Cities and the Organization of Manufacturing*

Emin M. Dinlersoz
Department of Economics
University of Houston
Houston, TX 77204
Phone: (713) 743 3839
Fax: (713) 743 3798
E-mail: edinlers@mail.uh.edu

This Version: March 2002

Abstract

This paper provides a recent account of the distribution of manufacturing activity across cities in the U.S. After years of nationwide decline in manufacturing employment, and migration of manufacturing plants to suburbs and rural areas, the following pattern emerges for overall manufacturing as of 1990: *i*) Number of manufacturing establishments increases more than proportionally with city population, *ii*) Manufacturing employment in a city increases in proportion to city population, and *iii*) Employee size distribution of establishments is stochastically decreasing as city population increases. While these results are in part driven by industry composition in cities, in many individual manufacturing industries larger cities tend to accommodate more employment through an expansion in number of establishments, but not always through an expansion of establishment size. Implications of these findings on the new economic geography and other theories relating city size and scale of production are discussed.

*Substantial improvements in this paper were made possible by the comments from the editor, John Quigley, and especially, from two referees, who were extremely patient and constructive. I am also grateful to Mark Bils, Jeffrey Campbell, Gordon Dahl, Thomas Holmes, Hugo Hopenhayn, Lance Lochner, Glenn MacDonald and seminar participants at the University of Rochester for comments and suggestions.

1 Introduction

Cities have been evolving over time, and so have the types and the sizes of economic activity across cities of different sizes. The Industrial Revolution initiated a trend of massive urbanization in the United States: increasing population, expanding manufacturing activity, and the "factory" concept increased the demand for higher agglomerations of workers, capital, and structure into cities. The transportation revolution made possible by the car reversed this trend in the earlier part of the 20th century, leading to suburbanization and a gradual movement of the manufacturing activity away from the city center. The decline in manufacturing activity itself in the later part of the 20th century further led to significant changes in the composition of economic activity within and across cities. As the service sector became increasingly dominant in the U.S. economy, the manufacturing sector further dwindled and has been, to some extent, pushed away from cities to suburbs. What remains of manufacturing in cities today is likely to be different from the patterns that prevailed in the early days of industrialization. In fact, we know little about how cities of different sizes accommodate manufacturing activity today: Do larger cities support proportionally more or less manufacturing employment? If so, does the increasing level of employment with city size predominantly lead to larger manufacturing establishments or to simply more establishments with little systematic change in establishment size distribution?[1]

The goal of this paper is to present recent empirical evidence on how manufacturing activity is related to city size in the post-industrial United States. The nature of this relationship has been the subject for a large amount of theoretical work. One strand of the literature, the new economic geography, relies heavily on variants of the monopolistic competition model parameterized using the constant elasticity of substitution (CES) utility and production functions due to Dixit and Stiglitz (1977). A prototype of such models is Krugman (1991). In a world where transportation costs matter, manufacturers concentrate geographically where there is a higher demand because this allows for majority of sales to be carried out without transportation costs. This can create a reinforcing mechanism: a large demand in a location means a large number of manufactured good producers locating there, and more producers create further demand that attracts even more producers. Consequently, given aggregate constraints on mobile factors of production in an economy consisting of many locations, locations with larger population support proportionally more manufacturing activity, both in terms of employment and output, a result labelled as the "home market effect". The strength of this effect depends in important ways on the extent of scale economies and transportation costs. Despite these predictions, however, the model, in its standard form, has trivial implications regarding the size of individual manufacturing establishments as it relates to city size. It assumes a high, symmetric elasticity of substitution among several differentiated products

[1] According to the Standard Industry Classification (SIC) system, an establishment is an economic unit at a single location where business is conducted or where services and industrial operations are performed. An establishment is classified into an industry on the basis of the primary activity of the establishment.

offered to the market by monopolistically competitive manufacturers. As a result, an increase in city size leads to an increase in the variety of products and the number of producers, but not to an increase in the scale of production of any particular differentiated good. These implications led to an increasing skepticism concerning the success of these models in describing the spatial organization of industries, especially in the context of manufacturing. For example, Krugman (1998) and more recently, Neary (2001), argue that the monopolistic competition model has been used mostly because of its tractability, without any serious attention to how industries are actually organized across the geography. While the predictions of this particular model are rather stark, less restrictive and more plausible implications can be obtained within the monopolistic competition framework. For instance, Holmes (1999) provides one modification that leads to an increase in the scale of individual products, as well as in the number of different products.

Another important strand of the literature in urban economics emphasizes the productivity enhancing role of technological externalities that stem from the agglomeration of inputs of production into close geographic quarters.[2] Unlike the pecuniary externalities involved in the monopolistic competition models above, these externalities directly influence the productivity of an establishment (i.e. learning from others, information spillovers, etc.), and/or the utility of consumers (i.e. pollution, congestion, etc.). While such externalities may play a role in the formation and growth of cities (e.g. Lucas (2001), Glaeser (1999), Glaeser, Kallal, Scheinkman and Shleifer (1992)), it is not so clear why and how they might matter for the sizes of individual establishments. In fact, as will be discussed, varying predictions can be obtained under several reasonable specifications. Essentially, considering the standard models where externalities are embedded in the production function of an establishment as a multiplicative productivity shifter, the way externalities matter for establishment size simply depends on how externalities alter the minimum efficient scale. Depending on the impact of externalities on fixed versus variable costs of a production unit, the scale of production can increase, decrease, or remain constant. Thus, models incorporating externalities in a standard way in a competitive framework provide little guidance.

Given a diverse set of predictions from the theoretical models in the literature, this paper aims to present a comprehensive first look at the patterns emerging. The following empirical regularities are observed: for manufacturing as a whole, the number of manufacturing establishments in a Metropolitan Statistical Area (MSA) tends to increase more than proportionally with MSA size, as measured by population.[3] Overall manufacturing employment in a MSA also increases with MSA population, but the rate of increase appears to be proportional to population. As a consequence, establishments tend to be smaller on average in larger MSA's. Furthermore, employee size distribution of establishments is stochastically decreasing as city population increases, consistent with the observation on average size. Because overall manufacturing consists of many different industries

[2] See, for example, Henderson (1986, 1997), Jacobs (1968), Glaeser, Kallal, Scheinkman, and Shleifer (1992), and Glaeser (1999)

[3] The terms "MSA" and "city" will be used interchangeably throughout the paper.

that differ in average establishment size and geographic concentration, a 2-digit breakdown of the manufacturing sector is also studied. This analysis reveals that the results for overall manufacturing are partly driven by industry composition. Nevertheless, in many 2-digit industries, larger cities tend to accommodate more manufacturing employment through an expansion in number of establishments, but not always through a systematic expansion in establishment size. The establishment size distribution is either invariant to MSA population or stochastically decreases with MSA population in many cases. The responsiveness of industry employment to MSA population varies considerably across industries. In about half of the industries investigated, manufacturing employment increases more than proportionally with MSA population. There is also some evidence, albeit weak, that the industries with high responsiveness of employment to local population appear to be the ones that produce goods with low transportation cost-to-value ratio.

The behavior of the manufacturing industries presented here contrasts with the patterns broadly observed in wholesale and retail industries. Holmes (1999a, 2000) finds that the total industry output is a convex function of city population in various wholesale industries, and presents a model where establishment scale, as well as variety, increases with city size. In the case of retail industries, Campbell and Hopenhayn (1999) find that average output of establishments in a city increases significantly with city size for a majority of the industries, while the number of establishments increases less than proportionally with city population. It appears that an increase in local market size seems to induce an expansion in the scale of an individual production unit in the wholesale and retail sectors, but not systematically so in the manufacturing sector.

The rest of the paper is organized as follows. Section 2 discusses the theoretical motivation for the empirical analysis to follow. The empirical methods to uncover the relationship between industry aggregates of interest and city size are described in detail in section 3. Section 4 describes the data used, followed by the presentation of the results in section 5. Section 6 discusses the results and relates them to the literature. Section 7 concludes.

2 Theoretical Motivation

The basic goal of this theoretical section is to discuss how the number of establishments and establishment scale are related to city size in models frequently used in the literature. The models should not necessarily be viewed as competing models. Both classes of models are in fact too simplistic in many dimensions. The aim is to obtain a feel for what the theoretical literature has focused on so far, and the empirical implications resulting from that literature. The literature has basically featured two main types of models relating the size of the economic activity in a locality to that locality's size. Fujita (1996) contains a detailed discussion of the structure of these two types of models. The first class of models uses the monopolistic competition framework based on Dixit and Stiglitz's (1977) work concerning CES utility functions (e.g. Abdel-Rahman and Fujita (1990)). These models have been the building block of the 'new economic geography', most notably exemplified by Krugman (1991). In these types of models, the market interaction between producers and consumers occurs through supply and demand linkages, i.e. through pecuniary externalities. The second type of models feature a competitive environment where market interaction of producers and consumers does not affect the actions of others directly. Instead, 'technological' externalities, such as knowledge spillovers, play a role in determining the productivity of the economic activity.

2.1 The Class of Monopolistic Competition Models

Krugman's (1991) model is the basic example of monopolistic competition models used in economic geography. The model features multiple locations (albeit only two) and tradeability of manufactured goods between these locations. This section will outline the model and its empirical implications for establishment size and number of establishments in a locality. Since the model is quite well-known and appears frequently in the literature, the discussion here will be brief and the reader is referred to Krugman (1980, 1991) and, especially, Neary (2001) for details.

Krugman's prototype model considers an economy with two locations that have different endowments of labor, which is mobile between the two locations. Workers are consumers at the same time. There are two sectors: agriculture, a constant returns to scale sector, which is tied to the land, and manufacturing, an increasing returns sector that can be located in either location. Agricultural goods can be transported costlessly across locations. On the other hand, manufactured goods can be traded between two regions subject to a transportation cost which takes the iceberg form; that is, a certain fraction of the good is lost during transportation. Many differentiated manufactured goods are each produced by identical monopoly producers using an increasing returns technology that involves a fixed cost and a constant marginal cost, both defined in units of labor. The high elasticity of substitution between differentiated products and increasing returns in the cost function ensures that each differentiated good is produced by only one producer and in any given location.

Consumers' preferences are symmetrically defined over the manufactured goods using the CES utility function. An important feature of the model is that, because of the structure of the CES specification and iceberg type transportation costs, the elasticity of demand facing any individual firm is a constant, regardless of the location of the demand for that firm's product (see Krugman (1980) or Neary (2001) for details on this result). This property leads to a constant output per firm and a constant firm size in terms of employees, irrespective of the wage rates, relative demand, etc. in the two locations.

Suppose now that wage and income levels are exogenously fixed across locations. If transportation costs are prohibitively high so as to make each location effectively a closed economy (as in the case of goods with a high transportation cost-to-value ratio), then the constant establishment size implies that the number of establishments should be proportional to the local demand. In other words, a 1% increase in the local demand should lead to exactly a 1% increase in the number of manufacturing establishments and manufacturing employment.

2.1.1 The Home Market Effect and the Role of Transportation Costs

When transportations cost for the manufactured goods are moderate, the 'home market effect' kicks in. The essence of this effect is the following. If a manufactured good is not produced in a given location, it has to be imported. Because imports incur transportation costs, and locally produced goods do not, this implies a lower manufacturing composite price index and a lower cost of living for the larger market. Since firm output and price of each variety are fixed, a lower price index can be sustained only through an expansion in the number of varieties produced, or equivalently, the number of producers. In other words, a larger market allows for a majority of sales to be carried out without incurring the transportation costs, so manufacturers would prefer to locate in a larger market, ceteris paribus. The implication of the model is that the market with a higher share of total demand should have a proportionally higher share of total manufacturing, or a 1% increase in local demand should yield a more than 1% increase in local manufacturing employment.[4] The magnitude of the home market effect depends on the magnitude of transportation costs, holding everything else constant. When the transportation costs are very high, and we are back in the case of a collection of closed economies, and each location produces every type of good. Furthermore, the home market effect becomes more pronounced as transportation costs decline within a certain range.[5] However, below some critical level of transportation costs, the home market effect disappears, as the nature of the equilibrium changes from one of a geographically dispersed economy to one where any one of the locations is potentially an agglomeration point. When transportation costs are exactly zero, being close to a larger market does not lead to any gains in terms of transportation costs. These implications follow because higher transportation costs should induce the production of the same type of good to be replicated by different producers in both locations, whereas lower transportation costs imply that a single producer of any given differentiated good located in the larger market can serve both markets and also eliminate transportation costs on a majority of shipments. The magnitude of the home market effect also depends on scale economies. Essentially, it arises as a result of the interaction between scale economies and transportation costs.[6]

It is important to note that the emergence of the home market effect depends crucially on

[4] A more technical statement of the home market effect can be found on page 540 of Neary (2001). Strictly speaking, the effect is stated in terms of shares with respect to total economic activity across all locations. That is, the market with a larger share of the total demand across all locations should have a higher than proportional share of total manufacturing output across all locations. It is easy to see that, when stated in elasticities (i.e. taking the logarithms of these shares), this is equivalent to saying that if the local demand increases by one percent without altering the total demand across locations, there would be a more than one percent increase in the local manufacturing output.

