How far do agglomeration effects reach?

Joel A. Elvery

Cleveland State University

j.elvery@csuohio.edu

Leo Sveikauskas

Bureau of Labor Statistics

Sveikauskas.Leo@bls.gov

July 2010

Abstract:

While there is a long literature on agglomeration effects, there is still a need for empirical evidence that helps us understand which sources of agglomeration benefits are most important. We argue that one way to gain such insight is to study agglomeration effects using detailed units of geography. This paper uses tract-level data from 377 metropolitan areas on employment and wages to determine the relative importance of agglomeration at different distances. We find the strongest agglomeration effects in the range of 5 to 20 miles, followed by metropolitan area, and then tracts. This suggests that labor market pooling and customer/supplier linkages are the leading source of agglomeration benefits, with idea exchange playing a role but being restricted to extremely short distances.

Three different recent reviews of the literature on agglomeration (Duranton and Puga (2004), Rosenthal and Strange (2004), Puga (2010)) each starts out with the observation that the concentration of production in cities and the economic dominance of urban production essentially rests on agglomeration economies. As Duranton and Puga (2004, page 2065) remark "We regard [agglomeration economies] as one of the fundamental quests of urban economics." Furthermore, "it is only by studying what gives rise to urban agglomeration economies … that we gain any real insight into why there are cities." In view of the importance of agglomeration economies, it is perhaps surprising that current understanding of agglomeration is, as each of the three reviews makes clear, still relatively modest. The quest for understanding agglomeration still has great distances to go.

One of the main developments in recent work on agglomeration has been the accumulating evidence that important agglomeration economies take place over remarkably short distances, often a mile or less (Rosenthal and Strange (2003); van Soest, Gerking, and van Oort (2006); Fu (2007); Arzaghi and Henderson (2008)). Evidence along these lines suggests that face-to-face contact and interaction with nearby colleagues is an important element in the overall advantages of cities. Such results also point the way towards a greater emphasis on learning and on the development and diffusion of ideas as a root cause of agglomeration benefits.

Nevertheless, most work on agglomeration still has been conducted on relatively large geographic areas. For example, Baumgardner (1988) examined division of labor effects at the county level. Holmes (1999) examined the effects of industrial concentration within circles with a 50 mile radius. Overman and Puga (2010) examine the effects of labor pooling within labor market areas.[1]

---

[1] Overman and Puga (2010) examine data within 308 Travel to Work Areas in the United Kingdom, which cover the entire country. They remark that these areas are similar to the areas the Bureau of Labor Statistics defines for the United States.

Similarly, Ellison, Glaeser, and Kerr (2007; 2010) study the effects of coagglomeration with data for metropolitan areas and counties.

The question then naturally arises as to how we are to integrate the new emphasis on short distance agglomeration economies with existing evidence which is typically based upon larger units of observation. To estimate agglomeration effects jointly over many different distances we adopt hierarchical linear models (HLMs) with random effects (Gelman and Hill (2007)). A HLM can be thought of as two or more separate least squares regressions combined into a single maximum likelihood model where each part of the model represents a different level of aggregation, such as tract and metropolitan area.

In addition to these agglomeration advantages which occur at the Census tract level or within other very small areas, agglomeration could also plausibly occur within urban subcenters (Redfearn (2007), Elvery and Sveikauskas (2009)), in circles of varying radius from 2.5 to 50 miles[2] or, more generally, within metropolitan standard areas. HLM can provide comparable estimates of the effect of employment density as measured within each of these different levels of geographic aggregation.

We motivate our work in terms of a recent influential contribution, Ellison, Glaeser, and Kerr (2007; 2010).[3] They distinguish between three separate influences on the coagglomeration of different industries. The first element is proximity to customers and suppliers, as measured by input-output linkages. The second is labor market pooling, as measured by the extent to which industries

---

As of March 2010, the Bureau defined 380 metropolitan areas, 590 micropolitan areas and 1,364 small market areas for the United States. These exhaust the entire geography of the United States.

