Productivity and the Geographic Concentration of Industry: The Role of Plant Scale

Authors	Christopher H. Wheeler
Working Paper Number	2004-024A
Creation Date	November 2003
Citable Link	https://doi.org/10.20955/wp.2004.024
Suggested Citation	Wheeler, C.H., 2003; Productivity and the Geographic Concentration of Industry: The Role of Plant Scale, Federal Reserve Bank of St. Louis Working Paper 2004-024. URL https://doi.org/10.20955/wp.2004.024

Published In	Regional Science and Urban Economics
Publisher Link	https://doi.org/10.1016/j.regsciurbeco.2005.10.004

Productivity and the Geographic Concentration of Industry: The Role of Plant Scale

Christopher H. Wheeler*
Research Division
Federal Reserve Bank of St. Louis
411 Locust Street
St. Louis, MO 63102
Christopher.H.Wheeler@stls.frb.org

September 27, 2004

Abstract

A large body of research has established a positive connection between an industry's productivity and the magnitude of its presence within locally defined geographic areas. This paper examines the extent to which this relationship can be explained by a micro-level underpinning commonly associated with productivity: establishment scale. Looking at data on two-digit manufacturing across a sample of U.S. metropolitan areas, I find two primary results. First, average plant size – defined in terms of numbers of workers – increases substantially as an industry's employment in a metropolitan area rises. Second, results from a decomposition of localization effects on labor earnings into plant-size and plant-count components reveal that the widely observed, positive association between a worker's wage and the total employment in his or her own metropolitan area-industry derives predominantly from the former, not the latter. Localization economies, therefore, appear to be the product of plant-level organization rather than pure population effects.

JEL Classification: J31, R12, R23
Keywords: Localization Economies, Establishment Size, Plant Size Wage Premium

*The views expressed herein are those of the author and do not represent the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

Productivity gains associated with the geographic concentration of industry are a long standing result in the urban economics literature. Yet, despite the presence of a substantial body of work documenting these 'localization economies' (e.g. Carlino (1979), Nakamura (1985), and Henderson (1986)), our understanding of their nature and causes remains somewhat limited.

There are, of course, a number of possible explanations which tend to fall into one of two basic categories: (*i*) productivity shifts that are external to firms, and (*ii*) efficiency gains tied directly to a plant's scale of production (i.e. internal economies of scale or increasing returns). Foremost among the theories belonging to the first group are Marshall's (1920) now famous three, which suggest that producers within the same industry agglomerate to take advantage of the spillover of industry-specific knowledge, the presence of a more extensive array of input providers, and/or economies of labor market search which facilitate the firm-worker matching process. With each of these mechanisms, an individual producer's efficiency is an increasing function of the (geographically proximate) extent of its industry. Over the past several decades, these three particular explanations have drawn an abundance of theoretical analysis (e.g. Henderson (1974), Abdel Rahman and Fujita (1990), Ciccone and Hall (1996), Black and Henderson (1999) to name just a few) and, more recently, have begun to receive some interesting empirical scrutiny (e.g. Dumais et al. (1997), Rosenthal and Strange (2001)).

In contrast, the second category of localization theories – internal productivity effects – has not attracted the same volume of attention (at least, as far as I am aware), possibly because the idea is so straightforward.[1] According to this line of reasoning, industrial concentration is merely the product of a firm's unwillingness (or inability) to produce in

[1] Dixit (1973) and Krugman (1991) are prominent examples of this approach.

many different locations simultaneously. This would happen, for example, in the presence of firm-level increasing returns to scale at a specific production site or the existence of fixed setup costs that must be incurred before producing at a particular location. Given variation in demand for a producer's output, differences in an industry's employment across markets may be tied to plant scale: some markets are populated by producers who operate on a large (productive) scale to meet a high demand, others are inhabited by firms producing on a smaller (less productive) scale to satisfy a lower demand. Productivity gains associated with the geographic concentration of industry, in this case, follow rather simply from plant-level productivity effects.

Although this second group of explanations has not received as much consideration as the first, a fair amount of evidence suggests that plant scale may actually be an important underpinning of localization economies. To begin, industries that exhibit greater spatial concentration in the U.S. also tend to be characterized by relatively large production units. Looking at data on U.S. manufacturing, for example, Kim (1995) and Holmes and Stevens (2002) find strong positive associations between localization – quantified by indexes that capture the degree to which an industry is over- or under-represented in an area relative to its national average – and the average number of employees per plant.[2]

What is more, large plants tend to be more productive (e.g. Idson and Oi (1999)) and pay higher wages (e.g. Brown and Medoff (1989), Troske (1999), Oi and Idson (1999)) than small ones, even after conditioning on a variety of observable producer and worker characteristics (e.g. capital intensity, education, experience). Based on U.S. manufacturing, for instance, Troske (1999) finds that wage earnings increase by approximately 3 to 4 percent as an establishment's total employment doubles. While the precise reasons for these producer-size productivity effects remain somewhat elusive, the regularity itself is strikingly robust.

[2]Holmes and Stevens (2002) note that, although this relationship is particularly strong for manufacturing, it also holds for a wide array of other industries.

Taken together, these two results suggest that the well-known positive association between productivity and the geographic concentration of industry may have a straightforward micro-level explanation: larger establishment size. This paper explores this conjecture.

To be sure, this is not the first paper to do so. Carlino (1979) and Henderson (1986), for example, both investigate the connection between productivity (defined in terms of returns to scale by the former, output per unit labor input by the latter) and average establishment size using data from the U.S. Census of Manufactures (CM). Their findings, unfortunately, are somewhat mixed: Carlino's evidence suggests some positive influence of plant size on city-industry productivity; Henderson's indicates that the association between the two is weak.

This paper re-visits this issue but takes a very different approach. To begin, neither Carlino (1979) nor Henderson (1986) explicitly investigates how plant size varies with city-industry employment. Doing so is one of the central aims of this paper. Moreover, to study the link between productivity and localization, I focus on individual-level wage earnings instead of measures of productivity derived from aggregate city-industry data such as that reported in the CM. There are at least two advantages to doing so. First, as noted by Wheaton and Lewis (2002), wage earnings are likely to involve less measurement error directly tied to agglomeration than city-level estimates of industry capital and output.[3] Second, unlike aggregate city-level data, individual-level observations allow me to control for the effects of numerous person-specific characteristics that likely influence a worker's efficiency. Inferences drawn from wages about the productivity effects of industrial concentration, therefore, should involve less bias than those derived from city-level industry aggregates.[4]

[3] Ciccone and Hall (1996) also express some skepticism about the usefulness of using CM data to study agglomeration effects on productivity.

[4] Perhaps for these reasons, wages have become a common object of analysis in studies of local market productivity (e.g. Rauch (1993), Glaeser and Mare (2001), Wheaton and Lewis (2002), Moretti (2003),

The results, which are based upon data from two-digit manufacturing in a sample of U.S. metropolitan areas over the period 1980-1990, indicate the following. First, an industry's total employment in a metropolitan area is strongly associated with the average size of its plants in that market. Defining average plant size as the ratio of workers to plants, the point estimates suggest an elasticity of approximately 0.65: that is, a 10 percent increase in city-industry employment is accompanied by a 6.5 percent increase in average employment per plant. Defining it in weighted terms (i.e. the average number of co-workers per employee), the elasticity is even higher: 0.93. These figures turn out to be remarkably consistent across all 20 two-digit industries and are highly robust to the inclusion of controls for a variety of industry, city, and time effects.

Second, estimates from standard hedonic wage regressions indicate that the positive association between a worker's wage and the total employment in his or her own city-industry is indeed highly significant. The findings imply an average elasticity of roughly 4 percent, which is similar to what previous research has documented (e.g. Henderson (1986)).[5] Given that total city-industry employment is merely the product of average establishment size and the total number of establishments, this localization effect can be decomposed into two terms: one tied to plant scale, the other plant counts. The results indicate that, overwhelmingly, the positive association between wages and city-industry employment operates through the former, not the latter. Interpreted literally, an increase in a city-industry's total employment stemming from an increase in the average size of a fixed number of plants is associated with significantly higher wage earnings. An increase in a city-industry's employment due to an increase in the number of plants of a fixed size, by contrast, generates little effect on wages.