[5] Technically, this can be seen from equation (12) of Neary (2001). The derivative of the proportionality term between industry size and demand with respect to transportation costs is negative.

[6] See, again, equation (12) in Neary (2001). The magnitude of the home market effect depends on the parameter that determines elasticity of demand, which, due to the special structure of the model, is also a scale parameter for individual establishments.

the assumed asymmetry of transportation costs for the manufacturing and agricultural sectors. In Krugman's model, agricultural goods can be costlessly traded. This assumption turns out to be not innocuous. Davis (1998) shows that when the transportation costs are equal across the two sectors, the home market effect disappears. Using data on internationally traded goods, he also finds no evidence that sectors characterized by high scale economies have unusually high trade costs. This finding raises a concern about the validity of the assumptions on the transportation costs and their relation to the home market effect.

In summary, the implications of this representative, albeit very restrictive, model are: *i*) Regardless of transportation costs, the establishment size as measured by either output or number of employees is a constant across locations, and *ii*) If the transportation costs are not prohibitively high but positive, the number of manufacturing establishments and manufacturing employment in a location are increasing more than proportionally with local demand. These implications will be investigated in the empirical work.

The unique establishment size implied by the model is clearly an oversimplification and it is counterfactual. Typically, both within and across many locations of an industry there is substantial heterogeneity in the size of establishments. The model above can be extended to accommodate a size distribution of establishments by introducing heterogeneity in the cost structure of establishments. For example, Holmes (1999a) analyzes a model of wholesale that allows tradeability across locations and introduces heterogeneity in the fixed costs of each differentiated product. That model implies a size distribution of establishments for each location that depends on the distribution of fixed costs across different goods and the size of the location. Individual establishments' sizes increase with local population because of the importance of local demand for the scale of wholesale establishments.

2.2 The Class of Competitive Models with Externalities

Another important strand of the literature in economic geography emphasizes the productivity enhancing role of technological externalities that stem from the agglomeration of inputs of production into close geographic quarters (e.g. Glaeser, Kallal, Scheinkman, and Shleifer (1992), Henderson (1986, 1987)). Unlike the pecuniary externalities involved in the monopolistic competition models above, these externalities directly influence the productivity of an establishment (i.e. learning from others, information spillovers, etc.), and/or the utility of consumers (i.e. pollution, congestion, etc.). If technological externalities are important, do they reflect themselves on the number and size of establishments in a city? If so, how? Unlike in the case of the monopolistic competition models, these issues have not been investigated in the framework of competitive models with externalities. The following model aims to provide a step in this direction. As in the monopolistic competition models, this model features multiple locations, tradeability of manufactured goods, and perfectly mobile factors of production. However, it differs from the monopolistic competition models outlined in the previous section by introducing a scarce resource (land), a perfectly competitive industry producing homogenous goods, and an explicit account of the establishment size distribution.

Consider an economy consisting of several cities. Take a city of population S. A typical manufacturing industry in the city is perfectly competitive with a large number of price taking establishments producing a homogenous good. The manufactured good is perfectly tradeable across cities and the transportation costs are negligible. In addition, labor and capital are perfectly mobile across cities. Land is the only scarce resource in the city.

Establishments and workers choose their locations simultaneously. Each establishment is run by a manager, whose managerial ability is summarized by a random parameter z, as in Lucas (1978). This parameter is independently and identically distributed across cities, that is, there are no systematic differences in the ability of managers across locations. An establishment owner first chooses a location, then hires a manager, who is a worker at the same time. The parameter z is then revealed.[7] An establishment then maximizes its profits using labor and capital

$$\pi(z;S) = zH(S)[n^{\alpha}k^{1-\alpha}]^{\delta} - w(S)n - rk \tag{1}$$

where $\alpha \in (0,1)$ and $H(\cdot)$ is an increasing, bounded function that describes the positive local externalities due to agglomeration. The externalities are embedded in a Hicks-neutral form in the production function, as frequently done in the literature.[8] The parameter $\delta \in (0,1)$ reflects decreasing returns to a fixed factor, such as the manager and/or the plant. The output is taken as the numeraire. The capital market is assumed to be competitive nationwide, so its rental rate, r, does not depend on city size. The wage, on the other hand, depends on city size, because the household's location problem introduces a link between local land prices and wages. This connection will become clear shortly.

The optimal choice of labor and capital by an establishment is given by the first order conditions

$$\begin{aligned} \delta z H(S)\alpha q &= wn \\ \delta z H(S)(1-\alpha)q &= rk \end{aligned}$$

where $q = [n^{\alpha}k^{1-\alpha}]^{\delta}$. Using these two conditions, the optimal choice of labor by the establishment can be expressed as

$$n^*(z) = \left(\frac{\phi z H(S)}{w(S)^{1-\delta(1-\alpha)}}\right)^{1/(1-\delta)} \tag{2}$$

where ϕ contains the parameters α, δ and the rent r. Equation (2) makes it clear that the employee size of an establishment depends on how $H(\cdot)$ and $w(\cdot)$ covary with S. Understanding this relationship further requires consideration of the household's problem.

[7] Also assume, for simplicity, that the lowest post-entry realization of z can is high enough to allow profitable operation in any location; that is, there is no exit once an entrepreneur chooses a location.

[8] See Henderson (1986), Glaeser, Kallal, Scheinkman, and Shleifer (1992), and Ciccone and Hall (1996).

Each individual in the city is a worker and maximizes the utility

$$u = x^{1-\beta} l^{\beta}$$

subject to the budget constraint

$$x + p(D)l = w(S)$$

where x is the consumption of the industry's product, and l is the land consumed for housing. Each worker is assumed to be endowed with one unit of time, so that income equals the wage rate. The price $p(D)$ is the relative price of land with respect to the price of industry's output, and it depends on the population density $D = S/A$, where A is the (fixed) land area of the city. The price of land is assumed to be an exogenously given, increasing function of density, as land is a scarce resource.

In equilibrium, neither establishments nor workers should have any incentive to relocate. For establishments, this requires expected profits net of fixed costs be equalized across locations. Moreover, free entry implies these profits are exactly zero. That is,

$$E[\pi(z; S)] = f(S)$$

where $f(\cdot)$ is a fixed cost which is allowed to vary across locations as a function of city size, and the expectation is taken over z. Inserting the post-entry choice of labor, n^*, into the profit function, and then taking the expectation with respect to z, the free entry condition at any location is given by

$$\psi w(S) \left(\frac{H(S)}{[w(S)]^{1-\delta(1-\alpha)}} \right)^{1/(1-\delta)} = f(S) \tag{3}$$

where ψ is a constant consisting of α, δ, capital rent r and the expected value of a function of z. This equation can be used to relate the externalities to wage and fixed costs as follows

$$H(S)^{1/(1-\delta)} = \frac{1}{\psi} f(S) w(S)^{\delta\alpha/(1-\delta)} \tag{4}$$

For workers, no incentive to move across cities implies that the utility must be the same across locations. With the Cobb-Douglas specification of utility, this implies

$$\frac{w(S)}{[p(D)]^{\beta}} = v \tag{5}$$

where v is the common indirect utility across cities. This equation implies that the higher costs of living, which is equivalent to the higher land prices in this model, must be exactly counterbalanced by an increase in wages. To close the model, the final requirement is that the market for the good and the labor market clear. For simplicity, assume an unlimited aggregate labor supply from the industry's point of view (e.g. the industry is small with respect to the overall manufacturing). This is not an implausible implication in the case of an individual industry. Then, given the mobility of labor, the scarcity of labor is not an issue.

It is now possible to analyze how establishment size changes with city size using the free entry condition in (3) and the constant utility condition in (5). We can express the establishment size conditional on z, after substituting for $H(S)$ using (4), as

$$n^*(z) = \theta z^{1/(1-\delta)} \frac{f(S)}{w(S)} \tag{6}$$

where θ is a constant. The average establishment size, $E[n^*(z)]$, is then determined by the relative magnitudes of the fixed costs and the wage. For the nature of the fixed costs, consider the following cases:

Case 1. Capital is the fixed factor, $f(S) = r$. In this case, average establishment size is proportional to $\frac{r}{w(S)}$, by (10). This implies that average establishment size decreases with S. This follows because, by (5), the density and the price of land increase with S, and so must the wage, while r is a constant across cities.

Case 2. A manager is the fixed factor, $f(S) = w(S)$. Note that a manager is assumed to be a worker at the same time, so the manager's compensation is equal to the wage. In this case, average establishment size is invariant to S, as can be seen from (6).

Case 3. Land is the fixed factor, $f(S) = p(D)$. In this case, average size is proportional to $\frac{p(D)}{w(S)}$. Solving for $w(S)$ in terms of $p(D)$ using (5), one obtains that average size is proportional to $p(D)^{(1-\beta)}$. This implies an increasing average size with S.

Now it is also easy to see the implications on the whole size distribution. Since the ex-ante distribution of z is assumed to be the same across all locations, the optimal establishment size in (6) implies that cases 1, 2, and 3 lead to stochastically decreasing, invariant, and stochastically increasing size distributions as S increases, respectively. Note further that the number of establishments in each location is implicitly determined by the free entry condition. In adjusting to equilibrium, since locations with higher externality imply greater post-entry profits for any given z, there would be higher entry in these locations, eventually driving expected profits at that location to zero. The specific rate of change in the number of establishments depends on the exact way the function $H(S)$ and $p(D)$ are specified. The profits for any given level of externality is increasing in z. Similarly, for any given level of z, the post-entry profit increases with the magnitude of externalities. The model does not incorporate any mechanism that would lead to concentration of high z entrepreneurs to high externality locations. This is partly because an entrepreneur does not observe his ability level before choosing a location. If that was possible or if the entrepreneur could relocate, higher z entrepreneurs would continue to locate at the highest externality location until the net profits are driven to zero at that location. This could lead to the sorting of entrepreneurs into locations with different externality levels according to their abilities, an extension not pursued here.

Finally, note that the model assumes away any impact of amenities in a city on the wages and the land costs. Externalities cause population, density, wages, and land costs to be perfectly correlated. This need not be the case when amenities affect the utility and/or production functions as in, for example, Roback (1982). Therefore, empirical work controls for the differences across cities in wage and land cost. The return from doing so is that wage and land costs will in part capture the differences in amenities.

3 Empirical Approach

As outlined in the previous sections, the two classes of models have implications on how the number of establishments, employment, and the size of establishments vary with city size. Since both classes of models are restrictive in many dimensions, empirical analysis will investigate these relationships without restricting the estimation procedures to a particular model's environment and functional forms. A deeper analysis of the models using specific structures of the models is left for future work. Econometric methods used to analyze each relationship are explained in more detail in the next section.

3.1 Employment and Number of Establishments vs. City Size

The models discussed relate the number of manufacturing establishments and manufacturing employment to city size and other city-specific characteristics that can influence local demand and cost structure for the industries. Let $Y(S)$ be the dependent variable, which is either the number of establishments or the total employment in a city of size S. The main model to be estimated is a log-linear model

$$\log Y(S) = \alpha + \beta \log S + \mathbf{x}'\boldsymbol{\delta} + \varepsilon \tag{7}$$

where $\mathbf{x}$ is a vector of city-specific variables. The error term ε is assumed to be distributed according to an $i.i.d.$ random variable across cities. In addition to this linear specification, a quadratic term in $\log S$ will be included to check for any additional non-linearities that remains after a log-linear specification.

