NBER WORKING PAPER SERIES

MARKET SIZE MATTERS

Jeffrey R. Campbell
Hugo A. Hopenhayn

Working Paper 9113
http://www.nber.org/papers/w9113

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2002

We thank Jim Levinsohn and Michael Mazzeo for their helpful discussions of an earlier draft of this paper. Emin Dinlersoz provided expert research assistance at an early stage of this project. Eric Heilman and Anna Levine entered much of the commercial rent and newspaper price data. Campbell also gratefully acknowledges research support from National Science Foundation grant SBR-9730442 to the NBER. The latest version of this paper is available on the World Wide Web at http://www.nber.org/~jrc. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Market Size Matters
Jeffrey R. Campbell and Hugo A. Hopenhayn
NBER Working Paper No. 9113
August 2002
JEL No. L11, L16, L81

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 nearly every industry we examine, establishments are larger in larger cities, and in four industries the dispersion of establishment sizes depends on market size. Models of competition in which individual producers' markups do not depend on the number of producers are inconsistent with these observations. Models in which competition is tougher in larger markets can reproduce the positive effect of market size on establishments' average size.

Jeffrey R. Campbell
Research Department
Federal Reserve Bank of Chicago
230 South LaSalle St.
Chicago, IL 60604-1413
and NBER
jeffrey_campbell@nber.org

Hugo A. Hopenhayn
University of Rochester
and Universidad Torcuato Di Tella

This paper empirically compares two approaches to modeling competition among large numbers of producers. The first approach simplifies the competitive process by removing face-to-face interactions between producers. Models of perfect competition, such as Hopenhayn's (1992), and models of monopolistic competition, such as Dixit and Stiglitz's (1977), exemplify this approach. In these models, a producer's actions impact its rivals only by influencing some aggregate statistic, such as the competitive price or the number of consumers per store. Simultaneously doubling the number of producers and consumers typically leaves this statistic and each producer's optimal actions and profits unchanged. The free-entry equilibrium number of producers is linear in the number of consumers; and the distributions of producers' actions, profits, and sizes are invariant to the number of consumers.

The second approach views competition among large numbers of producers as fundamentally similar to oligopoly. As in Prescott and Visscher's (1977) and Salop's (1979) models of spatial competition, one producer's actions have a large direct effect on a few of its nearby competitors and no direct impact on any others. These models inherit oligopoly theory's robust prediction that larger markets are more competitive. Increasing the number of consumers increases producers' sizes, because each one must recover its fixed cost by selling more at a lower markup.

A comparison of producers' sizes across large and small markets using readily available data can determine which of these two general approaches to large-group competition is more empirically promising. We conduct such a comparison using observations from thirteen narrowly-defined retail trade industries in 225 metropolitan statistical areas (*MSA*s), each of which we identify with a separate market. In all of the industries we consider, almost all *MSA*s 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 *MSA*s' factor prices and demographics, which covary with market size and can by themselves affect retailers' sizes. Economic theory imposes no particular functional form on these regressions, so we complement our linear regression estimates with Powell, Stock, and Stoker's (1987) nonparametric density-weighted average derivative estimates.

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 *MSA*s 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 0.104, and the slope of Figure 2's regression line is 0.064. 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, and in four industries market size affects the dispersion of establishments' sizes, measured with the fraction of establishments in either tail of the distribution.

Our investigation of market size's effects in industries with large numbers of competitors 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 by measuring the relationship between market size and the *number* of producers. We argue below that this procedure is biased toward finding tough competition when market size is poorly measured. Whether or not market size is measured with error, the elasticity of producers' average sales with respect to market size that we estimate provides a lower bound for the toughness of price competition. Davis (2001) and Mazzeo (2002) use direct measures of oligopolists' product characteristics 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 producers facing large numbers of competitors cannot use their product placement decisions to eliminate face-to-face competition.

The remainder of the paper is organized as follows. The next section motivates our empirical analysis by examining a simple model that nests both approaches to 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 Producers' Sizes and Market Size

Models of large-group competition differ widely in how they specify producers' strategic interactions. In models of perfect competition, all producers take the competitive output price as given. Monopolistic competition models share with perfect competition the assumption that an individual producer's decision problem only depends on other producers' actions through an aggregate statistic. In Dixit and Stiglitz (1977), this statistic is the number of customers per producer. In Hart's (1985) and Wolinsky's (1986) models, that statistic and the *distribution* of other stores' prices impact an individual's payoff. These models can be viewed as generalizations of perfectly competitive models, because no individual producer's actions impact any other producer's payoff. In contrast, models of oligopolistic competition among large numbers of producers, such as Prescott and Visscher's (1977) and Salop's (1979), assume that averages of other producers' actions are irrelevant or nearly so for any one producer. Instead, only the choices of a producer's close neighbors in product or geographic space impact its optimal action and post-entry profit.[1]

These two assumptions regarding producers' interactions have distinct implications for the relationships between market size and producers' sizes and numbers. The following symmetric free-entry model nests examples from both of these approaches to 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 on X_i and U_i.

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

[1] Grossman and Shapiro (1984) consider a model of spatial competition with advertising which is intermediate between these two extremes. In that environment, producers' advertisements are consumers' only source of information regarding product availability. Consumers' imperfect information implies that any pair of producers may end up competing for a particular consumer's business. However, some pairs of producers are more likely to compete than others, so the model displays market size effects similar to those in Salop's (1979) model.

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.

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.

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

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 unique 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

$$\eta(p, P, N') \geq \eta(p, P, N). \tag{3}$$

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.

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

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

[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, X_i and U_i.

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 $R(S_i) \equiv P(S_i) \times S_i \times q(P(S_i), P(S_i), N(S_i))$.

Our data set contains observations of the number of producers and their average revenue, $N(S_i)$ and $R(S_i)$.

1.1 Monopolistic Competition

Hart's (1985) definition of monopolistic competition requires that producers do not engage in face-to-face strategic interactions.[4] In his model's symmetric equilibrium, only the number of customers per producer and the common price charged by all other producers influence any one producer's demand. That is, for a given prices p and P, a producer's demand is homogeneous of degree zero in N and S_i jointly.

$$q\left(p, P, t \times N\right) = q\left(p, P, N\right)/t, \quad t > 0 \tag{5}$$

This is clearly a significantly stronger assumption than (1).