---

among many others).

[5]Interestingly, this figure is also very close to the estimated employer-size wage elasticities estimated by Brown and Medoff (1989) and Troske (1999).

Localization economies, therefore, do not appear to be pure population effects – that is, simply the product of more activity. Instead, they seem to be a function of how workers are organized into production units.

The remainder of the paper proceeds as follows. The next section provides a brief description of the data as well as a discussion of some measurement issues. Section 3 reports the results. Section 4 concludes.

2 Data and Measurement

The data used in the analysis below are drawn primarily from three sources. First, individual-level observations on the wage earnings of manufacturing workers are derived from two Census files: the 1980 and 1990 1 Percent Metro Samples of the Integrated Public Use Microdata Series (IPUMS).[6] In an effort to produce a sample of workers with a reasonably strong attachment to the labor force, I limit the analysis to individuals between the ages of 18 and 65, who report having usually worked at least 30 hours per week, and who were not in school at the time the Census was taken. I further limit the sample to workers who earned between 2 and 60 dollars per hour (in 1982 dollars) to eliminate the effects of outlier observations. After discarding all individuals for which either the metropolitan area of residence or any of the basic covariates used in the analysis were not reported (see below), I arrived at a sample consisting of 265403 observations across the two years. Additional details about these data appear in the Appendix.

Second, the USA Counties 1998 data file (U.S. Bureau of the Census (1999)) provides a variety of basic economic and demographic information over the years 1980 and 1990 for each county (and county-equivalent unit) in the country. From these data, I create city-level observations (for certain selected quantities listed below) by aggregating county-level

[6] See Ruggles and Sobek et al. (2003) at http://www.ipums.org.

observations into metropolitan statistical areas (MSAs) and either consolidated metropolitan statistical areas (CMSAs) or New England County Metropolitan Areas (NECMAs) if an MSA belongs to a CMSA or NECMA.[7] While CMSAs and NECMAs may seem rather large when considering local labor markets, using them greatly facilitates the creation of geographic areas with consistent definitions over time.[8] A total of 275 such metropolitan areas exist. Of these, the Census samples produce individual-level observations in both years for 200.[9]

Third, data used to calculate average establishment size and total employment among two-digit manufacturing industries across the sample of metropolitan areas are taken from the 1980 and 1990 County Business Patterns (CBP) files (U.S. Bureau of the Census (1982, 1992)). I consider two different measures of average establishment size: a 'simple' average and an 'employment-share weighted' average. The simple average is just the number of workers per establishment. That is, for industry i in city c,

$$\text{Simple Average}(i,c) = \frac{\text{Emp}(i,c)}{\text{Est}(i,c)} \tag{1}$$

where $\text{Emp}(i,c)$ and $\text{Est}(i,c)$ represent, respectively, employment and the total number of establishments in this city-industry.

Because this measure may not adequately capture the extent to which workers are

[7] Aggregation is based upon 1995 definitions. For expositional purposes, I use the terms 'city' and 'metropolitan area' interchangeably throughout the paper.

[8] In the IPUMS data, there are several instances in which the county-level composition of MSAs that belong to CMSAs changes between 1980 and 1990 (the county-level composition of each metropolitan area can be found in the IPUMS documentation). For example, some of the individuals assigned to the Dallas, TX MSA in 1980 would be assigned to the Ft. Worth-Arlington, TX MSA in 1990. Combining these two MSAs into the Dallas-Fort Worth CMSA mitigates this problem.

[9] In the 1980 data, the minimum, maximum, and mean number of observations per city are 24, 14149, and 721.5. In the 1990 data, they are 15, 11209, and 608.5. For the pooled sample: 49, 24516, and 1327.

concentrated in large plants (see Kumar et al. (1999)), I also calculate an employment-share weighted average which approximates the average number of co-workers that a typical worker has. By categorizing producers as belonging to one of K size categories, this measure of average establishment size follows as

$$\text{Employment-Share Weighted Average}(i,c) = \sum_{k=1}^{K} \frac{\text{Emp}(k,i,c)}{\text{Emp}(i,c)} \frac{\text{Emp}(k,i,c)}{\text{Est}(k,i,c)} \tag{2}$$

where $\text{Emp}(k,i,c)$ and $\text{Est}(k,i,c)$ are the number of employees and establishments, respectively, in establishment size category k for this city-industry.

The CBP data readily allow for the calculation of the simple average since total numbers of manufacturing establishments and workers are usually both reported. Where the total employment figures are reported as a range (due to disclosure restrictions), I estimate by taking midpoints.[10]

Constructing the weighted average, by contrast, is somewhat more difficult because, although counts of establishments falling into each of 12 size classes[11] are reported, total employment by size category is not. Therefore, I estimate the employment-share weighted average using the following procedure.

First, to estimate the average establishment size within each size class, $\frac{\text{Emp}(k,i,c)}{\text{Est}(k,i,c)}$, I use a simple method-of-moments procedure assuming that the distribution of establishment sizes is lognormal. Details regarding this procedure appear in the Appendix. Second, I estimate $\frac{\text{Emp}(k,i,c)}{\text{Emp}(i,c)}$ by multiplying each of these estimated means, $\frac{\text{Emp}(k,i,c)}{\text{Est}(k,i,c)}$, by the corresponding number of establishments to gain an estimate of $\text{Emp}(k,i,c)$. I then sum the estimated

[10] There are 12 employment ranges reported by the CBP: 0-19, 20-99, 100-249, 250-499, 500-999, 1000-2499, 2500-4999, 5000-9999, 10000-24999, 25000-49999, 50000-99999, and 100000 or more. The largest two categories did not appear for any of the county-industries considered here for either year.

[11] Establishment counts are given for the following 12 categories: 1-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000-1499, 1500-2499, 2500-4999, 5000 or more employees.

values of $\mathrm{Emp}(k, i, c)$ across the 12 size categories to gain an estimate of total city-industry employment, $\mathrm{Emp}(i, c)$, which permits for an estimate of $\frac{\mathrm{Emp}(k,i,c)}{\mathrm{Emp}(i,c)}$ to be constructed for each size category.

Summary statistics for many of the key variables used in the analysis below appear in Tables 1A and 1B. From them, a number of well-known trends can be seen. Notably, between 1980 and 1990, educational attainment increased – rising from 11.9 to 12.6 years of schooling for an average worker – while both own-industry manufacturing employment and average plant size (for a typical manufacturing employee in the sample) decreased, dropping from approximately 47000 own-industry workers to fewer than 38000; 172.6 workers per plant to 109.6 (1956.4 co-workers to 1283.9).[12]

3 Results

3.1 Localization and Plant Size

Two of the papers cited in the Introduction (Kim (1995) and Holmes and Stevens (2002)) have already established that localization and plant size are strongly associated in U.S. manufacturing. Yet, this conclusion is based upon the calculation of localization indexes which summarize the extent to which industries are disproportionately represented in total employment (relative to the national level) across a collection of local markets.

The approach taken here is somewhat different. In particular, since studies estimating localization effects commonly do so by correlating some measure of an industry's productivity with its overall scale (e.g. total employment) within a locally defined area,[13] I consider an analogous exercise by estimating how an industry's plant scale varies with its aggregate

[12]These last two observations are consistent with the drop off in both overall manufacturing employment and the average size of manufacturing plants described by Davis and Haltiwanger (1991) and Davis et al. (1996).

[13]See Eberts and McMillen (1999) for a summary of empirical work.

local market employment. Doing so should help to reveal how the plant-level organization of production varies as an industry's overall size changes, and, thus, may offer further insight into why productivity scales positively with employment.

To this end, I use the two measures of average establishment size for industry i, city c in year t, AES_{ict}, given by (1) and (2) to estimate the following:[14]

$$\log\left(\text{AES}_{ict}\right) = \beta \log\left(\text{Emp}_{ict}\right) + \gamma' \mathbf{z}_{ct} + \alpha_i + \delta_t + \mu_c + \epsilon_{ict} \tag{3}$$

where Emp_{ict} is the city-industry's total employment in year t; $\mathbf{z}_{ct}$ is a vector of city-time varying covariates (described below) that may influence plant scale; α_i, δ_t, and μ_c represent industry, time, and city-specific fixed effects influencing average establishment size; and ϵ_{ict} is a residual. Estimates from two different specifications of (3) appear in Table 2.