2 Rosenthal and Strange (2008) examine the effects of concentric circles extending out as far as 50 miles.

3 Studies of agglomeration can rest on the formation and growth of new firms, on wages and rents, or on measures of the coagglomeration of different industries. We mention Ellison, Glaeser, and Kerr (2007; 2010) here because this work clearly illustrates the types of concerns we are interested in. However, the empirical work of this paper examines wages rather than the coagglomeration of industries.

share a similar mix of occupations. The third factor is the role of ideas, as indicated by interindustry patent citations or measures of research and development spillovers across industries. Ellison, Glaeser, and Kerr test for such links within data for metropolitan areas or counties.

The problem is that contact with customers or suppliers is a classic source of new ideas and technical change. Similarly, workers in the same occupation are a frequent source of new ideas. Consequently, it is difficult to separate the effect of measures of input-output linkages or of similarities in occupational structure from the presence of ideas. The new empirical results mentioned above suggest that measures of such matters as input-output connections or occupational similarities will be especially closely intertwined with the role of ideas at the very short distances where ideas are likely to be exchanged. It is therefore necessary to determine how great these short distance effects are compared with the traditional longer distance effects.

To achieve this, we estimate tract-level median earnings regressions to estimate the returns to density at varying distances. Short distance effects, in a single Census tract or within a circle of radius 2.5 miles, have certain characteristics which point to the importance of information flow over very short distances. For example, an interaction term between employment density and the education of the work force has a positive impact over these short distances. Furthermore, short distance spillover effects are greater when the nearby employment which generates the spillover falls into the same occupational categories as the recipient of the spillover.[4]

Concentric circles illustrating the effect of further nearby employment increase up to a range of 10 miles. Even after 10 miles, circles of 20, 25, or 30 miles still consistently show positive effects. These findings corroborate the positive effects at long distance reported in Rosenthal and Strange

[4] Subcenter effects, based upon subcenter boundaries such as those established by Redfearn (2007) and Elvery and Sveikauskas (2009), often do not perform as well as concentric circles. There is little evidence that agglomeration effects are systematically stronger within subcenter boundaries.

(2008). One might consider such long-range effects, of employment within 25 miles, largely reflect the high commuting distances found in very large cities like New York, Chicago, Boston, and Washington. However, this is not the case; even in metropolitan areas with low population, measures of employment within circles of 20 or 30 miles still have a positive effect on observed wages.

One of the main findings of this paper is that, though the density of employment at distances below 2.5 miles generally has a positive impact on wages, employment density at substantially greater distances, such as 10, 20, or 25 miles, typically has a much greater influence on observed wages. Though our new estimates of short distance effects are usually significant, the summation of the different longer range effects is generally substantially greater.

Finally, measures of employment within a metropolitan area, such as the density of employment, also have a further impact on observed wages. Overall, the different effects observed at these different ranges demonstrate that it is possible to establish the effects of short distance information flows, intermediate distance economies, and broader metropolitan effects jointly in a single regression analysis.

I. Previous literature examining agglomeration effects at different distances

This section briefly reviews the relatively small number of studies which have evaluated agglomeration effects at various distances. Rosenthal and Strange (2003) is the first article to investigate the attenuation of agglomeration effects. This work examined how the founding and growth of new firms depended on employment in the same industry. The central result is that the effect of employment attenuates very sharply; employment within a mile is far more important than employment 2 to 5 miles away. Nearby employment is determined from concentric circles.

Employment is divided evenly across the area of each ZIP code, and concentric circles are drawn around the geographic centroid of each ZIP code.

van Soest, Gerking, and van Oort (2006) conduct an analysis of new firm formation in urban portions of the Netherlands. Their results confirm the findings of Rosenthal and Strange (2003) in that employment in the same industry in the same postal zone is by far the strongest influence on new firm formation.