Equation (5) immediately implies that (3) is an equality, so $P^{\star}(N)$ is a trivial function of N. Given this price function, (1) implies that $t \times N(S_i)$ is the unique solution to (4) for a market with $t \times S_i$ consumers. Therefore, we can conclude that

$$N\left(t \times S_i\right) = t \times N\left(S_i\right), \tag{6}$$

$$R\left(t \times S_i\right) = R\left(S_i\right). \tag{7}$$

Two otherwise identical markets with different numbers of consumers have the same equilibrium price and revenue per store, but the larger market has proportionally more stores.

[4]Formally, Hart (1985) requires that "each firm is negligible in that it can ignore its impact on, and hence reactions from, other firms;...." We interpret this as a statement that changing any one individual producer's action has no impact on any other producer's profits.'

Campbell (2002) analyzes a more general model of monopolistic competition that allows for 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. That more general environment displays the same linear equilibrium structure: Doubling the number of consumers doubles the number of producers and leaves their size distribution unchanged.[5]

1.2 Oligopolistic Competition

Adding a producer to an oligopoly generally increases all producers' residual demand elasticities and so reduces the equilibrium markup. To earn zero profits following additional entry, a producer must sell more at the lower markup. Sutton (1991) describes fiercer competition with more competitors as a robust theoretical prediction of oligopoly models. Models of competition among large numbers of producers that feature oligopolistic face-to-face interactions inherit this prediction. Within our simple framework, we interpret models of oligopolistic competition as implying that (3) is a *strict* inequality.

Under this assumption, $P^\star(N)$ is strictly decreasing in N. The homogeneity of demand, (1), and the zero profit condition, (4), together require that $N(S_i)$ is strictly increasing in S_i; so under oligopolistic competition

$$N(t \times S_i) < t \times N(S_i), \tag{8}$$

$$P(t \times S_i) < P(S_i), \tag{9}$$

$$R(t \times S_i) > R(S_i), \tag{10}$$

where $t > 1$. The first two results are immediate. That increasing the number of consumers in market i increases the revenues of each producer follows from (9) and rewriting the zero

[5]Although the size distribution is unchanged, the total variety of goods offered may be greater in the larger market.

profit condition as

$$\frac{P(S_i) - c}{P(S_i)} \times R(S_i) = \phi. \tag{11}$$

Together, (8) and (10) imply that a large market should have larger stores and fewer stores per customer than does a smaller but otherwise identical counterpart.

1.3 Empirical Implications

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

$$\ln R_i = m(S_i, X_i) + \epsilon_i, \tag{12}$$

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. Under monopolistic competition, (7) and the mutual independence between U_i and S_i imply that

$$\frac{\partial m(S, X)}{\partial \ln S} = 0. \tag{13}$$

That is, market size has no observable effect on producers' average size. Under oligopolistic competition, (10) implies that the partial derivative in (13) is positive, so the estimation of $m(S, X)$ provides one way of empirically distinguishing between these two approaches to large-group competition. Campbell (2002) shows that (13) also holds good if we replace $\ln R_i$ with any other statistic from the size distribution of a monopolistically competitive industry and measure size using either sales or employment. For that reason, we also estimate versions of (12) using employment-based producer size measures, which we describe in more detail below.

When the regression's dependent variable is $\ln R_i$, (11) implies that $\partial m(S,X)/\partial \ln S$ measures the 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 (11) with respect to $\ln S_i$ to get

$$\frac{d\ln R(S_i)}{d\ln S_i} = -\frac{\partial \ln\left(\left(P^\star(N(S_i)) - c\right)/P^\star(N(S_i))\right)}{\partial \ln N} \times \frac{d\ln N(S_i)}{d\ln S_i}. \tag{14}$$

The right-hand side of (14) 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. Under oligopolistic competition, $0 < d\ln N(S_i)/d\ln S_i < 1$, so 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 (14), explicitly recognizing dependence on X_i and U_i, and averaging over U_i yields

$$\frac{\partial m(S,X)}{\partial \ln S} < E\left[\left|\partial \ln\left(\frac{P^\star(N(S_i,X_i,U_i),X_i,U_i) - c(X_i,U_i)}{P^\star(N(S_i,X_i,U_i),X_i,U_i)}\right)/\partial \ln N\right|\right]. \tag{15}$$

That is, the derivative of our regression function 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.

1.4 Market Size and the Number of Producers

Our empirical work does not examine the potentially observable implications of (6) and (8) In this way, our work is very different from Bresnahan and Reiss's (1991) study of entry in oligopoly markets. In principle, we could follow Bresnahan and Reiss and also examine how the number of producers grows with market size for our industries. We do not do so because the nature of our sample differs in an important way from theirs. They examine retail and service industries in isolated and (they argue) relatively homogeneous small towns, each of which they take to be a market. Given this sample design, it is arguable that the town's population is a very accurate indicator of market size.

Because we are interested in competition in industries with large numbers of producers, Bresnahan and Reiss's sample design is inappropriate for our study. Instead, our sample is composed of U.S. cities with large numbers of consumers, who are likely to be much more heterogeneous in unobserved ways than the residents of Bresnahan and Reiss's towns. This heterogeneity implies that any observable measure of the city's size is likely to be an error-ridden measure of the true market size for that city's industry. It is well-known that measurement error biases linear regression coefficients toward zero, so regressing the number of producers on an error-ridden market size measure (in logarithms) can lead to the incorrect conclusion that the number of producers grows less than linearly with market size.[6] Hence, the appropriate empirical methodology for the study of market-size effects in small markets is inappropriate for the study of the same effects in larger markets.[7] Measurement error can make a true market-size effect on producers' sizes more difficult to detect, but it cannot by itself generate a spurious market-size effect in these regressions. In this sense, our procedure is conservative. The possibility that our linear regression estimates of $\partial \ln m(S, X)/\partial \ln S$ are biased toward zero reinforces our interpretation of them as lower bounds to the toughness of price competition.

