Urban Interactions: Soft Skills vs. Specialization

Marigee Bacolod
Department of Economics
University of California - Irvine
3151 Social Science Plaza
Irvine, CA 92697-5100 USA
mbacolod@uci.edu
(949) 824-1990

Bernardo S. Blum
Rotman School of Management
105 St. George St.
University of Toronto
Toronto, ON M5S 3E6
Canada
bblum@rotman.utoronto.ca
(416) 946-5654

William C. Strange
Rotman School of Management
105 St. George St.
University of Toronto
Toronto, ON M5S 3E6
Canada
wstrange@rotman.utoronto.ca
(416) 978-1949

January 31, 2008

Version: December 10, 2008

*We thank Henry Overman, Jason Faberman, and three anonymous referees for helpful suggestions. We are grateful to the Marcel Desautels Centre for Integrative Thinking and the Social Sciences and Humanities Research Council of Canada for research support. We also thank Philippe Roy for his exceptional work as research assistant.

Urban Interactions: Soft Skills vs. Specialization

Abstract

This paper considers the role of soft skills in cities and industry clusters. It begins by specifying a model of agglomeration economies where soft skills allow agents to interact more productively. The model exposes two conflicting forces: agglomeration allows opportunities to interact, but it also produces thick, specialized markets, and this specialization can be a substitute for interaction. In order to empirically evaluate the soft skills – agglomeration relationship, the paper matches data on the interaction requirements of occupations from the Dictionary of Occupational Titles to Census data. The within-industry average level of soft skills is found to be higher in cities but not in industry clusters. Workers at the top of the skill distribution in large cities typically have higher levels of soft skills than in small cities, while the least skilled workers are less skilled in large cities than in small cities. This pattern is reversed for industry clusters.

I. Introduction

Like a long list of papers before it, this paper will deal with the role of interactions in cities. This paper's innovation will be the focus on "soft skills," which we define as the worker skills needed for effective and productive interpersonal interactions. There is growing recognition among economists that non-cognitive skills such as these are highly and increasingly important (Bowles et al (2001), Autor et al (2003), and Heckman et al (2006)).

In considering skills, we build on an old literature that has considered how the development and use of skills contributes to the process by which cities and industry clusters create value. For instance, Marshall (1890) considers knowledge spillovers and the matching of skills to needs, while Jacobs (1969) considers the application of skills to "new work." Both Marshall and Jacobs recognize the importance of the specific class of skills that allow valuable interactions to take place. In other words, although neither uses the term, they both implicitly consider soft skills. It is typical for the interest of Marshall or Jacobs in an aspect of agglomeration to have launched a thousand intellectual ships, but it did not in this case. Although there has been a great amount of very valuable research on interactions, there has not been either theoretical or empirical work that has focused primarily on soft skills.

The purpose of this paper is to specify a model of soft skills in a system of cities and to empirically evaluate the relationship between soft skills and agglomeration. The model begins with the assumption that soft skills facilitate interaction, and it explicitly considers how such skills are involved in the generation of agglomeration economies. Surprisingly, there is not a monotonic equilibrium relationship between agglomeration and soft skills. To the extent that spatial concentration presents workers with more opportunities to interact, soft skills will be more valuable where activity is concentrated. We refer to this as the *opportunity effect.* This is not the only effect at work, however. One of the advantages of agglomeration is that it makes markets thicker. This allows better matching of workers to jobs (see Duranton-Puga (2004)). In this case, market thickness can potentially be a substitute for workers' soft skills. This *specialization effect* operates in the opposite direction of the opportunity effect.

We will investigate the empirical relationship between soft skills and agglomeration using the Dictionary of Occupational Titles (DOT) and the 5% sample of the 2000 Census. The DOT characterizes a number of skills that are required to perform more than 12,000 occupations. In particular, it characterizes different types of interpersonal interactions required by occupations. If one assumes that workers are assigned to jobs in a hedonic market clearing process, then one can infer a worker's skills from the occupation in which the worker is employed. We use this hedonic imputation to characterize a worker's level of several sorts of soft skills. These include the skills that allow interaction outside of

authority relationships (captured by the DOT variable *depl*), the capacity to direct, control, and plan activities (*dcp*), and the ability influence others (*influ*). The set of soft skill measures also includes the DOT measure of *people* skills, a ranking which goes from receiving instructions up to mentoring. Finally, we aggregate all these measures through factor analysis into a soft skills index (*peoindx*). Taken as a group, these variables characterize several dimensions of a worker's interactiveness.

The DOT data are matched with Census data to allow us to characterize the relationship between worker skills and agglomeration. We consider the two standard sorts of agglomeration, the formation of large cities (urbanization) and industry clusters (localization). We then estimate a range of models of the relationship between urbanization and localization and the soft skill measures described above. The most basic models are ordinary least squares (OLS) estimates of the relationship of mean skills to local characteristics controlling for industry fixed effects. More rigorous models involve a two-stage procedure where the effect of agglomeration on skill levels at various percentiles of a city-industry's skill distribution are calculated using feasible generalized least squares (FGLS).

The results strongly confirm the existence of a non-monotonic and nuanced relationship between agglomeration and soft skills. The OLS estimates of regressions of mean skill levels show a higher average level of all of the soft skills in larger cities. These relationships are statistically significant, but they are moderate in magnitude. The two-stage regressions of the entire skill distribution show an even clearer pattern. For most of the skill measures, an increase in city population is associated with a higher skill level at the 75th and 90th percentiles of the skill distribution, but lower skill level at the 10th and 25th percentiles. The increase in soft skills found in large cities is thus of a very particular form: the skills at the top end of the distribution. The low skill workers in large cities are by our measures even less skilled than the low skill workers in small cities.

The results are quite different for industry clusters. The OLS estimates of regressions of mean skill levels show that the average skill level of an industry's workers is insignificantly or, in some cases, negatively related to industry concentration. The two-stage regressions of the entire skill distribution show that in industry clusters the skill levels at the 75th and 90th percentiles tend to be lower, while the levels at the 10th and often the 25th are higher. This is opposite the pattern we found in looking at cities where the tails of the skill distribution became thicker. In industry clusters, there appears to be a tendency for workers to have moderate levels of soft skills.

This analysis builds on several strands of research on the theory of agglomeration. First, it builds on models of spatial interactions (see Beckman (1976) or Helsley-Strange (2007)). These models allow for interaction and look at the implications for urban spatial structure. They do not consider soft skills directly. Second, it builds on models of the microfoundations of agglomeration economies, especially matching. See Fujita and Thisse (2002) and Duranton and Puga (2004) for excellent surveys. Regarding

matching, we build primarily on Helsley-Strange (1990, 1991, 2002). Third, the paper builds on models of systems of cities. See Henderson (1974) for a model with a black box production function and Helsley and Strange (1990) for a strategic model with an explicit matching microfoundation. Finally, and most importantly, we build on research that directly address the role of flexibility in the generation of urban increasing returns, including Vernon (1960), Duranton-Puga (2001), and Strange et al (2006). As will be seen, our approach is very much consistent with the Storper-Chrisopherson (1987) analysis of "flexible specialization."

The analysis also builds on a growing literature on skills in cities. Glaeser and Mare (2001) and Wheeler (2001, 2006) document the existence of an urban wage premium. Bacolod et al (2007) consider the degree to which the urban wage premium is enjoyed by workers with different sorts of skills. Fallick et al (2006) document a positive relationship between worker turnover and the concentration of the computer industry in the Silicon Valley. Elvery (2007) constructs a measure of worker skills based on an occupation's mean wage in medium sized cities. Elvery's primary finding is that establishment skill intensity increases with city size. Lin (2007) uses the changes in the DOT to identify "new work" as in Jacobs (1969). None of these papers directly consider soft skills. Scott (2007) and Scott and Mantegna (2007) do consider soft skills, among other things. The former finds a positive correlation between the DOT measure *people* (mentioned above) and city size, while the latter shows a positive relationship between various measures of an occupation's behavioral requirements and city size.

The rest of the paper is organized as follows. Section II sets out a theoretical analysis of agglomeration that captures urbanization, localization, and both the opportunity and specialization effects. Section III discusses the matching of DOT and Census data to characterize the levels of a worker's soft skills. Section IV presents the results of the estimation of the within-industry relationship between soft skills and agglomeration. Section V concludes.

II. A model of soft skills and agglomeration

A. Basic model

This section carries out a theoretical analysis of soft skills and agglomeration. It has two key pieces. The first is a model of a system of cities where differences in size arise from the inherent advantages of different locations. The second is a model giving a microfoundation for a relationship between soft skills and agglomeration, which is then incorporated into the system model. Throughout, an agent's soft skills are those that give the ability to interact.

We begin with a basic model of a system of cities. There is one type of worker, which we will call a "production worker." These workers are assumed not to have soft skills. We conceive of these workers as engaging in the sort of activities that do not require interaction. The workers choose between

two cities, indexed j = {1, 2}, with the total number of production workers N being divided into city populations n_1 and n_2. The cities differ in inherent productivity, θ_j. This can be seen as capturing both natural advantages and productivity enhancing historical accidents. The latter could include, for instance, the presence of unusually able local entrepreneurs whose businesses spin off entrepreneurs who are themselves unusually able (Sorenson and Audia (2000) and Klepper (2007) for evidence of this sort of spinoff process for particular industries. We will adopt the notation that cities are ranked in decreasing inherent productivity, $\theta_1 > \theta_2$.

We suppose that the marginal product of a worker is decreasing in a city's scale, which can be interpreted as either its population or as the employment in a given industry. In the latter case, N should be interpreted as an industry's total employment and n_i as the industry's employment in city i. For convenience, we assume that a worker's marginal product is given by $f(n,\theta)$, with $\partial f/\partial n < 0$ and $\partial^2 f/\partial n^2 < 0$.[1] We have included the city heterogeneity parameter θ to generate asymmetry in the equilibrium system. We therefore assume $\partial^2 f/\partial n \partial \theta > 0$. Finally, regarding the costs of agglomeration, we will suppose that each worker bears cost c(n), where c(-) is increasing and convex. We have suppressed rents here. One could conceive of c(-) as capturing the net costs of inhabiting a city of population n, where differential rents have been redistributed.[2]

In this situation, the utility of a production worker in city j with population n_j will be

$$u = f(n_j,\theta_j) - c(n_j). \tag{1}$$

An interior equilibrium in this model is a pair of city populations such that utility is equal in the two cities and all production workers are located somewhere. Substitution of the population adding up constraint $n_1 + n_2 = N$ into (1) gives the conditions that define an interior equilibrium:

$$[f(n_1,\theta_1) - f(N-n_1,\theta_2)] - [c(n_1) - c(N-n_1)] = 0, \tag{2a}$$

$$n_2 = N - n_1. \tag{2b}$$

The key characteristic of the equilibrium in (2a) and (2b) is the asymmetric allocation of production workers in equilibrium. It is straightforward to show that $n_1 > N/2$ when $\theta_1 > \theta_2$ and that

[1] We could have instead assumed that there were agglomeration economies so that f(-) increased in n up to some sufficiently low level of population and decreased thereafter. In such a situation, as with all systems of cities models, both cities will be on the downward sloping side of the production worker marginal product curve. We make the global assumption to allow us to focus more tightly on soft skills.

[2] There is no reason that the asymmetry must be present only in f(-); it could be in c(-) also or instead, or there could be a separate term capturing the differences between locations.

$dn_1/d\theta_1 > 0$. The better location will have more production workers, and the gap between the locations will grow with the quality of the better location, *ceteris paribus*. We believe that this captures the actual pattern of systems of cities and of industrial localization. Regarding the latter, the Silicon Valley is the heart of the computer software industry, but it is not the only location where the activity occurs. Hollywood is the heart of the film and television production industry, but movies and TV programs are created at other locations. We will now consider how workers with soft skills are attracted to core and peripheral locations in industries in a model with an explicit microfoundation of agglomeration.[3]

Finally, it is worth making a few observations regarding the model's interpretation. We want to have a simple model that generates predictions about both urbanization and localization. The urbanization interpretation, where n_j represents population, is clear. Similarly, if one were to suppose, following Henderson (1974), that the cities specialize, then the model is, without modification, also a model of the clustering of individual industries. The model would obviously become much more complicated if there were multiple industries co-agglomerating in cities. In this case, in order to characterize a city's attractiveness, one would need to consider both the positive agglomeration effects that spill across industry boundaries and inter-industry congestion. It would still be true that for any given industry, in equilibrium there would be more activity in the location that was more attractive for that industry. This asymmetry is all that is required to motivate our empirical exercises.