Arzaghi and Henderson (2008) similarly look at the effect of existing firms on the location of new entrants, examining influences on the location of new single-unit firms in the advertising industry in Manhattan. Their central result is that the number of firms within 750 meters, and especially within 500 meters, is an important influence on the location of new firms. After 750 meters, no substantial effect remains.

The two studies which are most closely similar to our work, especially because they both analyze wages, are Rosenthal and Strange (2008) and Fu (2007). Rosenthal and Strange (2008) analyze wages on the basis of information on wages and employment for 1,239 different place of work locations gathered from responses to the 2000 United States Census of Population. The dependent variable is the logarithm of the wage of individual workers employed within the more metropolitan of these locations of work. The results show that the effect of nearby workers is considerably attenuated with distance. In the full sample, the number of workers employed within 0 to 5 miles of a worker has a four or five times greater impact than a worker employed at a 5 to 25 mile distance. Rosenthal and Strange (2008) further report that wages are consistently more responsive to the nearby presence of workers with a college degree; the effect of nearby college workers also attenuates considerably with distance.

Fu (2007) analyzes the effect of a remarkable range of influences on the basis of information gathered for the 1990 Census. Like Rosenthal and Strange, Fu analyzes the log wage of individual workers. However, Fu's sample contains information on individual Census tracts and blocks, and so rests on far more geographic detail than Rosenthal-Strange. On the other hand, Fu's work is based solely on the Boston Metropolitan Statistical Area.

Fu (2007, Section 6.3) reports the rate at which many different influences on wages attenuate with distance. For example, the average education of all workers is a strong effect for distances up to three miles, and thereafter declines sharply. Occupational and industrial density, occupational and industrial employment per kilometer, is a strong influence at very short distances, but insignificant at greater distances. Occupational specialization, the number of workers in a given occupation in a given tract as a proportion of total employment in that tract, decays very rapidly after 1.5 miles. Fu is able to provide such estimates for a broad range of different externality effects. However, Fu does not consider distances beyond 9 miles.[5]

Duranton and Puga (2004, page 37) state that "The forces driving agglomeration at small spatial scales may be quite different from those driving agglomeration at the city level." Rosenthal and Strange (2008) and Fu (2007) report a broad range of results which show how necessary it is to track agglomeration at different distances. Agglomeration economies are different in different locations, and different elements of agglomeration attenuate at different rates. It is crucial to be able to understand the effects of agglomeration at short distances and integrate such results with the long distance effects which have been the dominant theme in most work conducted to date. The

---

[5] Some of the differences between our results and Fu's conclusions may occur may occur because he analyzes only observations in the Boston metropolitan area. This excludes, for example, effects in Southern portions of the Boston area which may be affected by the presence of Providence, Rhode Island. In future work it may be useful to distinguish between agglomeration within one's own metro area and agglomeration beyond the boundaries of one's own area.

methods adopted here offer one way of organizing comparable information on the effect of employment density at different geographic scales.

Finally, Duranton and Puga (2004, pages 36 and 37) consider the issue of the micro-foundations of specifications such as those adopted in this empirical literature.[6] They discuss a spatial decay function (their equation (49)) in which the effect of firms declines with the distance between locations. Duranton and Puga mention several attempts to provide support for the decay of influence with distance and emphasize that it is important to provide deeper micro-foundations for this relationship throughout this literature.

II. The distances associated with different sources of aggolomeration benefits

The premise of this paper is that studying the partial effects of density for varying distances will improve our understanding of the relative importance of the different sources of agglomeration benefits. The theorized sources we consider are idea exchange, proximity to customers and suppliers, and labor market pooling. In this section we explain what distances are associated with each of these sources.

The gains from agglomeration that promote idea exchange and development are likely to occur at relatively short distances. Most ideas can be easily transmitted long distances through print and telecommunications. Such ideas are not likely to be a source of agglomeration benefits because there is no need to be in a dense area to gain from them. However, the kinds of idea and information exchanges that take place in downtowns (or even at conferences) occur over relatively small distances. Therefore, we think that the majority of agglomeration benefits from idea exchange come for within a tract or less than 5 miles away.