The log-linear specification makes it possible to assess how the dependent variable responds to city size in proportional terms. If the variable is increasing in proportion to city size, we expect the estimate for β to be close to 1. A more than proportional increase would imply an estimate greater than 1. Thus, a coefficient significantly higher than 1 indicates a convex relationship between city size and the dependent variable. A quadratic term in logarithm of population will also be added to check for any remaining non-linearities after log-linearization.

Both models involve taste and cost parameters that vary across locations. If there are systematic differences across cities in wages and fixed costs of establishments that are not fully related to population (for example, amenities in a location might drive up wages), then not including them as regressors would result in biased estimates. Therefore, variables that account for differences in production costs across cities will be included in $\mathbf{x}$. In addition, differences in tastes of consumers across cities might have an impact on the demand for the goods and, consequently, on the number of establishments. For example, a key parameter in Krugman's model is the weight placed on the manufactured goods in the utility function, which is expected to vary across locations. Thus, controls for demographic differences across cities will also be included in $\mathbf{x}$. The full set of regressors will be discussed in detail in the section that describes the data.

3.2 Establishment Size vs. City Size

The pattern of establishment size across locations can be analyzed using data either on the average establishment size and/or data on the establishment size distribution. The two measures are complementary and together provide a detailed account of the responsiveness of establishment scale to city size.

3.2.1 Establishment Size Distribution

The usage of the quantiles of the size distribution have some important advantages over the average size measure. First, the average size has the drawback that the existence of large establishment in a city might induce simultaneity bias; that is, the population and average size would be determined simultaneously in such cases. In contrast, the empirical cumulative distribution function (*c.d.f.*) is less sensitive to the existence of such large establishments. Second, while the average size can be influenced substantially by the existence of a very large establishment in a city, the empirical *c.d.f.* responds less to such outliers.

Suppose that, for any given city of size S, the counts of establishments of size less than E^j employees are available for $j = 1, ..., J$ different levels of E^j, and are denoted by $N^j(S)$, for $j = 1, ..., J$. Using this, the empirical *c.d.f.* of the size distribution can be calculated at all available J points as

$$F^j(S) = \frac{N^j(S)}{N(S)}$$

where $N(S)$ is the total number of establishments in the city. The behavior of the size distribution across cities can then be analyzed in a regression framework using the empirical *c.d.f.* as the dependent variable[9]

$$F^j(S) = \alpha^j + \beta^j \log S + \mathbf{x}'\boldsymbol{\delta}^j + \varepsilon^j \quad j = 1, ..., J \tag{8}$$

The J equations in (8) can be estimated for each equation by using OLS. One would like to see whether the *c.d.f.* changes systematically, e.g. whether the size distribution is increasing or decreasing stochastically with population or density. Recall that a distribution $F(\cdot)$ stochastically dominates (in the first order sense) another distribution $G(\cdot)$, if $F(x) \leq G(x)$, at each point x in the common support of the two distributions. An empirical counterpart to this definition can be

[9] Since the dependent variables are always in $[0, 1]$, another possible approach is to use a multinomial logit or probit. However, such methods assume functional forms for the error term that may be restrictive.

obtained by analyzing the signs and significance of the estimated coefficients, $\widehat{\beta}^j$. In particular, the following convention will be used in this paper. The size distribution is stochastically increasing (decreasing) if the following conditions hold: *i*) At least one coefficient is negative (positive) and significant at 5%, and *ii*) none of the coefficients is positive (negative) and significant at 5%, *iii*) all coefficients are jointly significant at 5%. The joint estimation of equation (8) for J different size levels is implemented using the seemingly unrelated regressions (SUR) framework. The coefficient estimates are the same as in separate OLS regressions. However, the joint significance tests take into account the correlation of error terms across equations, which are not likely to be independent.[10]

3.2.2 Average Establishment Size

The change in average establishment size across cities will be analyzed using the basic framework in (7) with several different specifications. The first one is the logarithm of the average size, where average size is defined in the usual way by dividing the manufacturing employment in the city by the number of manufacturing establishments. Call this specification I. This will be used to analyze average size in individual 2-digit industries as well. Specification II simply adds a second-order term in logarithm of population to check for any remaining non-linearities after the log-linear specification.

For overall manufacturing, the usual definition of the average size ignores differences across industries in average establishment size as well as the industry composition within a city. To account for the influence of these on average establishment size, specification III adds industry dummies to the right hand side of (7). This effectively constrains the coefficient for population to be the same across all industries. Clearly, this may be too restrictive. As an alternative, one can use the information on the relative importance of different industries in city employment. For a given industry i, denote the industry's employment share in city c by s_{ic}, the industry's average establishment size in city c by $\overline{n}_{ic}$, and the industry's average establishment size across all MSA's by $\overline{n}_i$.[11] Define the weighted average establishment size for overall manufacturing in city c as

$$\overline{n}_c = \sum_{i=1}^{I} s_{ic} \frac{\overline{n}_{ic}}{\overline{n}_i}$$

where I denotes the number of 2-digit industries analyzed. For a given city, this measure considers

[10] This joint test is a simple Wald test for the hypothesis that all estimated coefficients are equal to zero. An alternative is the Bonferroni approach. The criterion for significance in this test is far less strict than the one in the Bonferroni approach. The Bonferroni approach would require a significance level of $\frac{5\%}{J}$ for each of the J coefficients individually, for an approximate joint significance of 5%. The Wald test has the advantage that error terms across the J equations are not restricted to be independent.

[11] The average size is calculated by dividing the total employment of the industry within MSA's by the total number of establishments within MSA's. Note that it is important here to focus only on MSA's rather than the industry's nationwide average establishment size, because average size of an industry differs across urban and rural places, as analyzed in a later section.

the deviation of the average establishment size in that city from the overall average establishment size, and weights it by the relative importance of the industry for that city. Specification IV uses the logarithm of $\overline{n}_c$ as the dependent variable.

In all the econometric procedures discussed so far, potential endogeneity between manufacturing aggregates and city size is ignored. As such, the proposed regressions should be viewed as simple projections, rather than suggesting any structural connection between the variables considered. The endogeneity problem is mitigated to some extent if one considers the role of amenities in local factor prices, such as wage and rent. Thus, local factor prices and population will in part be determined by the amenities, and not solely by the extent of the industrial activity. Also, as mentioned before, the establishment size distribution is less prone to simultaneity problems, in the sense that a large establishment in a city leads to an abrupt increase in population, employment and average establishment size, but not a substantial shift in the size distribution.

Before closing this section, consider the empirical comparability between the pattern for the average establishment size and the pattern for the establishment size distribution. In theory, a stochastically increasing (decreasing) distribution implies an increasing (decreasing) mean, but not vice versa. However, this theoretical relationship is not always expected to hold here, because of the discrete nature of the size classes. Consider the following fictitious setup. Suppose that there are two cities and only three size classes: 0 to 10, 11 to 20, and 21 to 40 employees. In the smaller city, the industry has the following number of firms in the respective size classes: 1, 2, 2. Suppose that the actual sizes of establishments in this city are 10, 20, 20, 30 and 40 employees. In the larger city, the number of establishments in each size category are 1, 2, and 5, respectively, and the actual sizes of establishments are 5, 12, 14, 25, 25, 25, 25 and 30 employees. The values the *c.d.f* takes on in the small city are then 0.20, 0.60 and 1, whereas in the large city they are 0.125, 0.375, and 1. This implies a stochastically increasing size distribution with city size, as the large city has lower *c.d.f* values for the first two size classes. However, the average establishment size in the small city is 24 employees, whereas in the large city it is 20.1, implying a decreasing average establishment size. Thus, the distribution of individual establishments' sizes within employment size classes can potentially lead to discrepancies between the patterns exhibited by the establishment size distribution and the average establishment size.

4 Data

The main data sources are the 1990 edition of the County Business Patterns (CBP), and the 1994 edition of the County and City Data Book (CCDB), both available from the U.S. Census Bureau. The CBP provides, for each county and each manufacturing industry (classified by the Standard Industry Classification (SIC) Code), total number of establishments, total employment, payroll, and number of establishments in different size categories by employment. Employment and payroll are compiled from the payroll tax records of individual establishments, which makes the CBP a highly reliable source. The CCDB is an eclectic source of data obtained from several government institutions as well as private organizations. There is a wide range of demographic and economic variables available at the county level in the CCDB.

4.1 The Geographic Unit of Analysis

The choice of the geographic unit of analysis is not trivial. Are the models applicable to narrowly defined places, such as the most dense parts of a city or a county, or to larger units such as regions or countries? It is clear that political boundaries hardly coincide with the geographic extent of economic activity. For a first exploration, it seems reasonable to carry out the analysis at the Metropolitan Statistical Area (MSA) level. An MSA is an urbanized area consisting of a central city and contiguous counties which are economically and socially integrated with that city. As such, MSA boundaries are natural extents of urbanization and economic activity. Both models discussed earlier assume tradeability of goods across locations with perfectly mobile labor and/or capital. Considering that there is substantial trade of manufactured goods across the MSA's in the U.S. with little or no barriers to trade, the MSA level analysis provides a reasonable setup for investigating the validity of the models' implications. MSA level aggregation of data is not directly available. Rather, the data were assembled from the county level, using the constituent counties of MSA's. The Primary Metropolitan Statistical Areas (PMSA) were also treated as MSA's. PMSA's are collection of counties within large MSA's (called Consolidated Metropolitan Statistical Areas (CMSA)) and constitute separate entities that are economically and socially integrated, and relatively less dependent on the rest of the CMSA. New England County MSA's (NECMA's) were excluded from the analysis because their definitions many times include certain parts of counties that cannot be matched with the geographic detail available in the County Business Patterns Data Set. Finally, MSA's in Alaska and Hawaii were also excluded to focus only on the contiguous states. This selection resulted in a total of 297 MSA's.[12] These 297 MSA's altogether accommodated approximately 75% of all manufacturing employment in 1990.

4.2 Construction of Establishment Size Measures

As mentioned in the estimation methodology, there are two variables that can be constructed regarding the establishment size in a city: average establishment size and empirical *c.d.f.* of the establishment size distribution. While the total number of establishments in an industry is available for all counties, in many cases total manufacturing employment in a county is suppressed to prevent disclosure of individual establishments' employment levels, especially when there are only a few large establishments in a county. For manufacturing industries, the percent of counties which have data suppression range from a minimum of about 45% (in the Lumber and Wood Products industry)

[12] Further details on the 1990 standards for defining MSA's and other geographic units, visit the U.S. Census Bureau's website at *http://www.census.gov/population/www/estimates/mastand.html.*

to a maximum of about 90% (in the Leather and Leather Products industry). This suppression, however, is less severe for MSA counties, as these counties tend to have many establishments. To calculate the manufacturing employment and average establishment size in a city, data suppression has to be overcome some way. The methodology used in replacing the suppressed employment levels in a county is described in detail in the Appendix.

The empirical *c.d.f.* is constructed as follows. For each county-industry, CBP provides the employment, the number of establishments and the establishment counts in each of the 12 different size categories. The empirical *c.d.f.* of establishments is obtained by dividing the cumulative establishment counts at these categories by the total number of establishments in the county. The focus here is on 6 size categories: 0-19, 20-49, 50-99, 100-249, 250-499, and 500-999 employees. More than 95% of all establishments are covered for all industries with this choice of size groups. This leaves out establishments larger than 999 employees. At size classes with 1000 or more employees, the *c.d.f.* is equal to or very close 1, and there is little variation across cities in the tail probability. Considering the small changes in the *c.d.f.* at higher size classes, an alternative is to restrict the attention to only the first 3 or 4 size categories. This exercise does not lead to substantially different results in terms of the direction of change in the size distribution, but tends to increase the statistical significance of the shift in the *c.d.f.* in certain cases.

4.3 The Variables

A description of the regressors used in the analysis is provided in Table 1. Population is the sum of the populations of counties that constitute the MSA. The rest of the variables used in the analysis are controls for other local features that can potentially affect the size of establishments. To control for the impact of factor prices, county average manufacturing wage, median house rent in the county, and cost of electricity and natural gas at state level are included.[13] Median house rent is used as a proxy for cost of space for a manufacturing establishment. It clearly is not the ideal measure, which would be the cost per square foot of a manufacturing plant. Such a measure, unfortunately, is not available.