B. A simple model of interaction

There is a long tradition of research in urban economics that has focused on the role of cities as centers of interaction. While it is sensible to consider interactions that produce patentable processes and products, as many have, it is obvious that much interaction is of a more mundane sort. Vernon (1960) writes of high-fashion dressmakers who require an ever-changing range of buttons. Such producers benefit from being able to interact with local button producers. Jacobs' (1969) much repeated story of the invention of the brassiere also has interaction at its core. The inventor, Ida Rosenthal, was a dressmaker. Her interaction with customers led to her innovation. Interactions between agents have led to a range of innovations, as Jacobs writes:

> The process by which one sort of work leads to another must have happened millions of times in the whole history of human development...a cleaner of suede clothing is now starting to bottle and sell her cleaning fluid for people who want to clean their own suede; a chest and wardrobe manufacturer is starting, for a fee, to analyze what is wrong with one's household or office storage arrangements; a playground designer is starting to make and sell equipment for

[3] The generalization of (2a) and (2b) to an arbitrary number of populated cities or even potential city sites, J, is clear.

> playgrounds and nursery schools; a sculptor is starting a line of costume jewelry; a designer of theater costumes is launching himself as a couturier; a couturier is starting a boutique; an importer of Italian marble is starting to manufacture marble-top tables; a clothing store is starting classes in teen-age grooming and dieting. (Jacobs (1969), pp. 53-54)

We will present a model of adaptation that will emphasize the role of interaction and the soft skills that allow it. As in the above stories, adaptation nearly always involves working with other people. It thus requires the ability to interact (soft skills). We therefore introduce another sort of worker to the model, referred to as "analysts," a term that we intend to be taken quite broadly. Analysts are workers who have soft skills and so take on interactive tasks. This class of workers thus includes managers, entrepreneurs, and knowledge workers, among others. There are M of these workers, who will be divided between the two cities, with $M = m_1 + m_2$.

We assume that for analysts as for production workers, labor markets are competitive, and so analysts are paid their marginal products. We will suppose for simplicity that analysts' wages depend only on the ability to interact. Nothing would change were we to add to this a baseline of productivity in commonplace tasks. We also suppose that analysts must work as analysts rather than as production workers. Finally, we assume that all agents are risk neutral. Among other things, this implies that the uncertain elements of an analyst's marginal product are valued at their expectations.

In this model, we conceive of interactions as a kind of matching. Specifically, we suppose that analysts create value by responding to randomly arising opportunities to interact. The interaction may involve, among other things, two knowledge workers learning from each other, a manager contracting with an input supplier, or an entrepreneur responding to a customer. Suppose that each analyst has an address x on the unit circle that describes the activities that he or she is ideally suited to carry out. In interacting with others, the analyst may be called on to carry out other tasks. If an analyst carries out some other arbitrary activity y, there will be a loss in value that increases with the distance between the analyst's idiosyncratic abilities and the activity. We will suppose the loss to be linear, implying that the value of an analyst with skill x carrying out activity y is given by $a - b(s,z)*|x-y| = a - b(s,z)*d$. a is a positive parameter, b(s,z) is a function of the worker's soft skills (s) and other characteristics (z), and d has the natural interpretation of the distance in the space of activities.

In this model, b(s,z) captures an analyst's ability to performing activities that he or she is not ideally suited to carry out. An analyst with a lower value of b(s,z) suffers a lower adaptation cost when interacting with others. As soft skills allow valuable interactions, we assume that b(s,z) is monotonically

decreasing in s.[4] This interpretation fits well with managers, who must adapt the direction they give to other workers as the business environment changes. It fits also with entrepreneurs who must make choices of product positioning, production, and entry to best take advantage of fluctuating entrepreneurial opportunities. It fits as well with knowledge workers whose primary value-creating activity is in integrative thinking, the adaptation of a set of knowledge to new circumstances. The interpretation is also consistent with the stories of adaptation introduced above, both the adaptation carried out by Vernon's button makers and by Jacobs' synergistic creation of the brassiere.

The value accruing from this sort of activity depends on the arrival of opportunities to interact. We suppose that agglomeration increases an analyst's exposure to opportunities. Formally, we suppose that $\rho(n+m)$ represents the probability that an opportunity arises for a given analyst, where m is the number of analysts in the analyst's city, n is the number of production workers, and their sum is the degree of agglomeration. We further suppose that $\rho'(n+m) > 0$ and $\rho''(n+m) < 0$, implying that although agglomeration enhances opportunities for adaptation, it does so at a decreasing rate. It is worth reiterating that the arrival of opportunities might be related to the overall scale of population in the worker's city or instead to the scale of the analyst's industry.

C. Soft skills and agglomeration economies: the opportunity and specialization effects.

Under the above assumptions, an analyst's wage will equal the analyst's expected value arising from random opportunities. This depends on the expected adaptation distance associated with an opportunity. There are various ways that one might conceive of this. We will adopt a simple one that generates tractable closed form solutions. Suppose that analysts are evenly distributed on the unit circle. If there are m of them, then their addresses are $\{0,1/m, 2/m,...m-1,m\}$. Suppose that opportunities arise according to a uniform distribution on the unit circle. A given analyst with address x will be best-suited to respond to those opportunities whose address y is closer to x than to any other analyst. The expected distance will thus be the expectation of $|x-y|$ conditional on y being closer to x than to any other analyst's address. In this setup, the expected distance from a given random opportunity to the nearest analyst who will act on it will be 1/2m. The worker will be closest to 1/m of total opportunities. As with production workers, we suppose that the increasing and convex function $c(n+m)$ captures the costs of living in a concentrated environment. In these circumstances, analysts are specialized in the sense that they carry out a subset of all possible activities. As the market becomes thicker (m becomes bigger), the degree of specialization becomes greater in the sense that the analyst carries out a narrower set of activities.

---

[4] The ability to adapt to different activities may depend also on other characteristics. For example, cognitive skills may allow workers to better identify activities that he/she can perform. We explore the implications to some other skills in the empirical section.

We define an analysts' expected value from adaptation as ν:

$$\nu = \rho(n+m)/m * [a - b(s,z)*(1/2m)] - c(n+m). \quad (3)$$

This depends on the size of the market according to the probability of a worker having opportunities, $\rho(n+m)/m$, the value of a given interaction, $[a - b(s,z)*(1/2m)]$, and congestion costs, $c(n+m)$. An increase in the number of production workers creates more opportunities for adaptation, and this increases the value of analysts. The formal comparative static is

$$\partial\nu/\partial n = [(\rho'(n+m)/m]*[a - b(s,z)*(1/2m)] - c'(n+m) > 0. \quad (4)$$

The comparative statics for an increase in analysts are less straightforward and more interesting:

$$\partial\nu/\partial m = (\rho(n+m)/m) * [b(s,z)*(1/2m^2)] + [(\rho'(n+m)m - \rho(n+m))/m^2]*[a - b(s,z)*(1/2m)] - c'(n+m). \quad (5)$$

The congestion term is easy to understand and requires no further comment. The second term is parallel to the effect discussed above in the $\partial\nu/\partial n$ expression. The sign depends on the first term in brackets, equal to $(d/dm)(\rho(n+m)/m))$. If the flow of opportunities increases enough to outweigh the sharing of the opportunities more broadly --an effect absent in (4)-- then this term is positive.

The interesting and surprising result is found in the first term, which is unambiguously positive. A thicker market decreases the costs of a given random interaction because it means that the analyst is called on to perform a task for which he or she is better suited. This result is very much in the spirit of Storper and Christoperson's (1987) analysis of "flexible specialization." They write:

> In... flexibly specialized systems production is organized around the interactions of a network of small firms. These small firms specialize in batch or custom production of general classes of outputs... The production system as a whole is flexible because each production project can be organized with a different mix of specialized input-providing firms (p. 105).

Our model is a highly stylized representation of this, where an analyst's opportunity is met with a more inviting set of interaction partners in a thicker market. Formally, the first expression in (5) means that an increase in the number of other soft skill workers reduces the cost of interacting (captured by the distance $|x-y|$) because the market's thickness ensures that analysts are highly specialized in the sense that they are

closely matched to the tasks that they perform. This effect will be crucial below. The ambiguity of the second term and the opposite sign of the first term together mean that the overall effect is ambiguous.

To see how the value of agglomeration depends on the level of soft skills as measured by the parameter s, we obtain

$$\partial^2 \nu/\partial n \partial s = \partial b/\partial s\, [(\rho'(n+m)/m] * [-(1/2m)] > 0 \qquad (6)$$

and

$$\partial^2 \nu/\partial m \partial s = \partial b/\partial s(\rho(n+m)/m) * [(1/2m^2)] + [(\rho'(n+m)m - \rho(m))/m^2][-\partial b/\partial s(1/2m)] \qquad (7)$$

The positive sign of $\partial^2 \nu/\partial n \partial s$ in (6) means that with more opportunities to interact due to the presence of production workers, workers with more soft skills will have higher payoffs. This encourages agglomeration. As above, the situation is much more complicated when one considers the impact of the agglomeration of other analysts. If we suppose that $(d/dm)\,(\rho(n+m)/m) > 0$, the expression for $\partial^2 \nu/\partial m \partial s$ in (7) is ambiguous. The first term is positive. A larger market is thicker. Thickness allows specialization, which reduces the need to adapt. This, in turn, reduces the extra value that soft skills contribute. This *specialization effect* is opposed by the *opportunity effect* in the second term. When a larger market allows more opportunities, soft skills are more valuable since they increase the value of adapting to opportunities.

We have thus uncovered a surprisingly nuanced relationship between soft skills and agglomeration. Jacobs' (1969) analysis of new work in cities is filled with stories of fluid, unplanned, and informal interactions. Vernon (1960) likewise gives considerable emphasis to the "unstable" nature of the products in increasing returns industries. This is also the spirit of the elegantly formal treatments of agglomeration such as Ogawa and Fujita (1980) and Fujita and Ogawa (1982) and other papers of that sort. Although none of this work mentions soft skills explicitly, in every instance cities are presented as being centers of the sort of interaction that one would expect to be enhanced by soft skills. The microfoundations analysis that we have carried out here suggests that agglomeration is also important for a different reason: the specialization allowed by market thickness. In thicker markets, analysts end up only taking on tasks that are close to their ideal activities, and adaptation and interaction may be less necessary.

Our analysis shows, therefore, that soft skills and specialization are, in a sense, substitutes. Soft skills are valuable because they reduce the cost of the interaction between two agents at a given distance in the characteristic space. In contrast, specialization limits the distance over which the interaction takes

place. This substitutes relationship between soft skills and agglomeration means that workers with soft skills may not be unambiguously attracted to concentrations of economic activity. For some workers the specialization effect might be the dominant one while for others the opportunity effect might the strongest. Some forms of agglomeration might produce a strong specialization effect while others might lead to a strong opportunity effect. It is not certain at all, therefore, that there will be a monotonic relationship between agglomeration and soft skills. This will be the key issue addressed in our empirical analysis.

Before turning to the empirics, however, we must characterize the equilibrium of the system of cities with both analysts and production workers. Thus far, we have established the existence of conflicting tendencies for soft skills to be better rewarded in industry clusters. It should not be surprising that these conflicting tendencies will manifest themselves in the quantities of soft skills available in different places. To the extent that the opportunity effect dominates, we would expect to see a greater amount of soft skill in clusters. To the extent that the specialization effect dominates, we would expect the reverse.

Parallel to the case of production workers, we suppose that there are M analysts in total, and that each of them must be allocated between the two cities. An interior equilibrium in the system now requires that four conditions be met:

$$[f(n_1,\theta_1) - f(N-n_1,\theta_2)] - [c(n_1+ m_1) - c(N-n_1+M- m_1)] =0 \quad (8a)$$

$$n_2 = N - n_1 \quad (8b)$$

$$\nu(m_1,n_1) - \nu(M-m_1,N-n_1) =0 \quad (8c)$$

$$m_2 = M - m_1. \quad (8d)$$

The first two conditions are minor modifications of (2a) and (2b), the conditions governing equilibrium in the production-workers only model above. The second two are the parallel conditions for analysts.