Consider, first, the results from the baseline version, labeled I, which drops the vector of city-level variables, $\mathbf{z}_{ct}$, in an effort to focus on the plant size-industry employment relationship. The estimated coefficients clearly demonstrate that both measures of average plant size are strongly tied to overall city-industry employment. Each elasticity (0.65 for the simple average, 0.93 for the weighted average) is highly significant and suggestive of a reasonably large association. Using the mean values of the employment and plant size series (in logarithms), for example, they suggest that an employment increase of 52 workers within a metropolitan area is accompanied by increases of roughly 2.5 workers per plant and 10 co-workers per employee, on average.

Such figures are actually quite robust. To see this, consider, next, the results from the second specification (labeled II) in which the vector of city-time varying covariates is added back into the equation to account for the influence of various 'environmental'

[14] These figures are based upon all city-industry-year observations that could be identified from the CBP data, not just the 200 metropolitan areas covered by the IPUMS data.

features on plant size. In particular, this second specification includes the following eight characteristics: log population, log population density, log per capita income, the fraction of the adult population with a college degree, the proportions of the population under the age of 18 and over the age of 64, the fraction of population that is non-white, and the unemployment rate.[15] The first three are intended to capture the costs associated with overall urban scale, which Glaeser and Kahn (2001) and Dinlersoz (2004) have found to be important determinants of both the location and scale of manufacturing in U.S. cities. The education and demographic characteristics provide some basic information about the nature of the local labor force, including the local supply of human capital which the literature on firm size has long stressed as a key determinant of production scale (e.g. Lucas (1978) and Kremer (1993)). The unemployment rate is added to pick up any effects of the local business cycle (e.g. high unemployment may be associated with a lower average establishment size as plants lay off workers).

What the results show, however, is a general lack of significant coefficients for these city-time varying covariates, at least after having conditioned on all of the variables appearing in specification I. Only the logarithm of population and the fraction of residents over the age of 64 in the equation for the simple average enter significantly (and as one might expect, negatively).[16]

More importantly, the inclusion of these regressors does not change the estimated associations between average plant size and log industry employment. For both measures of establishment scale, the estimated localization parameters are identical across the two specifications.

While a strong producer size-industry employment connection emerges from the analysis

[15] Each of the quantities is derived from the USA Counties data file.

[16] Dinlersoz (2004) finds that plant size is negatively associated with city-level population in U.S. manufacturing. A larger fraction of the population that is beyond its prime working years (18 to 64) should reduce the size of labor pool (holding population constant) from which establishments can hire.

of single, summary plant size measures (i.e. averages), it should be noted that it also emerges from a more detailed examination of city-industry plant size distributions. To see this, let $F_{ict}(n)$ denote the empirical cumulative distribution function for plants of industry i in city c at time t, evaluated at employment level n (i.e. the fraction of establishments with n or fewer employees).[17] As in Dinlersoz (2004), I consider six values of n (19, 49, 99, 249, 499, and 999) and estimate the following analog to (3):

$$F_{ict}(n) = \beta \log\ (\text{Emp}_{ict}) + \gamma' \mathbf{z}_{ct} + \alpha_i + \delta_t + \mu_c + \epsilon_{ict} \tag{4}$$

where the regressors are the same as those described above.

Results are reported in Table 3. Because the estimated coefficients on log industry employment were essentially invariant with respect to including or dropping the vector of city-level variables, I have only reported results from the specification in which $\mathbf{z}_{ct}$ is added. On the whole, they indicate that, for each of the six employment levels considered, the empirical cumulative distribution function decreases significantly as industry employment rises. That is, significantly more probability mass falls on large establishments as total city-industry employment increases. Therefore, not only do average measures of plant scale increase with localization; the evidence also indicates that establishment size is stochastically increasing with employment.[18]

Similar results arise when (3) and (4) are estimated separately for each two-digit industry.[19] Those estimates, which appear in Table 4, generally reinforce the conclusions drawn

[17] In essence, one can interpret this quantity as representing the proportion of a city-industry's establishments accounted for by relativley small plants.

[18] That is, the plant-size cumulative distribution function shifts everywhere to the right as total industry employment increases.

[19] Doing so controls for city-industry fixed effects that are not considered in the pooled results described above. Given the rather broad industrial categorization used here (two digit), the types of establishments

above from the pooled sample. Both measures of average establishment size are significantly and positively associated with industry employment for each of the 20 industries, with the majority of the industry-specific elasticities falling relatively close to the figures documented in Table 2. Additionally, the point estimates from the empirical distribution function regressions suggest that, with the exception of Tobacco Products (SIC 21), establishment size stochastically increases with industry employment.

3.2 Decomposing Localization Effects: Plant Size vs. Plant Counts

Given that the average scale at which producers operate increases significantly with an industry's presence in a metropolitan area, I turn to this paper's fundamental question: might localization economies derive from plant-size productivity effects? To provide an answer, I consider the following characterization of the hourly wage earnings of worker j of city c in year t, w_{jct}:

$$\log(w_{jct}) = \beta_t' \mathbf{x}_{jct} + \gamma' \mathbf{z}_{ct} + \theta \log\left(\text{Emp}_{jct}\right) + \delta_t + \mu_c + \epsilon_{jct} \tag{5}$$

Here, the vector $\mathbf{x}_{jct}$ denotes a set of personal covariates including years of education completed, three educational attainment dummies – some or all high school completed, some college, college or more – and years of education interacted with each dummy; a quartic in potential experience; race, gender, and marital status dummies, fully interacted with one another; 7 one-digit occupation indicators; and 19 two-digit industry indicators. Each of these personal characteristics is specified with a time-varying coefficient to account for

belonging to, say, Food and Kindred Products (SIC 20) in one city may be quite different from those located in another. Estimating (3) and (4) separately by industry allows me to account for such differences (to the extent that they are reasonably fixed over the 10-year horizon) by correlating changes in average plant size and employment *within* city-industries.

changes in their 'prices' over time.[20] The vector $\mathbf{z}_{ct}$ contains a set of time-varying city-specific characteristics including log population, the proportion of the adult population with a college degree, the local unemployment rate, 8 Census division dummies (described in the Appendix), and an estimate of the local unionization rate;[21] $\log(\text{Emp}_{jct})$ is the logarithm of the individual's own-industry (two-digit) employment, designed to capture localization 'effects' on wages; δ_t denotes a year-specific intercept; μ_c is a city-specific term treated in various ways below; and ϵ_{jct} is an idiosyncratic term assumed to be uncorrelated across individuals, cities, and time.

The primary goal of this wage equation is to investigate the extent to which the localization effect, given by the parameter, θ, can be attributed to average establishment size. With this objective in mind, I estimate three specifications of (5).[22] In the first, I merely estimate the equation as written to evaluate the magnitude of industry localization effects. This is done primarily for the sake of comparison with previous work. In the next two specifications, I decompose log own-industry employment (log(Emp)) into the sum of the logarithm of average establishment size (log(AES)) and the logarithm of the total number of establishments (log(Est)), thereby replacing the term $\theta \log(\text{Emp}_{jct})$ with $\theta_1 \log(\text{AES}_{jct}) + \theta_2 \log(\text{Est}_{jct})$.[23] This allows me to evaluate the extent to which the localization effects estimated in the first

[20] I also performed the analysis using white males only. The resulting estimates were very similar to those reported here.

[21] The unionization data are based on Hirsch et al. (2001) who report state-level unionization rates (among non-agricultural wage and salary workers) for both 1980 and 1990. For each year, I impute a city's rate by taking a weighted average of the state-level rates across all states in which the city lies. The weights are given by the fraction of the city's Census observations falling into each state.

[22] Note, there may very well be an endogeneity problem associated with the estimation of this equation (e.g. wage levels may influence worker and producer location decisions, thereby affecting industry employment). Hence, even though I refer to the parameter estimates as 'effects,' the results should *not* be interpreted as causal.

