City Size and Skill Intensity

Joel A. Elvery*

This version: April 2007
Prior versions: August 2006, October 2006

Abstract:

Many have theorized that the productivity of human capital increases with city size. If this were true, one would expect that establishments in large cities are more skill intensive than those in small cities, even within industry. In this paper, I use data on the occupational mix at establishments to test whether the skill intensity of production methods varies with city size. Using data from the Occupational Employment Statistics survey and conditioning on detailed industry and establishment size, I show that establishments located in metropolitan areas with population below one million use a less skill intensive mix of workers than a comparable group of establishments in metropolitan areas with population above two million. The differences in skill intensity are not the same in all industries; establishments in relatively skill intensive industries are even more skill intensive in large MSAs. I interpret these results as evidence that the productivity of skilled labor is higher in large metropolitan areas than in small metropolitan areas.

JEL codes: J00, J24, R00

* The author is a research economist at the Bureau of Labor Statistics. The views expressed do not reflect the official positions or policies of the Bureau of Labor Statistics or the views of other staff members. This work has benefited greatly from comments from participants at the 2006 Comparative Analysis of Establishment Micro-data conference, Tony Barkume, Matt Dey, Keenan Dworak-Fisher, Judith Hellerstein, Lynn Phares, Anne Polivka, Laurie Salmon, Jeff Smith, George Stamas, and Leo Sveikauskas. My thanks go to them and to Ian Rucker for research assistance with the 2000 PUMS. Comments are very welcome. Prior to citing, please contact the author to see if there is an updated version of the paper. The author can be reached at jelvery(usual symbol)verizon.net.

1. Introduction

Labor and urban economists are interested in understanding why businesses are willing to pay workers in comparable jobs higher wages in large cities than in small cities. Many businesses find the benefits of locating in large cities sufficient to compensate for both higher land costs and higher nominal wages. This implies that many firms that produce goods and services for markets outside of their own city are more productive in a large city than they would be in a small city. If this were not the case, the firms would relocate to lower cost markets in order to increase profits. The theories as to why firms are more productive in large cities than in small cities focus on two types of productivity effects: general productivity effects and human capital productivity effects.[1]

The theories of how general productivity increases with city size tend to focus on gains from concentration.[2] For example, Krugman (1991) theorizes that businesses can reduce transportation costs by locating in cities with many potential customers and suppliers nearby. Large cities may reduce transportation costs by having transportation links that are more highly developed than those in smaller cities. In addition, the club theory literature suggests that large cities may provide a more comprehensive set of public goods since there are larger pools of firms and people to take advantage of public and quasi-public goods.[3]

The other broad class of theories center on ways in which large cities make human capital more productive. Building from Lucas's (1988) model, Rauch (1993) shows that cities can facilitate knowledge spillovers by putting people in close proximity to other people with valuable knowledge and that businesses that have human capital intensive production functions are more likely to benefit from these spillovers. Workers in large cities may also be higher ability than workers in smaller cities, either because they are attracted by urban amenities or because they selectively migrate to take

[1] Duranton and Puga (2003) provide a comprehensive and technical summary of the theories of urban agglomeration.
[2] This discussion of theories for why businesses are more productive in large cities closely follows the summary in Glaeser and Maré (1994).
[3] See Scotchmer (2002) for a survey of this literature.

advantage of higher wages available to high ability people in large cities (Borjas, Bronars, and Trejo 1992; Faberman 1998; Gould 2007). Glaeser and Maré (1994 and 2001) theorize that cities reduce the cost of acquiring human capital by facilitating interactions with people and therefore skilled workers will be more broadly available in large cities. Wheeler (2001) shows that cities can facilitate the job matching process, so forming worker-firm matches is less costly and more productive in large cities.

Oftentimes, cities' effects on productivity are thought of as some combination of general and human capital productivity effects. For example, agglomeration effects can be thought of as a combination of reduced transportation costs, increased knowledge spillovers, and more efficient industry specific input markets, including labor markets.[4] Theories of this sort explain the tendency of certain industries to be concentrated in one city, such as film production in Los Angeles or commodities trading in Chicago.

There have been a number of papers that shed light on how cities affect productivity by looking at the difference between the wages paid to comparable workers in cities of different sizes and types.[5] Rauch (1993) and Rosenthal and Strange (2005) show that wages increase with the education level of those nearby. Glaeser and Maré (2001), Wheeler (2001), and Kim (2003) provide evidence that the urban wage premium increases with education, suggesting that productivity effects are strongest for highly skilled workers. Similarly, Gould (2007) finds that the urban wage premium is much larger for white-collar workers than blue-collar workers. Even though he follows most of the literature and does not condition on the location specific cost of living, Gould finds that the urban wage premium for blue-collar workers can be largely explained by self-selection of high ability

[4] See Duranton and Puga (2003) for a survey of theories about the micro-foundations of agglomeration economies.
[5] See Rosenthal and Strange (2003) for a comprehensive survey of empirical research about the sources of agglomeration economies.

blue-collar workers to cities. These results suggest that cities do increase the productivity of human capital, but do not rule out an important role for general productivity effects.

While the evidence from the urban wage premium literature is valuable, many of the estimated wage effects are potentially biased due to selective migration and differences across cities in the cost of living. An extensive literature on comparing the quality of life across cities (e.g. Roback 1982; Blomquist, Berger, and Hoehn 1988; Eberts and Beeson 1989) shows that wages and rents are jointly determined and are affected by production and consumption amenities. Glaeser and Maré (2001) and Rosenthal and Strange (2003) note that it is difficult to estimate productivity effects with wage data when wages and rents are jointly determined.

A few recent papers have sought to address these problems by using establishment-level data to measure how establishment productivity varies across city types. Using longitudinal plant-level data from a subset of manufacturing industries, Henderson (2003) shows that plants' productivity is increasing in the presence of same-industry plants in the same county and finds little evidence of cross-industry effects. Morretti (2004) also focuses on manufacturing plants and estimates the productivity effects of human capital spillovers. Combining data on industry-city and city college education rates with plant-level data, Morretti shows that productivity is increasing in the share of city workforce that is college educated even when controlling for the share of own-industry workforce that is college educated. He finds that the productivity effects are even stronger for the share of similar industry workforce that is college educated. These results suggest that there are human capital externalities that increase productivity.

The establishment-level evidence on the productivity effects of urbanization and human capital spillovers has primarily focused on the manufacturing sector. This is largely due to data limitations. While longitudinal data on manufacturing plants became available in the early 1990's, it has been only very recently that longitudinal data is available for other sectors. Also, outside of the

manufacturing sector there is very little establishment-level data on capital and less precise output measures. The restriction to studying manufacturing is unfortunate. Large cities are increasingly specialized in service sectors, not manufacturing. Many of the theories about how urbanization increases labor productivity, such as knowledge spillovers or improved job matching, are more likely to be important outside of manufacturing.

In order to study a wider cross-section of the economy, I develop and implement a way to measure the skill intensity of firms' production methods using the occupational composition of their workforce. The mix of workers used by a firm provides information about the firm's production process. For example, firms with many workers in high skill occupations have skill intensive production processes. By comparing the occupational mix of establishments in small Metropolitan Statistical Areas (MSAs) to the occupational mix of a comparable group of establishments located in large MSAs, it is possible to test whether production processes differ in skill intensity across MSA size categories. It is important to look at a comparable pool of establishments because the industrial and establishment size composition of large and small MSAs are quite different and these factors affect occupational mix.

The urban wage premium literature consistently finds that the relative wage of skilled workers is highest is large cities. This implies that establishments in large MSAs should use less skill intensive production methods than those in small MSAs unless skilled labor is more productive in large MSAs. The idea is simple: if a factor is relatively more costly in a city and all other things are equal, then establishments in that city should substitute away from that factor. I find that the opposite is true; establishments in large MSAs are more skill intensive than comparable establishments in small MSAs. This suggests that all else is <u>not</u> equal – skilled labor is more productive in large MSAs.

Many of the theories regarding how cities affect human capital productivity suggest that the productivity effects would be strongest in skill-intensive industries. For example, Rauch (1993) theorized that knowledge spillovers are more likely to be capitalized by firms with knowledgeable workers than by other firms. Studying how skill intensity varies across city type by industrial sector can shed light on whether the differences are concentrated in certain sectors of the economy. Consistent with Rauch's hypothesis, I find that skill intensive sectors are more skill intensive in large cities. The differences in occupational mix across MSA size type in less skill intensive sectors are minor. Establishment-level regressions show that the gap in skill intensity between establishments in small and large MSAs increases with the industry's average skill intensity. In other words, the differences in skill intensity between establishments in small and large MSAs are greatest in skill intensive industries. This provides further support for theories that suggest that cities especially increase the productivity of skilled labor.

2. Theory of human capital productivity and skill composition

The endogenous technology adoption literature shows that establishments adopt different production techniques depending on the input markets they face [e.g. Acemoglu 1996 and 1998; Doms, Dunne, and Troske 1997; Nestoriak 2004; Beudry, Doms, and Lewis 2006].[6] This implies that an establishment's choice of factor mix provides information about the productivity of the factors in the establishment's market. The following representative firm model of labor demand develops that idea to illustrate the link between human capital productivity and occupational mix.

The model assumes that there are two types of workers, high skill and low skill, denoted by H and L. It is assumed that there is a fixed amount of capital required per worker of each type regardless of city size type, k^s for $s=\{H, L\}$. The cost per worker of type s, w_j^s, reflects this and

[6] Due to data limitations, most of the empirical evidence regarding technology adoption has focused on manufacturing.

$w_j^s = \omega_j^s + r_j k^s$, where *j* indexes city. Assume that there are two cities indexed with *j*=0,1. For simplicity, all firms produce the same good that sells for p_j in city *j* and there is no trade between the cities.

A firm in city *j* solves the following profit maximization problem:

$$\max_{H,L} \pi_j = p_j \varphi_j (\theta_j f(H) + g(L)) - (w_j^H H + w_j^L L) \tag{1}$$

where φ_j is the general productivity multiplier in city *j* and θ_j is the productivity multiplier specific to high skill workers in city *j*. The worker type specific production functions $f(H)$ and $g(L)$ do not vary across city type. Both $f(H)$ and $g(L)$ have positive first derivatives and negative second derivatives. I assume that $\frac{\partial f(n)}{\partial n} > \frac{\partial g(n)}{\partial n}$ and $\frac{\partial^2 f(n)}{\partial n^2} > \frac{\partial^2 g(n)}{\partial n^2}$, meaning that the marginal product of high skill workers is greater than the marginal product of low skill workers and the marginal product of high skill workers diminishes at a slower rate than that of low skill workers. Rearranging the first order conditions from (1) shows that:

$$\theta_j \frac{f'(H_j)}{g'(L_j)} = \frac{w_j^H}{w_j^L} \tag{2}$$

where $\{H_j, L_j\}$ is the profit maximizing mixture of high and low skill workers in city *j*. This shows that the firm chooses the mixture of workers such that the marginal benefit from hiring a high skilled worker relative to that from hiring a low skill worker is the same as the relative costs for the two types of workers. Note that neither the price of the output nor the general productivity multiplier enter into equation (2). Define the following ratios: $\Phi_j = f'(H_j) / g'(L_j)$ and $\Omega_j = w_j^H / w_j^L$. Comparing the maximizing worker mixes of firms in cities 1 and 0 shows that:

$$\left(\frac{\Phi_1}{\Phi_0}\right)^{-1} = \frac{\theta_1 / \theta_0}{\Omega_1 / \Omega_0}. \tag{3}$$

Given the assumptions stated above about $f(.)$ and $g(.)$, (3) implies that if $\Omega_1 / \Omega_0 > 1$ and $H_1 / L_1 > H_0 / L_0$ then $\theta_1 > \theta_0$.

Figure 1 illustrates the model in a case where $\Omega_1 / \Omega_0 > 1$ and $\theta_1 > \theta_0$. The convex lines are iso-quant curves. The lines $\overline{WY}$ and $\overline{WX}$ are the iso-cost lines for cities 0 and 1 respectively, with slopes $-1/\Omega_1$ and $-1/\Omega_0$. The optimal skill mix is at point A in city 0 and at point B in city 1 and $H_1 / L_1 > H_0 / L_0$. If the production function in city 1 were the same as that in city 0 (the iso-quant with quantity q') then the optimal skill mix in city 1 would be point C and $H_1' / L_1' < H_0 / L_0$. Alternatively, if the production function in city 0 were the same as that in city 1 (the iso-quant with quantity n'), then the optimal skill mix in city 0 would be point D and $H_1 / L_1 < H_0' / L_0'$.

The results of this model are intuitive. If an input is more heavily used by firms in city 1 than by comparable firms in city 0, then the input is either: relatively less expensive in city 1 than city 0, more productive in city 1 than in city 0, or both. This suggests that differences in skill mix in similar firms across city types can provide evidence as to whether cities increase the productivity of human capital, especially when combined with information on how costs per worker vary by skill type and city type.

