0.73	-0.26	0.38	-2.55***	-1.56***
	(0.84)	(0.76)	(0.42)	(0.31)	(0.31)	(0.14)
Grocery Stores	-1.42	-2.32***	-0.93**	0.96***	-1.05***	-0.72***
	(0.81)	(0.67)	(0.46)	(0.36)	(0.26)	(0.20)
New & Used Car Dealers	-4.56***	-5.96***	-3.55***	-7.36***	-11.48***	-9.49***
	(1.15)	(1.45)	(1.22)	(0.66)	(0.85)	(0.54)
Auto & Home Supply Stores	4.77***	2.64***	-0.35**	2.80***	1.89***	-0.22***
	(0.99)	(0.45)	(0.16)	(0.38)	(0.19)	(0.05)
Gasoline Service Stations	-4.44***	-0.39	-0.33***	-5.31***	-0.73***	-0.38***
	(0.88)	(0.25)	(0.09)	(0.42)	(0.11)	(0.04)
Women's Clothing & Specialty Stores	-1.61***	-1.70**	-0.01	0.77**	-1.40***	0.07*
	(0.73)	(0.38)	(0.10)	(0.33)	(0.12)	(0.03)
Shoe Stores	-1.38*	-0.49*	-0.03	-1.87***	-0.36***	0.02
	(0.71)	(0.30)	(0.10)	(0.29)	(0.10)	(0.02)
Furniture Stores	0.39	-2.07**	-0.14	-2.01***	-2.54***	-0.19**
	(1.19)	(0.83)	(0.18)	(0.43)	(0.28)	(0.08)
Homefurnishings Stores	-4.14***	0.05	-0.02	-5.48***	-1.03***	-0.18***
	(0.88)	(0.44)	(0.07)	(0.41)	(0.15)	(0.03)
Radio/TV/Computer/Music Stores	0.48	2.12***	-0.58***	0.25	1.96***	-0.67***
	(0.76)	(0.54)	(0.17)	(0.28)	(0.25)	(0.06)
Eating Places	2.47***	0.36	-2.23***	0.72***	-1.59***	-2.54***
	(0.71)	(0.60)	(0.32)	(0.26)	(0.25)	(0.18)
Drug & Proprietary Stores	-6.03***	4.90***	-0.06	-6.27***	4.39***	-0.44***
	(1.36)	(1.02)	(0.29)	(0.53)	(0.60)	(0.13)

Notes: (i) The table's entries are estimated density-weighted average derivatives, expressed in probability points, of the indicated variable with respect to the logarithm of *MSA* Population. Heteroskedasticity-consistent standard errors appear in parentheses. (ii) The superscripts ⋆, ⋆⋆, and ⋆ ⋆ ⋆ indicate statistical significance at the 10%, 5%, and 1% levels. (iii) In the column headings, $F(9)$, $F(19)$ and $F(49)$ refer to the empirical *c.d.f.* of the distribution of employment across an *MSA*'s establishments. "Controls Included" and "No Controls" refer to regressions with and without the control variables listed in Table 2.
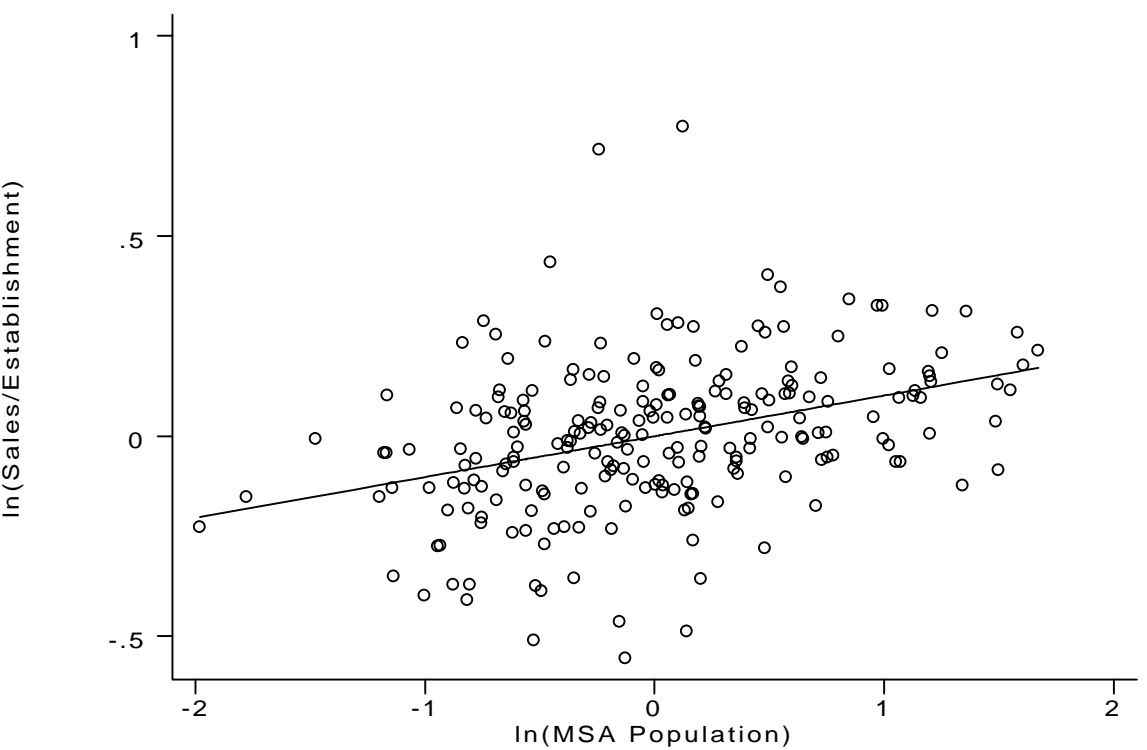
Figure 1: Average Sales versus Population for Women's Clothing and Specialty Stores[i]