A few final comments on specification are in order. First, and most important, the main result in the model is the demonstration that the relationship between soft skills and agglomeration may well be non-monotonic due to the conflicting opportunity and specialization effects. Having said that, we will not directly observe either opportunity or specialization effects. We observe their combined effects, combined again with numerous additional agglomeration effects that the model does not consider. The empirical work to follow should be taken as a reduced-form estimation of the soft skills - agglomeration relationship that is motivated by this section's model. Second, we have assumed that opportunities to adapt arise symmetrically from the population of production workers and other analysts. This is not necessary. However, if opportunities were related more strongly to the other analysts, then more weight

would be placed on the ambiguous $\partial E(\nu)/\partial m$ effect, increasing the likelihood that soft skills are not attracted to concentrations of activity. Third, we included the local attractiveness parameter θ in order to generate an asymmetric system of cities. Our empirical work will be based on the existence of cities of different sizes and on a given industry being allocated in clusters of different sizes as well. In the model, it is straightforward to obtain the comparative static result that more productive locations would tend to have more of both kinds of worker. In contrast, if all sites were identical, then the equilibrium would be symmetric as in Henderson (1974).

The fourth comment is more subtle. We have established the existence of the specialization effect where soft skills and agglomeration are substitutes using a matching model of the microfoundations of agglomeration economies. It is important to note that the substitutes relationship applies rather generally across agglomeration forces. For instance, the matching model can quite naturally capture labor market pooling (Helsley and Strange (1990)). It also can capture input sharing (Helsley and Strange (2002)) or knowledge spillovers (Berliant et al (2006)). The general result that agglomeration and soft skills are substitutes will be present in all of these cases. Of course, the negative side of the matching model's generality is that we will not be able to identify the microfoundations of agglomeration effects. See Duranton and Puga (2004) for more on this "Marshallian equivalence" issue.

The fifth and final comment concerns places where one might be most likely to find opportunity and specialization effects. There is no reason to believe that the opportunity and specialization effects will be the same for both urbanization and localization. Instead, it would be natural to suppose that the opportunity effect will tend to be more strongly associated with urbanization than with localization. This is clearly the spirit of the Jacobs' (1969) quote from earlier in the paper. It is also consistent with the "nursery city" theory and evidence in Duranton and Puga (2001). Their finding that relocations of industry tend to be towards locations where the industry is concentrated suggests that cities are about adaptation while clusters are about routine. In a similar vein, it would also be natural to suppose that the opportunity effect would be more important for workers near the top of an industry's skill distribution than for those near the bottom. Taken together, these speculations suggest that the strongest evidence of an opportunity effect would be found in cities and at the top of the skill distribution, while the strongest evidence of specialization effects would be in industry clusters and at the bottom of the skill distribution. We will therefore do more than simply look at mean skill levels in our empirical work. Before we can look for this sort of evidence, we must now describe the data that will allow our estimation.

III. Data

A. Dictionary of Occupational Titles

This paper's empirical work draws on data from the U.S. Census and the 1991 Revised Fourth Edition of the Dictionary of Occupational Titles (DOT). The DOT is a database that characterizes the skill requirements of occupations. In particular, it characterizes interpersonal skill requirements. Our approach to identifying a worker's soft skills follows the hedonic imputation approach taken in Autor et al (2003) and Bacolod and Blum (2005) by supposing that in a labor market equilibrium workers are matched to jobs that require skills that they have. To be concrete, a worker currently employed as a janitor lacks the soft skills that allow someone to carry direction, planning, and control (*dcp*), but a manager has this skill. Matching the DOT with the Census allows us to characterize the soft skills of workers by occupation, industry, and geographic location.

The Fourth Edition of the DOT released in 1977 provides measures of 44 different skills required to perform over 12,000 detailed occupations in the US labor market. These measures are the result of comprehensive studies by trained occupational analysts of how jobs are performed in establishments across the nation, and are composites of data collected from diverse sources. Primarily, US Department of Labor occupational analysts "go out and collect reliable data which is provided to job interviewers so they may systematically compare and match the specifications of employer job openings with the qualifications of applicants who are seeking jobs through its facilities." (United States Department of Labor Office of Administrative Law Judges Law Library). For the Fourth Edition of the DOT approximately 75,000 on site job analysis studies were conducted. These studies are then supplemented by information obtained through extensive contacts with professional and trade associations.[5]

The Revised Fourth Edition was released in 1991 and used data collected throughout the 1980s to revise the skill requirements of occupations as well as to include new occupations. As a result, the information on 1,692 of the 12,742 occupations was changed (including some occupations that disappeared from the US labor market) and 761 new occupations were included in the dictionary. These new DOT titles were mostly computer-related jobs. While the main use of DOT information has been for job matching, employment counseling, occupational and career guidance, and labor market information services, a few economists also have used the information in DOT, including, Autor et al. (2003), Bacolod and Blum (2005), Wolff (2003) and Ingram and Neumann (2005).[6]

Of the 44 different job characteristics available in the DOT, four capture distinct aspects of the interpersonal requirements of occupations (see Table 1, Panel A). The variable *dcp* assesses if an occupation requires direction, control, and planning of an activity. Clearly, this variable captures one

---

[5] For more information, see http://www.oalj.dol.gov/libdot.htm.

[6] Because of differences in occupation and industry classification codes, we were not able to match the newest implementation of DOT, called O*NET, to earlier years of the Census and to the NLSY. As will become clear in the next sections,,we want to be able to carry out analysis going back to the 1970s both to use the NLSY and for robustness exercises.

element of the ability to carry valuable interactions, the ability to manage. Similarly, the variable *influ* measures if an occupation requires exerting influence. It therefore captures a different type of valuable interaction that is also somewhat related to the ability to manage, although in this instance the "management" takes place outside of an authority relationship.

To be more concrete, it is useful to consider some specific occupations. To that end, Panels A and B of Table 2 list some occupations that do or do not require these various types of interactions. Beginning with the *dcp* and *influ* columns of Panel B, positions of authority such as financial managers and supervisors are required to engage in direction, control, and planning of activities (*dcp*=yes). However, financial managers and supervisors are not required to exert influence over others (*influ*=no). In contrast, teachers and lawyers are required to have influence over others (*influ*=yes), presumably over schoolchildren in the case of teachers and a jury or judge in the case of lawyers. But, while teachers are also required to direct, control, and plan activities (*dcp*=yes), lawyers are not (*dcp*=no).

The third measure of interpersonal interaction we use is the variable *depl*. It assesses an occupation's requirements of "adaptability to dealing with people beyond giving and receiving instructions." In our view, this variable captures the widest range of interactions among workers, and arguably is the one that fits best with the fluid, unplanned, and informal interactions usually thought to happen in cities and industry clusters. As the *depl* column in Panel B of Table 2 shows, the four occupations discussed in the previous paragraph require the ability to dealing with people beyond giving and receiving instructions, as do Physicians and Salespersons (*depl*=yes). Mathematicians, Insurance Underwriters, and Machine Operators do not require this skill (*depl*=no).

The last DOT measure of interaction required by occupations that we use is the *people* variable. Differently than the previous three variables, the *people* variable attempts to rank the *degree* of interpersonal interaction required by an occupation (see Table 1, Panel A). The ranking starts with mentorship being assigned more interpersonal interaction than negotiation, and then continues moving down to receiving instructions. The scale and structure of the ranking is intended to reflect a progression from simple to complex relations to people, such that each successive rank includes those that are simpler and excludes the more complex (Miller et al 1980). While we do not see the ranking as being beyond dispute -- does mentorship really require more or more complex interpersonal interaction than negotiation? -- we do view the arrangement of the people functions as being hierarchical in a more general sense. For instance, it seems hard to dispute that "instructing" people (*people*=7) involves a broader set of interpersonal interactions than "taking instructions" (*people*=1).

We also employ factor analysis to construct an index of the interpersonal interactions required by each DOT occupation by combining information from the four variables just described. The index *peoidx* is the first principal component formed from a weighted linear combination that maximizes the common

variation across all four soft skills measures, subject to the constraint that the sum of the squared vector weights is equal to one. The variable is constructed to have mean of 100 and standard deviation of 10.[7] Panel A of Table 2 lists the top and bottom occupations requiring the skills to carry the interactions defined by each of the measures we use. The occupations requiring the least soft skills include data-entry keyers and machine operators. The occupations requiring the most include therapists, physicians, dentists, administrators and lawyers. Clearly, the latter group includes occupations that involve more interaction than does the former group.

The analysis in the next sections uses the five measures of soft skills described above. In this way we aim at capturing, as much as the data available allow us, the multiple aspects of the skills that allow valuable interpersonal interactions that are useful in the US labor market.

B. Census

Our employment data come from the 2000 5% Census sample (IPUMS).[8] The sample we use includes employed individuals aged 25 to 70 not living in group quarters with positive weeks, hours, and wage income reports, whose occupational categories were merged with DOT information, and with non-missing or identifiable MSAs.

We then match DOT skill measures to workers in the IPUMS. There is no direct mapping of 1991 DOT occupational codes to the 2000 census occupation codes. There is, on the other hand, a mapping of 1991 DOT codes to 1990 census occupation codes from the National Crosswalk Service Center.[9] IPUMS Census samples from 1950 onward happen to be coded with a uniform occupation coding scheme where, in particular, 2000 census occupation codes are mapped to common 1990 definitions (variable *occ1990*). We then match DOT skill measures to workers in the 2000 Census using the mapping of 1991 DOT codes to 1990 occupation codes further aggregated to the uniform classification scheme *occ1990*.[10] There are fewer census occupational categories (469 in 1990) than the nearly 13,000 DOT occupational categories. Therefore to create soft skill measures for the coarser census occupation classifications we first average our skill measures across DOT occupations within each census occupation.

One important point to note is that, by computing these averages without using employment weights, we are in effect assuming that each DOT occupation within a census occupation is equally important across local labor markets in the US. In the best of worlds this just adds noise to our measures

---

[7] The index is constructed for occupations at the DOT level. See Bacolod and Blum (2005) or Bacolod et al (2007) for a somewhat more detailed discussion of the construction of the people index variable.
[8] We have also carried out all of the analysis using the 1990 Census. The results are consistent with those that will be reported below.
[9] http://www.xwalkcenter.org/index.html
[10] With this procedure we are able to obtain a match for over 99% of the workers in the 2000 Census.

and makes it less likely that we find statistically significant relationships between soft skills, agglomeration, and localization. If, however, the mismeasurements are correlated to city size or industry concentration, then they will actually bias the coefficients of any relationship we attempt to estimate. Fortunately, we can assess how relevant these mismeasurements are using a special version of the April 1971 CPS monthly file. This file was coded with both 1977 DOT and 1970 census occupation codes, and it was issued by the National Academy of Sciences (2001). In this file a committee of experts assigned individual DOT occupation codes to the 60,441 workers in the CPS sample. For the occupations in this special file we can compute our measures of soft skills for each census occupation using CPS sampling weights.[11] We can then compare the employment weighted and the unweighted measures of soft skills and test if they are statistically different. For our purposes, it is not sufficient to assess whether weighted and unweighted averages of soft skills are different, but also whether or not they vary systematically across SMSA status. For the 60,441 workers in that sample, we also know if they live in a SMSA or not. Therefore we can compute In/Out SMSA employment weighted measures of our soft skills across census occupations, and use these to test if any mismeasurement in our soft skill measures due to the lack of employment weights is correlated to urbanization. We perform this analysis (the details are provided in the Web Appendix) and strongly reject the hypothesis that the weighted and the unweighted measures are statistically different across occupations. We also reject the hypothesis that In SMSA and the Out SMSA employment weighted measures are statistically different.

As a final note on this issue, it is worth asking why the weighted and unweighted measures are so similar. Upon inspection, we find that, even though different DOT occupations within a Census occupation have different employment weights, they usually have very similar soft skill requirements. For instance, any of the different types of managers that are distinguished in the DOT but not in the Census is required to be able to deal with people beyond giving and receiving orders. Therefore, for many of the occupations employment weights do not matter at all, and for the remaining occupations they have a minor effect.

It is important to recognize, of course, that the hedonic equilibrium in labor markets has frictions. If the actual skills required by an occupation vary systematically with agglomeration, then the estimates of an agglomeration - soft skill relationship will be biased. For instance, it is possible that workers in large cities in a given occupation need to be more skilled than workers in the same occupation in small cities. Lawyers in large cities may be more likely to be involved in highly demanding corporate law, while lawyers in small cities might be more involved with routine law such as that involved in buying a house. We turn to the National Longitudinal Survey of Youth (NLSY79) to address this empirical

[11] For reference, 3,885 DOT occupations are represented in this special CPS sample.

concern. The NLSY79 has individual measures of worker abilities that the Census does not. Specifically, the Armed Forces Qualification Test (AFQT) measures cognitive ability, while the Rotter Index is a standard measure of non-cognitive skills. These two proxies for workers' skills have been shown in prior work to account for sizeable shares of wage variation.[12] If there is systematic within-occupation variation in the skills required to perform an occupation, the AFQT and Rotter variables should capture it. As seen in the Web Appendix, within occupations, neither the AFQT nor the Rotter Index has a clear pattern of being related to city size or industry clustering, where the former is captured by city population and the latter by the share of national employment in an industry found in a given city (i.e., employment in automobile manufacturing in Detroit divided by national employment in automobile manufacturing).