The urban wage premium literature provides estimates on how wages vary with city type and skill. These estimates can be transformed to provide estimates of relative wages across skill type and

city type. However, the relative wage figures do not include any information on how capital costs per worker varies across worker or city type. If I assume that capital costs per worker are ignorable, the estimates of Ω_1/Ω_0 from the urban wage premium literature range from 1.03 to 1.10, meaning that the relative wage of high skill workers is higher in large cities than in small cities.[7] The capital cost per worker almost certainly varies with city type, primarily due to differences in the cost of land. Given the concentration of high skill workers in expensive locations (e.g. downtown) and the concentration of low skill workers in less expensive areas (e.g. industrial parks), it is likely that high skill workers and expensive land are complements in production. This would reinforce the pattern in wages and suggests that, even including capital costs per worker, Ω_1/Ω_0 would be greater than one. If one takes as given that Ω_1/Ω_0 is at least one, then if firms in small cities are less skill intensive than those in large cities it is strong evidence that human capital is more productive in large cities. The rest of this paper is devoted to testing whether firms in small cities use a more or less skill intensive mix of workers than those in large cities.

3. Occupational Employment Statistics survey micro-data

The data used in this paper are the micro-data from the Bureau of Labor Statistics' (BLS) Occupational Employment Statistics (OES) surveys from 2001 through 2003. The OES survey measures occupational employment and wage rates for wage and salary workers in non-farm establishments in the 50 states and the District of Columbia. The OES survey covers all industries except for portions of the Agriculture sector and Private Households. The data is used to create the

[7] The 1.03 figure is derived from Table 1 of Wheeler (2001), using population of 5 million as the large city population, population of 500,000 as the small city population, people with 16+ years of school as high skill, and people with 9-12 years of school as low skill. The figure of 1.10 is derived from Table 4 in Glaeser and Maré (2001) and large cities are defined as metropolitan areas with a central city with population above 1 million, small cities are all other areas, high skill workers are those with 16 years of education, and low skill workers are those with 12 years of education.

annual publication *Occupational Employment and Wages*. The following description summarizes the information on survey methodology provided in *Occupational Employment and Wages, May 2003* (Bureau of Labor Statistics 2004).

Through mail-in surveys and telephone contact, responding establishments report the number of employees by occupation using the Standard Occupational Classification (SOC) system. The SOC system includes 801 detailed occupations. Covered employees include all full- and part-time wage and salary employees, including those who are temporarily absent.[8] Excluded from employment are contract workers, self-employed owners, partners in unincorporated firms, household workers, and unpaid family workers. For each occupation, establishments report the number of employees in each of 12 wage intervals. To create mean wage and other occupational wage statistics, the wage interval data is combined with data from other sources to generate information on wage distributions.[9]

The OES survey uses a survey design consisting of six probability sample panels of approximately 200,000 establishments that are selected bi-annually. The samples are constructed such that the six panels combined form a sample of establishments that is representative of the universe of establishments by substate area, industry, and establishment size class. The substate area definitions used in the survey are metropolitan statistical areas, primary metropolitan statistical areas, and non-metropolitan balance of state areas.[10] A full set of six panels provides a sample of approximately 1.2 million establishments, which means that 18 percent of in-scope establishments are surveyed once in a set of panels.[11] The use of biannual panels began in 2002 and the data used in

[8] Temporarily absent workers are only included if they are being paid during their absence.
[9] BLS's Employment Cost Index survey for nine major occupational groups is used to adjust for collecting the data at different times. BLS's National Compensation Survey is used to generate the mean wage for each {occupation, wage interval} cell in order to estimate mean wages.
[10] Depending on the state, there are one to six balance of state areas per state.
[11] Approximately 2 percent of the establishments in a set of six panels are certainty units, meaning that they are surveyed once in every six panel set.

this study are drawn from an annual panel in 2001, two biannual panels in 2002, and two biannual panels in 2003.[12]

One can think of the OES micro-data as linked employer-employee data where the only thing known about each individual employee is their occupation, their wage interval, their establishment, and the industry, location, and size of the establishment where they work. While the information on each worker is limited, the same information is known for all their co-workers at the same establishment at the same time. This makes the OES micro-data well suited to study how the occupational mix of comparable groups of establishments differs across city size types.

The published OES data has been used by other authors to study differences in occupational composition and wages across states and metropolitan areas. For example, Hajiha and Salmon (2005) provide maps that show how employment concentrations vary among states and metropolitan areas by major occupational group. In similar work, Cover (2005) compares the occupational distribution of metropolitan and non-metropolitan areas and shows that occupations' wages tend to be higher in whichever area type they are concentrated in. Watson (2005) finds that the occupational mix is different in fast, moderate, and slow growing metropolitan areas. Kilcoyne (2004) uses shift share analysis to show that the differences in mean wages across states is primarily explained by differences in wages within occupation, but that the occupation mix also plays a role. None of these studies control for differences in industrial composition or establishment size across different geographic areas.

[12] In the May 2003 survey, about 79 percent of establishments in the sample responded to the survey and these respondents represent 72 percent of pre-survey weighted employment (BLS 2004). For nonresponding establishments, occupational employment patterns are imputed using a hot-deck procedure where nonrespondents are given the occupational mix of the responding establishment that has the most similar combination of industry, employment size, and geographic area. Wage distributions are imputed using the empirical distribution of all responding establishments in the same geographic area, industry, and size class cell as the nonrespondent. In both of the imputation stages, if an insufficient number of responding establishments is found, the restrictions on area, industry, and size are loosened until imputation is possible.

4. Methodology

This paper uses two complementary methodologies. The first uses occupation-by-sector-by-MSA-type statistics to provide general descriptions of how skill mix varies across MSA types. The second methodology uses establishment level regression to test whether within establishment skill mix differs across MSA types and to test whether establishments in skill intensive industries are less skill intensive in small MSAs. I start with the shared features of the methodologies.

There is no established definition of what constitutes a small, large, or medium city. In our analysis, MSAs are treated as equivalent to cities and the definitions used are as follows. The categories are defined using 2000 population for MSAs as follows: small MSAs have population of no more than one million, medium MSAs have population between one and two million, and large MSAs have population of at least two million.[13] The large MSAs include the 22 largest MSAs, the medium MSAs include the next 28 largest MSAs, and the remaining 228 MSAs are small MSAs.[14] The definitions used were chosen based on looking at kink points in the distribution of population size across MSAs and the desire to use round numbers. Table 1 shows the largest and smallest MSAs for each size category.

To facilitate summarizing skill intensity across industries and MSA types, it is helpful to have a continuous measure of how skilled each occupation is. The challenge in constructing a broadly applicable skill index is that different occupations require very different sets of skills. While it is easy to compare skill level between similar occupations (e.g. chef versus short order cook), it is difficult to do so for dissimilar occupations (e.g. chef versus auto mechanic). Creating an index of educational and training requirements is one approach, but it would miss hard-to-measure skills such as sales talent or creativity. To deal with these challenges, the measure of occupational skill I use is

[13] MSAs are defined as either Consolidated MSA or MSAs, using the definitions of Consildated MSAs and MSAs that were used in collecting the 2000 Census.
[14] MSAs in Puerto Rico are excluded from the sample.

the mean wage paid to workers in an occupation at establishments located in the middle size category of cities. The premise is intuitive: occupations that are more highly paid are more skilled.

Mean wage in the medium MSAs is used to avoid biasing the skill intensity comparisons. Nominal wages are undeniably higher in large MSAs than in small MSAs. Using the national average of wages would cause the skill measure to be biased by the share of an occupation in each MSA size class. More precisely, the national average wage of occupation *j* would be $\overline{w}_j = \rho_{jS}\overline{w}_{jS} + \rho_{jM}\overline{w}_{jM} + \rho_{jL}\overline{w}_{jL}$, where $\overline{w}_j$ is the national mean wage, $\overline{w}_{jq}$ is the mean wage for occupation *j* in size class *q* (which is *S*, *M*, and *L* for small, medium, and large MSAs respectively), and ρ_{jq} is the proportion of employment in occupation *j* in MSA size class *q*. If $\rho_{jL} \neq \rho_{jS}$ and $\overline{w}_{jL} \neq \overline{w}_{jS}$, then $\overline{w}_j$ would be affected by what I want to measure, the distribution of occupations across MSA types. Another benefit of using $\overline{w}_{jM}$ is that the medium sized cities are the most homogenous of the three MSA types. The largest medium MSA is the Cincinnati--Hamilton, OH--KY--IN CMSA (population of 1.98 million) and the smallest is the Louisville, KY--IN MSA (population of 1.026 million). For these reasons, the measure of the skill of occupation *j* is the mean hourly wage for occupation *j* in medium sized MSAs, which I call the medium wage for occupation *j*. The medium wages are calculated using an adaptation of the computer programs used to create the published data in the 2003 Occupational Employment and Wages published data, so the method is similar to that described in the OES documentation (Bureau of Labor Statistics 2004).[15]

The next component of the methodology is using the medium wages and the occupational mix of firms to compare the skill intensity of their production processes. To show how occupational

[15] The only modification in the program is that mean wages are estimated by MSA type rather than another geographic unit. There is one occupation that has employment in both large and small MSAs but not in medium MSAs. Since it is impossible to calculate the medium wage for this occupation, it is dropped from the analysis and does not contribute to the figures. According to Table 1 of Occupational Employment and Wages, May 2003, there were less than 725 workers in the occupation in the United States.

mix provides information about production processes, consider the following example. Suppose there are two bolts manufacturers, Joe's Bolts and Precision Bolts, with the mix of workers described in Table 2. Joe's and Precision use the same mix of workers except that Joe's employs seven non-computer controlled machine operators and a machinist while Precision employs seven computer-controlled machine operators and a computer-controlled machine programmer. Based on the job duties of each of these occupations, I can infer that Joe's uses less automated, and probably older, machinery than Precision.

To come to this conclusion, not only do I need to know that Joe's and Precision are in the same industry, but I have to know something about the production technology choices available to them. It would require equivalent knowledge of all industries to make similar inferences for the economy as a whole. Lacking that knowledge, I limit the focus to comparing skill intensity, rather than production processes as a whole.

To quantify skill intensity, I first calculate the percent of employment at each establishment in each occupation and find the percentage point difference in occupational employment across the two establishments (see the fourth column of Table 2). Then I relate the percentage point difference in occupational employment to the medium wage. For this example, I use an occupation-level regression where the dependent variable is the percentage point difference in occupational employment and the independent variable is the natural log of medium wage. This produces the slope coefficient -0.0015 and, since this is negative, I conclude that Joe's production method is less skill intensive than Precision's. This approach is similar to the occupation-level analysis described below.

Another way to quantify skill intensity is to look at the differences between Joe's and Precision at the same percentiles of their within establishment skill distribution. These figures are shown at the bottom of Table 2. Using the occupations' medium wages as the measure of skill, the

skill level of the 25th percentile employee at Joe's is 9.9 percent less than that of Precision's 25th percentile employee.[16] Joe's 75th percentile skill is only 0.3 percent less than Precision's 75th percentile skill. These figures tell us that Joe's Bolts is less skill intensive than Precision Bolts and that the differences are larger at the 25th percentile than the 75th percentile. This approach to measuring differences in skill intensity is analogous to the establishment regressions presented below.

4.1 Occupation-level analysis

For the occupation-level analysis, I compare the occupation mix of similar groups of establishments. Small establishments have lumpy occupational distributions and comparing groups of establishments smoothes over this lumpiness. I form samples of establishments from large and small MSAs in such a way that the detailed industry and establishment size distribution of employment is the same in both samples. Occupations' shares of employment and the differences in the occupational distributions in large and small MSAs are calculated from these comparable samples. Finally, the differences in the distributions are related to occupational skill with least squares regressions.

To test whether establishments use different production processes in small MSAs than in large MSAs, it is necessary to use comparable sets of establishments. This can be done either by weighting establishments in large MSAs to be comparable to those in small MSAs or vice versa. The economies of large MSAs tend to be more diverse than those of small MSAs, both in terms of industry composition and employment size composition. This means that weighting establishments in large MSAs to be comparable to establishments in small MSAs requires fewer large weights than vice versa. There are 2,035 industry, employment size class pairs with positive employment in the OES sample used for this paper. Of those, 30 have zero employment only in small MSAs and five

[16] In this example and throughout the paper, if the number of employees is such that the percentile falls between two employees, the average of the skill of the employee immediately above the percentile and the worker immediately below the percentile is used as the skill of that percentile.

have zero employment only in large MSAs. This implies that weighting large MSA establishments to be comparable to small MSA establishments generates the most comparable set of establishments possible. Therefore, I weight large MSA establishments to be comparable to small MSA establishments.