[6]In a multiple regression context, the result that measurement error biases the coefficient toward zero requires that all of the regression's other right-hand side variables are measured without error.

[7]For the thirteen industries we examine in this paper, we have run regressions using the number of establishments' logarithm as the dependent variable and *MSA* population as the measure of market size, which are otherwise similar to regressions we report below, For ten of the thirteen industries, the estimated coefficients on market size are significantly below one.

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.[8] 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) 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 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.

We selected our industries from all those reported below the two-digit level for which the *CRT* reports data for all *MSA*s. Because our model characterizes 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 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

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

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.[9] The measure of an *MSA*'s 1992 population comes from the 1994 County and City Data Book (*CCDB*). Because the *CRT* and the *CCDB* use two different definitions of *MSA*s in the New England States, 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 inadequate information is available to construct the measure of commercial real estate rent that we describe below.

2.1 Observations from the Census of Retail Trade

For the resulting sample of 225, *MSA*s Tables 2 and 3 report summary statistics for our measures of establishments' average sales and employment. Table 2 reports sample quartiles (across *MSA*s) for these two variables, and Table 3 reports the estimated slope coefficients and their standard errors from regressing these variables' logarithms against the logarithm of the *MSA*'s 1992 population, our baseline measure of market size. All of the standard error estimates in this paper are robust to general forms of heteroskedasticity in the regression errors. 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. The regression coefficients in

[9] 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 from the data set. 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.

Table 3 show that for many industries, this variation is correlated with *MSA* population. In nine of the thirteen industries, the coefficients on population in both regressions are positive and statistically significant at the 5% level. Indeed, there is only one negative estimated slope coefficient, that for Refreshement Places' average employment.

2.2 Observations from the County Business Patterns

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 refer to these with $F(9)$, $F(19)$, and $F(49)$. The *CBP* does not report separate 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.

Table 4 reports summary statistics from these observations. For each industry, its first three columns report the average values across *MSA*s of $F(9)$, $F(19)$, and $F(49)$; and its last three columns report the coefficients from regressing these observations on the logarithm of *MSA* population. 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.

The bivariate regression coefficients reinforce the results from the bivariate average size regressions reported in Table 3. In every industry for which both average sales and average employment are positively correlated with *MSA* population, at least one of the regression coefficients in Table 4 is negative and statistically significant. For three of the industries, all three of the regression coefficients are negative and significant, as would be the case if the size distribution in a larger market stochastically dominates that from a smaller market. The regression coefficients for Eating Places suggest a more interesting and difficult to detect effect of market size on the size distribution. The slope coefficient from the regression of $F(9)$ on *MSA* population is *positive* and statistically significant, while the analogous slope coefficient for $F(49)$ is negative and statistically significant. That is, without controlling for variation in any other variable, the dispersion of establishments' sizes increases with market size.

2.3 Measures of Market Size

In our regressions, we use three distinct measures of market size. Our baseline measure is the simplest, *MSA* population. 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 *MSA*s that our control variables do not adequately measure. If every consumer's preferences can be represented by a homothetic utility function across composite goods, one for each industry, then a consumer's expenditure on a particular industry is a constant share of her income. In this case, the value of industry sales only depends on the joint distribution of consumers' preferences and endowments and is invariant to producer conduct.

2.4 Control Variables

Table 5 lists the independent variables we include in our baseline regression specification and their sources. To control for differences in retailers' costs of production across *MSA*s, 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.[10] These prices all appear in logarithms in our regressions. We also include demographic characteristics to control for differences across *MSA*s 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 6 reports summary statistics for our regressions' right-hand side variables, including their correlation 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 *MSA*s.

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

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

is we assume that

$$m(S_i, X_i) = \beta_S \ln S_i + \beta_X X_i \tag{16}$$

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 $m(S, X)$. The regression function's density-weighted average derivatives are defined as

$$\begin{aligned} \delta_S &\equiv \mathbf{E}\left[\frac{\partial m(S,X)}{\partial \ln S} f(\ln S, X)\right] / \mathbf{E}\left[f(\ln S, X)\right] \\ \delta_X &\equiv \mathbf{E}\left[\frac{\partial m(S,X)}{\partial X} f(\ln S, X)\right] / \mathbf{E}\left[f(\ln S, X)\right], \end{aligned} \tag{17}$$

where $f(\ln S, X)$ 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 (16) 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 78 regressions. To conserve space, we report complete results for only one industry (Women's Clothing and Specialty Stores) 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 7 reports the estimated regression coefficients for Women's Clothing and Specialty Stores, estimates of their standard errors, and the regressions' R^2 measures. The estimated elasticities of establishments' average sales and employment with respect to *MSA* population are 0.102 and 0.064. As noted in the introduction, both of these are statistically significant at the 1% level. They are also very close to the corresponding bivariate regression estimates reported in Table 3, 0.123 and 0.073. 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. Overall, the control variables do add explanatory power to our regressions, but the impression from the bivariate regressions in Table 3 that establishments' average size is higher in larger markets is unchanged.

For all of the industries we consider, Table 8 reports the coefficients on *MSA* population from the average sales and employment regressions. As with the bivariate coefficients in Table 3, only one estimate in Table 8 is negative. For most of the industries we consider, adding the control variables changes the estimates and their statistical significance very little. The three industries which are clear exceptions to this are New and Used Car Dealers, Refreshment Places, and Drug and Proprietary Stores. Adding the control variables reduces the estimated coefficients for New and Used Car Dealers by half, but they are still statistically significant at the 1% level. The multivariate estimates for Refreshment Places are larger than the corresponding bivariate estimates, and they are significant at the 10% level. The estimated coefficients for Drug and Proprietary Stores drop in magnitude and lose their statistical significance after adding the control variables. 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 significance at the 5% level is not the same across the two equations are Building Materials and Supplies, Gasoline Service Stations, and Furniture Stores.