Another set of variables is used to describe the city demographics. These variables include age composition, educational attainment, and race composition in the city. In addition to controlling for tastes that might affect the demand for a certain manufactured product in a city, these measures also control for the differences in worker skills, the differences in entrepreneurs' abilities, and, hence, the choice of technology across cities. Aggregation of these variables to the MSA level from the county level was done using the shares of MSA population in each of the constituent counties as the weights.

[13]For about 15% of all counties in U.S., average manufacturing wage in the county is not available due to data suppression. Whenever this poses a problem for a county that is part of an MSA, average wage in all industries in the county was used instead. The correlation between average manufacturing wage and average wage is 0.78 for those counties for which both wage measures are disclosed.

Table 2 provides descriptive statistics for the variables and their correlations with population. Note that income, wage, and rent are strongly correlated with both population, but the correlation is far from perfect. Other variables do not seem to be highly correlated with population.

5 Results

Following the empirical approach described earlier, the data described in the previous section is used to analyze the behavior of the number of establishments per capita, the employee size distribution, and the average establishment size across cities. Before going into more detailed empirical analysis, the next section highlights some important preliminary observations for overall manufacturing.

5.1 A Preliminary Look

Table 3 provides a summary of average establishment size for overall manufacturing and 2-digit components. Counties are classified into two groups, urban and rural, based on whether they are a part of an MSA or not. Using the standard measure of average size, a clear picture emerges: Establishments tend to be in general smaller in urban counties compared to rural counties. In 6 industries this result is reversed as emphasized in bold figures. Industries also differ substantially in the distribution of employment into urban and rural counties. In all but one of the industries, more than half of the industry employment is in urban counties. The Lumber and Wood Industry is the only exception. Industries also exhibit significant variation in their average share of total manufacturing employment in a city. The Food and Kindred, Industrial Machinery, and Transportation Equipment industries have the largest average share of about 10%. The Leather and Petroleum and Coal industries have the lowest average shares. But all shares exhibit substantial variation across cities, as indicated by relatively high standard deviations in many cases.

Figure 1 displays the relationship between manufacturing's share in MSA employment and MSA size. The average share across MSA's is 20% with a standard deviation of 9%. Las Vegas MSA in Nevada has the lowest share with only 3.1%, and Hickory-Morganton MSA in North Carolina has the highest at 56.1%. Overall, it appears that the share is declining with population, but there seems to be substantial variation left unexplained by population. A simple bi-variate regression of the share of MSA employment on the logarithm of MSA population yields a statistically significant (at 1%) coefficient of -0.034 with a t-statistic of -3.15. This implies that a 1% increase in MSA population is associated with -0.034 units of decline in manufacturing's share of employment. The R^2 value from this regression is only 0.026, suggesting that population cannot account for much of the variation in manufacturing's share of MSA employment. Clearly, other factors besides population are likely to contribute to this variation. Furthermore, no evidence of non-linearity was found, as adding a quadratic term in the logarithm of the population produced insignificant coefficients for both the linear and quadratic terms.

Figure 2 suggests that the average establishment size, calculated the conventional way, is declining with population, again subject to considerable variation. The average establishment size has a mean value of 58.3 employees with a standard deviation of 27.2 employees. Santa Fe MSA in New Mexico has the lowest average size: only 11.4 employees per establishment. The highest average size, approximately 229 employees per establishment, occurs in Kokomo MSA in Indiana. A regression of the average establishment size on the logarithm of the MSA population yields a statistically significant (at 1%) coefficient of -11.75 with a t-value of -3.67. This suggests that a 1% increase in population is associated with a decline of 11 employees per establishment. The R^2 value is only 0.036, again, indicating that factors beyond just population contribute substantially to variation in average size. As before, adding a quadratic term failed to produce significant estimated coefficients for the linear and the quadratic term. As shown in Table 4, using the logarithm of the average size as the dependent variable produces a coefficient of -0.052 significant at 5% with a t-statistic of -2.19. This coefficient implies a 0.05% decline in average size accompanies a 1% increase in MSA population. Using the weighted average size results in a coefficient that indicates a higher and more significant decline of about 0.06% as also shown in Table 4.

Figure 3 displays the relationship between the number of establishments and MSA population. The average number of establishments across MSA's is 898, with a standard deviation of 1873. The highest and lowest values are 43 and 19,649, occurring in Cheyenne MSA in Wyoming and Los Angeles-Long Beach PMSA in California. The overall relationship is convex in nature. The convex pattern is not driven by the fact that the horizontal axis is in log scale in this figure. In fact, a regression of the *level* of establishment count on the *level* of population and its square yields positive coefficients, both are significant at the 1% level. The regression using the *logarithm* of population and its square as explanatory variables similarly indicates a statistically significant convex relationship. These results suggest that the number of establishments increases more than in proportion to MSA population. When the logarithm of the number of establishments is regressed on the logarithm of population, one obtains a highly significant coefficient of 1.079 with a t-statistic of 57.37, as summarized in Table 4. The R^2 of this regression is 0.89. Thus, MSA population does a good job in explaining much of the variation in the number of establishments. A quadratic term in the logarithm of population turns out to be insignificant, suggesting that the assumed log-linear relationship is a reasonable approximation. Note that the estimated coefficient of population is also significantly different from 1, indicating that number of establishments increases more than proportionally with city population.

The pattern of manufacturing employment across cities is shown in Figure 4. As in the case of establishments, employment is increasing with MSA population. The lowest employment (953) is in Great Falls MSA in Montana, and the highest (875,837) is in Los Angeles-Long Beach PMSA in California. The convexity of the relationship is confirmed in a regression of employment *level* on population *level*. When logarithms, instead of levels, are considered, this convex pattern disappears, a likely indication that certain outliers drive the convexity in levels. The estimated coefficient of population in the log-linear case is 1.027 with a t-statistic of 31.96, as shown in Table 4. The estimated coefficient, however, is not significantly different from 1, indicating that one cannot reject the hypothesis that overall manufacturing employment is increasing proportionally with city size. A quadratic term in the logarithm of population turns out to be statistically insignificant. In a cross-equation test, the estimated coefficient for population is found to be significantly different and lower from the estimated coefficient of population in the establishment regression. The more-than-proportional increase in number of establishments and a proportional expansion in employment is consistent with the declining average establishment size.

Figure 5 provides an overview of the pattern for the size distribution. The empirical cumulative distribution function is calculated for all cities within 4 different population quartile ranges, and the figure plots the average *c.d.f.* within these ranges at 7 different size classes.[14] The plot makes it clear that there is a systematic shift in the *c.d.f.* as city population increases. Establishment size distribution stochastically decreases as one moves to higher quartile ranges for population. This picture is consistent with the declining average establishment size.

Finally, Table 5 reports the bi-variate regression results using equation (8) with the logarithm of the population and its square as the only explanatory variables. When only the logarithm of population is used, all estimated coefficients are positive, and all but the one for the smallest size class are significant at 1%. In addition, all coefficients were found to be jointly significant across equations at 1%. These results strongly point to a stochastically decreasing size distribution as the city size increases, in accordance with the convention adopted earlier. The quadratic term does not turn out to be significant in any of the equations, suggesting no strong non-linear relationship between the *c.d.f.* values at the size classes chosen here and the logarithm of the population. The stochastically decreasing size distribution is consistent with the decreasing average establishment size with population.

The following section extends this preliminary look in several dimensions as discussed earlier. First, the bi-variate regressions are extended to a multi-variate setting. Second, the possibility that the patterns for overall manufacturing could be driven by industry composition is investigated by considering the 2-digit breakdown of the industries classified under manufacturing.

5.2 Results for Overall Manufacturing

Table 6 reports the multi-variate regression results for overall manufacturing. Several specifications are used as described before. Note that specification III only applies to average establishment size, so it does not appear in the results for employment and establishment regressions. First, note that the signs of the estimated coefficients for population in specifications I, III, and IV are consistent with the signs of the coefficients in the bivariate case. Magnitudes of the coefficients

[14] This includes the 6 size classes described before, and the additional size class of 1000+ employees.

differ across specifications to some extent. Also note that co-variates have consistent sign and significance across all specifications for any given dependent variable.[15] In particular, wage rate has a positive association to all dependent variables, and rent has a negative association. Income seems to matter most for employment, and hence, for average establishment size. Natural gas and electricity costs have also uniform coefficients across all specifications for all dependent variables. Other controls have varying signs and magnitudes across different dependent variables. Also note that quadratic terms in specification II are almost never significant at 5% for any of the dependent variables, indicating that log-linearity is not a gross misspecification. Therefore, the rest of the discussion focuses on specifications I, III, and IV.

Consider specification I. The results indicate that number of establishments is increasing more than proportionally with city size: for a 1% increase in city population, there is an increase of 1.073% in number of establishments, controlling for observables. Some, but not all, controls have significant coefficients. Employment also responds positively to city size across all specifications, but the estimated coefficient is not significantly different from 1. Thus, employment appears to grow in proportion to city population. As a consequence, the average establishment size is declining at a rate of about 0.03% with a 1% increase in city population. However, average size seems to respond much more to some of the other controls in specification I, notably rent and income. Average size appears to be significantly lower in high rent cities, and higher in higher income cities. The negative and large coefficient of rent in average size and employment regressions lines up with an interpretation that cities where space is at a premium have smaller establishments on average, as well as lower manufacturing employment.

Specification III uses the weighted average size measure. The results are not different in terms of the signs of most coefficients compared to specification I. But the magnitudes of the estimates differ. The decline in average size is now somewhat more pronounced at 0.041%. Other coefficients also change in magnitude, but their signs and significance levels are not substantially different from those in specification I.

Finally, specification IV adds industry dummies to specification I. The decline in average size is robust to this specification. The coefficients of population in employment and establishment regressions both change in magnitude somewhat, but the main findings – number of establishments increases more than proportionally with city size and employment is proportional to city size– still hold under specification IV.

Next, consider the multi-variate regression results for the empirical *c.d.f.* of establishment size in Table 7. The notable pattern from this table is that the estimated coefficients of population for different size classes appear to be quite robust in sign, magnitude, and significance to the

[15] Note that for specifications I, II and IV, the difference between the estimated OLS coefficients for number of establishments and employment should be equal to the estimated coefficients for average establishment size, because of the log-linear specification and the definition of average establishment size. However, note also that this identity does not hold for LAD regressions used for individual 2-digit industries to reduce the effect of potential outliers.

addition of controls. The only exception is the estimated coefficient for the smallest size class (0-19 employees). This coefficient is much smaller and much less significant compared to the estimate for the bi-variate analysis in Table 5. Thus, it appears that the fraction of establishments that are 'small' respond little to city size. The cumulative fractions in other size classes increase significantly with city size. All coefficients are jointly significant at 1%, and all but one coefficient are significant individually at 1%. This clearly points to a stochastically decreasing size distribution following the convention described earlier. Note also that the sign, magnitude, and significance of coefficients for controls are broadly consistent with the estimates for average establishment size. As expected, the coefficients have opposite signs compared to those in the case of average size. For example, size distribution stochastically decreases (i.e. estimated coefficients are positive) with rent and average size decreases (i.e. estimated coefficient is positive).

5.3 Results for Individual Industries

Table 8 reports a summary of the multi-variate regression results for 2-digit manufacturing industries.[16] For clarity, only the coefficient estimates for population are reported, as these are the primary objects of interest. As the variation in the dependent variables appears to be higher for individual industries compared to overall manufacturing, least absolute deviation (LAD) regression results are reported along with ordinary least squares (OLS) regression, to check for the robustness of the results to potential outliers.

In 12 of the 18 2-digit industries, the number of establishments increases more than proportionally with city size (i.e. the estimated coefficient is greater than 1), but the coefficients are significant at 5% only for 9 of the industries. For 4 industries, the coefficients are significantly below 1, indicating a less than proportional increase with city population. For 3 industries, the number of establishments increases in proportion to city size. The rate of increase appears to be highest for the Apparel, Leather, Electronics, and Instruments industries, and lowest for the Textile and Lumber industries.