Two additional points are also worth making. First, while the hedonic imputation may be imperfect, there is no better alternative. In particular, there is no dataset with measures of individual workers' soft skills that is also large enough to allow for precise estimates of the models we specify below. Creating such a dataset would be very costly, and it is not at all clear that a survey of workers' self-reported soft skills would really produce a better measure of, say, a worker's ability to direct, control, and plan than would the judgment of an employer. Second, the construction of the DOT is national. It is intended to be a resource for job seekers. If there were believed to be significant spatial bias, then the US Department of Labor could easily correct this by creating more refined job categories (i.e., big city accountant vs. small city accountant). That it has not suggests that it is believed that any errors in skill requirements will not be substantial impediments to job seekers. Indeed, the DOT has been recently replaced by a new dataset, the O*NET, that also contains nation-wide measures of job requirements. For all these reasons, we believe that the hedonic approach to skills is both reasonable and an advance in the understanding of cities and skills. In any case, this approach is the best that can be done to allow one to analyze soft skills in large datasets.

C. Descriptive statistics

Using the matched dataset, Tables 3 and 4 describe some broad characteristics of the soft skills composition of industries and cities, respectively. Table 3 shows the top and bottom five industries ranked by the share of workers in occupations that require *depl*, *dcp*, and *influ*. In addition, Table 3 shows the top and bottom five industries ranked by workers' mean values of *people* and *peoidx*.

There are several noteworthy features of Table 3. First, the same industries show up as the top and bottom five industries across the different soft skill measures. For instance, within manufacturing, Apparel is the industry with the smallest share of workers in occupations that require *depl* (23% of the

[12] See for example, Neal and Johnson (1996) on the AFQT, and Bowles, Gintis, and Osborne (2001) on the Rotter score.

jobs in the industry require this skill), the least requiring *dcp* (16%), the lowest in *people*, and the lowest in *peoidx*. Similarly, Newspaper Publishing is the manufacturing industry with the highest share of jobs requiring *depl, influ, people,* and *peoidx*. However, Newspaper Publishing does not appear in the top 5 manufacturing industries requiring *dcp*. This is the second noteworthy point of Table 3: the various soft skill measures capture different things, and so are present in different levels across industries and cities. Finally, soft skills tend to be more concentrated in the industries we would expect. For instance, including all sectors of the economy, Child Care Services have a very high share of jobs requiring soft skills. So does the industry Elementary and Secondary Schools (top 5 in all).

Panel A of Table 4 shows the top and bottom 5 MSAs ranked by the share of employment in occupations that require interpersonal interactions. Similar to the first two points made above about Table 3, while the same MSAs show up as the top or bottom, there is also variation across measures. For instance, while Flint, MI is ranked among the cities requiring the least *dcp, influ, people,* and *peoidx*, it does not appear in the bottom 5 for *depl*. Also as expected, college towns (i.e., Bryan-College Station, TX, the home of Texas A&M) tend to have high concentrations of soft skills.[13]

Panel B of Table 4 repeats this exercise for cities with population above 1 million. It is clear that not all big cities are the same. Not surprisingly, both Boston and San Francisco appear among the top five cities for nearly every measure. It is also not surprising given the current state of its economy that Detroit appears in most of the bottom five lists, joined often by similar older manufacturing cities. It is notable that Greensboro-High Point -- the most important furniture cluster in the US -- is on every bottom five list. Not all clusters are the same either, apparently. It is important to point out, however, that even the cities near the bottom of the skill distribution for large cities have skill levels that are above the average for the entire county. The *peoidx* variable, for instance, is constructed to have a mean of 100, and the smallest value in the table is 106.8.

At this point, we turn to a more systematic analysis of the relationship between agglomeration and soft skills. We will use standard variables to capture the urbanization and localization dimensions of agglomeration. For urbanization, we use MSA population (*pop*). This is constant across industries. For localization, we use the share of the industry's national employment found in a particular MSA (*cluster).* These variables exhibit a moderate level of correlation, 0.38. Thus, although clusters are frequently found in large cities, the two agglomeration variables measure different things.

[13] It is worth pointing out that the variation in average levels of soft skills across cities is strongly correlated with the ratio of white to blue collar workers. We thank an anonymous reviewer for suggesting this. Interestingly, the correlation is much weaker between the average soft skill level by industry and the industry's white to blue collar ratio. These results are available in the Web Appendix, available at http://www.rotman.utoronto.ca/~wstrange/.

IV. Estimation

A. Overview

A natural way to begin is to consider the relationship between a given worker's skill and the presence of activity in the worker's industry and total city population. This involves estimating the following regression:

$$DOT_{ikm} = \delta_k + \alpha\ Cluster_{km} + \beta \log(Pop_m) + \nu_{ikm} \qquad (9)$$

DOT_{ikm} is the type of interactions (*depl*, *dcp*, etc.) required by the occupation of individual *i* working in industry *k* and living in MSA *m*. It is used as a proxy for individual *i*'s soft skills. The right-hand-side contains measures of national employment shares by industry and MSA (the *Cluster* variable) and MSA population.[14] Equation (9) includes controls for industry fixed-effects, 3-digit level, since the relationship predicted by the model operates within industries. Equation (9) is estimated for each of the soft skills discussed in Section III using Census individual weights. It is also estimated for measures of worker education like the number of years of schooling and the highest grade completed.[15] Because of the grouped nature of the data, the standard errors are clustered at the industry-MSA level.[16] The theory set out in Section II generates predictions of the location of a mobile industry. It is natural, therefore, to estimate (9) for manufacturing. The theory, of course, applies also to tradable service industries. We therefore also estimate (9) for select services whose outputs are plausibly at least partially tradable. We anticipate noisier estimates in these models.

The results for manufacturing sectors are presented in Table 5. Panel A presents the results on the education measures. In addition to being interesting by themselves, these results provide a benchmark against which we can compare the relationship between soft skills, agglomeration and localization. Larger cities have proportionally more workers with less than a high-school degree and with a college degree. Of course, they also have fewer workers with a high-school degree. This pattern echoes Berry and Glaeser (2005). In our data, we find that a doubling of MSA population corresponds to roughly 0.7 log points, which in turn is associated with a .027, -.032, and .015 point change in workers with less than high school, high school, and college degrees respectively. Given that by construction these education measures are {0,1}, one way to interpret these numbers is that doubling the MSA population raises the

---

[14] Specifically, $Cluster_{km}$ is defined as the share of industry k's national employment found in MSA m. The logarithm of MSA population is used in order to capture the apparent non-linear relation between population and soft skills we find in the raw data. All results of the paper hold when MSA population is used instead.

[15] Estimating (9) amounts to determining how an aggregate amount of national skills are allocated across locations. It does not directly address how workers invest in skills, learn, cope with skill deterioration, and choose occupations.

[16] The results are robust to a variety of approaches to clustering, including clustering at the MSA*industry*occupation level and clustering separately at the MSA, industry, and occupation levels.

probability that a worker in that MSA will have less than a high school degree by 0.027 points. In our sample 11% of the workers have less than a high school degree (see Table 1 Panel B). Thus, doubling the MSA population raises the probability of finding a worker with less than a high school degree by about 25%. Doubling an MSA's population also increases the probability of finding a worker with college degree, but by only 5% and lowers the probability of finding a worker with high school degree by about 13%. Overall, larger cities have workers with fewer years of education, although the effect is small. Doubling the population lowers the average education by .07 years.

The effects of localization on education measures are fundamentally different. Industry clusters also have more college graduates, but fewer workers with a high-school diploma or less - although the effect for workers with less than high-school is not statistically significant. The effects are small in magnitude. Moving from a location with 1% of an industry's employment to one with 2% is associated with a decrease of roughly .4% in the probability of finding a worker with a HS degree, and an increase in the probability of finding a worker with a College degree of .6%.

Panel B of Table 5 shows the results for the soft skills. On average, an industry's occupations that require soft skills are found in larger cities but not where the industry is clustered. A doubling of MSA population is associated with a .015 point change in *depl*. Given that 60% of workers are in occupations that require *depl*, a doubling of MSA population raises the probability of finding a worker in an occupation that requires *depl* by a about 2.5%. For the other {0,1} skill measures *influ* and *dcp*, the result of a population doubling is an increase in the analogous probability of about 5% and 2% respectively. These relationships are all statistically significant, but the magnitudes are relatively modest. For the skill variable *people*, which runs from 1 to 9, the effect is roughly 0.5 points. To see what this means, recall that this variable ranks the complexity of a job in relation to people. A doubling of population is associated with an average increase in skills that is half of the difference between two consecutive ranks, for example instructing and negotiating or negotiating and mentoring. For the index of interpersonal skills (*peoidx*), the effect of doubling an MSA's population is to increase the mean value of the index by .30. This is roughly three one-hundreds of one standard deviation of the distribution of this index. In sum, although population size and average soft skills are positively correlated, none of the individual measures increases especially strongly with MSA populationd

In contrast, the within-industry relationship between soft skills and the presence of own-industry employment is not positive. In fact, for three of our measures of soft skills (*depl*, *influ*, and *peoidx*) industry clusters have less soft skills on average - although only the effect on *influ* is statistically significant. For this skill, moving from a location with 1% of an industry's employment to one with 2% is associated with a decrease of roughly 0.002 in the probability that an occupation requires it. This is a weak correlation. The coefficients on *dcp* and on *people* are positive but only statistically significant for

dcp. Again, the magnitudes are small. In sum, the presence of own industry employment is only weakly related to the average level of soft skills. Sometimes the relationship is negative. Sometimes it is positive.[17]

These results establish the pattern that will appear in much of the subsequent estimation. Urbanization is associated with an increase in the average level of soft skills. Localization is not. This result is complementary to Duranton-Puga (2001) model of nursery cities. In this paper, theoretical analysis establishes the possibility that a large and diverse city is important in the development of prototypes, but the production of a mature product is more economically accomplished in a specialized industry cluster. Our result shows that large cities have higher levels of soft skills, which are presumably the sorts of interactive abilities that would be useful in the experimental phases of production. Industry clusters are not associated with soft skills, which is consistent with the Duranton-Puga model of the production of a mature product and with our notion of a specialization effect that is inimical to soft skills.[18]

It is worth noting that urbanization is positively related to the levels of all of five measures of soft skills. The variable *depl* captures interaction outside of authority relationships. The theoretical literature on spatial interactions (see Fujita and Thisse (2002)) is usually motivated by referring to this sort of interaction. In contrast, the variable *dcp* is about management. This suggests that in thinking about the sorts of interaction that are involved in agglomeration economies, it is important to consider also interactions that occur within organizations.

Given the importance of services for cities, we carry out a parallel analysis for selected service sectors. The choice of where to locate different parts of the production process of and industry is only available for tradable sectors. Thus, we focus on the subset of services sectors that are at least potentially tradable.[19] Table 6 presents the results. Panel A shows the results for worker education. The pattern found in manufacturing sectors is also present for these services sectors: larger cities have more workers with less than high school and college degrees and less with high school degrees. The magnitude of the effects is very different, though. The coefficient on workers with less than a high school education is less

---

[17] We have not made much of the pattern of results across different soft skills. This is because among the dummy variables (*depl*, *influ*, and *dcp*) the only pattern is that they are all related similarly to agglomeration. In some specifications, one may be larger. In others, a different one is larger.

[18] It is important to note that these results do not identify individual sources of agglomeration economies such as Marshall's labor market pooling, knowledge spillovers, or input sharing. Soft skills could potentially be involved in all three, and in other sorts of agglomeration economies as well. This difficulty has become known as the problem of "Marshallian equivalence." (See Duranton and Puga (2004)).