3.1.1 Control Variable Coefficients

Table 9 provides an overview of the control variables' estimated coefficients in our regressions. Each row corresponds to one of the variables included in X_i, and each column corresponds to one of the dependent variables. 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. Even in the simplest model of perfect competition in which all establishments have identical U-shaped average cost curves, the control variables' coefficients cannot be signed. In that model, changing consumers' demographic characteristics impacts producers' sizes only by changing the mix of products produced and their associated technology. Economic theory offers little guidance regarding the direction of such a change. Lowering a factor price unambiguously lowers producers' cost curves, but it could move the *location* of minimum average cost in either direction or leave it the same. Therefore, it is impossible to state whether or not the estimated coefficients' signs are surprising.

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 wage is clearly the most important, appearing significantly in five industries' average sales regressions and in six industries' average employment regressions. 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.

3.1.2 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.

Table 10 reports 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.

3.1.3 Alternative Market Size Measures

Tables 11 and 12 present OLS coefficient estimates and their standard errors from regressions that use our two alternative measures of market size. Replacing population with population density in our regressions leaves the coefficient estimates and the patterns of inference 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 industries, the concern that *MSA* population may be a noisy measure of market size seems warranted. 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.[11]

The estimated coefficients on the value of industry sales from the regressions that use it to measure market size are reported in Table 12. We do not attempt to measure the dependence of establishments' average sales on total sales, because this is equivalent to the potentially problematic exercise of measuring the dependence between the number of producers and market size. In the regressions of establishments' average employment on total industry sales, all of the estimated coefficients are positive. In all but one industry, Grocery Stores, they are statistically significant at the 1% level. For two of the industries, Gasoline Service Stations and Drug and Proprietary Stores, the analogous coefficients from the instrumental variables procedure are smaller and not significantly different from zero. However, for Auto and Home Supply Stores, Shoe Stores, and Refreshment Places, measuring market size with

[11] All other statistical inferences in Table 11 but one are robust to instrumenting commercial rent with residential real estate prices. This exception is the regression coefficient in Radio/TV/Computer/Music Stores' average employment regression, which increases and becomes statistically significant at the 10% level.

value of industry sales reverses the inference from Table 8 that market size has no impact on establishments' sizes.

3.1.4 Alternative Specifications

In addition to the regressions described above, we have also estimated versions of (16) using alternative definitions of the control variables and alternative samples. These estimates indicate that the results reported in Tables 8, 10, 11, and 12 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. Adding population growth to our set of control variables changes none of the point estimates substantially and alters no statistical inferences.

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.

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 8, 10, 11, and 12 are substantially unchanged.

3.2 Nonparametric Average Derivative Estimation Results

The relationship between ordinary linear regression estimates and density weighted average derivative estimation can be most easily understood by writing the right-hand side of (12) as a linear function of $\ln S_i$ and X_i and a residual term;

$$\ln R_i = \delta_S \ln S_i + \delta_X' X_i + u_i, \tag{18}$$

where δ_S and δ_X are defined in (17), and the residual term is related to the true regression function and regression disturbance by

$$u_i = m\left(S_i, X_i\right) - \delta_S \ln S_i - \delta_X X_i + \epsilon_i. \tag{19}$$

If the true regression function is nonlinear, then applying ordinary least squares to (18) will not consistently estimate δ_S and δ_X because the regressors will be correlated with the error term. Powell, Stock, and Stoker (1989) show that in this case a consistent instrumental variables estimator of δ_S and δ_X can be derived if one knows $f\left(\ln S, X\right)$, by proving that $\partial f\left(\ln S_i, X_i\right)/\partial \ln S$ and $\partial f\left(\ln S_i, X_i\right)/\partial X$ are legitimate instruments for the residual term defined in (19). Of course, prior knowledge of $f\left(\ln S, X\right)$ is rarely available, so Powell, Stock, and Stoker (1989) derive an estimator which replaces the derivatives of the actual *p.d.f.* with consistent kernel-based estimates. Implementing this estimator requires a choice of kernel and a bandwidth parameter. We detail these choices and our results' robustness to them in Appendix B.[12]

Table 13 reports the estimates of δ_S for the average sales and employment regressions using our baseline measure of market size. The estimates for most industries are positive,

[12] We know of no general characterization of measurement error's impact on the estimation of δ_S. However, just as with linear regressions, measuring market size with error does not bias the estimate of δ_S if (13) holds good.

statistically significant, and exceed the linear regression estimates reported in Table 8. 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.[13] 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 Table 14. For one industry, New and Used Car Dealers, the estimates of δ_S from all three regressions are negative 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. Market size has

[13]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.

a significant effect on the dispersion of establishments' sizes in that industry. The pattern of market size effects in Radio/TV/Computer/Music Stores and in Eating Places is similar. This finding of greater dispersion in larger markets is reminiscent of Syverson's (2001) finding that the productivity distribution of concrete production plants exhibits less dispersion in denser *MSA*s. However, our finding of greater dispersion in larger markets suggests that either the selection mechanism he emphasizes is not dominant in these three retail trade industries or that there is no tight relationship between an establishment's productivity and its size. Market size also significantly affects the dispersion of establishments' sizes in Drug and Proprietary Stores, but in that industry market size decreases dispersion.[14] 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

Our analysis of the effects of market size on the size distribution yields the strong conclusion that establishments are larger in larger markets for six industries: New and Used Car Dealers, Women's Clothing and Specialty Stores, Furniture Stores, Homefurnishings Stores, Radio/TV/Computer/Music Stores, and Restaurants. For these industries, we observe that establishments' average sales are larger in larger markets using a variety of estimation methods (both parametric and nonparametric) and measures of market size. With the possible exception of Radio/TV/Computer/Music Stores, the same is true for establishments' average employment. For the remaining industries, the effects of market size are more difficult to detect. In Building Materials and Supplies, Auto and Home Supply Stores, Shoe Stores, and Refreshment Places, using a more refined measure of market size than *MSA* population

[14] 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.

reveals effects of market size on either establishments' average sales or employment, while such effects are only evident in Grocery Stores using nonparametric regression techniques. In four industries, including one in which there are no significant effects of market size on establishments' average size, market size significantly affects the dispersion of establishments' sizes. Only one of the thirteen industries we consider, Gasoline Service Stations, displays a robust invariance of establishment sizes to market size.