[23] Formally, of course, this decomposition only strictly holds for the simple average plant size measure, not the weighted average. Nonetheless, for the sake of comparison, I utilize both measures in the analysis.

specification, θ, derive from a plant-size effect (i.e. larger plants) for a given number of producers, θ_1, or a plant-count effect (i.e. more establishments) for a given average plant size, θ_2.

Each specification of the model is estimated in two ways: random effects generalized least squares (GLS) and fixed effects. In the first procedure, the term μ_c is treated as a stochastic element assumed uncorrelated with the model's regressors. The fixed effects approach, by contrast, treats μ_c as a city-specific intercept to be estimated and, thus, does not rely on this particular assumption for consistency (Greene (2000, p. 576)).

Results from the specification of (5) in which the localization effect, θ, is constrained to be constant across two-digit industries, are given in the first row of estimates in Tables 5 (random effects) and 6 (fixed effects). For the sake of brevity, all other coefficient estimates have been suppressed, although a nearly complete list of them for this first specification appears in Table A1 of the Appendix.[24] Consistent with the findings of previous work (e.g. Henderson (1986)), the estimated values are significantly positive and suggest an elasticity in the neighborhood of 0.04 (i.e., a doubling of own-industry employment is associated with a 4 percent increase in average hourly earnings).

To what extent, then, can these effects be attributed to average plant size as opposed to the number of producers? Estimated values of the establishment-size component, θ_1, and the establishment-count component, θ_2, are given in the final four columns of results in the first rows of Tables 5 and 6. Overwhelmingly, they demonstrate that, between the two, the establishment-size effect is the more important piece. When using the weighted average, the implied firm-size wage elasticity is approximately 4 percent while that for the simple average lies between 7.5 and 8 percent. These figures, interestingly, are similar to those reported by Brown and Medoff (1989) and Troske (1999) whose estimates of plant-size

[24]Information about the numbers of individual- and city-level observations by two-digit industry used in the regression analysis is provided in Table A2 of the Appendix.

wage premia generally fall between 3 and 6 percent. All are highly significant.

At the same time, the results show that the establishment-count effect, holding plant size constant, is extremely small. Across the four estimates in the two tables, the largest is only 0.7 percent, and only one of the coefficients is statistically different from zero at conventional levels (i.e. at least 10 percent). Such results clearly suggest that, after conditioning on average establishment size, variation in the number of plants is not an important feature of the observed association between localization and wages.

Allowing the localization parameter, θ, and the decomposed effects, θ_1 and θ_2, to differ across two-digit industries produces qualitatively similar findings. The estimates, which appear in the remaining rows of Tables 5 and 6, indicate that, for each industry, localization effects are significantly positive at conventional levels. There is, to be sure, some difference across industries: the estimates, for instance, suggest that own-industry wage elasticities range from approximately 0.02 (for SIC 30, Rubber and Miscellaneous Plastics Products) to roughly 0.078 (for SIC 24, Lumber and Wood Products). Nevertheless, most lie reasonably close to the 4 percent benchmark derived in the pooled sample.

More importantly, the decomposed contributions of plant size and plant counts again demonstrate the significant role of average plant size in these localization terms. All of the coefficients on both measures of average establishment size are positive, and very nearly all of them are significant. Of the 20 coefficients, for example, the weighted average produces significantly positive coefficients in 18 cases using fixed effects; 19 using random effects GLS. Results for the simple average are similar: 19 are positive and significantly non-zero across both estimation techniques. Throughout, only one industry, SIC 31 (Leather and Leather Products) produces consistently insignificant (although positive) plant-size coefficients.

As for the plant-count effects, many of the estimated coefficients in Tables 5 and 6 are significant at conventional levels, unlike in the pooled results described previously. Three comments, however, are in order. First, of these significant coefficients, many are actually negative, suggesting that, for these industries, greater numbers of producers do not add to the wages of workers (and, thus, explain localization effects). Second, even among the positive plant-count coefficients, the magnitudes tend to be small, averaging roughly 1 to 1.5 percent. In fact, only one industry, SIC 24 (Lumber and Wood Products) consistently produces a coefficient in excess of 3 percent. The estimated plant-size effects, by contrast, average approximately 4.4 percent when considering the weighted measure, 7.5 percent when using the simple average. Third, following from this last point, direct comparison of the two effects within each industry indicates that the plant-size component is clearly the greater of the two. Only one industry (SIC 27, Printing and Publishing) shows any indication of a larger plant-count effect.

4 Concluding Comments

It is widely known that various measures of productivity, including wage earnings, rise as workers are organized into either larger production establishments or markets in which their industries are heavily represented. Yet, while a substantial body of research has explored these two empirical regularities, surprisingly little work has considered the possibility that they may be related. This paper offers evidence suggesting that they are.

Again, the findings documented here show that (i) establishment size increases substantially as an industry's total employment in a metropolitan area rises, and (ii) increases in hourly wage earnings tied to increases in city-industry employment operate primarily through plant scale, not the total number of plants. Localization effects on wages, therefore, seem to be plant-size effects, not plant-count effects.

Does this finding imply that localization economies are nothing more than a manifestation of plant-level scale economies and, thus, that the geographic concentration of industry, by itself, plays no role in raising productivity? That is, do the results imply that external productivity shifts are not an important aspect of industry clusters? A complete answer, naturally, is beyond the scope of this paper. However, from a theoretical perspective, there is no reason that this conclusion should follow. Indeed, it is certainly possible that, by generating positive externalities, the localization of industry creates an environment in which producers optimally choose to operate on a larger scale.

Take, for example, Marshall's (1920) three frequently-cited explanations for the geographic concentration of industry – technological externalities, increased intermediate-input variety, and economies of labor market search. Each implies that a producer's productivity increases as the size of its industry within a relevant local market expands. Higher levels of productivity may then translate into a larger average plant size either by increasing a producer's optimal scale of production (say, through an increase in a producer's marginal productivity of labor) or by attracting relatively large producers (possibly because such producers have more to gain from the increased productivity than small producers). Geographic concentration's role in generating localization effects may therefore take the form of attracting (or otherwise generating) larger, more productive establishments.

This line of reasoning suggests that the two traditional groups of explanations discussed in the Introduction – those relating to effects either external or internal to firms – should not be viewed independently from one another. Instead, future work should consider more carefully how externalities of various types, including Marshall's three, may influence the establishment-level organization of production.

Table 1A: Summary Statistics - 1980 Sample

Variable	Mean	Standard Deviation	Minimum	Maximum
Hourly Wage	10.01	5.87	2	59.86
Years of Education	11.91	2.96	0	20
Years of Experience	20.69	13.52	0	59
Female	0.31	0.46	0	1
Non-White	0.13	0.33	0	1
Female Non-White	0.047	0.21	0	1
Married	0.69	0.46	0	1
No High School	0.12	0.32	0	1
High School	0.6	0.49	0	1
Some College	0.15	0.36	0	1
College or More	0.13	0.34	0	1
Professional/Technical	0.11	0.32	0	1
Managers/Officials/Proprietors	0.08	0.27	0	1
Clerical Workers	0.13	0.34	0	1
Sales Workers	0.026	0.16	0	1
Craftsmen	0.2	0.4	0	1
Operatives	0.39	0.49	0	1
Service Workers	0.02	0.14	0	1
Laborers	0.04	0.2	0	1
College Fraction	0.17	0.035	0.077	0.35
Population	4723219	5247793	100376	17260490
Unemployment Rate	0.066	0.02	0.022	0.15
Unionization Rate	0.25	0.08	0.06	0.35
Own-Industry Simple Average Est. Size	172.6	278.2	2	3750
Own-Industry Weighted Average Est. Size	1956.4	2445.5	2.33	10353.4
Own-Industry Employment	46990.8	57489.8	6	210607
Own-Industry Establishments	688.4	1233.7	1	7299

Note: 144304 individual observations across 200 metropolitan areas.