The construction of comparable samples is discussed in detail in the Appendix. I use attribute matching and reweight establishments in large MSAs so that the share of employment in an industry-establishment size cell in large MSAs is the same as that in small MSAs. Industry-establishment size cells that are only found in one MSA type are excluded from the study. The industry codes used to form the comparable establishments samples are the same as are used in OES publications, which are a combination of four and five-digit NAICS2002 codes. The establishment size classes used are also the same as in OES publications. The establishment size class ranges are (in number of employees): (1) 1-9, (2) 10-19, (3) 20-49, (4) 50-99, (5) 100-249, and (6) 250 or more.[17]

Using the comparable samples of establishments, the percent of small MSA employment that is in occupation *j* is $p_{j,S}$ and the percent of large MSA employment in occupation *j* is $p_{j,S|L}$. For brevity, I use the term *j*'s occupation share interchangeably with the percent of employment in occupation *j*. Controlling for the industry and size composition of employment, the percentage difference between MSA size types in occupation *j*'s employment share is $\Delta_{j,L|S} = \ln(p_{j,S} / p_{j,L|S})$. I estimate the following model to show how skill intensity varies across the MSA types:

$$\Delta_{j,L|S} = \alpha + \beta_1 \ln(\overline{w}_{M,j}) + \beta_2 \ln(\overline{w}_{M,j})^2 + \beta_3 \ln(\overline{w}_{M,j})^3 + \varepsilon_j . \qquad \text{(R1)}$$

[17] The program used to generate the employment counts by industry, establishment size class, and MSA size type cells is adapted from the program used to create the Occupational Employment and Wages, May 2003. It is adapted to exclude the Agricultural sector and a single occupation that has no employment in medium MSAs. The program is also altered to produce counts for each MSA size type rather then more typical geographic units.

This regression is weighted by the occupation's employment in small MSAs, $e_{j,S|S}$; the results weighting by $e_{j,L|S}$ are nearly identical. If the skill mix of establishments in small and large MSAs is the same when conditioning on industry and establishment size composition, then $\alpha = \beta_1 = \beta_2 = \beta_3 = 0$.

We also estimate a model where the difference in skill intensity is allowed to vary by industrial sector. Define $\Delta_{jg,L|S} = \ln(p_{jg,S} / p_{jg,L|S})$ as the percent difference in occupation *j*'s share of sector *g* employment in small and large MSAs when conditioning on industry and establishment size composition. I_g is an indicator variable for sector *g*. The regression that allows differences in skill intensity to vary by sector is:

$$\Delta_{jg,L|S} = \sum_g I_g Z_j \theta_g + \mu_{jg}, \quad Z_j = \{1, \ln(\overline{w}_{M,j}), \ln(\overline{w}_{M,j})^2\}. \tag{R2}$$

This regression is weighted by employment in small MSAs in occupation *j* and sector *g*.

4.2 Establishment-level analysis

Another way to approach the question of whether or not establishments in small MSAs are less skill intensive than comparable establishments in large MSAs is to look at establishment-level measures of skill intensity. This has the benefit of avoiding the reweighting of the sample and focuses less attention on the tails of the occupation skill distribution. The establishment-level regressions also enable us to test whether differences in skill intensity across MSA types vary with an industry's skill intensity.

Define σ_{kl} as the skill level of the employee at the *k*th percentile of establishment *l*s skill distribution, where *k*=10, 25, 50, 75, and 90. Skill is again measured at the occupation level and is $\ln(\overline{w}_{M,j})$ for all workers in occupation *j*, regardless of industry or location. $\overline{\sigma}_{M,k,i}$ is the average of σ_{kl} for all establishments in industry *i* located in medium MSAs and $\overline{\sigma}_{M,k}$ is the average of $\sigma_{l,k}$ for

all establishments located in medium MSAs.[18] The skill intensity of establishment *l*s industry, *i*, is measured as $\varphi_{kl} = \bar{\sigma}_{M,k,i} - \bar{\sigma}_{M,k}$.[19]

To test whether skill intensity varies with MSA type and whether differences in skill intensity are related to industry-average skill intensity, I estimate the following regressions:

$$\sigma_{kl} = S_l \delta_0^k + S_l \varphi_k \delta_1^k + I_l \psi^k + V_l \xi^k + \mu_{kl} \tag{R3}$$

$$\sigma_{kl} = S_l \gamma_0^k + S_l \varphi_k \gamma_1^k + \varphi_k \gamma_2^k + V_l \xi^k + \nu_{kl} \tag{R4}$$

for *k*=10, 25, 50, 75, and 90. *S* is an indicator variable that equals one if an establishment is located in a small MSA, *I* is a vector of industry fixed effects, and *V* is vector of establishment size, state, and survey panel fixed effects. The sample for these regressions is restricted to establishments in small and large MSAs. All standard errors for the establishment-level regressions are clustered by survey strata. If $\hat{\delta}_0^k$ <0 ($\hat{\gamma}_0^k$ <0) it indicates that, on average, workers at the *k*th percentile of the establishment level skill distribution of establishments small MSAs are less skilled than corresponding workers employed in establishments in large MSAs. If $\hat{\delta}_1^k$ <0 ($\hat{\gamma}_1^k$ <0) it indicates that the skill gap between establishments in small and large MSAs is greater in more skill intensive industries.

4.3 Caveats

The methodology I use is well suited to the question and data at hand, but has some limitations. The analysis depends on comparing establishments that produce a similar set of goods and services. If there is heterogeneity in what establishments produce within an industry and establishment size class cell, then controlling for industry and size class composition may be insufficient to form comparable sets of establishments. Since the industry codes I use are unusually

[18] Industry is measured with 4-digit NAICS code.

[19] In employment weighted regressions, $\bar{\sigma}_{M,k,i}$ and $\bar{\sigma}_{M,k}$ are employment weighted means, where the weights are the product of establishment employment and the survey weight. Otherwise, only the survey weights are used.

detailed, within industry heterogeneity is likely to be small. In the OES survey, most industries are measured at the 4-digit NAICS level of detail, with a few industries measured at the 5-digit level of detail. The OES survey makes an effort to use 5-digit NAICS codes where there are major differences within 4-digit NAICS codes. As an example, movie theaters are given a different industry code from all other portions of the Motion Picture and Video Industries group, which primarily consists of motion picture and video production and distribution industries.

The OES survey treats part-time and full-time workers as equivalent. If establishments in small and large MSAs differ in their utilization of part-time workers, this would make their occupational distribution of employees different even if the occupational distribution of hours worked is the same. Using the 2000 Census Public Use Micro Sample (PUMS), I find that, within occupation, workers in small MSAs work about 2 percent fewer hours per week than those in large MSAs. The difference is larger for occupations with an average hourly wage below $7.00, falling to a difference of –7.0 percent.[20] This suggests that part-time work is only a factor in low wage occupations and that part-time work is more common in small MSAs.

Establishments with allocated data are kept in the sample when creating $p_{j,S}$ and $p_{j,L|S}$. By keeping them in the sample, the industry and establishment size distributions are made more representative than they would be without the establishments with allocated data. While imputations of occupation-level employment are done within state-industry-establishment size cell when possible, data from establishments in small MSAs may be used to impute for establishments in large MSAs, or vice versa.[21] Finally, this method can only pick up indirect effects of differences in the utilization of more or less skilled workers within an occupation. As an example, a low-fee general attorney contributes as much to the skill measure as a highly specialized merger and acquisitions

[20] These results are consistent with Rosenthal and Strange's work on hours and agglomeration (forthcoming).
[21] When possible, wage information is imputed within substate area-industry-establishment size class.

attorney because there is only one lawyer occupation. It is likely that this limitation biases the estimates of differences in skill intensity toward finding no differences. As with industry, the occupation codes I use are unusually detailed so within-occupation skill heterogeneity is less of an issue than it would be with most data sources.

5. Results

5.1 Occupation-level results

It is informative to look at the occupation distributions in different MSA size types at the major occupational group level of aggregation. Figure 2 shows the percent of employment in each of the major occupational groups for small and large MSAs, as well as for large MSAs adjusted to have the same industry and establishment size distribution as the small MSAs. For the remainder of the paper, the true figures for the large MSA sample will be called unadjusted and figures from the large MSA sample that has been weighted to be comparable to the small MSA sample will be called comparable. The occupation groups in Figure 2 are sorted such that those with the highest wages in medium MSAs are at the top of the graph and those with the lowest are at the bottom.

The most noticeable feature of Figure 2 is that the occupational distributions are fairly similar across MSA size types. This is especially true when comparing the occupational distribution of the small MSA sample to the comparable large MSA sample, indicating that industry and establishment size distribution has an effect on occupational mix. Looking at these comparable groups of establishments, the greatest percent difference between them is in Legal occupations, where employment is 14 percent lower in the small MSA sample than in the comparable large MSA sample.[22] The next largest difference is in the Computer and Mathematical occupation group (11

[22] Throughout the paper, percentage point differences are calculated as $\ln(X/Y)$. This method has the advantage that the percent differences are not affected by which group is used as the comparison group. The percent difference between X and Y is the negative of the percent difference in Y and X.

percent lower in small MSAs). Ignoring differences across MSA types in industry and establishment size distributions would make the differences in occupational distribution starker. For example, the percent difference between unadjusted large MSA and small MSA samples are 44 and 61 percent for Legal and Computer and Mathematical occupation groups, respectively.

Taken as a whole, Figure 2 shows that establishments in small MSAs use a less skilled mix of employees. For high wage occupations, the percent of employment in an occupation group is generally higher in the conditional large MSA sample than in the small MSA sample. The chief exception is Healthcare Practitioners and Technical occupations, where employment is three percent higher in small MSAs. All five of the lowest wage occupation groups are more prevalent in the small MSA sample than in the comparable large MSA sample.

Studying occupation groups rather than specific occupations can mask differences in skill intensity because of heterogeneity in skill level within occupational group. Next I use detailed occupational employment to look at differences in occupational mix across MSA size types. Table 3 provides medium wages, $p_{j,S}$, $p_{j,L}$, and $p_{j,L|S}$ for a few occupations. The occupational categories I use are quite detailed. For example, Credit Analysts and Management Analysts are separate occupations. The table is sorted by medium wage and the ordering of occupations is consistent with prior expectations about which occupations are most skilled. For example, Auto Mechanics and Machinists have the same medium wage.

For the remainder of the paper, "skill" will be used interchangeably with hourly wage in medium MSAs. Figure 3 is a graph of the predicted percentage point difference in occupations share of employment between small and large MSAs when estimated as a cubic function of skill (R1 above). In Figure 3, a positive (negative) difference indicates that occupations with that range of skill have a larger (smaller) share of employment in the small MSA sample. If the skill mix were the same across MSA types, then the line would be zero for every skill level. Note that very different

occupations can be adjacent to one another because the order in which occupations are shown is determined entirely by skill.

The solid line on Figure 3 shows the difference in occupational mix between the small MSA sample and the comparable large MSA sample. Establishments in small MSAs use a less skill intensive mix of employees than comparable establishments in large MSAs. Occupations with a medium wage below $20 have a slightly higher share of employment in small MSAs, although the predicted difference is always less than a five percent. Occupations with a medium hourly wage above $20 have a noticeably smaller share of small MSA employment, with a difference that ranges as large as approximately –20.0 percent. For most ranges of skill, the differences across MSA types are significantly different from zero. The dashed line shows the differences in skill mix between the small MSA sample and the unadjusted large MSA sample. Given the large differences in industry and establishment size composition, it is not surprising that establishments in the unadjusted large MSA sample use a much more skill intensive mix of workers than do the establishments the small MSA sample.

Next I turn to the question of whether differences in skill intensity across MSA types are concentrated in a few industrial sectors. Many of the theories of why human capital would be more productive in large MSAs suggest that the productivity gains should be strongest for establishments that use skill intensive production processes. According to these theories, human capital intensive industries would have larger differences in skill mix than other industries. To see if this is true, I repeat the process of creating occupation mix data by sector and estimate a regression that allows differences in skill intensity to vary by sector (R2 above). The results from this regression are in Table 5, but I find it more informative to look at the results graphically.[23]

[23] The summary statistics for the occupation-by-sector sample are in Table 4.

Figures 4a through 4c provide graphs of the predicted percent difference in occupations share of employment as a function of skill.[24] These graphs focus only on the difference between the small MSA sample and the comparable large MSA sample, meaning that the within-sector industry and establishment size distribution of employment is the same for the small and large MSA establishment samples. The graphs are grouped such that sectors with similar patterns are close together. The sectors in Figure 4a – Information; Finance and Insurance; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Manufacturing; and Administrative et al. Services – all have large and statistically significant differences in skill intensity across MSA types. For example, manufacturers in small MSAs are predicted to use approximately 10 percent fewer workers in occupations with a mean hourly wage in medium MSAs of 30 dollars than manufacturers in large MSAs do. The results also suggest that small MSA manufacturers use both fewer low skill and fewer high skill workers than do comparable large MSA manufacturers, though the differences at low skill levels are not significantly different from zero.