Turning to the results on employment, in 10 of the industries, employment appears to increase more than proportionally with city population. In the rest of the industries, the hypothesis that employment is proportional to city population cannot be rejected. The strongest increase in employment appears to be in Apparel, Chemicals, Transportation, and Instruments. The lowest increase is for Leather, but the coefficient is not significantly different from 1.

Results on average establishment size generally reveal positive coefficients for population, but the positive OLS coefficients are significant in 8 industries. When LAD regressions are considered, the positive coefficients are significant only in 6 cases. In some cases, magnitudes of coefficient

[16] Two industries, Tobacco Products (SIC 2100) and Miscellaneous Manufacturing (SIC 3900) were excluded. Tobacco Products industry has no employment in many MSA's, hence the sample size was very small for precise estimation. Miscellanous Manufacturing contains establishments from a diverse set of unclassified manufacturing activities and this makes it hard to interpret the patterns emerging for that industry.

estimates vary significantly between OLS and LAD regressions, indicating the influence of outliers in the observations. The highest increase appears to be in Transportation Equipment, Apparel, and Chemicals. Negative coefficients are observed in 2 and 5 industries considering OLS and LAD regression results, respectively. However, the coefficients are significant only for the Leather and Electronics industries. Overall, while there is substantial variation across industries, there appears to be no uniform evidence that average establishment size changes with city size in any particular direction.

Next, consider the results on the empirical cumulative distribution of establishment size. Rather than reporting the individual coefficient estimates for population for each of the J equations in (8), a compact summary of the findings is represented. The column labelled 'Ind.' summarizes the individual significance of the six coefficient estimates. If at least one coefficient is positive and significant at 5%, and no coefficient is significantly negative at 5%, then a (+) is assigned to that industry. As discussed earlier, this is one way of empirically identifying a stochastically decreasing size distribution. Similarly, if at least one coefficient is negative and significant at 5%, and no coefficient is significantly positive at 5%, then a (−) is assigned to that industry, indicating a stochastically increasing size distribution. If all coefficients are insignificant individually, or some of them are significantly negative while others are significantly positive (implying an ambiguous change in the size distribution), then the table entry is a (0). Columns labelled 'joint' report the test results for joint significance of all coefficients at 5% using the SUR framework. If coefficients are jointly significant at 5%, then the column entry is 'Y', otherwise it is 'N'. A joint significance of the coefficients means that the size distribution responds to population in some way. However, it does not provide information on the direction of change. The significance of individual coefficients helps determine the direction of change.

Under this convention, the pattern for population suggests that, in 10 of the industries, size distribution decreases stochastically as city population increases. For 2 industries, it increases stochastically, and for 6 industries, the change in size distribution either does not move in any particular direction or the pattern is ambiguous. In 15 of the industries, the coefficients are jointly significant, implying that the size distribution responds to city population significantly. Overall, results point to a stochastically decreasing size distribution in many cases. It is important to note that the results on the size distribution do not necessarily coincide with the results on average size. This is not surprising and several sources may contribute to this discrepancy. The first source is the discrete nature of the employment size classes used in the calculation of the $c.d.f.$ The sizes of individual establishments within size classes may lead to discrepancies between the two measures. This was demonstrated with a simple example earlier. The second source is the relatively high sensitivity of average size to outliers compared to the $c.d.f.$ An unusually large establishment in a city may affect the magnitude of the estimated coefficient for population substantially. Such an establishment would be counted as just another establishment in a large size class for the calculation of the $c.d.f.$, influencing the estimated coefficients much less. A third potential source of discrepancy is the estimation procedure used for the replacement of the missing employment observations. However, given the detailed procedure used for data replacement, the results are not likely to be different, at least qualitatively, from what could have resulted if the actual data were available.

Taken together, the results broadly suggest that an expansion in the number of establishments, rather than an expansion in the scale of individual establishments, characterizes the behavior of 2-digit manufacturing industries across cities in U.S. As population increases, the number of establishments in many manufacturing industries increases more than proportionally with city size, and the employee size distribution of establishments is either stochastically decreasing with city size or not changing substantially.

5.4 The Pattern of Transportation Costs

As discussed in the theoretical motivation, an important prediction of the models in the spirit of the new economic geography tradition is that the responsiveness of manufacturing employment to home market size is sensitive to transportation costs. At the extreme, if the transportation costs are prohibitively high, production of the good must be carried out in each market, and manufacturing employment should be proportional to local demand. For moderate, but non-zero, transportation costs, most of the producers locate in the larger market and export to the smaller market. Note that in the case of zero transportation costs location of an establishment is independent of the transportation cost savings considerations. Thus, the extent of the home market effect is directly related to the magnitude of transportation costs, and the effect is expected to be larger for low-transportation cost goods. A crude test for the relationship between the home market effect and the transportation costs can be carried out using the estimated coefficients for employment in Table 8. For high transportation cost goods, the estimated coefficient should be close to 1, and the lower the transportation costs, the higher should be the deviation of the estimated coefficient from 1 in the positive direction. For such an analysis, however, estimates of transportation costs for each 2-digit industry are needed.

Estimates of the ratio of transportation costs to value of the good are available by 2-digit industries from the Bureau of Economic Analysis (BEA). The main source of data used here is the 1992 release of the U.S. Transportation Satellite Accounts. The data provides, by different commodity groups, a breakdown of the personal consumption expenditures into producer's value, transportation costs, and wholesale and retail trade margins. There are certain advantages of the satellite accounts over other measures of transportation costs available in other public and private sources. For instance, the measured transportation costs include the transportation activity within the firm for which there are no observable prices. As an example, the transportation activities that are conducted by a manufacturer when moving semi-finished goods from one of its plants to another by using its own truck fleet are taken into account. The data, the methodology in constructing it, and its merits are available in more detail in the article by Fang, Han, Lawson, and Lum (1998) at the BEA's website: *http://www.bea.doc.gov/bea/0498io/maintext.htm.* Transportation costs constitute an important share of the value of manufactured commodities. For overall manufacturing, transportation costs accounted for 18.7% percent of producers' prices in 1992, the highest among all industries. That figure was 21.0% for for-hire transportation (transportation services purchased from outside parties), and 13.2% for own-account transportation (transportation services carried out by the producers themselves).

One way of measuring the importance of transportation costs in the total value of a good is to use the percentage share of transportation costs in total personal consumption expenditures on that good. But the wholesale and retail margins component of expenditures is related to considerations that are not addressed by the theoretical models discussed, such as local competition between retailers. To avoid such complications, the estimate used here is simply the transportation costs as a percentage of the producer's value. The commodity classification provided by the BEA does not exactly match the 2-digit SIC industry codes. In many cases, an industry is broken down into two or more commodity groups, and some commodity groups that are classified under an SIC code are left out in the BEA data. To aggregate the available commodity classifications to the industry level, the transportation cost-to-producer's values ratio for all commodities under a 2-digit industry were summed using the expenditure shares of each commodity as weights. This resulted in the estimates shown in the last column of Table 8. The highest ratios are those of the Rubber and Plastics and Primary Metal industries, and the lowest ones pertain to the Fabricated Metal, Leather, Instruments, and Apparel industries.

Figure 6 plots the OLS coefficients for population against the estimated transportation cost-to-value ratios. The simple correlation between the two measures is -0.37. A simple regression line is also added. The estimated coefficient is -0.023 with a t-statistic of -1.56. It is not highly significant (14%), but indicates a decline in the responsiveness of employment to local market size as the relative importance of transportation costs increases. Although the transportation cost measure used here is very rough, the results are broadly suggest that employment responds less to local market size if transportation costs are high.

Finally, Figure 7 looks at the relationship between the OLS coefficients for population in average establishment size regressions and the transportation cost-to-value ratios. While the two variables appear to be negatively associated, the relationship is much weaker than the case with employment in Figure 6. The simple correlation between the two variables is -0.21, and a simple bi-variate regression reveals a slope coefficient of -0.008 with no significance at conventional levels. The responsiveness of average establishment size to city population does not seem to be highly related to transportation costs.

6 Discussion and Some Remarks

Empirical findings broadly indicate that scale of establishments as measured by employment does not necessarily increase in response to an increase in local market size in many manufacturing industries. In many cases, the number of establishments increases more than proportionally with city size. Employment responds to city population significantly in some of the industries, and these tend to be the ones that manufacture lower transportation cost-to-value goods. The home market effect thus appears to be important for some industries. In an earlier study of manufacturing industries using Japanese regional output data, Davis and Weinstein (1999) found that for 8 of the 17 industries home market effect appears to be important. They concluded that, in contrast to their earlier findings with international data (Davis and Weinstein (1996)), economic geography appears to be more effective in explaining inter-regional patterns of production within a country. Of the industries they studied, Instruments, Transportation Equipment, Electrical Machinery, and Chemicals exhibit substantial home market effects. In the analysis here, employment in Instruments, Transportation and Chemicals also respond substantially to city population. While countries, level of geographic aggregation, and measure of industry size differ across the two studies, the findings here also point to a potentially important role of economic geography in these industries. This encourages further exploration of these effects using a more detailed data and more flexible models. For example, Holmes (2000) studies the role of home market effects by analyzing the location pattern of manufacturer's sales offices, an industry that feature the basic characteristics for home market effects to be important: the existence of scale economies, product differentiation, and transportation costs. It is also possible to extend the work here by including measures or scale economies, such as the average establishment size, and measures of product differentiation, for which crude proxy measures can be constructed, such as the number of different 4- or 5-digit industries classified under a 2-digit industry. This would be a step towards assessing the relative importance of the three factors involved in generating the economic geography effects.

The competitive model with externalities as presented here is also consistent with increasing number of establishments and employment as city size increases. However, the dependence of establishments' minimum efficient scale on how externalities effect fixed versus marginal costs does not lead to sharp enough predictions that are readily testable. The fact that, overall, there is tendency for establishment size to decline with city size implies that minimum efficient scale is negatively related to the externalities. But this conclusion is far from identifying the effects of externalities on scale. The model discussed in this paper is a step towards developing a sharper framework that may guide an investigation of externalities potentially using plant level data on different inputs. This would help identify which components of a plant's fixed and variable costs are most affected by the size of the economic activity in the plant's geographic neighborhood.[17]

[17] Identifying the plants that geographically neighbor a manufacturing plant is possible using data available from the Census Bureau. See Holmes (1999b) for details of this data.

It is important to stress the differences with respect to the other studies regarding the behavior of retail and wholesale industries (see Campbell and Hopenhayn (1999), Holmes (1999a, 2000)). In both of these sectors, the scale of establishments, as measured either by employment or output, appears to increase with city size. The impact of local market size on scale of establishments seems to be more important for retail and wholesale sectors, compared to manufacturing. Furthermore, in many retail industries, the number of establishments increases less than proportionally with city size, a finding that does not generally apply to the manufacturing industries studied here.[18] These different patterns deserve further attention. One possibility is that local strategic interaction between firms, which is absent in models of manufacturing discussed here, is important in retail and wholesale industries and may be more visible within the boundaries of cities, compared to manufacturing where firms usually compete at the national or regional level. Thus, the patterns emerging might point to the impact of local competition on the number and size of establishments in these industries, a direction that deserves further exploration.

Another question that needs to be addressed is how the behavior of final good producers versus intermediate good producers or input suppliers differ. At 2-digit level aggregation, final good and intermediate good producers are all lumped together. In view of the theories and empirical work on specialization and the extent of the market (e.g. Stigler (1951), Holmes (1999b)), it is important to understand how the organization of 'upstream' versus 'downstream' industries differ. Typically, transportation costs between stages of production are likely to be important in determining the joint location pattern of industries at different levels of the production ladder. If final good producers are considered as the 'market' for specialist producers, then one might expect that the scale of specialist producers responds to the scale of final good producers. This may amplify the home market effect. Deeper investigation of these issues requires disaggregated data that clearly distinguishes between final good producers and suppliers.

An important shortcoming of the analysis is the lack of reliable output measures for manufacturing industries at city level. Standard specifications of production functions, such as Cobb-Douglas, allow for a monotonic relationship between employment and output. However, in view of the facts that average employment in manufacturing establishments has been decreasing over time[19], and that certain manufacturing industries are highly capital-intensive, the results using employment may not be a precise reflection of the actual relationship between industry output and city size.