Table 1B: Summary Statistics - 1990 Sample

Variable	Mean	Standard Deviation	Minimum	Maximum
Hourly Wage	10.48	6.52	2	59.97
Years of Education	12.59	2.88	0	18
Years of Experience	21.1	12.1	0	59
Female	0.32	0.47	0	1
Non-White	0.13	0.33	0	1
Female Non-White	0.05	0.22	0	1
Married	0.67	0.47	0	1
No High School	0.07	0.26	0	1
High School	0.49	0.5	0	1
Some College	0.26	0.44	0	1
College or More	0.19	0.39	0	1
Professional/Technical	0.15	0.36	0	1
Managers/Officials/Proprietors	0.11	0.32	0	1
Clerical Workers	0.12	0.33	0	1
Sales Workers	0.036	0.19	0	1
Craftsmen	0.19	0.4	0	1
Operatives	0.32	0.47	0	1
Service Workers	0.02	0.13	0	1
Laborers	0.04	0.19	0	1
College Fraction	0.22	0.05	0.1	0.37
Population	5008215	5668820	108711	17830586
Unemployment Rate	0.06	0.01	0.03	0.14
Unionization Rate	0.17	0.07	0.046	0.29
Own-Industry Simple Average Est. Size	109.6	167.3	1	2047.2
Own-Industry Weighted Average Est. Size	1283.9	1792.5	2.3	8794.8
Own-Industry Employment	37554.9	49457.8	3	223972
Own-Industry Establishments	749.9	1257.4	1	6442

Note: 121099 individual observations across 200 metropolitan areas.

Table 2: Localization and Plant Size

Average Plant Size Measure Results

	Simple Average		*Weighted Average*	
Variable	*I*	*II*	*I*	*II*
Log Industry Employment	0.65	0.65	0.93	0.93
	(0.005)	(0.005)	(0.006)	(0.006)
Log Population	–	-0.9	–	-0.4
		(0.23)		(0.4)
Log Population Density	–	0.17	–	0.02
		(0.22)		(0.38)
Log Per Capita Income	–	0.17	–	-0.02
		(0.2)		(0.26)
College Fraction	–	-0.16	–	-0.11
		(0.97)		(1.24)
Fraction Under 18	–	1.93	–	-0.15
		(1.24)		(1.54)
Fraction Over 64	–	-2.76	–	-2.16
		(1.3)		(1.7)
Fraction Non-white	–	-0.9	–	-0.15
		(1.21)		(1.5)
Unemployment Rate	–	0.7	–	0.35
		(0.58)		(0.74)
R^2	0.83	0.84	0.86	0.86

Note: Results from estimates of (3). 9722 city-industry-year observations. Each specification also includes a time effect (for the year 1980), city-specific effects, and industry-specific effects. Heteroskedasticity-consistent standard errors appear in parentheses.

Table 3: Localization and Plant Size

Empirical Distribution Function Results

Variable	*Share of Plants with Employment*					
	1-19	1-49	1-99	1-249	1-499	1-999
Log Industry Employment	-0.11	-0.1	-0.08	-0.05	-0.03	-0.013
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log Population	0.19	-0.11	0.12	0.02	0.03	0.02
	(0.18)	(0.17)	(0.09)	(0.06)	(0.03)	(0.02)
Log Population Density	-0.08	0.16	-0.08	-0.02	-0.004	-0.005
	(0.17)	(0.17)	(0.08)	(0.06)	(0.03)	(0.015)
Log Per Capita Income	-0.06	0.06	0.02	0.09	0.02	0.01
	(0.07)	(0.07)	(0.06)	(0.05)	(0.04)	(0.02)
College Fraction	0.13	-0.39	-0.16	-0.5	-0.24	-0.04
	(0.36)	(0.33)	(0.3)	(0.26)	(0.2)	(0.12)
Fraction Under 18	-0.56	-0.08	-0.17	-0.22	-0.38	-0.15
	(0.46)	(0.42)	(0.36)	(0.28)	(0.23)	(0.13)
Fraction Over 64	0.29	0.27	0.74	0.14	0.05	0.17
	(0.51)	(0.45)	(0.39)	(0.29)	(0.22)	(0.13)
Fraction Non-white	-0.98	-0.15	-0.14	0.14	0.33	0.23
	(0.45)	(0.41)	(0.37)	(0.31)	(0.28)	(0.19)
Unemployment Rate	0.22	0.06	-0.02	-0.09	-0.16	-0.1
	(0.21)	(0.19)	(0.17)	(0.14)	(0.12)	(0.06)
R^2	0.56	0.57	0.54	0.4	0.26	0.18

Note: Results from specification II of (4). 9722 city-industry-year observations. Also included are a time effect (for the year 1980), city-specific effects, and industry-specific effects. Heteroskedasticity-consistent standard errors appear in parentheses.

Table 4: Localization and Plant Size

Industry-Specific Results

SIC	Simple Average	Weighted Average	Share 1-19	Share 1-49	Share 1-99	Share 1-249	Share 1-499	Share 1-999
20	0.77	0.91	-0.13	-0.09	-0.08	-0.03	-0.02	-0.009
	(0.05)	(0.07)	(0.02)	(0.02)	(0.02)	(0.007)	(0.005)	(0.003)
21	0.59	0.71	0.06	0.1	-0.12	-0.08	-0.1	-0.013
	(0.13)	(0.15)	(0.08)	(0.09)	(0.06)	(0.06)	(0.05)	(0.02)
22	0.74	0.91	-0.15	-0.1	-0.11	-0.09	-0.04	-0.026
	(0.05)	(0.07)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
23	0.76	1	-0.14	-0.09	-0.07	-0.03	-0.008	-0.0004
	(0.04)	(0.05)	(0.02)	(0.02)	(0.01)	(0.008)	(0.003)	(0.0003)
24	0.73	1.04	-0.1	-0.07	-0.05	-0.015	-0.002	-0.0004
	(0.04)	(0.09)	(0.02)	(0.01)	(0.01)	(0.004)	(0.001)	(0.0003)
25	0.74	0.93	-0.11	-0.09	-0.08	-0.023	-0.01	-0.002
	(0.04)	(0.06)	(0.01)	(0.01)	(0.01)	(0.007)	(0.005)	(0.001)
26	0.8	0.97	-0.18	-0.16	-0.16	-0.075	-0.07	-0.007
	(0.05)	(0.09)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.005)
27	0.78	1.19	-0.054	-0.06	-0.05	-0.03	-0.019	-0.002
	(0.04)	(0.15)	(0.027)	(0.02)	(0.01)	(0.007)	(0.004)	(0.001)
28	0.77	0.99	-0.12	-0.09	-0.07	-0.045	-0.025	-0.002
	(0.04)	(0.07)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.001)
29	0.79	0.79	-0.19	-0.08	-0.04	-0.007	-0.004	-0.005
	(0.05)	(0.07)	(0.03)	(0.02)	(0.015)	(0.007)	(0.007)	(0.003)

Table 4 Continued

SIC	Simple Average	Weighted Average	Share 1-19	Share 1-49	Share 1-99	Share 1-249	Share 1-499	Share 1-999
30	0.84	0.84	-0.13	-0.11	-0.1	-0.075	-0.04	-0.037
	(0.03)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.015)	(0.015)
31	0.78	0.93	-0.16	-0.15	-0.11	-0.06	-0.022	-0.001
	(0.06)	(0.08)	(0.02)	(0.02)	(0.02)	(0.02)	(0.013)	(0.001)
32	0.8	1.05	-0.1	-0.08	-0.05	-0.027	-0.016	-0.004
	(0.04)	(0.09)	(0.02)	(0.01)	(0.01)	(0.006)	(0.005)	(0.001)
33	0.81	1.05	-0.15	-0.14	-0.11	-0.056	-0.044	-0.012
	(0.03)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.005)
34	0.85	1.07	-0.1	-0.07	-0.07	-0.043	-0.017	-0.006
	(0.04)	(0.06)	(0.02)	(0.01)	(0.01)	(0.007)	(0.004)	(0.002)
35	0.83	1.19	-0.07	-0.06	-0.05	-0.033	-0.026	-0.017
	(0.02)	(0.08)	(0.01)	(0.01)	(0.006)	(0.005)	(0.006)	(0.006)
36	0.85	1.1	-0.11	-0.1	-0.1	-0.07	-0.042	-0.02
	(0.03)	(0.04)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.007)
37	0.81	1.05	-0.14	-0.12	-0.1	-0.05	-0.03	-0.02
	(0.03)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.007)
38	0.83	1.1	-0.09	-0.08	-0.066	-0.05	-0.03	-0.018
	(0.04)	(0.06)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.006)
39	0.82	1.06	-0.09	-0.07	-0.066	-0.03	-0.01	-0.001
	(0.04)	(0.06)	(0.02)	(0.01)	(0.01)	(0.007)	(0.004)	(0.001)