The sectors in Figure 4b – Retail Trade; Real Estate and Rental Leasing; Health Care and Social Assistance; Arts, Entertainment, and Recreation Services; Accommodation and Food Services; and Other Services – have very similar skill mixes in small and large MSAs. In fact, Health Care and Social Assistance is the only one of these sectors for which it is not possible to reject the hypothesis that the skill mix is the same in both MSA types. For this sector, absolute value of the predicted percent difference in occupation's share of employment is always less than 5 percent. It is interesting to note that the sectors that have similar skill mixes in large and small MSAs produce goods and services that are largely not traded across MSAs.

The last set of sector graphs, Figure 4c, is a mixed bag. All of the sectors in the top row – Construction; Educational Services; and Wholesale Trade – have statistically significant differences

[24] Occupations with mean hourly wage in medium MSAs above $60.00 are excluded from the graphs to make the graphs easier to read. Fewer than ten occupations are excluded from the graphs.

in skill intensity across MSA types and use a less skilled mix of workers in small MSAs than in large MSAs. The Utilities sector appears to be more skill intensive in large MSAs, but the differences are not statistically significant. Mining and Transportation and Warehousing have inverted u-shaped curves, meaning that establishments in these sectors in small MSAs employ a smaller fraction of workers in either tail of the skill distribution than do establishments in these sectors in large MSAs. Since these two sectors are land intensive, it is not surprising that their skill mix patterns are different from those of the other sectors.

The regressions and graphs discussed above provide compelling evidence that, even when conditioning on detailed industry and establishment size, establishments in small MSAs use a less skilled mix of workers than establishments in large MSAs. The differences in skill intensity are small or not statistically significant for about half of the sectors of the economy. These include some very large sectors, such as Health Care and Social Assistance and Retail Trade.

5.2 Establishment-level results

Another way to approach the question of whether establishments in large MSAs adopt more skill intensive production techniques is to look at how the skill level at various percentiles of establishments' skill distributions varies across MSA types. I do this with establishment level regressions where the dependent variable is the skill level of the 10th, 25th, 50th, 75th, or 90th percentile of the establishment's skill distribution. These regressions can capture differences in skill intensity that would too small to see in Figure 4. This approach also makes it possible to test whether the skill gap between establishments in large and small MSA is larger in more skill intensive industries.

As detailed in Section 4, I calculate the kth percentile of the establishment's skill distribution for all establishments, using the log of medium wage to measure an occupation's skill. Industry-level skill intensity is measured as the average of the kth percentile skill for establishments in the industry that are located in medium MSAs; it is demeaned so that the average across industries is zero. By

interacting industry-level skill intensity with an indicator for being located in a small MSA, it is possible to test whether the difference in skill at the *k*th percentile between establishments in large and small MSAs varies with industry-level skill intensity. The sample for the establishment regressions is restricted to establishments located in small and large MSAs. All models include state, survey panel, and establishment size fixed effects.[25] All of the estimates discussed below are significantly different from zero at at least the 5 percent level of significance.[26]

The establishment-level regressions are reported in Table 7.[27] The set of regressions that is most comparable to the occupation-level results (panel a) use 4-digit NAICS industry fixed effects and an indicator for being in a small MSA. Looking first at the employment weighted results, the 10th skill percentile is 1.1 percent lower in small MSAs and the gap increases monotonically so that the 90th skill percentile in small MSAs is 2.4 percent lower than in large MSAs. These results confirm the occupation-level results in Figure 3: when conditioning on industry and establishment size, establishments in small MSAs use a less skilled mix of workers than do establishments in large MSAs and the differences are greater higher in the skill distribution. The results from equivalent regressions that are not employment weighted show that the 10th percentile skill level in small MSAs is 1.9 percent lower than that of large MSAs; the comparable figure for the 90th percentile is 1.7 percent.

The next set of regressions (panel b) adds an interaction between the small MSA indicator and the industry-level skill measure. The coefficients for the small MSA indicator from the employment-weighted regressions are essentially unchanged by adding this interaction, which is not surprising since the industry skill measures are demeaned. At all skill percentiles, the coefficients on

[25] I have also estimated all of the models without state or establishment size fixed effects and the results are nearly the same.
[26] The standard errors for all establishment-level regressions are clustered by survey strata: state, MSA, 4/5-digit NAICS code, and establishment size.
[27] The summary statistics for the establishment-level regression sample are in Table 6.

the interaction of industry-level skill and the small MSA indicator are negative and statistically significant, showing that differences in skill intensity between large and small MSAs are greater in skill intensive industries than in less skill intensive industries. For example, the predicted skill gap at the 10th percentile from the employment-weighted regressions is –0.5 percent for Retail Trade establishments and –1.5 percent for Finance and Insurance establishments.[28] At the 90th percentile, the corresponding gaps are –2.2 percent and –3.8 percent, respectively. The estimates are similar from the regressions that are not employment weighted, though as before the small MSA coefficient is more similar across skill percentiles than in the employment weighted regressions.

As a robustness check, the last set of regressions replace the industry fixed effects with the industry average of the skill percentile (panel c). The results are robust to using this alternate method to control for industry effects. The resulting coefficients for the small MSA indicator and the interaction between industry level skill and the small MSA indicator are always within one standard error of the estimates from the model that includes industry fixed effects. Other robustness checks include redefining small MSAs as being MSAs with population of less than 500,000 rather than 1,000,000. The estimated skill gaps grow, but by less than one standard error. I also estimate the establishment regressions with a sample that excludes establishments with allocated employment data. Again, the results are robust to this change and, as would be expected with random non-response, the estimated differences in skill intensity are larger.

6. Measuring skill with education

The results above use occupations' mean wages in medium MSAs to measure their skill levels. However, there are a number of reasons to think that wages are not determined solely by skill

[28] The average of φ_{10} is -0.2374 for Retail Trade establishments and 0.1045 for Fire and Insurance establishments. The corresponding figures for φ_{90} are –0.1850 and 0.2026 respectively.

– compensating wage differentials, efficiency wages, and monopsonistic competition, to name but a few. It is interesting to know whether the variation in skill intensity across MSA types is similar when using a different measure of skill. This section presents results using within-occupation education distribution as an alternative way to measure the skill level of an occupation. The results are consistent with the wage-based results: even when controlling for industry composition, establishments in small MSAs use a less skill intensive mix of workers than those in large MSAs.

I use the 2000 PUMS to calculate the share of medium MSA workers in each occupation whose highest completed degree is: less than high school, high school, some college, a Bachelor's, and more than a Bachelor's. The occupations' education distributions are then combined with the occupational employment distribution by MSA type from the OES used in the occupation-level regressions above. This gives the expected education distribution as a function of the occupational distribution of employment by MSA type. Part B of the Appendix gives more details on the construction of the expected education distributions by MSA type.

The variation in the expected education distributions by MSA type in Table 8 and Figure 5 is consistent with the occupation wage results. For all sectors pooled together, occupations that are common for people with less than a high school degree make up 0.4 percent points more of employment in small MSAs than in large MSAs, even after conditioning on the distribution of industry and establishment size. The corresponding figure for occupations that are common for people with more than a Bachelor's degree is -0.5 percentage points, a difference of -6.6 percent. The difference in the percent of employment in occupations common for people with the remaining education categories follows the same pattern: occupations that are held by people with more education make up a lower fraction of employment in establishment in small MSAs than in a comparable pool of establishments in large MSAs.

The education results by sector mimic the wage results by sector: the differences in skill intensity are highest in sectors that employ more skilled workers and/or produce traded goods. Table 8 also has the expected educational distribution of workers in Manufacturing, Retail Trade, and Professional, Scientific, and Technical Services. As with the wage results, manufacturers in small MSAs use slightly more low skilled workers and many fewer high skill workers than a comparable group of manufacturers in large MSAs. The difference in the expected education distribution across comparable MSA types in Retail Trade is small, ranging from a high of 0.3 percent for some college to a low of -0.8 percent for less than high school. For Professional, Scientific, and Technical Services, the differences in the expected distribution of education are large and monotonically fall with education level. Relative to a comparable set of establishments in large MSAs, establishments from this sector located in small MSAs employ 12.2 percent more workers in occupations common for people with less than a high school degree and 13.5 percent less workers with more than a Bachelors degree.

While I do not have standard error estimates for the figures in Table 8, the differences in the expected education distribution correspond closely with the results that use occupations' mean wages as a measure of skill. Skill intensity is lower for establishments in small MSAs than for comparable establishments in large MSAs. The sectoral patterns are similar, as well. For the sectors examined, the percent difference in the expected percent of workers with a high school degree or less is of a similar magnitude as the percent difference in the percent of workers in occupations with hourly wages below 20 dollars. The percent difference of workers with a Bachelor's or higher is comparable to the difference in the percent of workers in occupations with hourly wages of at least 30 dollars.

7. Conclusions

Using data from a large, representative survey of establishments, I find that establishments in small MSAs use a less skilled mix of employees than do comparable establishments in large MSAs. The differences vary by industrial sector, with some sectors showing large differences and others using the same skill mix in both MSA types. However, there are no sectors that are significantly more skilled in small MSAs than in large MSAs. The sectors that have small or no differences in skill mix are less skill intensive and/or non-traded sectors, such as Retail Trade and Accommodation and Food Services. Using establishment regressions that control for detailed industry and establishment size, I show that the skill gap between establishments in small and large MSAs is larger in industries that are more skill intensive.

These results suggest that skilled labor is more productive in large MSAs than in small MSAs. The urban wage premium literature provides evidence that the relative wages of high skilled workers are on average higher in large MSAs than in small MSAs. If the relative productivity of high skilled workers is the same in large and small MSAs, then establishments in large MSAs should substitute away from high skilled workers. I find that they do exactly the opposite, indicating that the relative productivity of skilled work is higher in large MSAs than in small MSAs. The fact that the industries where the differences in skill intensity are largest are the most skilled industries provides further evidence that skilled labor is more productive in large MSAs than in small MSAs. These results are consistent with urbanization theories that suggest that human capital productivity is higher in large cities.

It is not possible to tell from the results what causes skilled labor to be more productive in large MSAs. It could be that large MSAs attract highly productive workers who would be equally productive in all MSAs. Firms in large MSAs may react to the availability of these highly productive workers by hiring more workers in the occupations in which they are concentrated. If the ability

differences are particularly strong for high skill workers, this could lead to the same pattern regarding skill utilization that would come from large MSAs having higher human capital productivity than small MSAs. While there is evidence that high ability workers are more likely to migrate (Borjas, Bronars, and Trejo 1992) and are drawn to large MSAs (Faberman 1998), it is not clear that they are drawn for reasons other than increasing net expected lifetime income. If, as found by Borjas, Bronars, and Trejo, high ability workers migrate because they are able to receive higher real wages, then ability differences are largely a by-product of the productivity differences that lead to the urban wage premium. The fact that the urban wage premium remains large even when conditioning on hard to observe measures of ability (Glaeser and Maré 2001; Gould 2007) also suggests that ability differences only partially explain differences in labor markets in large and small MSAs. Therefore, it is unlikely that differences in ability fully explain the differences in skill intensity.

An interesting extension of the work in this paper would be to look at how differences in skill intensity by MSA size type vary across traded and non-traded sectors. Differences in productivity across MSA size type may be most prevalent in the traded sector. This is because costs are higher in large MSAs than small MSAs and output prices are set by the national or international market in the traded sector and in the local market for non-traded goods. In the absence of productivity gains from locating in large MSAs, establishments in the traded sector would locate in small MSAs, which tend to have lower costs. The skill intensity results for Retail Trade, a non-traded good sector, and Manufacturing, a traded good sector, suggest that this may be important. Looking at Table 8, establishments in Manufacturing and Retail Trade in small MSAs have similar expected education distributions. However, manufacturers have large differences in skill intensity across MSA type while retailers do not.

The final conclusion is that there is great potential for other research using the OES micro-data. The occupational employment data in the OES data provides a way to measure production

processes at the establishment level, which is unusual outside of the manufacturing sector. This information can enable one to look at a number of research questions, such as quantifying the degree of heterogeneity in production processes, looking at how firm survival and expansion probabilities are related to skill intensity, and studying how the mix of occupations at establishments adjust to the skill mix of the population. The Bureau of Labor Statistics has a system that allows outside researchers access to micro-data files; more information on that program is available at http://www.bls.gov/bls/blsresda.htm.

Appendix

A. Construction of occupation shares

This appendix details the construction of occupation shares for comparable groups of establishments in large and small MSAs. MSA size type, indexed with *q*, is S for small MSAs and L for large MSAs. Q_l is an indicator variable that equals one if establishment *l* is located in MSA size type *q* and zero otherwise. D_{liz} equals one if establishment *l* is in industry *i* and employment size class *z* and equals zero otherwise. M_{iz} equals one if there is positive employment in industry *i* and employment size class *z* in both MSA size types; otherwise M_{iz} equals 0.

Employment in occupation *j* in establishment *l* is e_{jl} and is weighted using the survey weights to make the sample of establishments match the universe of establishments. Define the total employment in industry *i*, employment size class *z*, and MSA size type *q* as $E_{iz,q} = \sum_j \sum_l Q_l D_{liz} e_{jl}$. Total employment in MSA size type *q* is $N_q = \sum_i \sum_z M_{iz} E_{iz,q}$. Note that N_q excludes employment for industry and employment size class pairs that have zero employment in either MSA size type.