[19]The list of service sectors is: motion pictures and video industries, internet publishing and broadcasting, data processing, legal, accounting, advertising, architectural and engineering, design, computer system design, scientific, and technical consulting, scientific research and development, management services and FIRE (banking, savings institutions, non-depository credit, securities, commodities, funds, trusts, and other financial investments, insurance carriers, and real state).

than one-tenth of the one in manufacturing and the one on workers with a high school education is about one-half. On the other side of the distribution, the coefficient on workers with college degrees is almost twice as large. Finally, larger cities have workers with more years of education in services sectors, but the difference is small. The relationship between localization and education measures in the services sectors is also similar to the ones in manufacturing.

Panel B shows the results for soft skills. As with manufacturing, large cities also contain services sectors occupations that require soft skills in larger cities. The coefficients here are smaller than for manufacturing. In the case of *depl*, for example, it is 75% smaller than the coefficient for manufacturing, which was itself not large. The effect of industry concentration is somewhat different than the patterns found in manufacturing sectors. For all of our measures of soft skills, industry clusters have less soft skills on average. Having said this, the magnitudes continue to be small. As a group, these results are consistent with the key patterns from manufacturing in the soft skill - agglomeration relationship. There is evidence consistent with both opportunity and specialization effects. The former appear to dominate for urbanization, while the latter appear to be strongest for localization.

It is helpful to compare the soft skill results to results for some "harder" cognitive, motor, and physical skills. Table 7 presents estimates for three cognitive skills (general educational development (GED-) in math(M), reading(R), and language(L)), three physical skills (ability to work with "things" (THINGS), strength, and the ability to work to standards and tolerances(STS)), and two indices of occupations' cognitive and physical requirements.[20] The cognitive skills and the measure of physical strength are straightforward. The *things* measure ranges from high-level activities such as setting up production processes and precision manufacturing down to feeding and handling. The standards and tolerance variable, *sts*, addresses the occupation's requirements regarding working to standards such as ensuring that bottles produced meet specific tolerances.

It is immediately apparent that the results for cognitive and physical skills are quite different than the previously reported results for soft skills. All of the cognitive skills have mean values that rise with both the degree of urbanization <u>and</u> localization. Thus, both cities and clusters are associated with cognitively demanding occupations. The physical skills all fall with city population. The *things* variable and the *sts* variable both rise with localization. *Strength*, in contrast, has a negative coefficient. Since high levels of *things* and *sts* are associated with precision manufacturing, this suggests that clusters are associated with advanced activities. The simplest physical activities are thus negatively associated with cities and clusters, while more advanced physical activities are positively associated with clusters but not with cities.

[20] See Bacolod et al (2008) for the details of the construction of the indices and on the skills.

C. The distribution of skills

Although illustrative, the evidence from means does not paint a complete picture. Section II's theory gives us reason to wonder whether the upper and lower tails of an industry's skill distribution are related to urbanization and localization in the same way that the mean is. Does a line worker in manufacturing have the same opportunities for interaction that a manager does in a large city or industry cluster? Do the tasks performed by both sorts of worker become highly specific as the division of labor becomes more refined in a city or cluster? These questions can only be answered by considering the entire distribution of worker skills.

In this section, we consider the distribution of skills by industries across locations. As noted above, the predictions of our model are about the allocation of soft skills within industries. It is therefore essential that any estimates of this distribution must consider the tendency of the industry to employ workers with soft skills. The mean model in (9) does so by including fixed effects. Estimating a similar model for the percentiles of the skill distribution would have the usual incidental parameters problem of estimating fixed effects in a nonlinear model (see Arellano and Honore 2001 and Koenker 2004). Because of this problem, we do not estimate a quantile model. Since the explanatory variables of interest only vary at the MSA-industry level, and because we want the industry fixed effects to have a pure location shift effect on the skill distribution, we use a 2-stage procedure that groups the individual data at the MSA-industry level.[21] Under the usual assumptions of exogeneity of the regressors, the estimates of the model described below can be shown to be consistent.

In the first stage, we net out the industry-specific component of the distribution of the soft skill and calculate percentile values of the skills by industry and MSA:

$$\mathrm{DOT}_{ikm} = \delta_k + \sum_q \alpha^q_{km} d^q_{km} + \eta_{ikm}. \qquad (10)$$

In (10), d^q_{km} equals one if the individual is at the qth percentile of the skill distribution in industry k and MSA m and zero otherwise. In the second stage we then use $\hat{\alpha}^q_{km}$ to estimate the relationship between the percentile values of the distribution of skills and industry concentration and agglomeration at the MSA-industry level. For q={10, 25, 50, 75, 90}, we estimate:

[21] The 2-stage approach used in Moretti (2004) is similar to the one employed here, except that it aggregates up to the MSA level and focuses only on mean values.

$$\begin{bmatrix} \hat{\alpha}_{km}^{10} \\ \hat{\alpha}_{km}^{25} \\ \hat{\alpha}_{km}^{50} \\ \hat{\alpha}_{km}^{75} \\ \hat{\alpha}_{km}^{90} \end{bmatrix} = \begin{bmatrix} \kappa^{10} \\ \kappa^{25} \\ \kappa^{50} \\ \kappa^{75} \\ \kappa^{90} \end{bmatrix} + \begin{bmatrix} \gamma^{10} \\ \gamma^{25} \\ \gamma^{50} \\ \gamma^{75} \\ \gamma^{90} \end{bmatrix} * \text{Cluster}_{km} + \begin{bmatrix} \lambda^{10} \\ \lambda^{25} \\ \lambda^{50} \\ \lambda^{75} \\ \lambda^{90} \end{bmatrix} * \log(\text{Popn}_m) + \begin{bmatrix} \varepsilon_{km}^{10} \\ \varepsilon_{km}^{25} \\ \varepsilon_{km}^{50} \\ \varepsilon_{km}^{75} \\ \varepsilon_{km}^{90} \end{bmatrix} \tag{11}$$

The system of equations above is estimated jointly by feasible GLS to allow for correlation of the residuals. The standard errors are bootstrapped to take into account the fact that the LHS variables are estimated parameters. This approach also allows for clustering of the standard errors at the MSA level. An additional issue is that in some industry/MSA-pairs the number of workers sampled is very small because employment in these industries and MSAs is itself small. In these cases the estimated percentile values will be noisy and not carry much information. We deal with this by restricting the sample to industry/MSA-pairs for which there are at least 20 workers sampled.[22]

This quantile approach imposes very little structure on the relationship between the skill distribution and agglomeration. As an alternative, we also estimate the effects of urbanization and localization on the 90th-10th and the 75th-25th percentile differences:

$$\begin{aligned} \left(\hat{\alpha}_{km}^{90} - \hat{\alpha}_{km}^{10}\right) &= \kappa^{90,10} + \gamma^{90,10} * \text{Cluster}_{km} + \lambda^{90,10} * \log(\text{Popn}_m) + \varepsilon_{km}^{90,10} \\ \left(\hat{\alpha}_{km}^{75} - \hat{\alpha}_{km}^{25}\right) &= \kappa^{75,25} + \gamma^{90,10} * \text{Cluster}_{km} + \lambda^{75,25} * \log(\text{Popn}_m) + \varepsilon_{km}^{75,25} \end{aligned} \tag{12}$$

The equations above are estimated separately and, as before, the standard errors are bootstrapped and clustered at the MSA level. The sample is also restricted to industry/MSA-pairs for which there are at least 20 workers sampled.

Before presenting the results, it is worth noting that the interpretation of the analysis of the percentiles of the skill distributions for variables that can only assume values {0,1} – *depl, influ, dcp* - is slightly different than for continuous variables, but carries the same message. For instance, a positive and significant coefficient for the 75th percentile of *dcp* with respect to city population indicates that the 75th percentile of the distribution of *dcp* is more likely to be an occupation that requires direction control and planning in a large than in a small city.

Table 8 reports estimates for the manufacturing sector. The first panel shows the coefficients of urbanization and localization on the different percentiles of the distribution of years of education. Larger

[22] None of the results are driven by this restriction.

cities' most educated workers clearly have more years of education. The coefficients are all highly significant. A similar pattern emerges for industry clusters, with locations with greater shares of an industry's activity having more educated workers. However, the cluster estimates are quite noisy, and many are insignificant. Table 8 also shows the effect of urbanization and localization on the inter-percentile values of the distribution of years of education. Larger cities are more unequal in the way education is distributed across workers. For instance, doubling a city's population is associated to an increase of .5 years in the difference between the number of years of education of the 90th percentile and the 10th percentile worker. Industry concentration has a significant and positive effect on the 75th-25th percentile difference, but an insignificant positive effect on the 90th-10th difference.

The remaining panels of Table 8 relate to soft skills. The clear pattern in the estimates is that the distribution of soft skills is wider in large cities and more concentrated in industry clusters. For MSA population (urbanization), the *depl, influ,* and *dcp* skill values are significantly greater for both the 75th and 90th percentiles and lower for the 10th and 25th percentiles. The results are mixed for the 50th percentile, with a significant positive coefficient for *depl* and *dcp,* and a significant negative coefficient for *influ*. In contrast, for the Cluster variable (localization), the coefficients are negative and significant for the 75th and 90th percentiles and positive for the 10th and 25th for the three variables in every instance but one (the 90th percentile for *dcp*), and that coefficient is still negative but insignificant. For the 50th percentile, the coefficient for *depl* is negative and significant, while for *dcp* and *influ* the coefficients are positive and significant. The entire pattern makes it very clear that soft skills are related to urbanization very differently than to localization. While we are not able to identify the opportunity and specialization effects themselves, the results are consistent with the discussion at the end of Section III.

The results for *people* and *peoidx* are similar. The only difference is that the tendency for urbanization to be associated with a decrease in soft skills at the bottom of an industry's skill distribution is less pronounced, affecting the tenth percentile of the distributions but not the twenty-fifth percentile.

We also performed the analyses for the same sample of potentially tradable service sectors discussed above. Together with our other robustness checks, the results are available in the Web Appendix. The pattern in this case is less clear than the pattern in Table 8. The distribution of education and soft skills in these service sectors has a much weaker relationship to industry concentration and urbanization. If we focus on the effects on the 90th-10th difference, we find that the same patterns found in manufacturing hold for services. For all five measures of soft skills, urbanization increases the spread of the distribution, with the estimated parameters statistically significant in 3 of the 5 cases. For four of the soft skill measures localization reduces the spread of the skill distribution (all except *depl*), although

in this case the parameter estimates are not statistically significant. Overall, the results for service sectors are similar to manufacturing, but are less sharp.[23]

As with the results for mean levels of skills, it is also interesting to compare our soft skill distribution results to parallel results for cognitive and physical skills. The latter are presented in Table 9. It is immediate in Panel A that an increase in population is associated with an increase in the level of cognitive skills at every point in the skill distribution. As with Table 7's mean results, big city occupations are more cognitively demanding. Interestingly, the results diverge somewhat from Table 7 for *cluster*. The cognitive skill levels at the 75th and 90th percentiles increase with *cluster*. The cognitive skill levels at the 10th and 25th tend to fall. For all of the cognitive skill measures, the 90th-10th percentile difference increases with both *population* and *cluster*. It is clear from all of this that the effects of agglomeration on the distribution of cognitive skills are quite different than the effects on soft skills. For physical skills, the tendency for skill levels across the distribution is either to fall with *population* (for the 10th, 25th, and 50th percentiles) or to be insignificantly different (the exception is the effect on *sts*, standards and tolerances, which increases). For *cluster*, physical skill levels increase at lower points of the distribution (10th and 25th percentiles). The effects at higher levels are mixed. A consistent pattern is that at the 90th percentile, the coefficients are statistically insignificant. The clearest message from these results is that cities and clusters are both centers of cognitive skills, at least at the top of the skill distribution. Regarding physical skills, the least skilled occupations are more demanding in cities and clusters, but there is no consistent evidence regarding the skill levels of the most skilled. It is important to be clear that this does not mean that cities and clusters do not carry out advanced precision manufacturing. They clearly do. The results mean that they are also carried out outside of cities and clusters. And it is important also to recall that the mean levels of the coefficients on the more advanced physical skills (*things* and *sts*) both increase with *cluster*.

D. Robustness

The pattern of results documented above proves to be robust. A very similar pattern of results holds for models estimated using 1990 Census data and using a sample including both the manufacturing industries and the selected service industries together. The results are available in the Web Appendix.