Note: Coefficients on log industry employment from specification II (see Tables 2 and 3) of (3) and (4) estimated separately for each industry. Numbers of city-year observations by industry are 550 (SIC 20), 101 (SIC 21), 406 (SIC 22), 541 (SIC 23), 546 (SIC 24), 526 (SIC 25), 484 (SIC 26), 550 (SIC 27), 537 (SIC 28), 427 (SIC 29), 531 (SIC 30), 320 (SIC 31), 549 (SIC 32), 478 (SIC 33), 548 (SIC 34), 550 (SIC 35), 521 (SIC 36), 519 (SIC 37), 496 (SIC 38), 542 (SIC 39). Heteroskedasticity-consistent standard errors appear in parentheses.

Table 5: Estimated Localization, Plant-Size, and Plant-Count Effects

Random Effects Estimates

		Simple Average		*Weighted Average*	
SIC	Emp, $\hat{\theta}$	AES, $\hat{\theta}_1$	Est, $\hat{\theta}_2$	AES, $\hat{\theta}_1$	Est, $\hat{\theta}_2$
All	0.041 (0.001)	0.075 (0.001)	0.003 (0.002)	0.04 (0.001)	-0.0003 (0.002)
20	0.066 (0.003)	0.071 (0.008)	0.018 (0.003)	0.044 (0.005)	0.022 (0.003)
21	0.058 (0.008)	0.091 (0.01)	-0.018 (0.019)	0.078 (0.008)	-0.057 (0.02)
22	0.021 (0.003)	0.033 (0.006)	0.004 (0.003)	0.008 (0.004)	0.007 (0.003)
23	0.038 (0.002)	0.05 (0.008)	0.011 (0.002)	0.041 (0.006)	0.014 (0.002)
24	0.078 (0.005)	0.079 (0.015)	0.035 (0.006)	0.05 (0.009)	0.04 (0.006)
25	0.038 (0.003)	0.068 (0.007)	0.00001 (0.004)	0.042 (0.005)	0.004 (0.004)
26	0.026 (0.004)	0.1 (0.009)	-0.008 (0.004)	0.052 (0.006)	-0.013 (0.003)
27	0.061 (0.002)	0.12 (0.009)	0.021 (0.002)	0.02 (0.005)	0.024 (0.003)
28	0.04 (0.002)	0.09 (0.005)	0.023 (0.007)	0.049 (0.003)	-0.01 (0.003)
29	0.061 (0.006)	0.058 (0.008)	0.023 (0.007)	0.043 (0.007)	0.016 (0.008)
30	0.018 (0.006)	0.12 (0.01)	-0.008 (0.006)	0.061 (0.007)	-0.027 (0.005)
31	0.022 (0.006)	0.017 (0.014)	-0.009 (0.006)	0.01 (0.01)	-0.002 (0.006)
32	0.041 (0.004)	0.072 (0.01)	-0.009 (0.004)	0.041 (0.006)	-0.003 (0.004)
33	0.056 (0.002)	0.11 (0.005)	0.009 (0.003)	0.071 (0.003)	-0.019 (0.003)
34	0.036 (0.002)	0.053 (0.009)	-0.001 (0.003)	0.016 (0.004)	0.004 (0.003)
35	0.045 (0.002)	0.084 (0.006)	0.009 (0.002)	0.034 (0.002)	0.003 (0.002)
36	0.031 (0.002)	0.051 (0.005)	-0.002 (0.002)	0.032 (0.003)	-0.007 (0.002)
37	0.049 (0.002)	0.075 (0.003)	0.002 (0.002)	0.048 (0.002)	-0.005 (0.003)
38	0.049 (0.003)	0.084 (0.005)	0.007 (0.003)	0.05 (0.004)	-0.0001 (0.003)
39	0.025 (0.002)	0.034 (0.007)	-0.01 (0.003)	0.017 (0.005)	-0.003 (0.003)

Note: Coefficients on the log of an individual's own-industry employment (Emp), log own-industry average establishment size (AES), and log own-industry number of establishments (Est) from estimation of (5). 265403 observations. Standard errors are reported in parentheses.

Table 6: Estimated Localization, Plant-Size, and Plant-Count Effects

Fixed Effects Estimates

		Simple Average		*Weighted Average*	
SIC	Emp, $\hat{\theta}$	AES, $\hat{\theta}_1$	Est, $\hat{\theta}_2$	AES, $\hat{\theta}_1$	Est, $\hat{\theta}_2$
All	0.041 (0.001)	0.075 (0.001)	0.002 (0.002)	0.04 (0.001)	-0.0007 (0.002)
20	0.065 (0.003)	0.076 (0.008)	0.018 (0.003)	0.046 (0.005)	0.022 (0.003)
21	0.057 (0.008)	0.09 (0.01)	-0.018 (0.019)	0.078 (0.009)	-0.056 (0.02)
22	0.021 (0.003)	0.031 (0.006)	0.003 (0.003)	0.006 (0.004)	0.008 (0.003)
23	0.038 (0.002)	0.054 (0.008)	0.011 (0.002)	0.04 (0.006)	0.014 (0.002)
24	0.077 (0.005)	0.078 (0.015)	0.034 (0.006)	0.05 (0.009)	0.04 (0.006)
25	0.037 (0.003)	0.068 (0.007)	-0.001 (0.004)	0.042 (0.005)	0.004 (0.004)
26	0.026 (0.004)	0.1 (0.009)	-0.008 (0.004)	0.053 (0.006)	-0.012 (0.004)
27	0.059 (0.002)	0.11 (0.01)	0.019 (0.002)	0.021 (0.005)	0.022 (0.003)
28	0.04 (0.002)	0.088 (0.005)	-0.003 (0.003)	0.047 (0.003)	-0.009 (0.003)
29	0.059 (0.006)	0.051 (0.008)	0.026 (0.007)	0.038 (0.007)	0.022 (0.008)
30	0.018 (0.006)	0.12 (0.01)	-0.009 (0.006)	0.06 (0.007)	-0.026 (0.005)
31	0.022 (0.006)	0.019 (0.014)	-0.009 (0.006)	0.013 (0.011)	-0.002 (0.006)
32	0.041 (0.004)	0.072 (0.01)	-0.009 (0.004)	0.041 (0.006)	-0.003 (0.004)
33	0.056 (0.002)	0.11 (0.005)	0.009 (0.003)	0.07 (0.003)	-0.017 (0.003)
34	0.036 (0.002)	0.053 (0.009)	-0.002 (0.003)	0.016 (0.004)	0.004 (0.003)
35	0.044 (0.002)	0.077 (0.006)	0.007 (0.002)	0.032 (0.002)	0.003 (0.002)
36	0.03 (0.002)	0.05 (0.005)	-0.004 (0.002)	0.032 (0.003)	-0.007 (0.002)
37	0.049 (0.002)	0.077 (0.003)	0.002 (0.002)	0.048 (0.002)	-0.006 (0.003)
38	0.047 (0.003)	0.084 (0.005)	0.006 (0.003)	0.05 (0.004)	-0.001 (0.003)
39	0.024 (0.002)	0.033 (0.007)	-0.01 (0.003)	0.017 (0.005)	-0.003 (0.003)

Note: Coefficients on the log of an individual's own-industry employment (Emp), log own-industry average establishment size (AES), and log own-industry number of establishments (Est) from estimation of (5). 265403 observations. Standard errors are reported in parentheses.