In order to generate comparable sets of establishments across MSA size types, establishments are weighted such that the industry and employment size class distribution of employment is the same for both sets of establishments. The MSA size type that the data is adjusted to match is indexed with *r* =*S,L*. The weight used to make the employment distribution of MSA size type *q* comparable to that of MSA size type *r* is:

$$\omega_{iz,q|r} = \frac{(E_{iz,r} / N_r)}{(E_{iz,q} / N_q)}.$$

The count of employment in occupation *j* in MSA size type *q* and weighted to be comparable to MSA size type *r* is $e_{j,q|r} = \sum_i \sum_z \omega_{iz,q|r} M_{iz} \sum_l Q_l D_{liz} e_{jl}$. When weighting such that the industry and

employment size class distribution of employment in MSA size type *q* is comparable to that of MSA size type *r*, the percent of MSA size type *q* employment that is in occupation *j* is

$p_{j,q|r} = 100 \frac{e_{j,q|r}}{\sum_j e_{j,q|r}}$. I simplify the notation for cases where *q*=*r* such that $p_{j,q|r} = p_{j,q}$. For brevity, I use the term *j*'s occupation share interchangeably with the percent of employment in occupation *j*. The method used to generate the occupation shares for a single industrial sector is the same as above, except that the sample is first restricted to only establishments in the sector of interest.

B. Construction of Education Variables

As an alternate measure of the skill level of an occupation, I use the 2000 Census 5% Public Use Micro Sample (PUMS) to calculate the education distribution for each occupation. The education measure I use is the percent of people in the 2000 PUMS within an occupation whose highest level of completed education is: less than high school, high school, some college, bachelor's degree, and more than a bachelor's degree. Only people working in medium MSA's are used to calculate occupations' education distributions so that the measures will be independent of within occupation variation in education across small and large MSAs. Observations are weighted by their survey weight and all workers are included in the sample, including the self-employed, government workers, and workers of all ages.

The PUMS uses a different occupation coding scheme than does the OES. The education measures are calculated for occupations defined with the PUMS occupation code OCCSOC5, which are the PUMS occupation codes most similar to the SOC codes used in the OES. I create a crosswalk between SOC codes and OCCSOC5 codes. Some similar SOC occupations are combined into fewer OCCSOC5 occupations and, based on the similarity in occupational titles, the occupations that are combined are similar to one another. Some SOC occupations are linked to

multiple OCCSOC5 occupations, usually a similar specific occupation and at least one other less detailed category. In these cases, I use the OCCSOC5 occupation that is the same as the SOC occupation to the highest digit. For example, the SOC occupation 19-3041 (Sociologists) is only given OCCSOC5 code 19-3041 (Sociologists) even though it is included in 19-30XX (Miscellaneous Social Scientists, Including Sociologists).

The process for calculating the expected education distributions by MSA type are analogous to the process used to get the wage distributions used in the occupation-level regressions. Define the share of workers working in medium MSAs in occupation *j* with schooling level *l* as $\bar{G}_{jl}$. This is combined with the occupation employment data by MSA size type from the OES to get the percent of workers in MSA type q with education level l when conditioning the industry and establishment size distribution of employment to match that of MSA type r is:

$$G_{l,q|r} = \sum_j p_{j,q|r} \bar{G}_{jl}$$

where $p_{j,q|r}$ is as defined in part A of this Appendix. The sector level statistics are calculated in the same way except that the sample used to calculate $p_{j,q|r}$ is restricted to establishments in the specific sector. $\bar{G}_{jl}$ varies only across occupations, not sectors.

References

Acemoglu, Daron. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *The Quarterly Journal of Economics*. Pages 779-804, v. 111(3). 1996.

_____. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics*. Pages 1055-1089, v. 113(4). 1996.

Beeson, Patricia E. and Randall W. Eberts. "Identifying Productivity and Amenity Effects in Interurban Wage Differentials." *The Review of Economics and Statistics*. Pages 443-452. v. 71(3). 1989.

Beaudry, Paul, Mark Doms, and Ethan Lewis. "Endogenous Skill Bias in Technology Adoption: City-Level Evidence from the IT Revolution." Federal Reserve Bank of San Francisco Working Paper. August 2006.

Blomquist, Glenn C., Mark C. Berger, and John P. Hoehn. "New Estimates of Quality of Life in Urban Areas." *The American Economic Review*. Pages 89-107, v. 78(1). 1988.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. "Self-selection and Internal Migration in the United States." NBER Working Paper No. 4002. February 1992.

Bureau of Labor Statistics. "Appendix B. Survey Method and Reliability Statement for the May 2003 Occupational Employment Statistics Survey." *Occupational Employment and Wages, May 2003*. Bureau of Labor Statistics Bulletin 2567. Pages 226-34. September 2004.

Cover, Benjamin. "A Comparison of Occupational Employment and Wages in Metropolitan Areas and Nonmetropolitan Areas." *Occumpational Employment and Wages, May 2004*. Bureau of Labor Statistics Bulletin 2575. Pages 21-4. September 2005.

Doms, Mark, Timothy Dunne, and Kenneth R. Troske. "Workers, Wages, and Technology." *The Quarterly Journal of Economics*. Pages 253-290, v.112(1). 1997.

Duranton, Gilles and Diego Puga. "Micro-foundations of urban agglomeration economies." Draft of chapter eventually published in *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-Francois Thisse. Dated June 11, 2003.

Faberman, R. Jason. "Wages, Migration, and Urban Amenities: Do Smarter People Move to Better Places?" Unpublished mimeograph. May 2000.

Glaeser, Edward L. and David C. Maré. "Cities and Skills." *Journal of Labor Economics*. Pages 316-342. v.19(2). 2001.

_____. "Cities and Skills." NBER Working Paper 4728. May 1994.

Gould, Eric D.. "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium." *The Review of Economic Studies*. Pages 477-506. Vol. 74, No. 2. April 2007.

Hajiha, Fatemeh and Laurie Salmon. "Employment Composition: Variation Across States and Metropolitan Areas." *Occumpational Employment and Wages, May 2004*. Bureau of Labor Statistics Bulletin 2575. Pages 25-59. September 2005.

Henderson, J. Vernon. "Marshall's Scale Economies." *Journal of Urban Economics*. Pages 1-28, v.53. 2003.

Kilcoyne, Patrick. "The Role of Occupational Composition in State Wage Differentials." *Occumpational Employment and Wages, May 2003*. Bureau of Labor Statistics Bulletin 2567. Pages 8-13. September 2004.

Kim, Bonggeun. "The Wage Gap between Metropolitan and Non-metropolitan Areas." Unpublished working paper dated April 30, 2003.

Krugman, Paul. *Geography and Trade*. Cambridge, MA: MIT Press, 1991.

Moretti, Enrico. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *The American Economic Review*. Pages 656-690, v. 94(3). 2004.

Nestoriak, Nicole. *Labor Market Skill, Firms and Workers*. University of Maryland dissertation. 2004.

Rauch, James E. "Productivity Gains from Geographic Concentrations of Human Capital: Evidence from the Cities." *Journal of Urban Economics*. Pages 3-33, v. 34. 1993.

Roback, Jennifer. "Wages, Rents and Quality of Life." *Journal of Political Economy*. Pages 1257-1278, v. 90 (Dec.). 1982.

Rosenthal, Stuart S. and William C. Strange. "Agglomeration and Hours Worked." *Review of Economics and Statistics*. Forthcoming.

Rosenthal, Stuart S. and William C. Strange. "The Attenuation of Human Capital Spillovers: The Mahattan Skyline Approach." Unpublished working paper dated October 13, 2005.

Rosenthal, Stuart S. and William C. Strange, "Evidence on the Nature and Sources of Agglomeration Economies." Draft of chapter eventually published in *Handbook of Regional and Urban Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-Francois Thisse. Dated August 24, 2003.

Scotchmer, Suzanne. "Local public goods and clubs." In Alan J. Auerback and Martin Feldstein (eds.) *Handbook of Public Economics*, volume 4. Amsterdam: North-Holland. Pages 1997-2042. 2002.

Watson, Audrey. "Occupational Structure and Urban Growth." *Occumpational Employment and Wages, May 2004*. Bureau of Labor Statistics Bulletin 2575. Pages 10-20. September 2005.

Wheeler, Christopher. "Search, Sorting, and Urban Agglomeration." *Journal of Labor Economics*. Pages 879-99 v. 19(4). 2001.

Figure 1: Illustration of relationship between skill mix, relative wage, and skilled labor productivity