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. However, 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.[15] Determining the relevance of such competition for the industries we consider is on our agenda for future research. Whether our finding reflects oligopolistic competition between a large number of producers or quality competition between true oligopolists, we conclude that face-to-face strategic interactions are an important component of competition in these industries.

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

4 Conclusion

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 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. Our observations also suggest that the extremely localized strategic interactions emphasized in models of spatial competition are relevant for most of the industries we consider. Measuring their importance requires both observations of individual producers and further progress in modeling their strategic interactions.

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 9, 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 $f(\ln S, X)$. 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, $K(u)$, to be

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

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" $K(u)$ equal zero.[16] The order of $K(u)$, 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 $m(S, X)$ 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

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

stage estimation of $f(\ln S, X)$ 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. 2002. A Robust Prediction of 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.

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

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.

Grossman, Gene M. and Carl Shapiro. 1984. Informative Advertising with Differentiated Products. *Review of Economic Studies,* **51**(1), 63–81.

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.

Prescott, Edward C. and Michael Visscher. 1977. Sequential Location Among Firms with Foresight. *Bell Journal of Economics,* **8**(2), 378–393.

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.

Syverson, Chad. 2001. Market Structure and Productivity: A Concrete Example. Center for Economic Studies Working Paper 01-06, U.S. Census.

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

Table 1: Retail Industries with Large Establishment Numbers(i)

Industry	SIC Code(s)(ii)
Building Materials and Supplies	521,3
Grocery Stores	541
New and Used Car Dealers	551
Auto and Home Supply Stores	553
Gasoline Service Stations	554
Women's Clothing and Specialty Stores	562,3
Shoe Stores	566
Furniture Stores	5712
Homefurnishings Stores	5713,4,9
Radio/TV/Computer/Music Stores	573
Restaurants	Part of 5812(iii)
Refreshment Places	Part of 5812(iii)
Drug and Proprietary Stores	591

Notes: (i) For these industries, the fifth percentile of the number of establishments across all *MSA*s equals or exceeds 10. (ii) When multiple SIC codes are given, the industry is defined as the union of those SIC industries. (iii) Each of these industries is a subset of SIC 5812, Eating Places. Restaurants are those establishments which provide table service.

Table 2: Sample Quartiles from Census of Retail Trade Data

	Average Sales(i,ii,iii)			Average Employment(i,ii)		
Industry	$Q\,(1)$	Median	$Q\,(3)$	$Q\,(1)$	Median	$Q\,(3)$
Building Materials & Supplies	1900	2235	2602	11.4	13.3	15.9
Grocery Stores	2354	2820	3516	18.5	22.9	28.0
New & Used Car Dealers	11074	13984	18284	32.5	38.8	47.5
Auto & Home Supply Stores	612	701	795	6.1	6.9	7.8
Gasoline Service Stations	1106	1287	1422	5.9	6.6	7.6
Women's Clothing & Specialty Stores	431	493	563	6.6	7.5	8.4
Shoe Stores	401	444	500	4.6	5.0	5.5
Furniture Stores	746	911	1074	6.3	7.4	8.8
Homefurnishings Stores	474	559	642	4.6	5.3	6.2
Radio/TV/Computer/Music Stores	641	825	990	5.4	6.3	7.6
Restaurants	433	494	555	17.0	19.3	21.6
Refreshment Places	462	502	547	16.5	18.3	20.0
Drug & Proprietary Stores	1275	1500	1880	10.5	12.3	14.7

Notes: (i) The headings $Q\,(1)$ and $Q\,(3)$ refer to the first and third sample quartiles. (ii) The entries in each column are the sample quartiles, across *MSA*s, of establishments' average sales and employment for that industry. (iii) Average sales is reported in thousands of 1992 dollars. See the text for further details.

Table 3: Bivariate Regression Coefficients on *MSA* Population[i,ii]

Industry	Average Sales	Average Employment
Building Materials & Supplies	0.065***	0.077***
	(0.022)	(0.020)
Grocery Stores	0.015	0.013
	(0.022)	(0.023)
New & Used Car Dealers	0.178***	0.128***
	(0.022)	(0.019)
Auto & Home Supply Stores	0.009	0.013
	(0.017)	(0.014)
Gasoline Service Stations	0.095***	0.040**
	(0.017)	(0.016)
Women's Clothing & Specialty Stores	0.123***	0.076***
	(0.014)	(0.012)
Shoe Stores	0.047***	0.010
	(0.015)	(0.015)
Furniture Stores	0.122***	0.040**
	(0.024)	(0.020)
Homefurnishings Stores	0.083***	0.079***
	(0.019)	(0.019)
Radio/TV/Computer/Music Stores	0.175***	0.078***
	(0.021)	(0.018)
Restaurants	0.093***	0.051***
	(0.015)	(0.015)
Refreshment Places	0.011	-0.011
	(0.010)	(0.011)
Drug & Proprietary Stores	0.105***	0.075***
	(0.022)	(0.018)

Notes: (i) The table's entries are estimated slope coefficients from bivariate regression of the indicated variable's logarithm on the logarithm of *MSA* population. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. (ii) The superscripts $\star$, $\star\star$, and $\star\star\star$ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 4: Sample Statistics from County Business Patterns' Empirical Size Distribution[i,ii]

Industry	Mean[iii] $F(9)$	$F(19)$	$F(49)$	Regression Slope[iv] $F(9)$	$F(19)$	$F(49)$
Building Materials & Supplies	0.66	0.82	0.94	0.006	-0.006***	-0.012***
Grocery Stores	0.62	0.75	0.85	0.012	0.006	-0.002
New & Used Car Dealers	0.22	0.35	0.73	0.005	-0.012	-0.056***
Auto & Home Supply Stores	0.75	0.94	1.00[v]	-0.006	0.002	0.001
Gasoline Service Stations	0.81	0.97	0.99	-0.018***	-0.002	0.000
Women's Clothing & Specialty Stores	0.70	0.94	1.00[v]	-0.017***	-0.015***	-0.002 ***
Shoe Stores	0.90	0.99	1.00[v]	0.004	-0.001	0.000
Furniture Stores	0.74	0.92	0.99	0.006	-0.005	-0.005***
Homefurnishings Stores	0.85	0.97	1.00[v]	-0.015***	-0.007***	-0.002***
Radio/TV/Computer/Music Stores	0.83	0.95	0.99	-0.012**	-0.008***	-0.004***
Eating Places	0.42	0.64	0.91	0.013***	0.001	-0.005**
Drug & Proprietary Stores	0.51	0.82	0.97	-0.027***	-0.024***	-0.003