We have also estimated specifications including a range of individual and MSA controls. The empirics thus far have considered the question of how industries allocate jobs requiring soft skills between large and small cities and between industry clusters and other locations. We believe that this is

[23] One possible explanation for this is that many of the service sectors do not tend to cluster. If this is because agglomeration economies are weak, then we would expect to find a weaker soft skills - agglomeration relationship. As a robustness check, we re-estimated Table 8 for a more restrictive set of service industries that exhibit tendencies to agglomeration. We obtain strong results confirming that urbanization widens the distribution of soft skills.

the question posed by the theory. However, it is worth pushing as far as we can by adding controls to the standard models for other worker characteristics. Education is one obvious control. This amounts to considering the related question: within the jobs of college educated workers, do industries allocate jobs that require soft skills in big cities and/or clusters? To answer this and related questions, we have estimated models with individual controls (i.e. education attainment, age, race, and marital status). We have also estimated models with MSA controls (i.e. percentage of population with college degree, mean income level, and air quality index). Finally, we have also estimated models including both individual and MSA controls. The inclusion of such a long list of controls substantially reduces the amount of variation from which we are able to estimate our models. Despite this, the results, reported in the Web Appendix, are actually stronger than those in the standard models. We find increases in the average levels of soft skills with city size and a widening of the skill distribution. We do not find increases in average soft skill levels in industry clusters.

As a final robustness check, we also estimate models based on an individual attribute, the Rotter Index. As noted briefly above, the index measures the degree to which an individual believes him- or herself to be in control of life circumstances, rather than being at the mercy of external forces. This is referred to as the locus of internal control. See Rotter (1966). Heckman et al (2006) make a strong case for the importance to labor markets of non-cognitive skills such as those measured by the Rotter Index. The Rotter Index is not a measure of interactive skills *per se*, but it has been shown to be correlated with the individual's social skills (Lefcourt et al (1985)). In the Rotter Index, a lower score corresponds to a greater sense of control and thus better social skills. We therefore estimate models for the inverse of the Rotter Index so that the results can be readily compared to the DOT results (with a higher score indicating a greater level of non-cognitive skills). Using the NLSY-79 panel, we estimate both mean and skill distribution models for the Rotter Index. We use a confidential geocode version of the NLSY in order to identify county of residence, with counties converted to MSAs using the Census correspondence.[24] Because the NLSY is much smaller than the Census, we estimated both mean and distribution models for the manufacturing sample and also on a sample of all industries. Again, these results are available in the Web Appendix. We do not find a statistically significant relationship between the mean Rotter Index and either the city size or the industry cluster variable. The distribution results, however, display the same pattern that we found using the imputed DOT variables. In larger cities, the spread of soft skills is greater.

---

[24]We thank the Bureau of Labor Statistics for making the confidential geocode version available. Following Moretti (2004), we exclude the military supplemental samples from our analyses. Our data is then an unbalanced panel spanning the years 1979-1996, with a total of 110,659 individual-year observations with non-missing values for all the relevant variables.

V. Conclusion

This paper has considered the role of soft skills in cities and industry clusters. The analysis has focused on a tension between two ways that agglomeration can produce increasing returns. One is that agglomeration presents opportunities for valuable interactions between agents. Since soft skills are likely to facilitate this sort of interaction, cities and industry clusters may prove to be attractive for workers who possess these skills. However, the other sort of increasing return is that cities allow a highly refined division of labor. In this situation, a worker's role in the industrial system is specialized, and it is possible that interaction skills are less important. With the opportunity and specialization effects operating in the opposite direction, we show that there is no reason to expect a monotonic relationship between agglomeration and soft skills. While for some workers the opportunity effect may be the most important, for others the specialization effects can be the relevant one. Also, one dimension of agglomeration (i.e., the concentration of an industry) may lead to a strong specialization effect while the aggregation of population may generate a strong opportunity effect.

Having set out a theoretical model that establishes the existence of this tension between cities as centers of soft-skill enabled interaction and cities as refined divisions of specialized labor, the paper then carries out an empirical analysis of the relationship between soft skills and agglomeration. To do so, the paper matches data from the Dictionary of Occupational Titles to Census data. The within-industry average level of soft skills is found to be higher in cities but not in centers of an industry's activity. Furthermore, the skill level at of the most skilled workers is higher for large cities than for small cities. The skill level for the least skilled workers is lower. In contrast, mean soft skill levels tend to be insignificantly related to industry clustering, and there is greater concentration around the mean when an MSA's share of the national employment in an industry rises. Although these relationships are consistent with conflicting opportunity and specialization effects, as in the model, there are obviously many effects that might potentially be at work.

All of this leads to the paper's main conclusion: the relationship between agglomeration and soft skills is a nuanced one. Although cities have higher mean levels of soft skills, this effect is driven by the top of the skill distribution. Workers with low levels of soft skills tend to have even lower levels in large cities. Thus, cities are not unambiguously centers of soft-skill enabled interaction. Industry clusters are even less so.

References

Arellano, M., and B. Honore (2001), *Panel Data Models: Some Recent Developments. In Handbook of Econometrics*, Volume 5, ed. By J. J. Heckman, and E. Leamer, North-Holland.

Autor, D.H., F. Levy, and R. Murnane (2003), "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics* 118(4), 1279-1333.

Bacolod, M. and B. Blum (2005), "Two sides of the same coin: U.S. 'residual' inequality and the gender gap," Working Paper.

Bacolod, M., B. Blum, and W. Strange (2007), "Skills and the City," Working Paper.

Beckmann, M.J. (1976), "Spatial Equilibrium in the Dispersed City" in Y. Papageorgiou (ed.), *Mathematical Land Use Theory* (Lexington: Lexington Books), 117-125.

Berliant, M. P. Wang and R. R. Reed (2006), "Knowledge Exchange, Matching, and Agglomeration," *Journal of Urban Economics* 60, 69-95.

Berry, C.R., and E. L. Glaeser (2005), "The divergence of human capital levels across cities," *Papers in Regional Science* 84:3 407.

Bowles, S., H. Gintis, M. Osborne, 2001. The determinants of earnings: a behavioral approach, *Journal of Economic Literature* 39, 1137-76

Duranton, G. and D. Puga (2001), "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products" *The American Economic Review*, 91 (5), 1454-1477.

Duranton, G. and D. Puga (2004), "Micro-foundations of urban agglomeration economies," in: J. V. Henderson and J.-F. Thisse (Eds.), *Handbook of Urban and Regional Economics*, Volume 4, North Holland, Amsterdam, 2004, 2063-2118.

Elvery, J. A. (2007), "City size and skill intensity", Working Paper.

Fallick, B., C. A. Fleishman, and J. B. Rebitzer (2006), "Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster," *Review of Economics and Statistics* 88, 472-481.

Fujita, M. and H. Ogawa (1982), "Multiple equilibria and structural transition of non-monocentric urban configurations," *Regional Science and Urban Economics* 12, 161-196.

Fujita, M. and J. Thisse (2002), *The Economics of Agglomeration* (Cambridge: Cambridge University Press).

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

Heckman, J., J. Stixrud, and S. Urzua. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior" NBER Working Paper 12006. Februrary 2006.

Helsley, R. W. and W. C. Strange (1990), "Agglomeration Economies and Matching in a System of Cities." *Regional Science and Urban Economics* 20: 189-212.

Helsley, R. W. and W. C. Strange (1991), "Agglomeration Economies and Urban Capital Markets." *Journal of Urban Economics*, 29, 1991, 96-112.

Helsley, Robert W. and William C. Strange (1994), "City Formation with Commitment." *Regional Science and Urban Economics* 24, 373-390.

Helsley, R. W. and W. C. Strange (2002), "Innovation and Input Sharing, " *Journal of Urban Economics* Volume 51, Issue 1, January 2002, Pages 25-45

Helsley, R. W. and W. C. Strange (2007), "Urban Interactions and Spatial Structure," *Journal of Economic Geography*, 7, 119-138.

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

Ingram, B. and G. Neumann (2005), "The Returns to Skill," forthcoming, *Labour Economics*.

Jacobs, J. (1969), The Economy of Cities (New York: Vintage).

Klepper, S. (2007), "Disagreements, Spinoffs, and the Evolution of Detroit as the Capital of the U.S. Automobile Industry," *Management Science* 53(4), 616-631.

Koenker, R. (2004), "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74-89.

Lefcourt, H. M., R A. Martin, C. M. Fick, and W. E. Saleh (1985), "Locus of Control for Affiliation and Behavior in Social Interactions, Journal of Personality and Social Psychology," 48(3), 755-759.

Lin, J. (2007), "Innovation, cities, and New Work," Working Paper.

Marshall, A. (1890), Principles of Economics (London: MacMillan).

Miller, A.D.T., P. Cain, and P. Roose eds. (1980), Work Jobs and Occupations: A Critical Review of the Dictionary of Occupational Titles (Washington D.C.: National Academy Press).

Moretti, E. (2004), "Estimating the Social Return to Higher Education: Evidence From Longitudinal and Repeated Cross-Sectional Data," *Journal of Econometrics*, 121 (1-2), 175-212.

Ogawa, H. and M. Fujita (1980), "Equilibrium land use patterns in a non-monocentric city." *Journal of Regional Science* 20, 455-475.

Rosenthal, S. S. and W. C. Strange (2004), "Evidence on the Nature and Sources of Agglomeration Economies", in Henderson, J.V. and J.-F. Thisse, eds., *Handbook of Urban and Regional Economics*, Volume 4. Amsterdam: Elsevier, 2119-2172.

Rotter, J.B. (1966), "Generalized expectancies for internal versus external control of reinforcement," *Psychological Monographs* 80(1), 1-28.

Scott, A.J. (2005), On Hollywood: The Place, The Industry, Princeton University Press: Princeton, NJ.

Scott, A. J. (2007), "Production and Work in the American Metropolis: A Macroscopic View" Working Paper.

Scott, A. J. and A. Mantegna (2007), "Human capital assets and the structure of work in metropolitan areas: a preliminary exploration of the O*Net data base," Working Paper.

Sorenson, O. and P. G. Audia (2000), "The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940-1989," *The American Journal of Sociology,* 106 (2), 424-46.

Storper, M. and S. Christopherson (1987), "Flexible specialization and regional industrial agglomerations: The case of the U.S. motion picture industry," *Annals of the Association of American Geographers* 77(1), 104-117

Strange, W. C., W. Hejazi and J. Tang, "The Uncertain City: Competitive Instability, Skills, Innovation, and the Strategy of Agglomeration," *Journal of Urban Economics* 59(3), 2006, 331-351.

Vernon, R. (1960), *Metropolis 1985*, Harvard University Press, Cambridge, MA, 1960.

Wheeler, C. (2001), "Search, Sorting, and Urban Agglomeration," *Journal of Labor Economics* 19(4), 880-898.

Wheeler, C. (2006), "Cities and the growth of wages among young workers: Evidence from the NLSY," *Journal of Urban Economics* 60, 162-184.

Wolff, Edward E. (2003), "Skills and Changing Comparative Advantage," *Review of Economics and Statistics* 8 (1), 77-93.

Table 1. Data Description

Panel A. Description of Soft Skill Measures from the Dictionary of Occupational Titles

DOT VARIABLES	DESCRIPTION
depl	adaptability to *dealing with people* beyond giving and receiving instructions
dcp	adaptability to accepting responsibility for *d*irection, *c*ontrol or *p*lanning of an activity
influ	adaptability to *influ*encing people in their opinions, attitudes or judgments about ideas or things
people	complexity at which worker performs job in relation to human beings; also animals dealt with on an individual basis as if they were human. From highest to lowest:
	9. Mentoring: Dealing with individuals in terms of their total personality in order to advise, counsel, and/or guide them with regard to problems that may be resolved by legal, scientific, clinical, spiritual, and/or other professional principles.
	8. Negotiating: Exchanging ideas, information, and opinions with others to formulate policies and programs and/or arrive jointly at decisions, conclusions or solutions.
	7. Instructing: Teaching subject matter to others, or training others (including animals) through explanation, demonstration, and supervised practice; or making recommendations on the basis of technical disciplines.
	6. Supervising: Determining or interpreting work procedures for a group of workers, assigning specific duties to them, maintaining harmonious relations among them, and promoting efficiency. A variety of responsibilities is involved in this function.
	5. Diverting: Amusing others, usually accomplished through the medium of stage, screen, television, or radio.
	4. Persuading: Influencing others in favor of a product, service or point of view.
	3. Speaking-Signaling: Talking with and/or signaling people to convey or exchange information. Includes giving assignments and/or directions to helpers or assistants.
	2. Serving: Attending to the needs or requests of people or animals or the expressed or implicit wishes of people. Immediate response is involved.
	1. Taking Instructions-Helping: Helping applies to "non-learning" helpers. No variety of responsibility is involved in this function.