A Appendix

A.1 Census Data Details

All individual observations used in the wage regressions are derived from the 1980 ('B') and 1990 1 Percent Metro Samples of the Integrated Public Use Microdata Series. As stated in the text, the samples are limited to individuals employed in manufacturing who are 18 to 65 years of age, are not in school, report having usually worked at least 30 hours per week, and earn between 2 and 60 dollars per hour (in real, 1982 dollars). The bottom figure lies slightly above one half of the 1982 minimum wage (3.35 dollars per hour). The top figure is the same cutoff as the one used by Moretti (2003). The estimated localization effects (including the decompositions) were not sensitive alternative values (e.g. 70 or 80 dollars per hour). Hourly wages are computed as annual wage and salary earnings divided by the product of usual weekly hours and the number of weeks worked. Following previous research using these Census data (e.g. Autor et al. (1998)), topcoded wage and salary earnings are imputed as 1.5 times the topcode for 1980, and as 210000 dollars for 1990. Given the trimming of the sample, this transformation affected very few observations: only 0.1 percent of the total. These figures are then deflated using the Personal Consumption Expenditures Chain-Type Price Index of the National Income and Product Accounts.

Because the 1990 Census does not report years of schooling completed for all individuals, I impute years of education for each individual in this year using the figures reported in Table 5 of Park (1994). Potential experience is then calculated as the maximum of (age - years of education - 6) and 0.

A.2 Calculating the Weighted Plant Size Measure

County Business Patterns (CBP) reports total establishment counts within 12 employment size classes at the county level. In order to estimate the mean number of workers per plant for each of these categories, I begin by calculating the fraction of all establishments (across all industries and counties) that fall into each size class. These shares then allow me to estimate 11 quantiles characterizing the distribution of plant sizes by taking cumulative shares. For example, in the 1990 CBP data 27.4 percent of all establishments have between 1 and 4 workers. I use this information to approximate the 0.274 quantile of the distribution as 4. Label these quantiles X_α. In addition, I use these 11 cumulative percentages to find the corresponding quantiles from a normal (0,1) distribution. Label these quantiles U_α. Assuming a lognormal plant-size distribution, X_α and U_α are related as follows:

$$X_\alpha = \exp(\zeta + U_\alpha \sigma)$$

where ζ and σ are the mean and variance parameters characterizing the lognormal distribution (see Johnson and Kotz (1970, p. 117)). These parameters can be obtained rather simply by taking logarithms and estimating by OLS.[25] With the estimated parameters $\hat{\zeta}$ and $\hat{\sigma}$ in hand, size category means are found by evaluating

[25]Because each of the 11 observations used to estimate ζ and σ represents a distributional feature to be

$$\left(\int_a^b \left(x\sqrt{2\pi}\hat{\sigma}\right)^{-1} \exp\left(\frac{-(\log(x)-\hat{\zeta})^2}{2\hat{\sigma}^2}\right) dx\right)^{-1} \left(\int_a^b \left(\sqrt{2\pi}\hat{\sigma}\right)^{-1} \exp\left(\frac{-(\log(x)-\hat{\zeta})^2}{2\hat{\sigma}^2}\right) dx\right)$$

for each closed bin [a,b] (i.e. 1-4, 5-9, ..., 2500-4999). The mean size for the open interval, 5000 or more, is found by taking the difference between total employment and the estimated total employment across all of the closed bins implied by these estimates. The resulting estimates (size class) for the 1980 data are: 2.33 (1-4), 6.84 (5-9), 13.99 (10-19), 31.7 (20-49), 70.3 (50-99), 154.9 (100-249), 346.2 (250-499), 687.2 (500-999), 1209.9 (1000-1499), 1894.6 (1500-2499), 3374.4 (2500-4999), 10412.6 (5000+). The resulting estimates (size class) for the 1990 data are: 2.31 (1-4), 6.82 (5-9), 13.96 (10-19), 31.6 (20-49), 70.05 (50-99), 154.03 (100-249), 345.1 (250-499), 684.9 (500-999), 1208.5 (1000-1499), 1891 (1500-2499), 3362.8 (2500-4999), 8944.9 (5000+). From these averages, the weighted average establishment size is calculated for each city-industry-year as described in the paper.

A.3 Composition of U.S. Census Divisions

Pacific: Washington, Oregon, California, Alaska, Hawaii

Mountain: Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico

West North Central: North Dakota, South Dakota, Minnesota, Nebraska, Iowa, Kansas, Missouri

West South Central: Oklahoma, Arkansas, Texas, Louisiana

East North Central: Wisconsin, Illinois, Michigan, Indiana, Ohio

East South Central: Kentucky, Tennessee, Mississippi, Alabama

New England: Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island

Middle Atlantic: New York, New Jersey, Pennsylvania

South Atlantic: Delaware, Maryland, District of Columbia, West Virginia, Virginia, North Carolina, South Carolina, Georgia, Florida

fitted, I refer to this procedure as a 'method-of-moments' approach. The resulting goodness-of-fit statistics from these regressions, incidentally, are extremely high. For both years of data, the R^2 exceeds 0.999.

Table A1: Additional Wage Regression Parameter Estimates

Variable	Random Effects	Fixed Effects
Intercept	0.71 (0.04)	-0.56 (0.25)
Non-White	-0.06 (0.01)	-0.07 (0.01)
Non-White*80	0.00002 (0.01)	0.004 (0.01)
Female	-0.17 (0.005)	-0.17 (0.005)
Female*80	-0.06 (0.006)	-0.06 (0.006)
Female*Non-White	0.03 (0.01)	0.03 (0.01)
Female*Non-White*80	0.02 (0.01)	0.02 (0.01)
Married	0.15 (0.003)	0.15 (0.003)
Married*80	0.003 (0.005)	0.001 (0.005)
Married*Non-White	-0.04 (0.01)	-0.04 (0.01)
Married*Non-White*80	-0.005 (0.01)	-0.005 (0.01)
Married*Female	-0.15 (0.006)	-0.16 (0.006)
Married*Female*80	-0.02 (0.008)	-0.02 (0.01)
Married*Female*Non-White	0.04 (0.01)	0.04 (0.01)
Married*Female*Non-White*80	0.02 (0.02)	0.02 (0.02)
Some/All High School	-0.35 (0.02)	-0.35 (0.02)
Some/All High School*80	0.1 (0.03)	0.1 (0.03)
Some College	-0.07 (0.09)	-0.07 (0.09)
Some College*80	0.09 (0.01)	0.08 (0.1)
College	-0.75 (0.07)	-0.73 (0.07)
College*80	0.6 (0.08)	0.58 (0.08)
Education Years	0.02 (0.002)	0.02 (0.002)
Education Years*80	0.004 (0.002)	0.004 (0.002)
Education Years*Some/All High School	0.04 (0.002)	0.04 (0.002)
Education Years*Some/All High School*80	-0.02 (0.003)	-0.02 (0.003)
Education Years*Some College	0.02 (0.007)	0.02 (0.007)
Education Years*Some College*80	-0.02 (0.008)	-0.015 (0.008)
Education Years*College	0.07 (0.004)	0.07 (0.004)
Education Years*College*80	-0.05 (0.005)	-0.05 (0.005)

Table A1 Continued

Variable	Random Effects	Fixed Effects
Experience	0.05 (0.002)	0.05 (0.002)
Experience*80	-0.004 (0.002)	-0.004 (0.002)
Experience^2	-0.2 (0.01)	-0.2 (0.01)
Experience^2*80	-0.002 (0.02)	-0.003 (0.02)
Experience^3	0.03 (0.004)	0.03 (0.004)
Experience^3*80	0.004 (0.005)	0.004 (0.005)
Experience^4	-0.002 (0.0004)	-0.002 (0.0004)
Experience^4*80	-0.0005 (0.0005)	-0.0006 (0.0005)
Professional	0.29 (0.01)	0.29 (0.01)
Professional*80	-0.1 (0.01)	-0.09 (0.01)
Manager	0.39 (0.01)	0.39 (0.01)
Manager*80	-0.07 (0.01)	-0.06 (0.01)
Clerical	0.11 (0.01)	0.11 (0.01)
Clerical*80	-0.07 (0.01)	-0.06 (0.01)
Sales	0.32 (0.01)	0.31 (0.01)
Sales*80	-0.1 (0.01)	-0.09 (0.01)
Craftsman	0.18 (0.01)	0.18 (0.01)
Craftsman*80	-0.05 (0.01)	-0.05 (0.01)
Operative	0.03 (0.01)	0.03 (0.01)
Operative*80	-0.02 (0.01)	-0.02 (0.01)
Service	-0.05 (0.01)	-0.05 (0.01)
Service*80	0.01 (0.01)	0.01 (0.01)
Log Resident Population	-0.03 (0.003)	0.034 (0.017)
College Fraction	0.95 (0.06)	2.74 (0.11)
Unionization Rate	0.63 (0.04)	0.57 (0.06)
Unemployment Rate	0.76 (0.1)	0.3 (0.12)