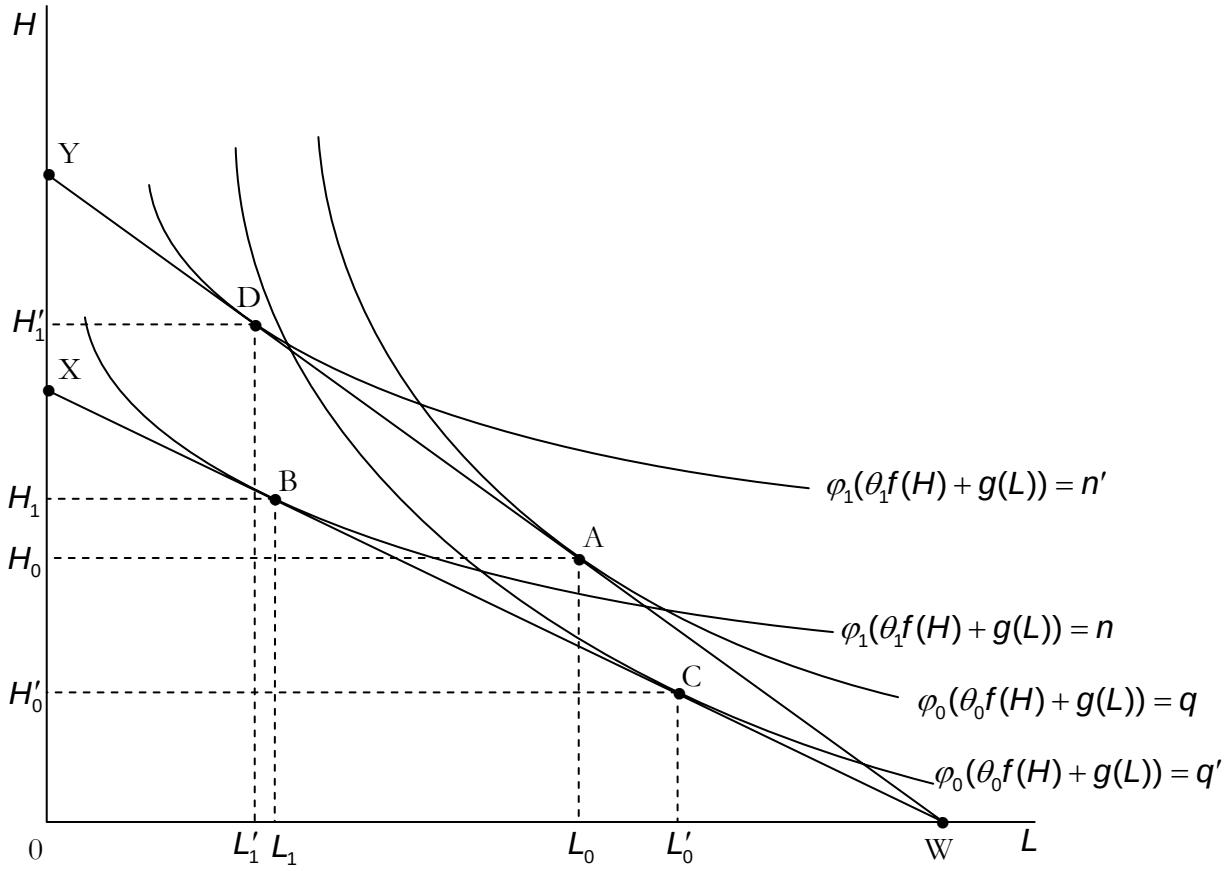
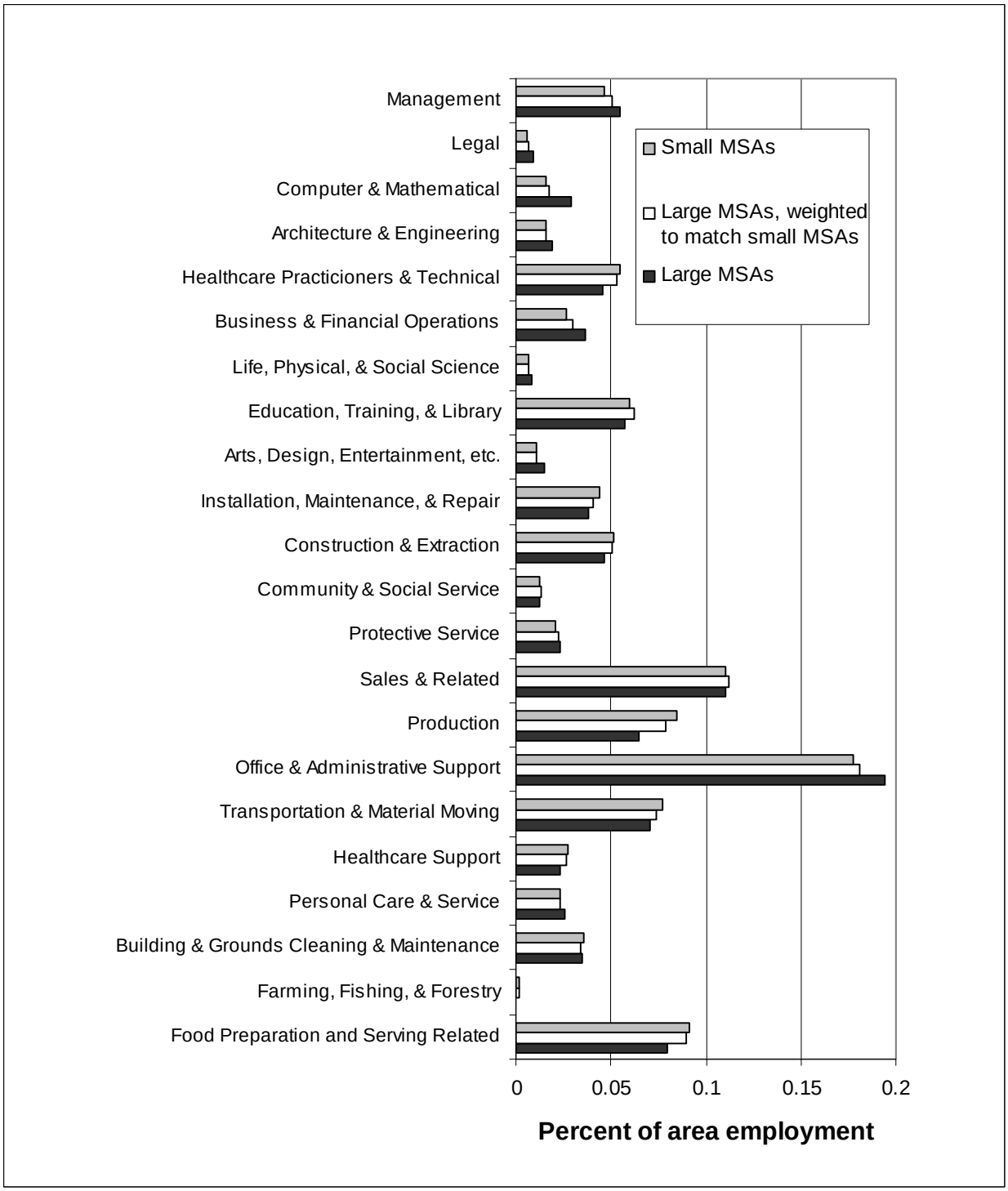
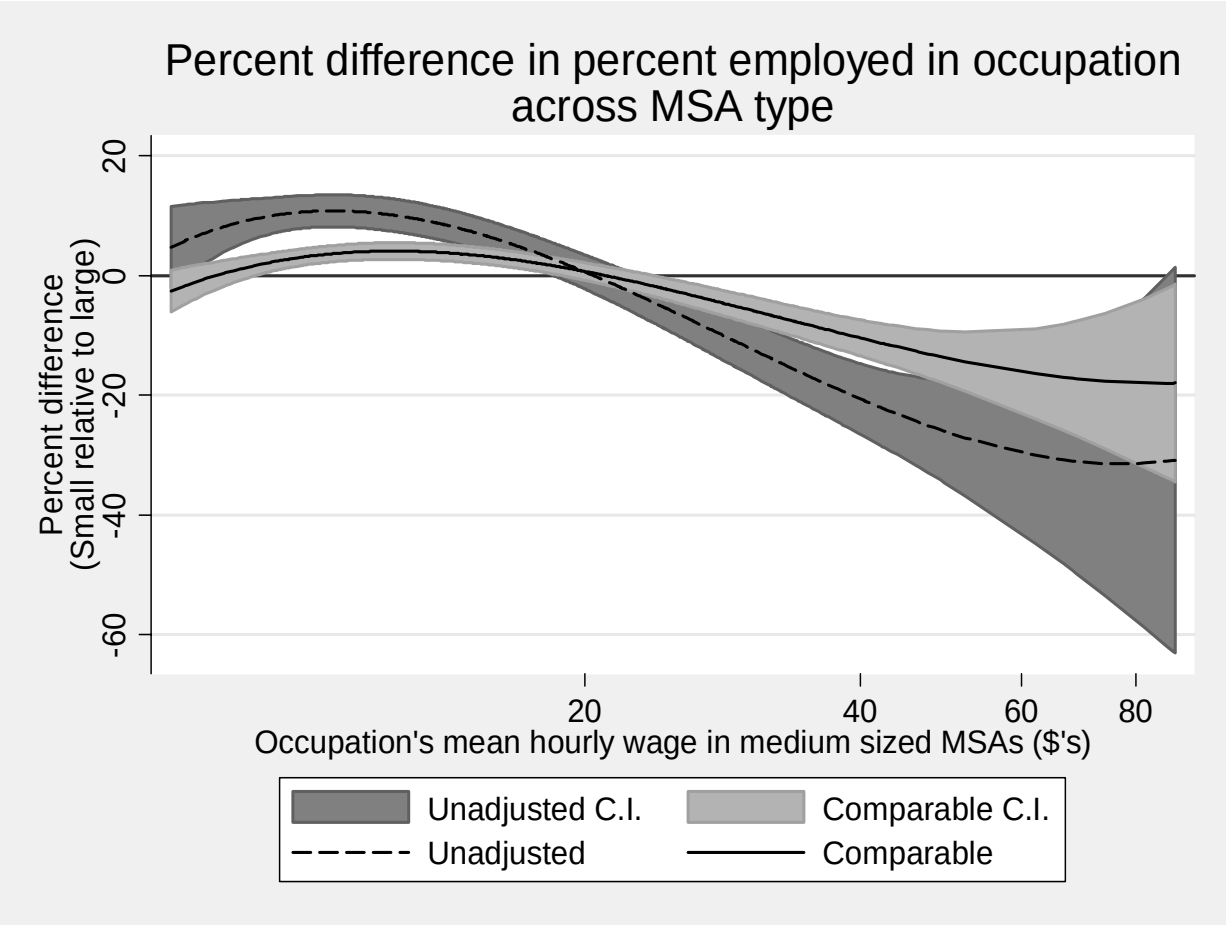


Figure 2: Occupation group employment share by MSA size category



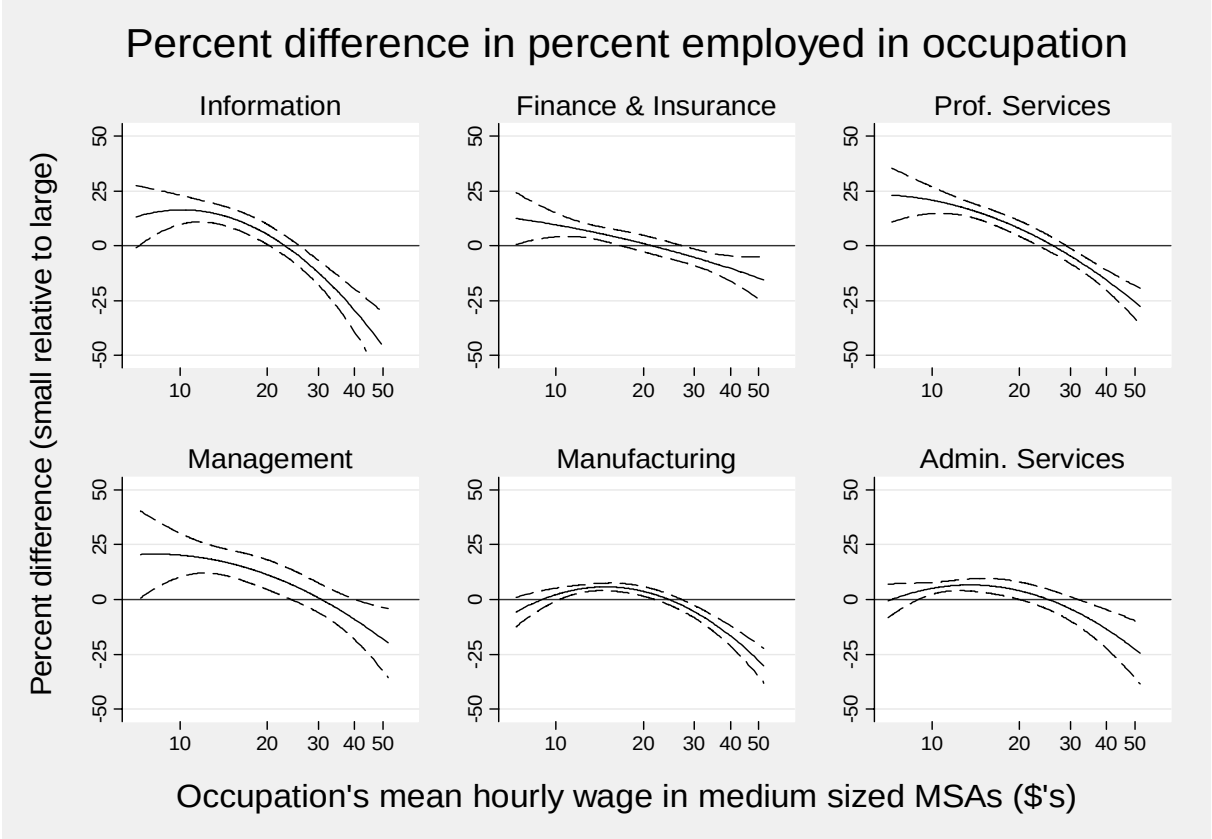
Source: Author's calculations from OES micro-data
Notes: Large MSAs are MSAs and CMSAs with population over 2 million in April 2000. Small MSAs are MSAs and CMSAs with population below 1 million in April 2000. The "Large MSAs, weighted to match small MSAs" is the sample of establishments of establishments in large MSAs weighted to have the same industry and establishment size composition as is in small MSAs. See paper for details.

Figure 3: Comparison of occupational mix in small and large MSAs (all sectors pooled)



Source: Author's calculations from OES micro-data
Notes: Shaded areas represent 95% confidence intervals around the predicted percent difference in percent of employment in occupation across MSA type. The models predict the percent difference as a cubic of occupations mean hourly wages in medium sized MSAs. See equation R1 in the paper. The "Comparable" figures adjust for differences in industry and establishment size composition while the unadjusted figures do not adjust for either industry or establishment size.

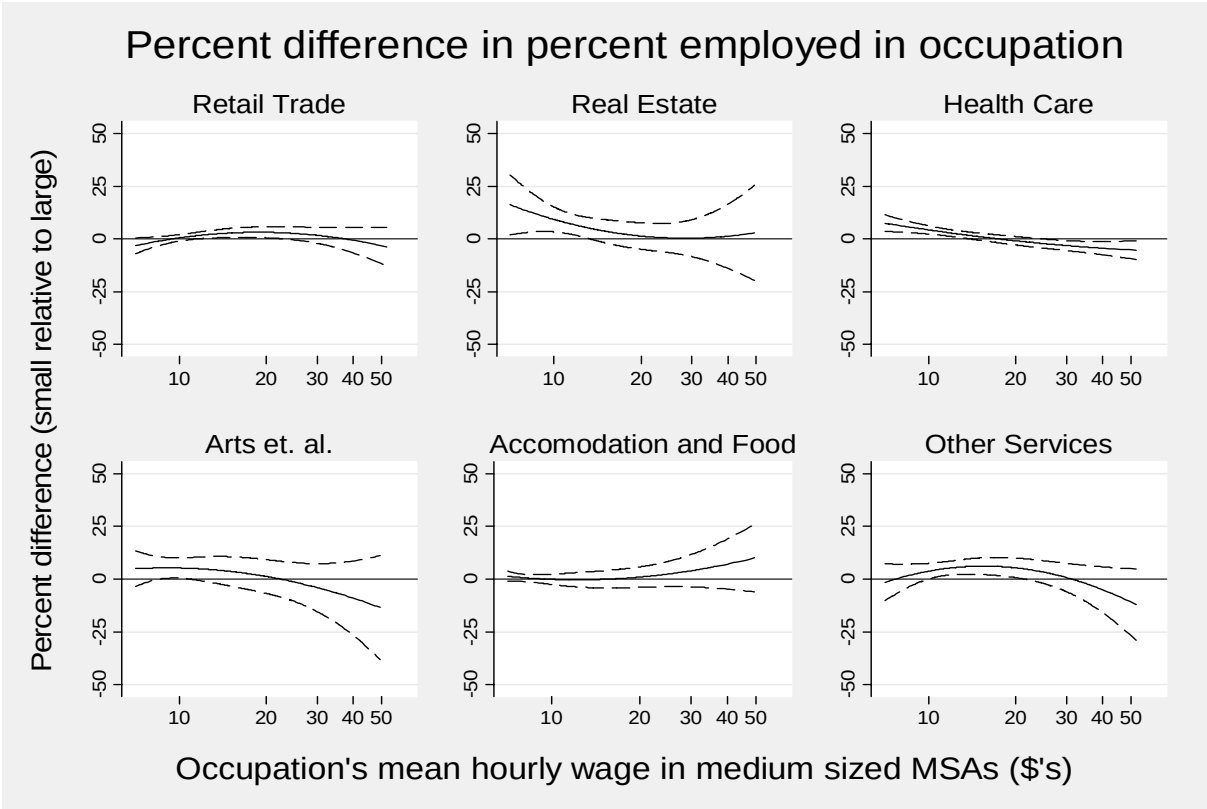
Figure 4a: Comparison of occupational mix in small and large MSAs by sector



Source: Author's calculations from OES micro-data

Notes: Comparison of predicted percent difference in occupation's share of employment by sector when using the comparable sample of large MSA establishments. The solid line is the predicted percent difference and the dashed lines are 95 percent confidence interval bounds from regression in Table 5. Occupations with a mean hourly wage in medium MSAs above $60.00 are excluded from the graphs but included in the regression sample.