Notes: (i) 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. (ii) The superscripts $\star$, $\star\star$, and $\star\star\star$ indicate statistical significance at the 10%, 5%, and 1% levels based on t-statistics constructed using unreported heteroskedasticity-consistent standard errors. (iii) The entries in these columns are the means across *MSA*s of the indicated statistic. (iv) The entries in these columns are the slope coefficients from bivariate regressions of the indicated statistics, in levels, against *MSA* population, in logarithms. (v) These averages are strictly below one. All statistics in this column are rounded to two significant digits. See the text for further details.

Table 5: Independent Variables Used in the Regressions

Variable	Description	Source[i]
Population	Total MSA Residents	CCDB
Retail Wage	First Quarter Retail Payroll/March Employment	CRT
Commercial Rent	Median Rent per Square Foot for Strip Malls	Authors' Calculations[ii]
Advertising Cost	Cost of Standard Ad in Sunday Newspaper	Authors' Calculations[iii]
Income	Per Capita Personal Income	BEA
Percent Black	% of Population that is Black	CCDB
Percent College	% of Population over 25 with a College Degree	CCDB
Vehicle Ownership	Vehicles per Household	CCDB

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) Our observations of rent per square foot for strip malls comes from the 1993 Shopping Center Directory. (iii) 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 6: Summary Statistics for Independent Variables[i,ii]

Variable	$Q\,(1)$	Median	$Q\,(3)$	Correlation[ii].
Population	136764	254861	471837	1.00
Retail Wage[iii]	2484	2585	2724	0.39
Commercial Rent[iv]	7.00	8.00	9.50	0.33
Advertising Cost[iii]	0.43	0.51	0.58	-0.40
Income[iii]	17376	18668	20407	0.37
Percent Black	2.72	7.11	16.22	0.09
Percent College	14.83	18.04	21.62	0.10
Vehicle Ownership	1.66	1.71	1.8	-0.16

Notes: (i) The headings $Q\,(1)$ and $Q\,(3)$ refer to the first and third sample quartiles of the indicated variable across *MSA*s. (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. See the text and Appendix A for more details regarding the variables' definitions and construction.

Table 7: OLS Coefficients for Women's Clothing & Specialty Stores[i]

	Dependent Variable	
	Average Sales	Average Employment
Population	0.102***	0.064***
	(0.015)	(0.015)
Retail Wage	-0.077	-0.462***
	(0.214)	(0.171)
Commercial Rent	-0.051	-0.034
	(0.062)	(0.050)
Advertising Cost	-0.025	-0.020
	(0.050)	(0.049)
Income	0.194	0.273***
	(0.160)	(0.106)
Percent Black	0.000	-0.001
	(0.001)	(0.002)
Percent College	0.005**	0.005**
	(0.002)	(0.001)
Vehicle Ownership	-0.476***	-0.375***
	(0.126)	(0.114)
R^2	0.32	0.26

Note: (i) The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 8: Multivariate Regression Coefficients on *MSA* Population[i,ii]

Industry	Average Sales	Average Employment
Building Materials & Supplies	0.034	0.062***
	(0.025)	(0.022)
Grocery Stores	0.008	0.011
	(0.023)	(0.025)
New & Used Car Dealers	0.081***	0.064***
	(0.026)	(0.023)
Auto & Home Supply Stores	-0.007	0.015
	(0.020)	(0.018)
Gasoline Service Stations	0.055***	0.025
	(0.019)	(0.018)
Women's Clothing & Specialty Stores	0.102***	0.064***
	(0.014)	(0.015)
Shoe Stores	0.019	0.015
	(0.014)	(0.014)
Furniture Stores	0.114***	0.051*
	(0.032)	(0.026)
Homefurnishings Stores	0.047**	0.052**
	(0.021)	(0.022)
Radio/TV/Computer/Music Stores	0.156***	0.069***
	(0.024)	(0.021)
Restaurants	0.050***	0.054***
	(0.017)	(0.017)
Refreshment Places	0.020*	0.022*
	(0.011)	(0.012)
Drug & Proprietary Stores	0.028	0.030
	(0.024)	(0.022)

Notes: (i) The table's entries are estimated coefficients on the logarithm of *MSA* population from the multivariate regression described in the text. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. (ii) The superscripts $\star$, $\star\star$, and $\star\star\star$ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 9: +/- Table for Control Variables' Coefficients(i)

	Dependent Variable	
	Average Sales	Average Employment
Retail Wage	5/0	2/4
Commercial Rent	0/0	0/0
Advertising Cost	0/1	0/1
Income	3/1	5/0
Percent Black	4/5	2/4
Percent College	9/0	10/0
Vehicle Ownership	3/3	1/1

Note: (i) 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.

Table 10: Instrumental Variables Estimates[i,ii]

Industry	Average Sales	Average Employment
Building Materials & Supplies	0.021	0.049⋆
	(0.032)	(0.022)
Grocery Stores	0.005	0.026
	(0.026)	(0.030)
New & Used Car Dealers	0.149⋆⋆⋆	0.115⋆⋆⋆
	(0.048)	(0.037)
Auto & Home Supply Stores	-0.010	0.027
	(0.040)	(0.023)
Gasoline Service Stations	0.022	-0.009
	(0.027)	(0.027)
Women's Clothing & Specialty Stores	0.101⋆⋆⋆	0.091⋆⋆⋆
	(0.027)	(0.027)
Shoe Stores	0.004	0.011
	(0.024)	(0.024)
Furniture Stores	0.175⋆⋆⋆	0.103⋆⋆⋆
	(0.048)	(0.040)
Homefurnishings Stores	0.085⋆⋆	0.088⋆⋆
	(0.037)	(0.038)
Radio/TV/Computer/Music Stores	0.186⋆⋆⋆	0.091⋆⋆⋆
	(0.035)	(0.026)
Restaurants	0.050⋆⋆⋆	0.069⋆⋆⋆
	(0.019)	(0.020)
Refreshment Places	0.014	0.035⋆
	(0.015)	(0.019)
Drug & Proprietary Stores	-0.043	-0.021
	(0.044)	(0.036)