[18] The two results – an increase in average establishment size and a less than proportional increase in number of establishments as city size increases – can be obtained in many standard models of imperfect competition, as discussed further in Campbell and Hopenhayn (1999).

[19] See Davis (1990) for a documentation of this observation.

7 Conclusion

The central finding of this paper is that in many manufacturing industries the employee size distribution of establishments appears to be stochastically decreasing with city population, whereas the number of establishments is increasing more than proportionally with city size. In some industries, employment increases with city size, in some others, no significant association is present between employment and city size. This is a general indication that larger cities tend to specialize in the production of certain manufactured goods, whereas in some industries employment is increasing in proportion to local market size. The class of monopolistic competition models often used in the economic geography have predictions that are in line with the patterns in some of the industries analyzed here. In particular, the home market effect appears to be important for some, but not all, manufacturing industries. There also appears to be a tendency for home market effect to get larger as transportation costs decline. While this evidence is weak and the measure of transportation costs used is very crude, further investigation of these patterns is promising. More detailed joint investigation of the patterns emerging for manufacturing, retail, and wholesale, as well as services, might prove to be informative. The relation of transportation costs to the patterns found encourages the development of models that provide a comprehensive account of the way the transportation costs matter for industry output and establishment scale in wholesale, retail, and manufacturing sectors. Such models can then be analyzed using data. This would eventually lead to a more comprehensive understanding of the role of cities in the economy.

References

[1] Abdel-Rahman, Hesham and Fujita, Masahisa (1990) "Product Variety, Marshallian Externalities, and City Sizes", *Journal of Regional Science,* **30**: 165-183.

[2] Campbell, Jeffrey R, and Hopenhayn, Hugo A. (1999) "Market Size Matters", Unpublished Manuscript, University of Chicago and University of Rochester.

[3] Davis, Donald R. (1998) "The Home Market, Trade, and Industrial Structure", *American Economic Review,* **88:** 1264-1276.

[4] Davis, Donald R,. and Weinstein, David (1996) "Does Economic Geography Matter for International Specialization?", *NBER Working Paper #5706.*

[5] Davis, Donald R,. and Weinstein, David (1999) "Economic Geography and Regional Production Structure: An Empirical Investigation", *European Economic Review,* **43:** 379-407.

[6] Davis, Steven J. (1990) "Size Distribution Statistics from County Business Patterns Data", Unpublished Manuscript, University of Chicago.

[7] Dixit, Avinash K, and Stiglitz, Joseph E. (1977) "Monopolistic Competition and Optimum Product Diversity", *American Economic Review,* **67:** 297-308.

[8] Fang, Bingsong, Han, Xiaoli, Lawson, Ann M., and Lum, Sherlene K.S. (1998) "U.S. Transportation Satellite Accounts for 1992", from April 1998 Survey of Current Business, Bureau of Economic Analysis.

[9] Fujita, Masahisa (1996) "On the Self-organization and Evolution of Economic Geography", *Japanese Economic Review,* **47:** 34-61.

[10] Glaeser, Edward (1999) "Learning in Cities", *Journal of Urban Economics,* **46:** 254-277.

[11] Glaeser, Edward, Kallal, Hedi, Scheinkman, Jose, and Andrei Shleifer (1992) "Growth in Cities", *Journal of Political Economy,* **100:** 1126-1152.

[12] Henderson, Vernon (1986) "Efficiency of Resource Usage and City Size", *Journal of Urban Economics,* **19**: 47-70.

[13] Henderson, Vernon (1997) "Externalities and Industrial Development", *Journal of Urban Economics,* **42**: 449-470.

[14] Holmes, Thomas J. (1999a) "Scale of Local Production and City Size", *American Economic Review,* **89:** 317-320.

[15] Holmes, Thomas J. (1999b) "Localization of Industry and Vertical Disintegration", *Review of Economics and Statistics,* **81:** 314-325.

[16] Holmes, Thomas J. (2000) "The Home Market Effect and City Specialization: Estimates for Sales Offices", Mimeo, University of Minnesota.

[17] Krugman, Paul (1980) "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review,* **70:** 950-959.

[18] Krugman, Paul (1991) "Increasing Returns and Economic Geography", *Journal of Political Economy*, **99:** 483-499.

[19] Krugman, Paul (1998) "Space: The Final Frontier", *Journal of Economic Perspectives,* **12:** 161-174.

[20] Lucas, Robert E. (1978) "On the Size Distribution of Business Firms", *Bell Journal of Economics,* **9:** 508-523.

[21] Lucas, Robert E. (2001) "Externalities and Cities", *Review of Economic Dynamics,* **4:** 245-274.

[22] Neary, Peter (2001) "Of Hype and Hyperbolas: Introducing the New Economic Geography", *Journal of Economic Perspectives,* **39:** 536-561.

[23] Roback, Jennifer (1982) "Wages, Rents, and the Quality of Life", *Journal of Political Economy,* **90:** 1257-1277.

[24] Stigler, George (1951) "The Division of Labor is Limited by the Extent of the Market", *Journal of Political Economy,* **59:** 185-193.

A Procedure for Estimating the Non-disclosed Data

Suppose that the total employment in industry i in county c, E_{ic}, is not disclosed. This typically results from the suppression of employment level, E^j_{ic}, for a subset A of the size classes used to construct the empirical $c.d.f$. If one has access to an estimate of the average establishment size, $\widehat{e}^j_{ic}$, for each of the size categories $j \in A$, then it is possible to obtain an estimate of total employment E_{ic}. Denote the number of establishments in each size category by m^j_{ic}. An estimate of E_{ic} is then given by

$$\widehat{E}_{ic} = \sum_{j \notin A} E^j_{ic} + \sum_{j \in A} \widehat{e}^j_{ic} m^j_{ic}$$

Using this estimated employment, one can proceed to calculate an estimate of the average establishment size in an MSA by summing the estimated employment levels across counties in the MSA and then dividing it by the total number of establishments in the MSA.

The following procedure is used to obtain the estimate $\widehat{e}^j_{ic}$ for a 2-digit industry i in a given county c :

1. If employment is disclosed for size class j for the state in which county c is located, then the average size in size class j for that state is used as the estimate $\widehat{e}^j_{ic}$, (using state level data whenever available is superior to using data at the national level, because (i) state-specific factors might influence employment and number of establishments in a given industry, (ii) the state level average establishment size contains more information for the average sizes in MSA's in that state),

2. If there is non-disclosure at the state level for industry i, then, whenever disclosed, state employment levels for the 3-digit components of industry i in size class j are used to estimate the average for class j for the 2-digit industry. The estimate $\widehat{e}^j_{ic}$ is calculated as the weighted average of "average size in class j" for all constituent 3-digit industries, where the weights are the employment shares for these industries in size class j,

   2a. If there is non-disclosure at the national level for a constituent 3-digit industry in size class j in performing step 2, then the national average size for class j at the 2-digit level is used for that industry, whenever disclosed,

   2b. Finally, if the 2-digit employment is suppressed for class j at the national level in performing step 2a, then the mid-point for that size class is used as the estimate of the average size for class j. For example, if the size class is 50 to 99 employees, then the estimate of average size is set to 75.

This 4-step procedure was also applied to the case of overall manufacturing, where, in each step, 2-digit industries were used instead of 3-digit ones, and overall manufacturing was used instead of 2-digit industries. In most cases, steps $1, 2$, and 2a were sufficient to obtain an estimate. Step 2b was used only for a few cases, almost all of them occurring in the Leather industry.

Table 1: Description of the variables used in the analysis

Variable	Description	Source
POPULATION	MSA population, 1990	Census
WAGE	Average first quarter manufacturing wage, 1990	CBP
RENT	Median gross rent in renter occupied housing units, 1990	CCDB
ELECTRICITY	State average cost of electricity, 1995	DOE
NATURAL GAS	State average cost of natural gas, 1995	DOE
INCOME	Median family income, 1990	CCDB
COLLEGE	Persons 25 years and over, percent with bachelor's degree or higher, 1990	CCDB
AGE25U	Percent of population under 25 years of age, 1990	CCDB
WHITE	Percent of population that is white, 1990	CCDB

Notes: Abbreviations stand for: CCDB (County and City Data Book), CBP (County Business Patterns) and DOE (Department of Energy)

Table 2: Descriptive statistics for the variables used in the analysis

Variable (Unit)	Mean	Standard Deviation	Correlation with Population
POPULATION (Persons)	598,405.9	1,007,078	1.00
WAGE (Dollars)	6,520.7	9,468.6	0.64
RENT (Dollars)	416.8	98.5	0.53
ELECTRICITY (Dollars/Million BTU)	20.4	5.4	0.11
NATURAL GAS (Dollars/Thousand cubic feet)	4.1	0.8	0.03
INCOME (Dollars)	34,378.2	6,099.0	0.49
COLLEGE (Percent)	0.19	0.06	0.13
AGE25U (Percent)	0.63	0.04	0.23
WHITE (Percent)	0.84	0.11	-0.30

Notes: Reported correlations are for logarithms in base 10, except for those variables that are in percentages.

Table 3: Patterns of average establishment size and employment share for manufacturing industries

Industry	Average Establishment Size			Share of Employment in Urban Counties (%)	Avg. Share of MSA Mfg. Employment (%)
	All Counties	Urban Counties	Rural Counties		
Overall Manufacturing	50.7	50.5	52.8	80.0	100.0
Food and Kindred	70.9	67.0	74.0	70.1	10.0 (9.2)
Textile Mill Products	105.4	77.2	171.8	54.2	2.8 (6.8)
Apparel and Misc. Textile	43.0	32.8	100.2	65.3	3.8 (5.5)
Lumber and Wood	20.3	**20.0**	**19.8**	46.2	3.9 (6.5)
Furniture and Fixtures	42.0	34.4	70.1	68.8	2.2 (3.4)
Paper and Allied	98.8	83.6	166.1	73.0	4.0 (5.5)
Printing and Publishing	24.5	**25.0**	**22.3**	86.8	8.7 (6.0)
Chemicals and Allied	70.2	65.3	81.4	83.0	5.4 (9.1)
Petroleum and Coal	50.1	**51.1**	**35.2**	85.1	1.4 (4.0)
Rubber and Misc. Plastics	57.6	51.5	83.2	76.5	4.6 (5.5)
Leather	57.9	45.7	80.2	67.4	0.6 (1.5)
Stone, Clay, and Glass	32.3	**33.8**	**29.1**	73.8	3.6 (4.1)
Primary Metal	105.3	100.0	117.2	80.6	5.2 (7.9)
Fabricated Metal Products	39.3	36.9	50.1	83.0	7.5 (5.4)
Industrial Machinery	37.1	35.5	43.3	81.4	10.2 (8.0)
Electronic Equipment	90.4	82.7	145.2	83.5	7.9 (8.9)
Transportation Equipment	166.6	**182.6**	**114.5**	85.3	9.1 (13.0)
Instruments	94.1	**96.0**	**85.9**	93.0	4.4 (6.6)

Notes: A bold indicates a reversal in the general pattern that average establishment size is lower for urban counties. Standard deviations are in parentheses.

Table 4: Bi-variate OLS regression results for overall manufacturing

	Dependent Variable:							
	log(Average Size)		log(Weighted Avg. Size)		log(No. of Estab.)		log(Employment)	
POPULATION	-0.052*	0.642	-0.059*	0.724	1.079*†	0.746*	1.027*	1.388*
	(-2.19)	(1.21)	(-2.46)	(1.48)	(57.34)	(2.06)	(31.96)	(1.97)
POPULATION^2	–	-0.061	–	-0.044	–	0.029	–	-0.031
		(-1.35)		(-1.01)		(0.93)		(-0.52)
R^2	0.02	0.02		0.04	0.89	0.89	0.74	0.74

Notes: Heteroskedasticity-consistent t-statistics in parantheses. A (*) denotes significance at 5% or less. A (†) indicates that the coefficient is different from 1 at 5%.