6 In addition to the empirical work discussed in this section, Lucas and Rossi-Hansberg (2002) conduct various simulations based on similar measures of distance.

It is more difficult to tie the benefits of being close to customers and suppliers to a specific range of distances. For example, some lawyers locate near court houses to minimize the transporation costs of going to court. Other lawyers locate near corporate centers to have better access to their main customer base. Manufacturers often have international customer bases and look to have supply chains with 200 miles. It seems likely that some customer-supplier linkages occur over very short distances while most operate within the range of 5 to 20 miles.

Labor market pooling is unlikely to have short distance effects. In 2003, only 29% of workers in the US commuted 5 miles or less (Bureau of Transportation Statistics, 2003). 33% commute further than 15 miles. Some labor markets are clearly national and international in scope, while most probably fall within a 10 to 20 mile diameter.

Notice that none of these theories are closely tied to the metropolitan area level of geography. That is one of the motivations for our research. It makes sense to study agglomeration effects using a unit of geography that is connected to theory, but the majority of studies in the literature use metropolitan area as the unit of analysis. It may be that any metro effects above and beyond the ring effects represent the transmission of higher costs from the areas that experience the greatest agglomeration benefits to other nearby areas through the land, housing, and labor markets. However, we see little connection between the theories regarding the sources of agglomeration and the metro area or distances beyond 20 miles.

III. Data and methodology

The 2000 Census Transportation Planning Package (CTPP) contains detailed information on the characteristics of workers employed in each location, as gathered from the 2000 Census of Population. Giuliano and Small (1991) and McMillen and Smith (2003) have used early versions of the CTPP to study employment in subcounty areas. The CTPP provides information on the characteristics of all workers employed in a given area; no information is available for specific individuals.

We use tract-level data from the CTPP because this is the finest level of geographic detail which is available for every metropolitan area. Variables in the CTPP include total employment, mean and median annual earnings, the occupational distribution of employment, the industry distribution of employment, the age of workers, and the types of employers. We supplement the CTPP with further information on the area and centroid of each tract and the distance between nearby tracts.

One limitation of the CTPP is that it does not report the level of education for workers employed in each tract. The inclusion of occupational data mitigates this limitation, since occupation is arguably a less coarse measure of human capital than education. Because it is convenient to characterize the human capital observed in a given tract, we calculate the education expected in a given area from information on occupational employment in 24 distinct occupation groups and further data on the national distribution of education within each of these groups. [7] This procedure assumes that occupations have the same distribution of education in each Census tract as they have nationally.

The metropolitan areas covered in this study are those Core Based Statistical Areas (CBSAs) with a population of 150,000 or greater in 2000.[8] This provides a larger number of metropolitan

7 The distribution of education in each occupational group is obtained from the 5% Public Use Micro Sample of the 2000 Census of Population.

8 Due to data problems, the following CBSAs are excluded from the sample: Boulder, CO; Colorado Springs, CO; Denver-Aurora-Broomfield, CO; Fort Collins – Loveland, CO; Greeley, CO; Heber, UT; Lynchburg, VA; New Orleans-Metairie-Kenner, LA Ogden-Clearfield, UT; Richmond, VA; Roanoke, VA; Rochester, MN; and Salt Lake City, UT.

areas than Rosenthal and Strange (2008), who essentially study large metropolitan areas with 2 or more Place of Work PUMAs, or Fu (2007), whose study is limited to Boston. We use data on CBSAs because their boundaries are based on population and commuting data for 2000.