Panel B. Descriptive Statistics from the 2000 5% IPUMS merged with 1991 DOT

Variable	Mean	Std Dev	p10	p25	p50	p75	p90
Years of Schooling	13.59	2.78	12	12	13	16	16
Less than HS	0.11	0.31	0	0	0	0	1
HS Grad	0.24	0.43	0	0	0	0	1
Some College	0.32	0.47	0	0	0	1	1
College+	0.34	0.47	0	0	0	1	1
depl	0.60	0.40	0	0.2	0.78	0.97	1
dcp	0.36	0.40	0	0	0.17	0.83	1
influ	0.15	0.27	0	0	0	0.15	0.6
people	3.51	1.85	1.37	2	3	4.53	6.52
peoidx	107.95	9.93	94.68	98.32	106.67	116.20	121.61
Industry Share	0.015	0.026	0.001	0.003	0.007	0.016	0.032
ln(MSA pop'n)	14.21	1.16	12.49	13.34	14.32	15.03	15.92
No. of Workers	3,556,261						
No. of MSA's	297						
No. of Industries	264						

Note: Sample includes only workers aged 25-70 with positive hours, weeks, wage income, with identifiable MSA and occupation codes matched to the DOT. Statistics are weighted to reflect U.S. population.

Table 2. Soft Skill Content of Various Occupations

Panel A. Top and Bottom 5 Occupations for Soft Skill Measures

DCP*		INFLU*	
Low	**High**	**Low**	**High**
Janitors	Financial managers	Geologists	Subject instructors (HS/college)
Stevedores	Chief executives	Chemical engineers	Primary schoolteachers
Packers	Speech therapists	Optometrists	Door-to-door sales vendors
Typists	Primary schoolteachers	Industrial engineers	Public relations managers
Dental technicians	Secondary schoolteachers	Bartenders	Advertising sales jobs

DEPL*		PEOPLE	
Low	**High**	**Low**	**High**
Postal Mail Carriers	Speech therapists	Data entry-keyers	Therapists
Movie Projectionists	Salespersons	Machine operators	Secretaries
Ship crews	Funeral directors	Asssemblers	Social Workers
Miners	Subject instructors (HS/college)	Packers	Aministrators
Engravers	Receptionists	Car washers	Salespersons

People Skills Index	
Low	**High**
Bakers	Speech therapists
Oil well drillers	Primary schoolteachers
Upholsterers	Chief executives
Garbage collectors	Physicians
Stevedores	Educational counselors

Panel B. Soft Skills for Select Occupations

	DEPL		INFLU		DCP		PEOPLE
	YES	NO	YES	NO	YES	NO	
Financial Managers	**x**			**x**	**x**		Supervising
Chemical Engineers		**x**		**x**	**x**		Speaking-Signaling
Insurance underwriters		**x**		**x**		**x**	Speaking-Signaling
Mathematicians		**x**		**x**		**x**	Speaking-Signaling
Physicians	**x**			**x**	**x**		Mentoring
Subject Instructors (HS/College)	**x**		**x**		**x**		Instructing
Social Workers	**x**		**x**		**x**		Supervising
Lawyers	**x**		**x**			**x**	Negotiating
Technicians		**x**		**x**		**x**	Servicing
Salesperson	**x**			**x**		**x**	Persuading
Secretaries	**x**			**x**		**x**	Speaking-Signaling
Supervisors	**x**			**x**	**x**		Supervising
Cook		**x**		**x**		**x**	Servicing
Waiter/Waitress	**x**			**x**		**x**	Servicing
Machine Operators		**x**		**x**		**x**	Taking Instructions
Promotion Agents	**x**		**x**		**x**		Instructing

Note: *There are actually more than 5 occupations in the top and bottom, in that several have values equal to 0 or 1. See text for more description of data.

Table 3. Industry Concentration of Soft Skills

DEPL		DCP		INFLU	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**	**BOTTOM 5**	**TOP 5**
MANUFACTURING		**MANUFACTURING**		**MANUFACTURING**	
Apparel, except Knit (23%)	Newspaper Publishing (62.8%)	Apparel, except Knit (16%)	Guided Missiles (46.7%)	Logging (2.1%)	Newspaper Publishing (30.7%)
Tires and Inner Tubes (24.8%)	Computer Equipment (56%)	Carpets and Rugs (16.7%)	Drugs (42.2%)	Ship and Boat Building (3.4%)	Publishing, Except Newspaper (16.3%)
Knitting Mills (25%)	Drugs (53.6%)	Meat Products (17.4%)	Agricultural Chemicals (42%)	Iron and Steel Foundries (4%)	Beverage Industries (14.4%)
Meat Products (26.7%)	Publishing, Except Newspaper (49%)	Knitting Mills (18.8%)	Radio, TV & Communication (41%)	Railroad Equipment (4.1%)	Drugs (14.2%)
Yarn and Fabric Mills (27.4%)	Radio, TV & Communication (49%)	Yarn and Fabric Mills (20.7%)	Computer Equipment (41%)	Aircraft and Parts (4.2%)	Paints, Varnishes & Related (12.8%)
ALL INDUSTRIES		**ALL INDUSTRIES**		**ALL INDUSTRIES**	
Shoe Repair Shops (18.7%)	Barber Shops (98%)	Barber Shops (2.5%)	Elem. and Sec. Schools (74.3%)	Barber Shops (0.3%)	Elem. and Sec. Schools (55.5%)
Fishing, Hunting, Trapping (21.1%)	Offices of Dentists (93%)	Taxicab Service (7.4%)	Educational Services (64.8%)	Taxicab Service (0.9%)	Direct Selling Establishments (43.3%)
Miscellaneous Repair Services (22%)	Beauty Shops (90%)	Beauty Shops (7.5%)	Child Care Services (64.5%)	Offices of Dentists (1%)	Colleges and Universities (39.4%)
Apparel, except Knit (23%)	Child Care Services (89%)	U.S. Postal Service (9.9%)	Colleges and Universities (58.1%)	Beauty Shops (1.1%)	Drugs (39.3%)
Automotive Repair (24%)	Elem. And Sec. Schools (86%)	Shoe Repair Shops (11%)	Management Services (57.8%)	U.S. Postal Service (1.4%)	Religious Organizations (36.9%)

PEOPLE		PEOPLE INDEX	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**
MANUFACTURING		**MANUFACTURING**	
Apparel, except Knit (2.15)	Newspaper Publishing (3.52)	Apparel, except Knit (99.8)	Newspaper Publishing (108.2)
Meat Products (2.3)	Drugs (3.4)	Meat Products (100.5)	Drugs (107.5)
Knitting Mills (2.31)	Agricultural Chemicals (3.31)	Knitting Mills (100.6)	Computer Equipment (107.4)
Tires and Inner Tubes (2.31)	Computer Equipment (3.3)	Tires and Inner Tubes (100.7)	Agricultural Chemicals (106.7)
Carpets and Rugs (2.35)	Paints, Varnishes and Related (3.2)	Carpets and Rugs (101)	Guided Missiles and Parts (106.4)
ALL INDUSTRIES		**ALL INDUSTRIES**	
Taxicab Service (2)	Religious Organizations (5.8)	Shoe Repair Shops (98.9)	Elem. and Sec. Schools (119.4)
Shoe Repair Shops (2.1)	Elem. & Sec. schools (5.76)	Taxicab Service (99)	Educational services (116.4)
Barber Shops (2.1)	Education Services (5.3)	Miscellaneous Repair Services (99.7)	Religious Organizations (115.3)
Apparel, except Knit (2.15)	Colleges and Universities (5)	Apparel, except Knit (99.8)	Colleges and Universities (115.2)
Miscellaneous Repair Services (2.2)	Legal services (4.9)	U.S. Postal Service (100.5)	Child Care Services (114.7)

Note: Percentages reflect percent of workers in that industry requiring that skill; other figures in parentheses are means. Statistics are weighted to add up to the U.S. population.

Table 4. Location Concentration of Soft Skills
Panel A. Among ALL MSA's

DEPL		DCP	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**
Kokomo, IN (46.4%)	Stamford, CT (70.6%)	Williamsport, PA (25.9%)	Bryan-College Station, TX (45.7%)
Hickory-Morgantown, NC (46.7%)	Barnstable-Yarmouth, MA (67.7%)	Flint, MI (26.1%)	Boulder-Longmont, CO (44.6%)
Janesville-Beloit, WI (47.2%)	Tallahassee, FL (67.3%)	Danville, VA (26.8%)	Stamford, CT (44%)
Danville, VA (47.3%)	Boulder-Longmont, CO (65.9%)	Waterbury, CT (27.3%)	Santa Fe, NM (43.6%)
Decatur, AL (48.3%)	Boston, MA (65.8%)	Hickory-Morgantown, NC (27.6%)	Danbury, CT (43.3%)

INFLU		PEOPLE	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**
Kokomo, IN (9.1%)	Stamford, CT (20.9%)	Kokomo, IN (2.9)	Stamford, CT (3.97)
Danville, VA (10%)	Bloomington, IN (20.7%)	Flint, MI (2.97)	Bryan-College Station, TX (3.96)
Flint, MI (10.2%)	Bryan-College Station, TX (20.5%)	Danville, VA (2.98)	Iowa City, IA (3.95)
Jacksonville, NC (10.5%)	State College, PA (19.9%)	Hickory-Morgantown, NC (3)	Gainesville, FL (3.94)
Hickory-Morgantown, NC (10.8%)	Champaign-Urbana-Rantoul, IL (19.4%)	Janesville-Beloit, WI (3.1)	Columbia, MO (3.89)

PEOPLE INDEX	
BOTTOM 5	**TOP 5**
Kokomo, IN (104.6)	Stamford, CT (110.68)
Danville, VA (104.7)	Bryan-College Station, TX (110.2)
Hickory-Morgantown, NC (104.8)	Iowa City, IA (110)
Flint, MI (104.97)	Gainesville, FL (110)
Janesville-Beloit, WI (105)	Tallahassee, FL (109.9)

Note: Percentages reflect percent of workers in that industry requiring that skill; other figures in parentheses are means. Statistics are weighted to add up to the U.S. population.

Table 4. Location Concentration of Soft Skills (continued)

Panel B. Among MSA's of At Least 1 Million People

DEPL		DCP	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**
Greensboro-Winston Salem-High Point, NC (55.7%)	Boston, MA (65.8%)	Las Vegas, NV (30%)	San Jose, CA (40.9%)
Norfolk-VA Beach-Newport News, VA (56.1%)	Nassau Co, NY (65.6%)	Greensboro-Winston Salem-High Point, NC (32.8%)	Boston, MA (40.8%)
Detroit, MI (56.7%)	San Francisco-Oakland-Vallejo, CA (65.4%)	Cleveland, OH (32.9%)	Washington, DC/MD/VA (40.8%)
Riverside-San Bernardino, CA (56.9%)	Monmouth-Ocean, NJ (65%)	Providence-Fall River-Pawtucket, MA/RI (32.9%)	Austin, TX (40.3%)
Providence-Fall River-Pawtucket, MA/RI (58%)	New York-Northeastern NJ (65%)	Pittsburgh-Beaver Valley, PA (33.3%)	San Francisco-Oakland-Vallejo, CA (40.1%)

INFLU		PEOPLE	
BOTTOM 5	**TOP 5**	**BOTTOM 5**	**TOP 5**
Las Vegas, NV (11.8%)	Nassau Co, NY (17.7%)	Las Vegas, NV (3.25)	Boston, MA (3.76)
Greensboro-Winston Salem-High Point, NC (13.2%)	Austin, TX (17.1%)	Greensboro-Winston Salem-High Point, NC (3.33)	Nassau Co, NY (3.76)
Detroit, MI (13.3%)	Boston, MA (17%)	Detroit, MI (3.34)	Washington, DC/MD/VA (3.74)
Norfolk-VA Beach-Newport News, VA (13.4%)	San Francisco-Oakland-Vallejo, CA (16.8%)	Riverside-San Bernardino, CA (3.35)	San Francisco-Oakland-Vallejo, CA (3.71)
Jacksonville, FL (13.6%)	Orange County, CA (16.7%)	Providence-Fall River-Pawtucket, MA/RI (3.38)	Monmouth-Ocean, NJ (3.66)

PEOPLE INDEX	
BOTTOM 5	**TOP 5**
Las Vegas, NV (106.8)	Boston, MA (109.5)
Greensboro-Winston Salem-High Point, NC (106.8)	Nassau Co, NY (109.3)
Detroit, MI (107)	Washington, DC/MD/VA (109.3)
Riverside-San Bernardino, CA (107.1)	San Francisco-Oakland-Vallejo, CA (109.2)
Providence-Fall River-Pawtucket, MA/RI (107.2)	Monmouth-Ocean, NJ (108.9)

Note: Percentages reflect percent of workers in that industry requiring that skill; other figures in parentheses are means. Statistics are weighted to add up to the U.S. population.