Table 5: Bi-variate OLS regression results for overall manufacturing: Establishment Size Distribution

	Dependent Variable: C.D.F. of Establishment Size											
Size Class:	0-19		20-49		50-99		100-249		250-499		500-999	
POPULATION	0.013	-0.112	0.027*	-0.019	0.028*	0.032	0.015*	-0.035	0.007*	-0.11	0.002*	-0.004
	(1.48)	(-0.55)	(3.96)	(-0.12)	(5.91)	(0.30)	(5.68)	(-0.53)	(5.34)	(-0.36)	(3.58)	(-0.31)
POPULATION^2	–	0.011	–	0.004		-0.0003		0.004		0.001		0.0006
		(0.63)		(0.31)		(-0.03)		(0.78)		(0.61)		(0.48)
R^2	0.005	0.007	0.040	0.040	0.085	0.085	0.080	0.082	0.070	0.071	0.028	0.029

Notes: Heteroskedasticity-consistent t-statistics in parantheses. A (*) denotes significance at 5% or less.

Table 6: OLS regression results for overall manufacturing using covariates

	Dependent Variable:									
	log(Average Size)				log(No. of Establishments)			log(Employment)		
	I	**II**	**III**	**IV**	**I**	**II**	**IV**	**I**	**II**	**IV**
POPULATION	-0.034*	0.843	-0.041*	-0.039*	1.073*†	0.553	1.068*†	1.039*	1.397*	1.029*
	(-2.01)	(1.47)	(-2.16)	(-2.11)	(50.29)	(1.62)	(51.50)	(31.24)	(2.13)	(33.16)
POPULATION^2	–	-0.077	–	–	–	0.045	–		-0.032	–
		(-1.63)				(1.55)			(-0.55)	
WAGE	0.035	0.035	0.041	0.031	0.052*	0.052*	0.048*	0.087*	0.087*	0.079*
	(1.48)	(1.51)	(1.56)	(1.52)	(2.76)	(2.75)	(2.69)	(2.60)	(2.71)	(2.65)
RENT	-0.877*	-0.925*	-0.966*	-0.912	-0.028	-0.001	-0.029	-0.905*	-0.925*	-0.941*
	(-3.53)	(-3.72)	(-3.81)	(-3.48)	(-0.16)	(0.01)	(-0.25)	(-3.12)	(-3.19)	(-3.23)
INCOME	0.826*	0.862*	0.811*	0.818*	-0.074	-0.096	-0.077	0.751*	0.766*	0.741*
	(2.77)	(2.93)	(2.46)	(2.66)	(-0.30)	(-0.39)	(-0.35)	(2.11)	(2.15)	(2.18)
COLLEGE	-0.349**	-0.323	-0.301	-0.323**	0.188	0.173	0.183	-0.161	-0.150	-0.140
	(-1.72)	(-1.58)	(-1.44)	(-1.67)	(1.07)	(0.97)	(1.02)	(-0.62)	(-0.58)	(-0.53)
WHITE	-0.227**	-0.238*	-0.216**	-0.238**	0.139	0.145**	0.141**	-0.088	-0.093	-0.097
	(-1.86)	(-1.92)	(-1.68)	(-1.88)	(1.60)	(1.69)	(1.65)	(-0.62)	(-0.65)	(-0.77)
AGE25U	-0.116	-0.125	-0.111	-0.118	0.644*	0.649*	0.634*	0.528	0.524	0.516
	(-0.34)	(-0.37)	(-0.25)	(-0.31)	(2.38)	(2.40)	(2.41)	(1.24)	(1.22)	(1.34)
NATURAL GAS	0.567*	0.558*	0.542*	0.555*	0.389*	0.394*	0.378*	0.956*	0.952*	0.933*
	(3.93)	(3.86)	(3.78)	(3.91)	(3.28)	(3.33)	(3.88)	(5.06)	(5.00)	(5.25)
ELECTRICITY	-0.248**	-0.238**	-0.254**	-0.261**	-0.157	-0.163	-0.161	-0.406*	-0.401*	-0.422*
	(-1.85)	(-1.75)	(-1.72)	(-1.82)	(-1.55)	(-1.60)	(-1.59)	(-2.37)	(-2.32)	(-2.42)
Industry Dummies	–	–	–	YES	–	–	YES	–	–	YES
R^2	0.16	0.17	0.22	0.31	0.91	0.91	0.94	0.80	0.81	0.88

Notes: Heteroskedasticity-consistent t-statistics in parantheses. (*) and (**) denote significance at 5% and 10%, respectively. (†) indicates that the coefficient is significantly different from 1 at 5%. See text for the details of specifications I-IV.

Table 7: OLS regression results for overall manufacturing with covariates: Establishment Size Distribution

	Dependent Variable: C.D.F. of Establishment Size					
Size Class:	0-19	20-49	50-99	100-249	250-499	500-999
POPULATION	0.007	0.024*	0.027*	0.014*	0.006*	0.002*
	(0.73)	(3.17)	(5.01)	(4.72)	(4.30)	(3.33)
WAGE	-0.023*	-0.020*	-0.013*	-0.005**	-0.0008	-0.0001
	(-2.83)	(-3.22)	(-3.17)	(-1.86)	(-0.61)	(0.00)
RENT	0.281*	0.192*	0.123*	0.076*	0.041*	0.023*
	(3.29)	(2.91)	(2.46)	(2.92)	(2.79)	(3.47)
INCOME	-0.227*	-0.125	-0.051	-0.023	-0.033**	-0.025*
	(-2.19)	(-1.56)	(-0.84)	(-0.70)	(-1.83)	(-2.95)
COLLEGE	0.349*	0.244*	0.130*	0.036	0.008	-0.005
	(4.74)	(4.79)	(3.48)	(1.57)	(0.61)	(-0.79)
WHITE	0.100*	0.089*	0.068*	0.037*	0.016**	0.007**
	(2.34)	(2.54)	(2.56)	(2.13)	(1.65)	(1.82)
AGE25U	0.084	0.072	0.022	-0.016	-0.004	0.002
	(0.72)	(0.85)	(0.35)	(-0.47)	(-0.18)	(0.21)
NATURAL GAS	-0.293*	-0.233*	-0.158*	-0.073*	-0.029*	-0.008**
	(-6.29)	(-5.91)	(-5.29)	(-4.14)	(-2.97)	(-1.65)
ELECTRICITY	0.057	0.069*	0.049*	0.024**	0.012**	0.003
	(1.31)	(1.91)	(1.88)	(1.60)	(1.70)	(0.81)
R^2	0.29	0.30	0.29	0.21	0.14	0.09

Notes: Heteroskedasticity-consistent t-statistics in parantheses. (*) and (**) denote significance at 5% and 10%, respectively.

Table 8: Summary of all regression results for 2-digit industries

Industry	Size Distribution		Average Size			No. of Estab.			Employment			Transport. Cost to Value Ratio (%)
			Population		R^2	Population		R^2	Population		R^2	
	Ind.	Joint	OLS	LAD	(OLS)	OLS	LAD	(OLS)	OLS	LAD	(OLS)	
Food and Kindred	+	Y	0.079	0.066	0.09	0.902*†	0.901*†	0.65	0.981*	0.968*	0.36	2.72
			(1.54)	(1.35)		(33.63)	(19.94)		(18.23)	(16.44)		
Textile Mill	+	Y	0.076	0.121	0.08	0.838*†	0.881*†	0.49	0.915*	0.964*	0.31	3.87
			(0.86)	(0.80)		(12.30)	(16.77)		(7.05)	(5.65)		
Apparel and Textile	0	N	0.222*	0.123	0.14	1.207*†	1.143*†	0.73	1.429*†	1.377*†	0.62	0.78
			(3.53)	(1.71)		(25.36)	(18.35)		(16.54)	(13.85)		
Lumber and Wood	0	N	0.098*	0.072*	0.10	0.859*†	0.816*†	0.61	0.958*	0.785*†	0.54	2.64
			(2.76)	(2.03)		(22.79)	(14.12)		(15.00)	(11.42)		
Furniture and Fixtures	+	Y	0.161*	0.163*	0.12	1.127*†	1.137*†	0.75	1.288*†	1.264*†	0.44	0.75
			(2.72)	(2.18)		(34.12)	(27.29)		(17.51)	(15.63)		
Paper and Allied	+	Y	-0.075	-0.017	0.06	1.006*	1.083*	0.66	0.931*	0.912*	0.40	2.60
			(-1.43)	(-0.36)		(22.10)	(17.02)		(11.51)	(6.82)		
Printing and Publishing	0	Y	0.102*	0.170*	0.12	1.063*†	1.101*†	0.95	1.165*†	1.181*†	0.58	4.87
			(2.21)	(2.24)		(61.21)	(36.74)		(44.03)	(30.20)		
Chemicals and Allied	0	Y	0.202*	0.174*	0.11	1.159*†	1.130*†	0.77	1.362*†	1.334*†	0.31	2.32
			(2.70)	(2.35)		(33.00)	(22.01)		(18.00)	(12.48)		
Petroleum and Coal	0	N	0.043	0.084	0.09	0.883*†	0.985*	0.66	0.926*	0.912*	0.24	5.81
			(1.25)	(1.46)		(19.95)	(9.33)		(7.98)	(6.46)		
Rubber and Plastics	+	Y	0.083	0.037	0.07	1.169*†	1.188*†	0.78	1.252*†	1.189*	0.45	9.16
			(1.49)	(1.01)		(32.95)	(25.09)		(16.98)	(14.56)		
Leather	+	Y	-0.093*	-0.105*	0.12	1.394*†	1.332*†	0.53	1.301*†	1.244*†	0.29	0.81
			(-2.80)	(-2.42)		(10.83)	(7.38)		(8.88)	(7.46)		
Stone, Clay, and Glass	–	Y	0.125*	0.152*	0.09	0.938*†	0.942*†	0.85	1.064*	1.122*	0.48	4.14
			(3.81)	(3.28)		(41.00)	(30.47)		(23.74)	(18.17)		
Primary Metal	+	Y	0.045	-0.058	0.07	0.951*	1.017*	0.67	0.997*	0.945*	0.30	9.34
			(0.67)	(-0.77)		(21.46)	(16.42)		(10.74)	(6.16)		
Fabricated Metal	+	Y	0.012	-0.035	0.06	1.149*†	1.143*†	0.80	1.161*†	1.162*†	0.44	0.27
			(0.32)	(-0.95)		(32.27)	(24.49)		(18.97)	(16.15)		
Industrial Machinery	+	Y	0.008	0.018	0.05	1.107*†	1.119*†	0.77	1.115*	1.070*	0.32	1.41
			(0.20)	(0.41)		(29.72)	(24.55)		(18.40)	(15.19)		
Electronic Equip.	+	Y	0.029	-0.064*	0.04	1.233*†	1.239*†	0.81	1.263*†	1.128*†	0.39	2.90
			(0.50)	(-1.68)		(35.35)	(36.67)		(16.07)	(12.00)		
Transportation Equip.	0	Y	0.378*	0.351*	0.13	1.015*	1.020*	0.73	1.393*†	1.330*†	0.38	2.50
			(2.31)	(2.24)		(23.40)	(14.87)		(16.38)	(9.60)		
Instruments	–	Y	0.106*	0.117	0.12	1.208*†	1.222*†	0.82	1.314*†	1.327*†	0.41	0.59
			(2.02)	(1.52)		(34.91)	(22.62)		(18.84)	(12.75)		

Notes: A + (–) indicates at least one coefficient is positive (negative) and significant at 5%, and no coefficient is negative (positive) and significant at 5%. A (0) indicates all coefficients are insignificant individually or signs are mixed. 'Y' means coefficients are jointly significant at 5% with a Wald test. 'N' denotes otherwise. A (*) indicates significance at 5%. A (†) indicates that the coefficient is significantly different from 1 at 5%.