Note: (i) In the Figure, the logarithms of both Average Sales and *MSA* Population are defined as residuals from regressions against the control variables listed below Population in Table 5. The control variables enter these regressions as described in Section 2. See the text for further details

35
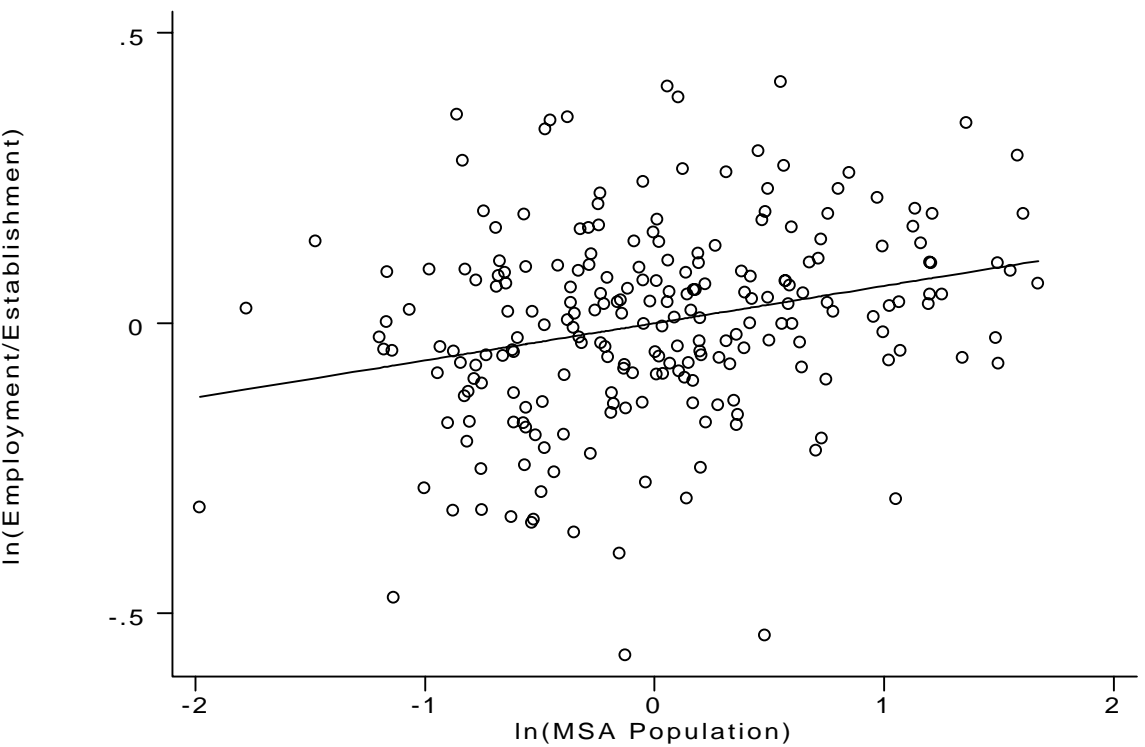
Figure 2: Average Employment versus Population for Women's Clothing and Specialty Stores[i]

Note: (i) In the Figure, the logarithms of both Average Employment and *MSA* Population are defined as residuals from regressions against the control variables listed below Population in Table 5. The control variables enter these regressions as described in Section 2. See the text for further details

Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*

Working Paper Series *(continued)*