Table 5: Mean Regressions: Manufacturing Sectors - 2000

Panel A: Education Measures

	Less Than HS	HS Grad.	Some Coll.	Coll. Grad.	Years Educ.
Log(Pop)	0.028	-0.032	-0.011	0.016	-0.105
	[0.002]***	[0.001]***	[0.001]***	[0.002]***	[0.017]***
Cluster	-0.194	-0.382	-0.017	0.594	2.979
	[0.126]	[0.055]***	[0.083]	[0.118]***	[1.261]**
R-squared	0.07	0.04	0.01	0.09	0.11

Panel B: Soft Skills

	depl	influ	dcp	peoidx	People
Log(Pop)	0.021	0.011	0.008	0.438	0.069
	[0.001]***	[0.001]***	[0.001]***	[0.032]***	[0.005]***
Cluster	-0.068	-0.176	0.263	0.334	-0.198
	[0.115]	[0.046]***	[0.088]***	[2.826]	[0.464]
R-squared	0.04	0.02	0.04	0.04	0.03

Notes: Standard errors in brackets are clustered by industry/MSA. * significant at 10%; ** significant at 5%; *** significant at 1%. Dependent variables in Panel A for first four columns are indicator variables of a worker's highest education level. Years of schooling is created from reported categorical schooling. Regressions also include industry fixed effects. No of observations: 521344. See text for further details.

Table 6: Mean Regressions: Selected Services Sectors - 2000

Panel A: Education Measures

	Less Than HS	HS Grad.	Some Col.	Col. Grad.	Years Educ.
Log(Pop)	0.002	-0.015	-0.016	0.028	0.101
	[0.001]***	[0.001]***	[0.003]***	[0.003]***	[0.012]***
Cluster	0.006	-0.026	-0.331	0.351	1.378
	[0.029]	[0.062]	[0.169]*	[0.171]**	[0.663]**
R-squared	0.02	0.03	0.02	0.07	0.11

Panel B: Soft Skills

	depl	influ	dcp	Peoidx	People
Log(Pop)	0.006	0.003	0.009	0.198	0.032
	[0.001]***	[0.001]***	[0.001]***	[0.028]***	[0.005]***
Cluster	-0.133	-0.086	-0.096	-3.585	-0.634
	[0.035]***	[0.042]**	[0.051]*	[1.014]***	[0.19]***
R-squared	0.11	0.11	0.07	0.05	0.08

Notes: Standard errors in brackets are clustered by industry/MSA. * significant at 10%; ** significant at 5%; *** significant at 1%. Dependent variables in Panel A for first four columns are indicator variables of a worker's highest education level. Years of schooling is created from reported categorical schooling. Regressions also include industry fixed effects. No of observations: 535342. See text for further details.

Table 7: "Hard" Skills Mean Regressions: Manufacturing Sectors - 2000

Panel A: Cognitive Skills

	Cog Index	GED-M	GED-R	GED-L
Log(Pop)	0.003	0.028	0.037	0.047
	[0.0004]***	[0.005]***	[0.004]***	[0.005]***
Cluster	0.112	1.282	1.006	1.279
	[0.032]***	[0.353]***	[0.315]***	[0.408]***
R-squared	0.13	0.13	0.12	0.13

Panel B: Physical Skills

	Mot Index	Things	Strength	STS
Log(Pop)	-0.003	-0.057	-0.031	-0.012
	[0.0002]***	[0.005]***	[0.002]***	[0.001]***
Cluster	0.062	0.771	-0.531	0.158
	[0.011]***	[0.244]***	[0.178]***	[0.071]**
R-squared	0.04	0.03	0.1	0.03

Notes: Standard errors in brackets are clustered by industry/MSA. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions also include industry fixed effects. No of observations: 521344. Variables under cognitive and physical skills are further described in the Web Appendix.

Table 8: Skill Distribution Regressions for Manufacturing – 2000

	Years of Education						
	p10	p25	p50	p75	P90	P90-P10	P75-P25
Log(Pop)	-0.493	-0.133	0.066	0.192	0.219	0.713	0.325
	[0.111]***	[0.06]**	[0.023]***	[0.028]***	[0.022]***	[0.118]***	[0.048]***
Cluster	-2.904	-2.952	2.393	2.270	1.016	3.919	5.222
	[4.647]	[2.475]	[1.26]*	[1.245]*	[1.078]	[4.746]	[2.153]**
	DEPL						
	p10	p25	p50	p75	P90	P90-P10	P75-P25
Log(Pop)	-0.006	0.000	0.054	0.081	0.021	0.027	0.081
	[0.001]***	[0.001]	[0.007]***	[0.007]***	[0.002]***	[0.002]***	[0.005]***
Cluster	0.216	0.178	-0.781	-1.410	-0.187	-0.403	-1.588
	[0.068]***	[0.073]**	[0.250]***	[0.237]***	[0.086]**	[0.062]***	[0.238]***
	INFLU						
	p10	p25	p50	p75	P90	P90-P10	P75-P25
Log(Pop)	-0.002	-0.002	-0.002	0.017	0.076	0.078	0.019
	[0.0004]***	[0.0004]***	[0.0003]***	[0.002]***	[0.008]***	[0.008]***	[0.002]***
Cluster	0.102	0.102	0.070	-0.401	-2.039	-2.141	-0.503
	[0.021]***	[0.021]***	[0.023]***	[0.059]***	[0.225]***	[0.24]***	[0.083]***
	DCP						
	p10	p25	p50	p75	P90	P90-P10	P75-P25
Log(Pop)	-0.003	-0.003	0.009	0.060	0.026	0.029	0.064
	[0.001]***	[0.001]***	[0.003]***	[0.008]***	[0.003]***	[0.003]***	[0.008]***
Cluster	0.135	0.157	0.242	-0.984	-0.192	-0.327	-1.141
	[0.078]*	[0.073]**	[0.142]*	[0.339]***	[0.119]	[0.111]***	[0.403]***
	PEOPLE VARIABLE						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Pop)	-0.007	0.022	0.141	0.268	0.108	0.116	0.246
	[0.004]*	[0.007]***	[0.018]***	[0.025]***	[0.015]***	[0.016]***	[0.022]***
Cluster	0.674	0.584	-1.919	-5.986	-2.118	-2.792	-6.569
	[0.221]***	[0.329]*	[0.650]***	[0.761]***	[0.676]***	[0.699]***	[0.735]***
	PEOPLE INDEX						
	p10	p25	P50	p75	p90	P90-P10	P75-P25
Log(Pop)	-0.086	0.076	0.992	1.705	0.704	0.791	1.630
	[0.023]***	[0.046]*	[0.134]***	[0.167]***	[0.071]***	[0.076]***	[0.126]***
Cluster	4.566	5.412	-12.877	-33.973	-5.315	-9.881	-39.384
	[1.713]***	[1.968]***	[4.326]***	[4.843]***	[2.877]*	[2.799]***	[5.398]***

Note: Boostrapped standard errors in brackets clustered at the MSA level. * significant at 10%; ** significant at 5%; *** significant at 1%. No of observations: 4852 industry-MSA's. Dependent variables are adjusted for industry fixed effects and are industry-MSA percentiles. That is, we first regress individual worker skills on industry fixed effects and indicators for worker's percentile on that skill's industry-MSA distribution. The second stage (reported above) is an FGLS regression of the percentile coefficients on population, cluster, and a constant. See text for further details.

Table 9: "Hard" Skills Distribution Regressions: Manufacturing Sector - 2000

Panel A: Cognitive Skills

	Cognitive Index						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	0.003	0.005	0.009	0.009	0.006	0.003	0.004
	[0.001]***	[0.001]***	[0.001]***	[0.001]***	[0.001]***	[0.0006]***	[0.001]***
Cluster	0.014	-0.061	-0.065	0.026	0.133	0.119	0.087
	[0.026]	[0.028]**	[0.056]	[0.052]	[0.037]***	[0.035]***	[0.039]**
	GED-M						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	0.024	0.049	0.081	0.075	0.054	0.030	0.026
	[0.006]***	[0.009]***	[0.012]***	[0.014]***	[0.011]***	[0.010]***	[0.008]***
Cluster	-0.017	-0.417	-0.624	0.713	1.788	1.806	1.130
	[0.228]	[0.330]	[0.611]	[0.549]	[0.593]***	[0.572]***	[0.352]***
	GED-R						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	0.041	0.049	0.092	0.093	0.064	0.023	0.044
	[0.007]***	[0.009]***	[0.013]***	[0.014]***	[0.008]***	[0.006]***	[0.007]***
Cluster	-0.029	-0.657	-0.645	0.184	1.135	1.165	0.841
	[0.271]	[0.239]***	[0.513]	[0.479]	[0.312]***	[0.341]***	[0.437]*
	GED-L						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	0.030	0.056	0.115	0.134	0.083	0.054	0.078
	[0.007]***	[0.010]***	[0.015]***	[0.015]***	[0.009]***	[0.009]***	[0.011]***
Cluster	0.194	-0.475	-0.803	-0.202	0.915	0.721	0.274
	[0.33]	[0.433]	[0.702]	[0.585]	[0.405]**	[0.367]**	[0.409]

Notes: Boostrapped standard errors in brackets clustered at the MSA level. * significant at 10%; ** significant at 5%; *** significant at 1%. Dependent variables are further described in the Web Appendix. No of observations: 4852 industry-MSA's. Dependent variables are adjusted for industry fixed effects and are industry-MSA percentiles. That is, we first regress individual worker skills on industry fixed effects and indicators for worker's percentile on that skill's industry-MSA distribution. The second stage (reported above) is an FGLS regression of the percentile coefficients on population, cluster, and a constant. See text for further details.

Table 9: "Hard" Skills Distribution Regressions: Manufacturing Sector – 2000 (continued)

Panel B: Physical Skills

	Motor Index						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	-0.002	-0.006	-0.004	-0.001	0.000	0.002	0.005
	[0.0004]***	[0.001]***	[0.0006]***	[0.0005]**	[0.0006]	[0.0007]***	[0.001]***
Cluster	0.082	0.102	0.044	0.033	0.037	-0.045	-0.069
	[0.023]***	[0.023]***	[0.030]	[0.025]	[0.028]	[0.035]	[0.039]*
	THINGS						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	-0.022	-0.206	-0.107	0.000	0.015	0.037	0.206
	[0.006]***	[0.022]***	[0.016]***	[0.009]	[0.011]	[0.014]***	[0.024]***
Cluster	0.679	3.049	1.725	-0.487	0.114	-0.565	-3.536
	[0.354]*	[0.738]***	[0.844]**	[0.308]	[0.486]	[0.593]	[1.029]***
	STRENGTH						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	-0.040	-0.079	-0.064	-0.031	-0.019	0.021	0.048
	[0.004]***	[0.007]***	[0.008]***	[0.006]***	[0.004]***	[0.004]***	[0.006]***
Cluster	0.394	0.854	-0.148	-0.492	0.025	-0.369	-1.346
	[0.266]	[0.396]**	[0.283]	[0.212]**	[0.167]	[0.359]	[0.515]***
	STS						
	p10	p25	p50	p75	p90	P90-P10	P75-P25
Log(Popn)	-0.021	-0.039	-0.033	-0.002	0.006	0.027	0.036
	[0.002]***	[0.004]***	[0.003]***	[0.001]*	[0.001]***	[0.002]***	[0.003]***
Cluster	0.278	0.694	0.748	0.067	-0.006	-0.285	-0.628
	[0.100]***	[0.123]***	[0.137]***	[0.054]	[0.063]	[0.085]***	[0.104]***

Notes: Boostrapped standard errors in brackets clustered at the MSA level. * significant at 10%; ** significant at 5%; *** significant at 1%. Dependent variables are further described in the Web Appendix. No of observations: 4852 industry-MSA's. Dependent variables are adjusted for industry fixed effects and are industry-MSA percentiles. That is, we first regress individual worker skills on industry fixed effects and indicators for worker's percentile on that skill's industry-MSA distribution. The second stage (reported above) is an FGLS regression of the percentile coefficients on population, cluster, and a constant. See text for further details.