Figure 1: Share of MSA employment in manufacturing industries versus MSA population in U.S., 1990

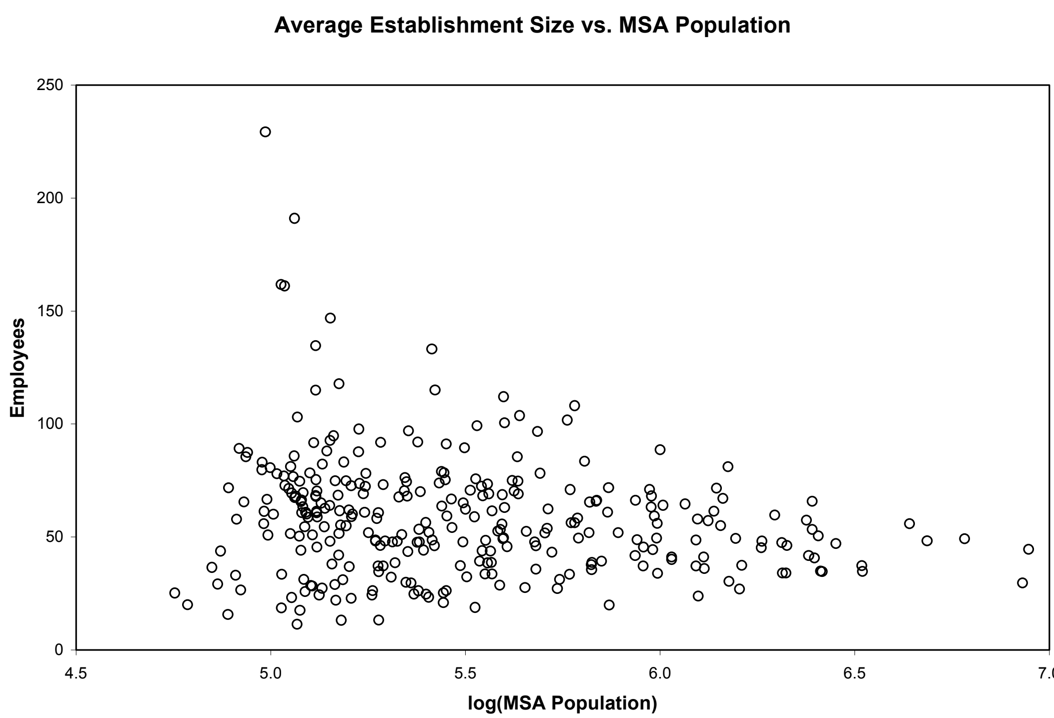

Figure 2: Average manufacturing establishment size versus MSA population in U.S., 1990.

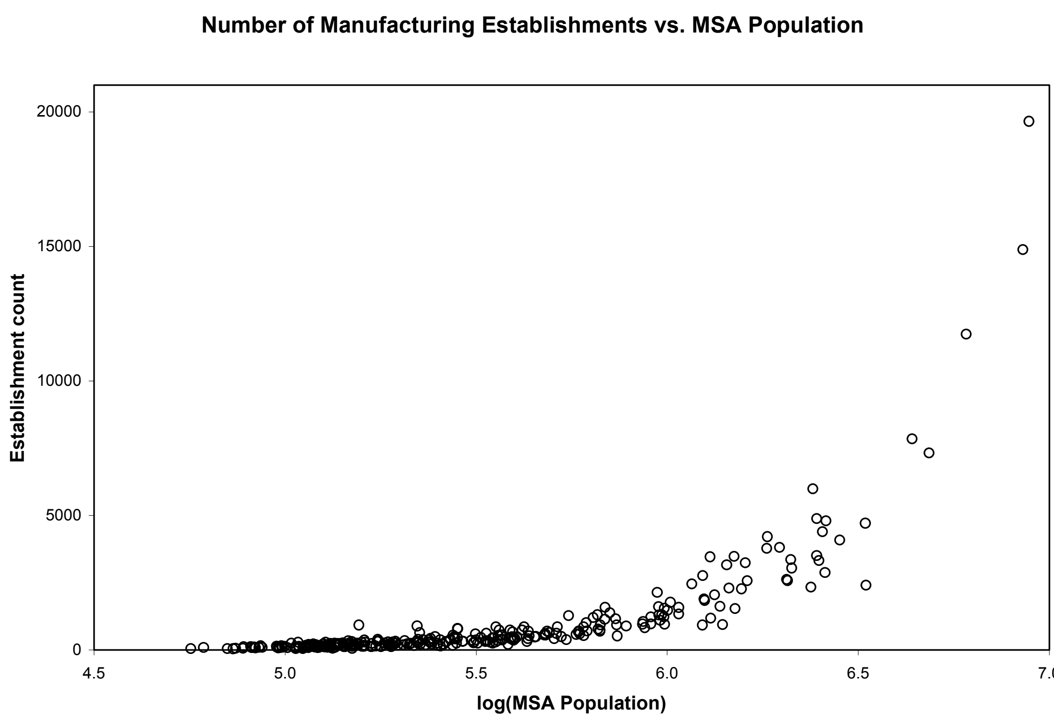

Figure 3: Number of manufacturing establishments versus MSA population in U.S., 1990

Figure 4: Manufacturing employment versus MSA population in U.S., 1990

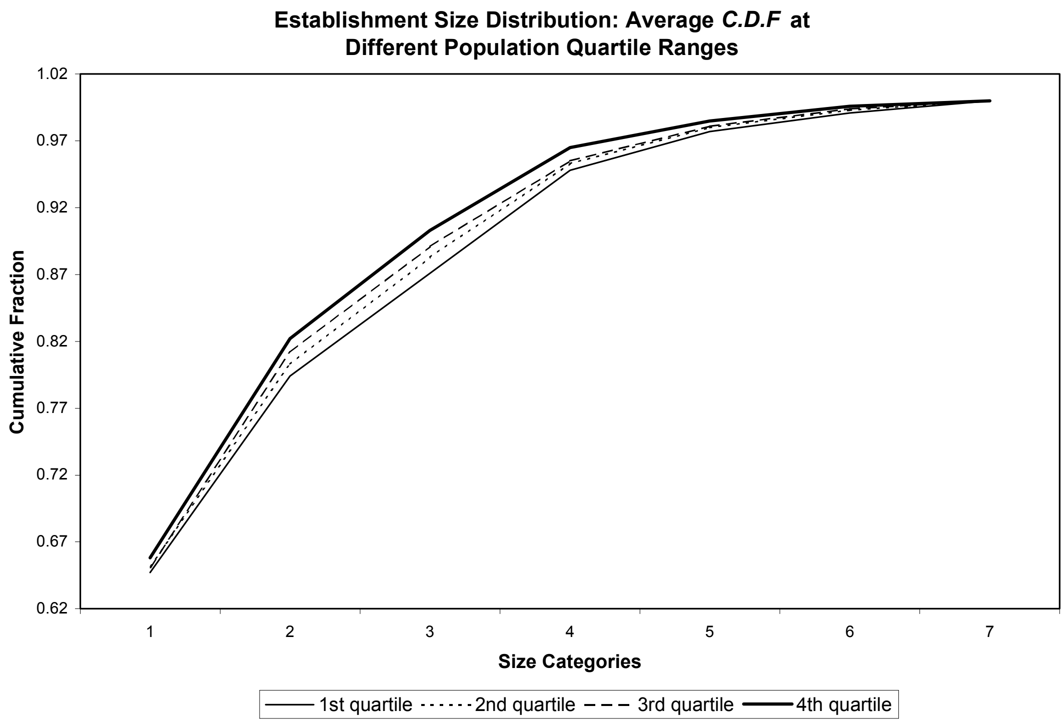

Figure 5: Average cumulative distribution function ($c.d.f$) of manufacturing establishment size within population quartile ranges, 1990

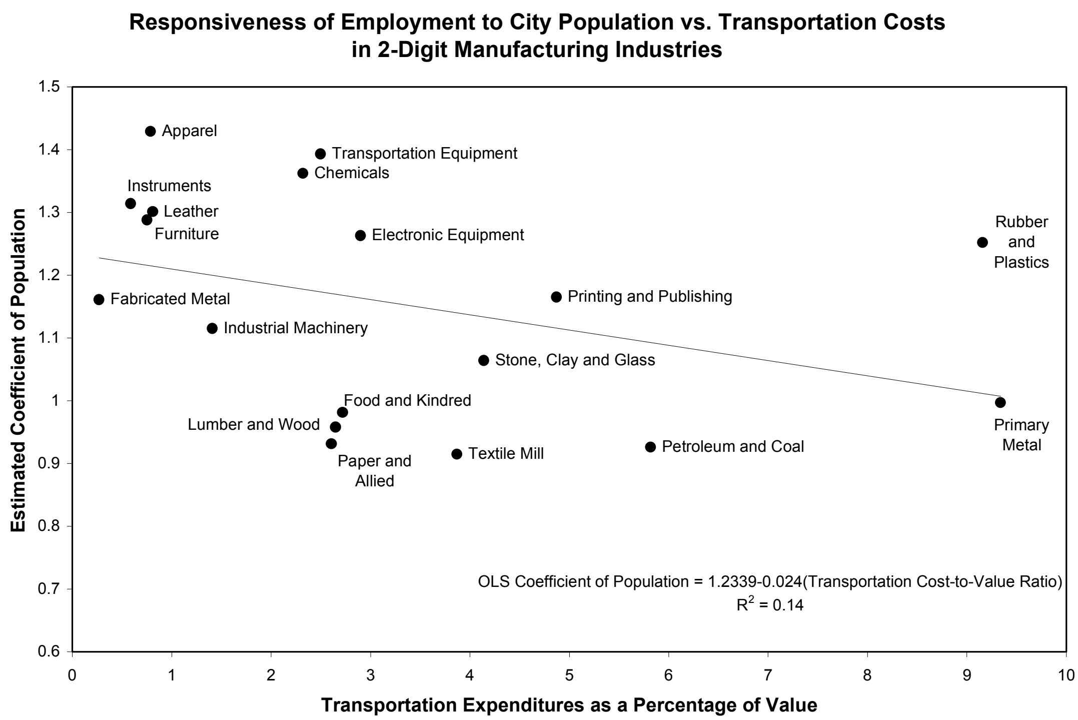

Figure 6: The relationship between responsiveness of employment to city population and transportation costs across 2-digit manufacturing industries

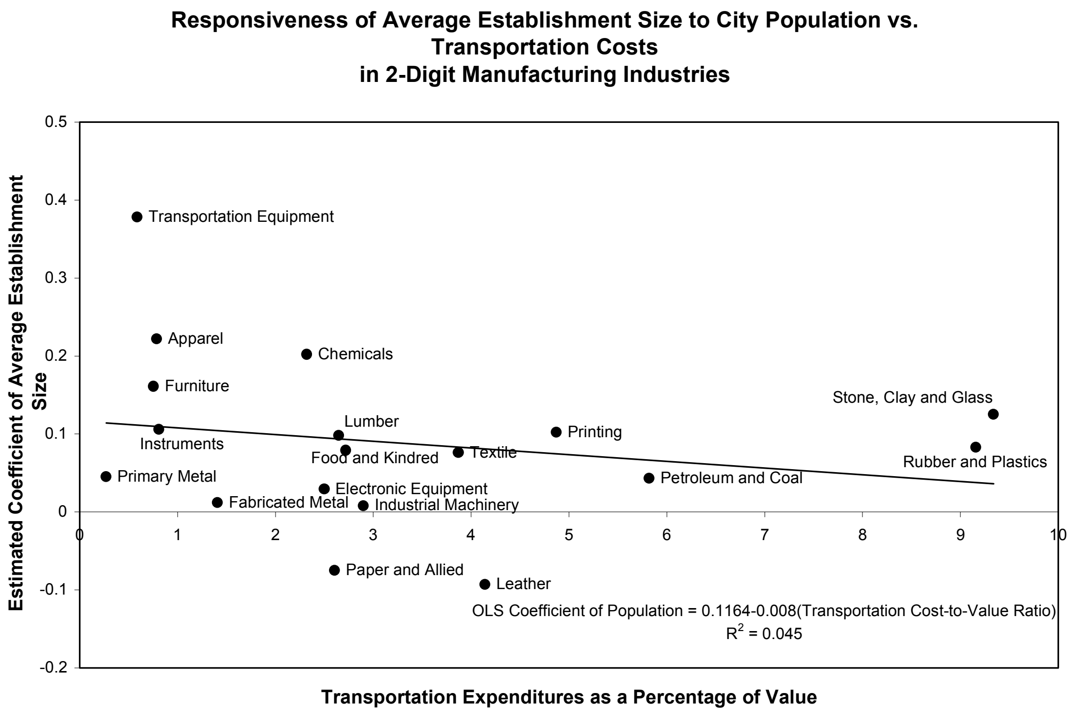

Figure 7: The relationship between responsiveness of average establishment size to city population and transportation costs across 2-digit manufacturing industries