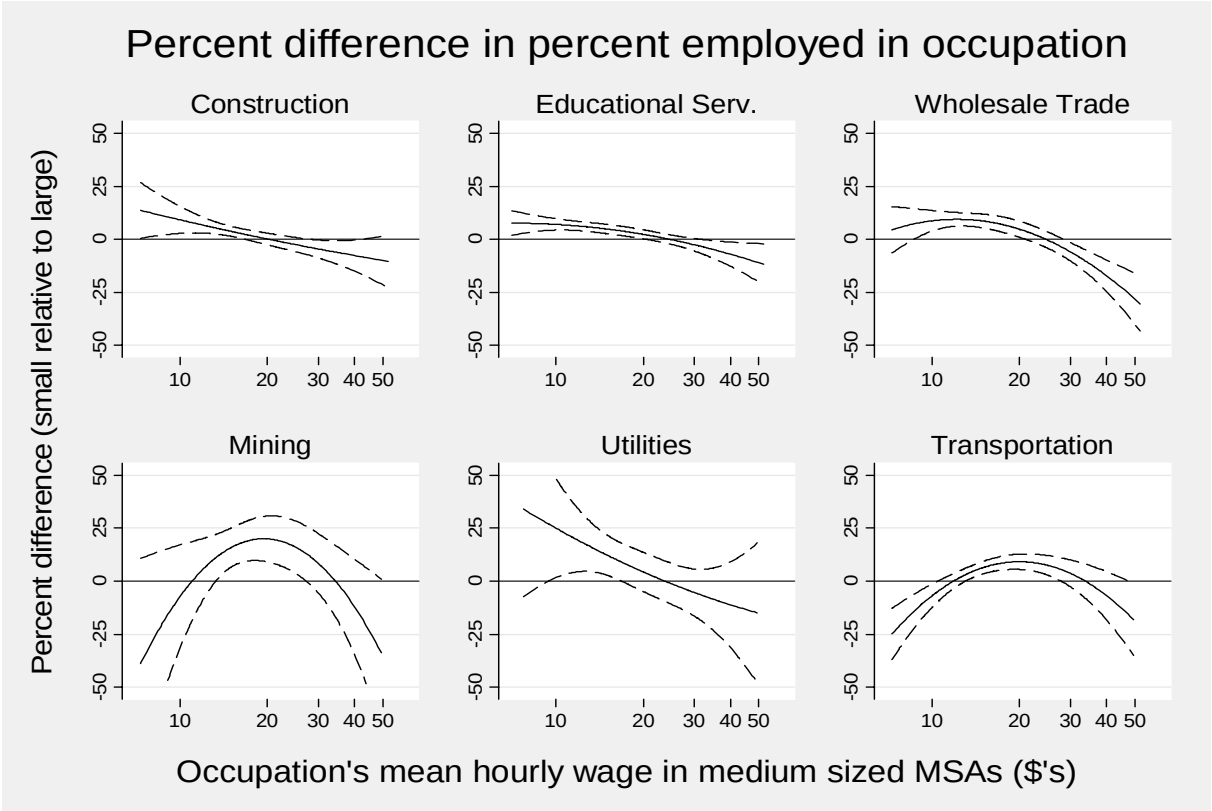
Figure 4b: Comparison of occupational mix in small and large MSAs by sector



Source: Author's calculations from OES micro-data

Notes: Comparison of predicted percent difference in occupation's share of employment by sector when using the comparable sample of large MSA establishments. The solid line is the predicted percent difference and the dashed lines are 95 percent confidence interval bounds from regression in Table 5. Occupations with a mean hourly wage in medium MSAs above $60.00 are excluded from the graphs but included in the regression sample.

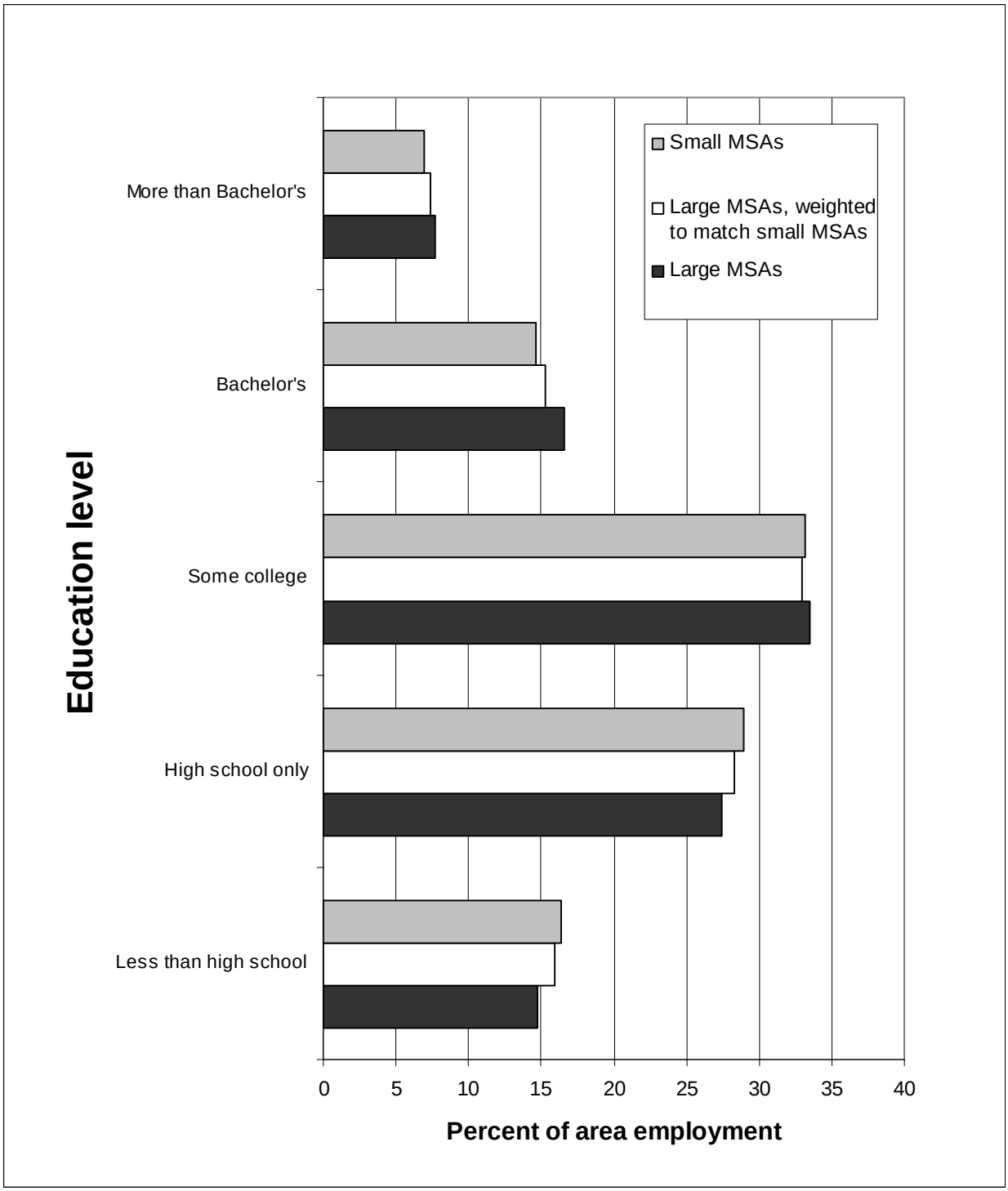
Figure 4c: Comparison of occupational mix in small and large MSAs by sector



Source: Author's calculations from OES micro-data

Notes: Comparison of predicted percent difference in occupation's share of employment by sector when using the comparable sample of large MSA establishments. The solid line is the predicted percent difference and the dashed lines are 95 percent confidence interval bounds from regression in Table 5. Occupations with a mean hourly wage in medium MSAs above $60.00 are excluded from the graphs but included in the regression sample.

Figure 5: Expected educational distribution of employment



Sources: Author's calculations from OES micro-data and 2000 5% PUMS
Notes: The expected education distribution of employment is the sum across occupations of the percent of workers in medium MSAs with the specified education level in an occupation multiplied by the percent of workers in the MSA type indicated in the occupation. See Part B of Appendix for more details.

Table 1: Classification of Metropolitan Statistical Areas into size groups

Name of Metropolitan Statistical Area	Population	Size rank	Size group
New York--Northern New Jersey--Long Island, NY--NJ--CT--PA CMSA	21,199,865	1	Large
Los Angeles--Riverside--Orange County, CA CMSA	16,373,645	2	Large
…			
Pittsburgh, PA MSA	2,358,695	21	Large
Portland--Salem, OR--WA CMSA	2,265,223	22	Large
Cincinnati--Hamilton, OH--KY--IN CMSA	1,979,202	23	Medium
Sacramento--Yolo, CA CMSA	1,796,857	24	Medium
…			
Oklahoma City, OK MSA	1,083,346	48	Medium
Louisville, KY--IN MSA	1,025,598	49	Medium
Richmond--Petersburg, VA MSA	996,512	50	Small
Greenville--Spartanburg--Anderson, SC MSA	962,441	51	Small
…			
Casper, WY MSA	66,533	275	Small
Enid, OK MSA	57,813	276	Small

Source: 2000 Census of Population and Housing

Table 2: Expository example of construction of skill intensity measures

Faux firms in NAICS 3327: Machine Shops; Turned Product; and Screw, Nut, and Bolt Manufacturing

	Number employed at		% of employment at		Difference in % of empl.	Hypothetical medium wage
	Joe's	Precision	Joe's	Precision	(3) - (4)	(in $'s)
	(1)	(2)	(3)	(4)	(5)	(6)
11-1021: General and Operations Managers	1	1	8.3	8.3	0.0	40.00
43-5071: Shipping, Receiving, and Traffic Clerks	2	2	16.7	16.7	0.0	11.25
43-6014: Secretaries, Except Legal and Medical, and Executive	1	1	8.3	8.3	0.0	12.25
51-4011: Computer-Controlled Machine Tool Operators, Metal and Plastic	0	7	0.0	58.3	-58.3	15.65
51-4012: Numerical Tool and Process Control Programmers	0	1	0.0	8.3	-8.3	19.75
51-4031: Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic	3	0	25.0	0.0	25.0	13.00
51-4034: Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic	4	0	33.3	0.0	33.3	15.60
51-4041: Machinists	1	0	8.3	0.0	8.3	17.00
Total employment	12	12				
Slope of difference in % on log medium wage						-0.002

	Joe's	Precision	Difference	% difference
25th percentile of establishment skill distribution	12.63	13.95	-1.32	-9.94
75th percentile of establishment skill distribution	15.60	15.65	-0.05	-0.32

Table 3: Occupational employment by MSA type for selected occupations

Occupation	Mean wage in medium sized MSAs (1)	Percent of employment in: Small MSAs (2)	Large MSAs before conditioning (3)	Large MSAs after condtioning (4)	Percent difference: (2) vs. (3) (5)	(2) vs. (4) (6)
Cashiers	8.10	3.04	2.66	3.06	13.6	-0.7
Construction Laborers	12.82	0.71	0.70	0.75	1.4	-5.9
Automotive Service Technicians and Mechanics	16.65	0.62	0.54	0.64	13.9	-2.8
Machinists	16.65	0.30	0.27	0.32	10.7	-5.2
Medical and Public Health Social Workers	19.27	0.09	0.09	0.10	4.4	-7.7
Electrical and Electronic Engineering Technicians	20.93	0.14	0.17	0.13	-16.8	11.3
Public Relations Specialists	22.56	0.12	0.15	0.12	-22.5	0.5
Credit Analysts	26.37	0.04	0.08	0.05	-64.6	-26.7
Architects, Except Landscape and Naval	28.61	0.06	0.10	0.07	-44.8	-3.7
Management Analysts	32.75	0.23	0.45	0.29	-65.1	-22.2
Economists	34.05	0.01	0.01	0.01	-35.3	-11.7
Electrical Engineers	34.47	0.12	0.15	0.11	-25.7	3.9
Lawyers	49.42	0.30	0.56	0.39	-62.1	-25.6
Internists, General	81.89	0.04	0.05	0.06	-13.2	-34.8

Source: Authors calculations from OES micro-data.