Note: Coefficient estimates for selected regressors from specification of equation (5) in which localization effects are captured by the log of own-industry employment, constrained to be equal across industries. Not listed are the estimated coefficients for 8 Census division indicators, a dummy for the year 1980, 19 industry indicators, and interactions between these industry dummies and the year dummy. A *80 suffix represents the interaction of a variable with the year dummy. Coefficients and standard errors on experience^2 and experience^2*80 have been multiplied by 100; experience^3 and experience^3*80 by 1000; experience^4 and experience^4*80 by 10000. 265403 observations. Standard errors in parentheses.

Table A2: Observations By Two-Digit Industry

SIC	Industry	1980 Census Obs.	1990 Census Obs.	1980 City Obs.	1990 City Obs.
20	Food and Kindred Products	9507	7664	198	199
21	Tobacco Products	499	372	46	41
22	Textile Mill Products	4140	3357	133	145
23	Apparel and Other Textile Products	7047	5141	180	175
24	Lumber and Wood Products	2000	2096	180	179
25	Furniture and Fixtures	2732	3009	173	171
26	Paper and Allied Products	4168	3482	174	169
27	Printing and Publishing	11637	15698	200	200
28	Chemicals and Allied Products	9028	7808	185	191
29	Petroleum and Coal Products	1590	1163	110	103
30	Rubber and Miscellaneous Plastics Products	1213	1172	132	134
31	Leather and Leather Products	1146	624	91	93
32	Stone, Clay, Glass, and Concrete Products	4063	3093	186	187
33	Primary Metal Industries	9487	4797	188	180
34	Fabricated Metal Products	9922	6893	196	194
35	Industrial Machinery and Equipment	18572	14693	200	198
36	Electrical and Electronic Equipment	15385	12734	192	197
37	Transportation Equipment	20430	17314	194	194
38	Instruments and Related Products	4916	4456	167	172
39	Miscellaneous Manufacturing Industries	6822	5533	191	190

Note: Number of individual observations and metropolitan area observations used in the estimation of (5).

References

Adbel Rahman, H. and M. Fujita. (1990) "Product Variety, Marshallian Externalities, and City Sizes." *Journal of Regional Science*. 30, 165-183.

Autor, D., L. Katz, and A. Krueger. (1998) "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*. 113, 1169-1213.

Black, D. and V. Henderson. (1999) "A Theory of Urban Growth." *Journal of Political Economy*. 107 (2), 252-284.

Brown, C. and J. Medoff. (1989) "The Employer Size-Wage Effect." *Journal of Political Economy*. 97 (5), 1027-1059.

Carlino, G. (1979) "Increasing Returns to Scale in Metropolitan Manufacturing." *Journal of Regional Science*. 19 (3), 363-373.

Ciccone, A. and R. Hall. (1996) "Productivity and the Density of Economic Activity." *American Economic Review*. 86, 54-70.

Davis, S., and J. Haltiwanger. (1991) "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-86." *Brookings Papers on Economic Activity: Microeconomics*. 115-180.

Davis, S., J. Haltiwanger, and S. Schuh. (1996) *Job Creation and Destruction*. Cambridge, MA: MIT Press.

Dinlersoz, E. (2004) "Cities and the Organization of Manufacturing." *Regional Science and Urban Economics*. 34 (1), 71-100.

Dixit, A. (1973) "The Optimum Factory Town." *Bell Journal of Economics and Management Science*. 4 (2), 637-651.

Dumais, G., G. Ellison, and E. Glaeser. (1997) "Geographic Concentration as a Dynamic Process." NBER Working Paper No. 6270.

Eberts, R. and D. McMillen. (1999) "Agglomeration Economies and Urban Public Infrastructure." *Handbook of Regional and Urban Economics*. Volume 3. P. Cheshire and E. Mills eds. New York: Elsevier North Holland.

Glaeser, E. and M. Kahn. (2001) "Decentralized Employment and the Transformation of the American City." NBER Working Paper No. 8117.

Glaeser, E. and D. Mare. (2001) "Cities and Skills." *Journal of Labor Economics*. 19 (2), 316-342.

Greene, W. (2000) *Econometric Analysis*. 4th Edition. Upper Saddle River, NJ: Prentice Hall.

Henderson, V. (1974) "The Sizes and Types of Cities." *American Economic Review.* 64 (4), 640-656.

Henderson, V. (1986) "The Efficiency of Resource Usage and City Size." *Journal of Urban Economics.* 19(1), 47-70.

Hirsch, B., D. Macpherson, and W. Vroman. (2001) "Estimates of Union Density by State." *Monthly Labor Review.* 124 (7), 51-55.

Holmes T. and J. Stevens. (2002) "Geographic Concentration and Establishment Scale." *Review of Economics and Statistics.* 84 (4), 682-690.

Idson, T. and W. Oi. (1999) "Workers Are More Productive in Large Firms." *American Economic Review.* 89(2), 104-108.

Johnson, N. and S. Kotz. (1970) *Continuous Univariate Distributions.* Boston: Houghton Mifflin.

Kim, S. (1995) "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends inn U.S. Regional Manufacturing Structure, 1860-1987." *Quarterly Journal of Economics.* 110 (4), 881-908.

Kremer, M. (1993) "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics.* 108(3), 551-575.

Krugman, P. (1991) *Geography and Trade.* Cambridge, MA: MIT Press.

Kumar, K., R. Rajan, and L. Zingales. (1999) "What Determines Firm Size?" NBER Working Paper No. 7208.

Lucas, Robert E. (1978) "On the Size Distribution of Business Firms." *Bell Journal of Economics.* Autumn, 508-523.

Marshall, A. (1920) *Principles of Economics.* London: Macmillan.

Moretti, E. (2003) "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics.* Forthcoming.

Nakamura, R. (1985) "Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities." *Journal of Urban Economics.* 17, 108-124.

Oi, W. and T. Idson. (1999) "Firm Size and Wages." in *Handbook of Labor Economics*, vol. 3B, O. Ashenfelter and D. Card eds. New York: Elsevier.

Park, J. (1994) "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and New CPS Measures of Educational Attainment." Princeton Industrial Relations Section Working Paper No. 338.

Rauch, J. (1993) "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics*. 34, 380-400.

Rosenthal, S. and W. Strange. (2001) "The Determinants of Agglomeration." *Journal of Urban Economics*. 50, 191-229.

Ruggles, S. and M. Sobek et al. (2003) *Integrated Public Use Microdata Series: Version 3.0*. Minneapolis: Historical Census Projects, University of Minnesota. http://www.ipums.org.

Troske, K. (1999) "Evidence on the Employer Size-Wage Premium From Worker-Establishment Matched Data." *Review of Economics and Statistics*. 81(1), 15-26.

U.S. Bureau of the Census. (1982) *County Business Patterns, 1980*. [machine readable data file]. Washington, D.C.: The Bureau.

U.S. Bureau of the Census. (1992) *County Business Patterns, 1990*. [machine readable data file]. Washington, D.C.: The Bureau.

U.S. Bureau of the Census. (1999) *USA Counties 1998 on CD-ROM*. [machine readable data file]. Washington, D.C.: The Bureau.

Wheaton, W. and M. Lewis. (2002) "Urban Wages and Labor Market Agglomeration." *Journal of Urban Economics*. 51, 542-562.