Notes: (i) The table's entries are estimated coefficients on the logarithm of *MSA* population from the multivariate instrumental variables regression described in the text. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. (ii) The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 11: Multivariate Regression Coefficients on Population Density(i,ii)

Industry	Average Sales	Average Employment
Building Materials & Supplies	0.064***	0.083***
	(0.022)	(0.020)
Grocery Stores	0.000	0.011
	(0.023)	(0.024)
New & Used Car Dealers	0.063***	0.054**
	(0.025)	(0.022)
Auto & Home Supply Stores	0.007	0.031*
	(0.021)	(0.018)
Gasoline Service Stations	0.020	0.008
	(0.020)	(0.018)
Women's Clothing & Specialty Stores	0.095***	0.065***
	(0.015)	(0.016)
Shoe Stores	0.046***	0.060***
	(0.016)	(0.015)
Furniture Stores	0.116***	0.081***
	(0.027)	(0.025)
Homefurnishings Stores	0.035*	0.047**
	(0.019)	(0.019)
Radio/TV/Computer/Music Stores	0.090***	0.040
	(0.029)	(0.026)
Restaurants	0.042***	0.053***
	(0.017)	(0.016)
Refreshment Places	0.040***	0.024**
	(0.011)	(0.011)
Drug & Proprietary Stores	0.049**	0.040**
	(0.022)	(0.020)

Notes: (i) The table's entries are estimated coefficients on the logarithm of population density from the multivariate regression described in the text. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. (ii) The superscripts ⋆, ⋆⋆, and ⋆⋆⋆ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 12: Multivariate Regression Coefficients on Industry Sales[i,ii]

Industry	Average Employment
Building Materials & Supplies	0.138***
	(0.027)
Grocery Stores	0.009
	(0.026)
New & Used Car Dealers	0.082***
	(0.022)
Auto & Home Supply Stores	0.055***
	(0.019)
Gasoline Service Stations	0.063***
	(0.018)
Women's Clothing & Specialty Stores	0.085***
	(0.014)
Shoe Stores	0.063***
	(0.018)
Furniture Stores	0.127***
	(0.025)
Homefurnishings Stores	0.101***
	(0.023)
Radio/TV/Computer/Music Stores	0.130***
	(0.018)
Restaurants	0.070***
	(0.016)
Refreshment Places	0.037***
	(0.012)
Drug & Proprietary Stores	0.075***
	(0.020)

Notes: (i) The table's entries are estimated coefficients on the logarithm of total industry sales from the multivariate regression described in the text. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. (ii) The superscripts $\star$, $\star\star$, and $\star\star\star$ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 13: Estimates of *MSA* Population's Density-Weighted Average Derivative[i,ii]

Industry	Average Sales	Average Employment
Building Materials & Supplies	0.193***	0.140***
	(0.036)	(0.030)
Grocery Stores	0.125***	0.153***
	(0.027)	(0.029)
New & Used Car Dealers	0.124***	0.145***
	(0.031)	(0.026)
Auto & Home Supply Stores	0.041	-0.052**
	(0.034)	(0.026)
Gasoline Service Stations	0.163***	0.105***
	(0.025)	(0.022)
Women's Clothing & Specialty Stores	0.217***	0.131***
	(0.022)	(0.021)
Shoe Stores	0.189***	0.139***
	(0.032)	(0.031)
Furniture Stores	0.130***	0.094***
	(0.033)	(0.031)
Homefurnishings Stores	0.061*	0.106***
	(0.032)	(0.031)
Radio/TV/Computer/Music Stores	0.245***	0.034
	(0.036)	(0.025)
Restaurants	0.167***	0.084***
	(0.024)	(0.021)
Refreshment Places	0.059***	0.083***
	(0.015)	(0.016)
Drug & Proprietary Stores	-0.005	0.036
	(0.027)	(0.024)

Notes: (i) The table's entries are estimated density-weighted average derivatives of the indicated variable with respect to the logarithm of *MSA* Population. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. See the text for further details. (ii) The superscripts $\star$, $\star\star$, and $\star\star\star$ indicate statistical significance at the 10%, 5%, and 1% levels.

Table 14: Estimates of *MSA* Population's Density-Weighted Average Derivative(i,ii,iii)

Industry	$F(9)$	$F(19)$	$F(49)$
Building Materials & Supplies	0.0209**	-0.0073	-0.0026
	(0.0084)	(0.0076)	(0.0042)
Grocery Stores	-0.0142	-0.0232***	-0.0093**
	(0.0081)	(0.0067)	(0.0046)
New & Used Car Dealers	-0.0456***	-0.0596***	-0.0355***
	(0.0115)	(0.0145)	(0.0122)
Auto & Home Supply Stores	0.0477***	0.0264***	-0.0035**
	(0.0099)	(0.0045)	(0.0016)
Gasoline Service Stations	-0.0444***	-0.0039	-0.0033***
	(0.0088)	(0.0025)	(0.0009)
Women's Clothing & Specialty Stores	-0.0161***	-0.0170**	-0.0001
	(0.0073)	(0.0038)	(0.0010)
Shoe Stores	-0.0138*	-0.0049*	-0.0003
	(0.0071)	(0.0030)	(0.0010)
Furniture Stores	0.0039	-0.0207**	-0.0014
	(0.0119)	(0.0083)	(0.0018)
Homefurnishings Stores	-0.0414***	0.0005	-0.0002
	(0.0088)	(0.0044)	(0.0007)
Radio/TV/Computer/Music Stores	0.0048	0.0212***	-0.0058***
	(0.0076)	(0.0054)	(0.0017)
Eating Places	0.0247***	0.0036	-0.0223***
	(0.0071)	(0.0060)	(0.0032)
Drug & Proprietary Stores	-0.0603***	0.0490***	-0.0006
	(0.0136)	(0.0102)	(0.0029)

Notes: (i) The table's entries are estimated density-weighted average derivatives of the indicated variable with respect to the logarithm of *MSA* Population. Heteroskedasticity-consistent standard errors appear below each coefficient in parentheses. See the text for further details. (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.