Table 4: Occupation-by-sector summary statistics

Variable	Mean (1)	Standard deviation (2)
Percent difference in occupation's share of employment in small and large MSAs (Comparable)	0.025	(0.238)
Percent difference in occupation's share of employment in small and large MSAs (Unadjusted)	0.047	(0.317)
Log of occupation's average wage in medium MSAs	-0.580	(0.053)
(Log of occupation's average wage in medium MSAs) squared	1.321	(0.217)
(Log of occupation's average wage in medium MSAs) cubed	-0.194	(0.012)
Industrial sectors' shares of small MSA employment		
Mining	0.004	(0.064)
Utilities	0.005	(0.070)
Construction	0.062	(0.241)
Manufacturing	0.129	(0.335)
Wholesale Trade	0.041	(0.198)
Retail trade	0.143	(0.350)
Transportation and Warehousing	0.040	(0.195)
Information	0.021	(0.145)
Finance and Insurance	0.043	(0.203)
Real Estate and Rental and Leasing	0.015	(0.123)
Professional, Scientific, and Technical Services	0.042	(0.201)
Management of Companies and Enterprises	0.012	(0.107)
Admin. & Support and Waste Management & Remediation Serv.	0.059	(0.236)
Educational Services	0.101	(0.302)
Health Care and Social Assistance	0.139	(0.346)
Arts, Entertainment, and Recreation	0.014	(0.118)
Accomodation and Food Services	0.098	(0.298)
Other Services	0.032	(0.175)
Number of occupation-by-sector observations	7,642	
Total employment in small MSA sample	25,581,574	

Source: Author's calculations from the OES micro-data.
Notes: Each observation is weighted by employment in the occupation-sector cell in small MSAs.

Table 5: Percent difference in occupational employment regression

Sector	Log of medium MSA wage (1)	(Log of med. MSA wage) squared (2)	Intercept (3)	Number of occupations (4)	Can reject that (1)=(2)=0? (5)
Mining	3.661	-0.617	-5.229	316	Yes
	(1.417)	(0.231)	(2.148)		
Utilities	-0.580	0.053	1.302	361	Yes
	(1.321)	(0.217)	(1.994)		
Construction	-0.194	0.012	0.475	465	Yes
	(0.391)	(0.065)	(0.586)		
Manufacturing	1.226	-0.227	-1.595	602	Yes
	(0.234)	(0.039)	(0.342)		
Wholesale Trade	0.940	-0.188	-1.080	569	Yes
	(0.397)	(0.067)	(0.575)		
Retail trade	0.390	-0.066	-0.539	540	No
	(0.179)	(0.032)	(0.245)		
Transportation and Warehousing	1.997	-0.333	-2.900	525	Yes
	(0.437)	(0.076)	(0.628)		
Information	1.128	-0.244	-1.143	417	Yes
	(0.458)	(0.078)	(0.662)		
Finance and Insurance	0.038	-0.030	0.168	422	Yes
	(0.386)	(0.064)	(0.571)		
Real Estate and Rental and Leasing	-0.545	0.081	0.916	495	No
	(0.572)	(0.100)	(0.800)		
Professional, Scientific, and Technical	0.406	-0.112	-0.134	612	Yes
Services	(0.348)	(0.056)	(0.527)		
Management of Companies and	0.500	-0.119	-0.320	603	Yes
Enterprises	(0.605)	(0.099)	(0.904)		
Admin. & Support and Waste	0.894	-0.172	-1.098	639	Yes
Management & Remediation Serv.	(0.324)	(0.057)	(0.454)		
Educational Services	0.191	-0.049	-0.111	620	Yes
	(0.228)	(0.040)	(0.318)		
Health Care and Social Assistance	-0.164	0.017	0.331	571	Yes
	(0.123)	(0.021)	(0.180)		
Arts, Entertainment, and Recreation	0.287	-0.065	-0.263	469	No
	(0.471)	(0.086)	(0.627)		
Accomodation and Food Services	-0.267	0.053	0.331	392	No
	(0.241)	(0.046)	(0.304)		
Other Services	0.683	-0.125	-0.874	604	No
	(0.362)	(0.064)	(0.504)		

Source: Author's calculations from the OES micro-data.

Notes: The dependent variable is percent difference in occupation's share of Small MSA employment and occupation's share of Large MSA employment when Large MSAs data weighted to match the industry and size class distribution of Small MSAs. The dependent variable is scaled so one percent equals 0.01. There are 7,642 occupation-by-sector observations and the R-squared of the regression is 0.071. Each observation is weighted by employment in the occupation-sector cell in small MSAs. The model is fully interacted with sector.

Table 6: Establishment level summary statistics

	Survey weighted MSA Type		Employment weighted MSA Type	
Variable	Small	Large	Small	Large
10th percentile skill level	2.506	2.552	2.364	2.402
	(0.364)	(0.401)	(0.256)	(0.282)
25th percentile skill level	2.553	2.598	2.459	2.510
	(0.372)	(0.409)	(0.295)	(0.334)
50th percentile skill level	2.665	2.713	2.642	2.700
	(0.403)	(0.439)	(0.369)	(0.407)
75th percentile skill level	2.823	2.865	2.834	2.896
	(0.461)	(0.488)	(0.424)	(0.453)
90th percentile skill level	2.983	3.002	3.076	3.127
	(0.503)	(0.521)	(0.447)	(0.468)
Demeaned industry average:				
10th percentile skill level	-0.020	0.002	-0.007	0.015
	(0.223)	(0.231)	(0.179)	(0.182)
25th percentile skill level	-0.021	0.001	-0.007	0.019
	(0.237)	(0.244)	(0.207)	(0.211)
50th percentile skill level	-0.022	0.002	-0.002	0.025
	(0.265)	(0.272)	(0.269)	(0.267)
75th percentile skill level	-0.021	0.003	-0.001	0.028
	(0.272)	(0.278)	(0.295)	(0.296)
90th percentile skill level	-0.015	0.002	0.000	0.024
	(0.240)	(0.244)	(0.266)	(0.269)
Share in sector:				
Mining	0.004	0.001	0.004	0.002
Utilities	0.002	0.001	0.005	0.004
Construction	0.102	0.081	0.060	0.056
Manufacturing	0.051	0.049	0.131	0.104
Wholesale Trade	0.068	0.078	0.040	0.052
Retail trade	0.155	0.126	0.139	0.120
Transportation and Warehousing	0.029	0.022	0.039	0.042
Information	0.015	0.016	0.022	0.033
Finance and Insurance	0.062	0.055	0.043	0.057
Real Estate and Rental and Leasing	0.047	0.056	0.015	0.020
Professional, Scientific, and Technical Services	0.105	0.152	0.042	0.074
Management of Companies and Enterprises	0.005	0.004	0.012	0.018
Admin. & Support and Waste Management & Remediation Serv.	0.052	0.058	0.061	0.071
Educational Services	0.019	0.016	0.106	0.096
Health Care and Social Assistance	0.101	0.108	0.139	0.120
Arts, Entertainment, and Recreation	0.014	0.014	0.014	0.016
Accomodation and Food Services	0.083	0.080	0.095	0.081
Other Services	0.085	0.080	0.032	0.034
Number of employees at establishment:				
1 to 9	0.694	0.731	0.130	0.125
10 to 19	0.140	0.115	0.103	0.086
20 to 49	0.102	0.090	0.167	0.150
50 to 99	0.035	0.034	0.133	0.127
100 to 249	0.020	0.021	0.162	0.176
250 or more	0.008	0.008	0.305	0.335
Number of establishments	348,790	322,191	348,790	322,191
Sum of weights	1,456,387	2,976,889	26,717,964	54,039,629

Table 7: Establishment-level skill percentile regressions

	Employment weighted regressions					Survey weighted regressions				
	Percentile of within establishment skill distribution					Percentile of within establishment skill distribution				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a. With 4-digit industry fixed effects										
Small MSA	-0.011	-0.018	-0.021	-0.021	-0.025	-0.019	-0.020	-0.023	-0.020	-0.018
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
R-squared	0.56	0.57	0.62	0.62	0.53	0.44	0.46	0.50	0.46	0.39
b. With 4-digit industry fixed effects										
Small MSA	-0.011	-0.018	-0.021	-0.021	-0.025	-0.020	-0.021	-0.024	-0.020	-0.017
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Small*Industry average of skill percentile in medium MSAs	-0.040	-0.056	-0.038	-0.023	-0.018	-0.048	-0.043	-0.024	0.010	0.048
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.017)	(0.015)	(0.013)	(0.017)	(0.022)
R-squared	0.56	0.57	0.62	0.62	0.53	0.44	0.46	0.50	0.46	0.39
c. Without industry fixed effects										
Small MSA	-0.010	-0.018	-0.023	-0.024	-0.028	-0.017	-0.018	-0.022	-0.021	-0.021
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)
Small*Industry average of skill percentile in medium MSAs	-0.035	-0.055	-0.043	-0.032	-0.030	-0.043	-0.037	-0.020	0.011	0.043
	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)	(0.018)	(0.016)	(0.014)	(0.019)	(0.029)
Industry average of skill percentile in medium MSAs	0.979	1.006	1.006	0.999	1.005	0.881	0.919	0.948	0.965	0.985
	(0.006)	(0.007)	(0.007)	(0.007)	(0.009)	(0.016)	(0.014)	(0.012)	(0.017)	(0.027)
R-squared	0.55	0.56	0.62	0.61	0.52	0.42	0.44	0.49	0.44	0.37

Source: Author's calculations from OES micro-data

Notes: The estimation sample has 670,981 establishments. All models include state, survey panel, and establishment size fixed effects. Standard errors (in parentheses) are clustered by survey strata (State, MSA, 4/5-digit NAICS industry, and establishment size category). A "*k* th skill percentile" is the skill level of the occupation of the employee at the *k* th percentile of an establishment's skill distribution, where an occupations skill is measured as the log of the mean hourly wage paid to employees in that occupation at establishments located in medium sized MSAs. Establishments are weighted by the product of the survey weights and establishments employment for columns 1 to 5 and by the survey weights for columns 6 to 10. The "industry average" variables are calculated with the same sets of weights used in the regressions.

Table 8: Expected education distribution by MSA type for selected sectors

Education level	Percent of employment in			Percent difference	
	Small MSAs (1)	Large MSAs before conditioning (2)	Large MSAs after condtioning (3)	(1) vs. (2) (4)	(1) vs. (3) (5)
a. All sectors					
Less than high school	16.4	14.7	16.0	10.4	2.4
High school only	28.9	27.4	28.3	5.3	2.1
Some college	33.1	33.5	33.0	-1.2	0.4
Bachelor's	14.7	16.6	15.3	-12.3	-4.3
More than Bachelor's	6.9	7.7	7.4	-10.8	-6.6
b. Manufacturing					
Less than high school	19.2	17.1	18.5	12.0	3.7
High school only	37.6	34.7	36.7	8.0	2.6
Some college	29.9	30.7	30.1	-2.7	-0.6
Bachelor's	10.2	13.2	11.3	-25.5	-9.8
More than Bachelor's	3.1	4.3	3.5	-35.1	-12.6
c. Retail Trade					
Less than high school	20.5	20.3	20.6	1.0	-0.8
High school only	32.4	32.1	32.3	1.0	0.2
Some college	33.5	33.6	33.4	-0.3	0.3
Bachelor's	11.1	11.5	11.2	-3.2	-0.3
More than Bachelor's	2.5	2.6	2.5	-3.7	0.2
d. Professional, Scientific, and Technical Services					
Less than high school	4.5	4.0	4.0	13.5	12.2
High school only	18.7	16.8	17.3	10.6	7.9
Some college	37.0	34.9	35.4	5.8	4.5
Bachelor's	25.2	27.7	26.6	-9.5	-5.6
More than Bachelor's	14.6	16.6	16.7	-13.0	-13.5

Source: Authors calculations from OES micro-data and 2000 Census 5% PUMS.

Notes: Expected education distribution is the expected percent of workers in each education category as a function of the education distribution of occupation and the occupational distribution of MSA types. See Appendix for more detail on the construction of the education variables. Column (3) reweights the Large MSA data so that the industry-establishment size distribution of employment is the same in Small and Large MSAs. Column (2) uses the unadjusted Large MSA employment distribution.