Because we use data on tracts and cover a large number of metropolitan areas, our work contains information on wages in a large number of distinct geographic locations. We utilize data on 41,824 different tracts in 377 CBSAs. In contrast, Fu (2007) studied 621 Census tracts and 11,395 blocks in the Boston Metropolitan Statistical Area and Rosenthal and Strange (2008) considered workers in 1,239 Place of Work PUMAs, though a smaller number of these geographic areas was actually used in their analysis. Of course, Fu and Rosenthal-Strange also had access to data on individual workers, so that they considered a large number of data points within their relatively small number of geographic locations.

In order to find the relative importance of agglomeration at different distances, we calculate the attributes of rings of tracts at varying distances. Following Fu (2007), we use centroid-to-centroid distances to define the rings. The rings we use are: 0 to 2.5 miles (excluding target tract), 2.5 to 5 miles, 5 to 10 miles, 10 to 15 miles, …, and 45 to 50 miles. One strength of our work is that most rings consist of a number of tracts. For example, 84% of the tracts in the sample have 4 or more tracts in the smallest ring (0 to 2.5 miles); 57% have 10 or more tracts in this ring. This eliminates some of the measurement error inherent in using a larger geography, such as place of work PUMA, to measure small-area employment levels or density.

Duncan dissimilarity indexes are used to measure the similarity of the economic activity in a tract relative to that in the rings of the tract (Duncan and Duncan, 1955). The occupation dissimilarity index indicates the percent of workers in the ring who would have to change occupations for the occupation mix to be the same in the tract and the ring. A parallel occupation dissimilarity index is used to compare tracts with metro areas.

Because our regressions include some CBSA-level variables and we want to simultaneously account for CBSA heterogeneity, we use hierarchical linear models (HLMs) with random effects. In a standard OLS regression with CBSA fixed-effects it is impossible to include CBSA-level covariates due to singularity. However, a regression that does not account for CBSA-level heterogeneity can suffer from misspecification bias. We use HLM with random effects because it is the most straightforward model that lets us include CBSA-level covariates and account for heterogeneity across CBSAs. Furthermore, HLMs give standard errors that account for the fact that the CBSA-level variables are not independent across tracts.[9] A HLM can be thought of as two or more separate least squares regressions combined into a single maximum likelihood model, where each part of the model represents a different level of aggregation. In our case, the model has the form:

$$y_i \sim N\left(\alpha_{j[i]} + \beta X_i, \sigma_y^2\right), \text{for } i = 1, \dots, n$$

$$\alpha_j \sim N\left(\gamma_o + \gamma Z_j, \sigma_\alpha^2\right), \text{for } j = 1, \dots, J$$

where y_i is the log of tract *i* 's median wage, X_i is the set of tract-level explanatory variables (including ring attributes), and Z_j is the set of CBSA-level explanatory variables. The parameters estimated by maximum likelihood are γ_o, σ_α , and the matrices β and γ.

In order to make the results more comparable to what would be found with individual level data, the regressions are weighted by the square root of total employment in a tract.

[9] Gelman and Hill (2007) is a good introduction to HLM.

IV. Results

Table 1 reports baseline results and illustrates the effect of controlling for tract level characteristics. The first column shows baseline results with a complete set of demographic, industry, and occupation controls. (Table A1 of the Appendix reports the full results.) The elasticity of median annual earnings with respect to density is highest for CBSA density, 0.013. The elasticity with respect to tract density is relatively low at 0.004. The first ring, 0 to 2.5 miles, has an insignificant elasticity of 0.001. The elasticities start to increase at the 2.5 to 5 miles range (0.004). After the CBSA effect, the next highest elasticities are at 5 to 10 miles, 10 to 15 miles, and 15 to 20 miles (0.010, 0.008, and 0.008). The elasticities for the rings beyond 20 miles are smaller and decline steadily, becoming insignificant at 40 to 45 miles.