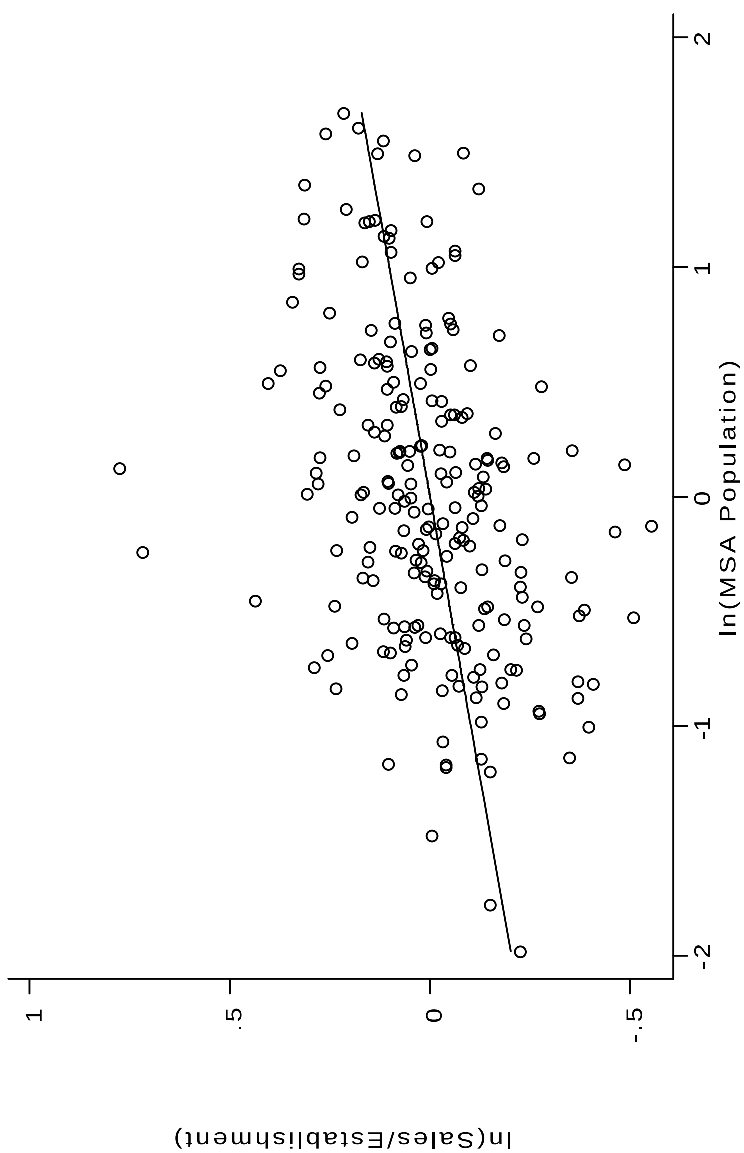

Figure 1: Average Sales versus Population for Women's Clothing and Specialty Stores(i)

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

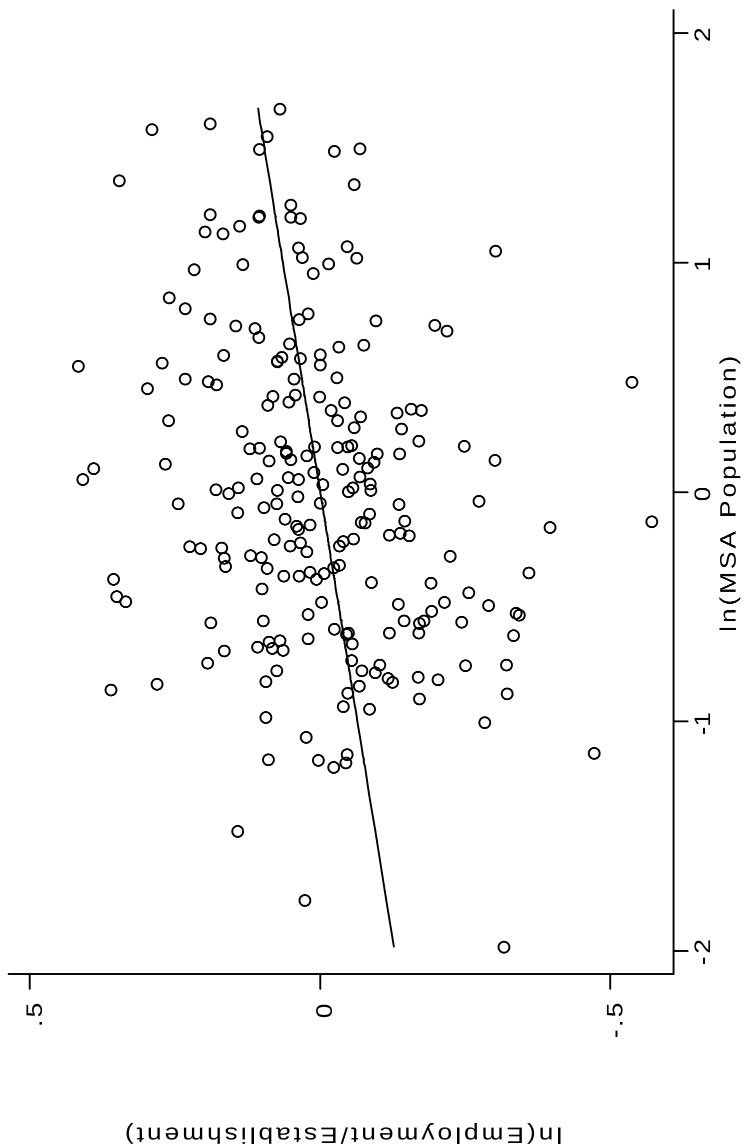

Figure 2: Average Employment versus Population for Women's Clothing and Specialty Stores[i]

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