The low elasticities of median earnings with respect to the density of the tract and the two rings below 5 miles is surprising, especially in view of the findings to the contrary in Rosenthal and Strange (2008) and Fu (2007). This may be due to the presence of industry and occupation mix controls. The CTPP gives us geographically detailed information on employment, making it well suited to study agglomeration. However, its chief limitation is that it is not microdata so we can only observe means and medians for the tract. Using tract level statistics is not directly comparable to controlling for individual level characteristics. For example, with individual data one would have a dummy for working in a legal occupation; this would serve as an attribute of the individual, and not as a feature of the area in which people work. With tract level data, we control for the percent of people who are in legal occupations. That means that the control variables account for the type of activity occurring in the tract. Therefore, with tract data it is difficult to distinguish between direct human capital effects and spillover effects due to the composition of the workforce of the tract.

Column 2 of Table 1 provides regression results when we omit the occupation and industry mix control variables and add expected education, which is derived from occupation mix. Figure 1 depicts these results, along with the results from columns 1 and 3. The elasticity of earnings with respect to tract density is 0.008, more than double that from the regression with complete industry and occupation mix controls. However, the elasticities with respect to the density of the two rings covering from 0 to 5 miles both decline, with the elasticity of the 0 to 2.5 mile ring becoming negative and significant. The regression in the third column of Table 1 omits the expected education mix variables as well as the occupation and industry mix variables. The tract density coefficient from this regression is more than four times greater than the estimate from column (1) which controls for occupation and industry mix. Omitting expected education also increases the elasticity estimates for the rings from 2.5 miles to 20 miles, though the changes are not as large as at the tract level. While overcontrolling for tract industry and occupation could explain the relatively low effects of tract density, it cannot explain the low effects for rings within five miles.[10]

Table 2 reports similar regression results using less detailed rings to allow comparison between our results and those of Rosenthal and Strange (2008). The first column includes the target tract in the 0 to 5 mile ring, which is most analogous to Rosenthal and Strange (2008). The basic pattern found with the more detailed rings still holds. The elasticities of earnings with respect to the densities of the CBSA and the rings are positive and significant. The highest elasticities are those for the 5 to 25 mile ring (0.031) and the CBSA as a whole (0.017). The regression in the next column separates the target tract from the 0 to 5 mile ring. Unlike in the prior model with a more complete

[10] Wheaton and Lewis (2002) demonstrated the substantial effect of industrial and occupational concentration and Fu (2007) showed that these influences were very important at small scale. It is therefore quite plausible that the entire set of industry and occupational controls removes an important element of localized spillovers. Note that the coefficient of density of one's own tract increases from .0036 in column (1) to .0080 in column (2) when the implied level of education is used instead of industry and occupation dummies to correct for labor quality. Even this doubling of the tract effect still leaves short distance effects below the cumulative impact of longer distance effects.

The .0208 coefficient for tract density, with no controls, probably greatly overstates the true impact of employment density in one's own area.

set of rings, the elasticity of earnings with respect to the density of the 0 to 5 mile ring is larger than that of the tract (0.007 vs. 0.003). The last two columns use increasingly fewer control variables and show that one reason for the low tract density estimates is that we are controlling for industry and occupation mix.

For the remainder of the paper, we include the occupation and industry mix variables because we are more concerned about omitted variable bias than about the risk of underestimating the agglomeration benefits due to controlling for attributes which may themselves be a function of agglomeration.

Our sample covers nearly all tracts in metropolitan areas with population above 150,000. Tracts that have no other tracts within 2.5 miles are excluded from the sample. These are typically semi-rural tracts at the urban fringe. This means that our sample covers a more complete set of areas than Rosenthal and Strange (2008) or Fu (2007). This allows us to examine whether the elasticity of earnings with respect to each density varies with CBSA size. CBSAs are placed in three categories based on their total employment: Small (employment ≤ 250,000), Medium (250,000 < emp. ≤ 750,000), and Large (750,000 < emp.).

Table 3 reports the regression with interactions of the density measures with CBSA size. The table has coefficients from a single model, but puts the interactions in separate columns. Figure 2 depicts the elasticity of earnings with respect to density for the three CBSA size categories. While not statistically significant at the 5% level, the most noticeable difference is that the elasticity with respect to CBSA density in medium CBSAs is about 40% lower than in large CBSAs. In small CBSAs, the elasticity of earnings with respect to CBSA density is about 0.0003, which is 90% lower than in large CBSAs. This may reflect the fact that small CBSAs tend to have high proportions of rural and semi-rural areas, meaning that the observed CBSA density understates the density most workers in the CBSAs experience. It may also be that tracts in small CBSAs are more economically linked to areas outside their CBSA than tracts in large CBSAs. This interpretation is supported by the fact that the elasticities with respect to distances beyond 25 miles are generally higher for medium and small CBSAs than for large CBSAs; 8 out of the 10 differences in this ranges are statistically significant at at least the 10% level.

Another interesting pattern is that the ring with the peak elasticity is the 10 to 15 mile ring in large CBSAs and the 5 to 10 mile ring in small and medium CBSAs. Since these medium range distance effects are in part due to labor market effects, the fact that the distance peaks at a greater distance in large CBSAs may be due to the longer commutes in those CBSAs. Perhaps the most surprising result is that the elasticity of earnings with respect to tract density is effectively zero in medium CBSAs.

We argue that the density of tracts and nearby rings help capture the importance of knowledge spillovers relative to other theories regarding why there are agglomeration effects. If these short distance effects are driven by knowledge spillovers, then we would expect the effect of these densities to be greater when there are more educated workers because educated workers are more likely to produce and benefit from knowledge (Rauch 1993). Similarly, we would expect knowledge spillovers from nearby tracts to be higher when economic activity in nearby rings is more similar to the activity in the tract. The last two regressions test whether these indicators of knowledge spillovers are present.

Table 4 presents results from a regression of the employment densities interacted with the expected percent of the employees in each area who possess at least a bachelor's degree. Table 4 reports three coefficients for each of the units of geography, with one column for each type of coefficient, with the coefficients for the interactions in the last column. The results show that the elasticities of earnings with respect to density of tract and the 0 to 2.5 mile ring are significantly higher when workers in the tract or ring are more educated. This supports the idea that these short-distance rings capture knowledge spillovers. The elasticities of earnings for the rings from 10 to 25 miles are also significantly higher when workers in these areas are more educated.

The next regression includes interactions of the densities with measures of how dissimilar the occupation mix in the ring or CBSA is to the occupation mix in the tract. The Duncan index ranges from 0 to 100 and is the percent of workers in the ring or CBSA who would have to change occupations in order to have the same occupation mix as the tract. The higher the index, the less similar the activity. Table 5 presents these results and has the same structure as Table 4. The elasticity of earnings with respect to the density of the 0 to 2.5 mile ring is significantly higher when the occupation mix in the ring is more similar to the occupation mix in the tract. This finding supports the idea that the closest ring helps to capture knowledge spillovers because it is logical to expect that the opportunity for knowledge spillovers from nearby firms is greater when the occupation mix is more similar. It is interesting to note that the elasticity of earnings with respect to the densities of the 10 to 15 and 35 to 40 mile rings are higher the less similar the occupation mix of these rings is to the tract's.

We have estimated a number of regressions as robustness checks that we do not include here. These include using log of employment rather than log of density as our agglomeration measure, substituting log density with a weighted mean of density for the ring and CBSA density to measure the density experienced by the average worker in the area, dropping tracts with fewer than 4 tracts in any rings, and dropping tracts in the upper 5% tail of the distribution of ring area. The first two checks examined the effect of using different measures of agglomeration and the last two examined the possibility that measurement error is driving our results. With each of these checks, the core patterns we find in the data remain unchanged. We still find that the units of geography that matter most are the CBSA and the rings from 5 to 20 miles. We also still find that agglomeration within the tract is important, but that agglomeration from 0 to 5 miles is relatively unimportant.

V. Conclusions

This paper analyzes the relationship between annual earnings within 41,824 tracts and the density of the areas surrounding these tracts. The data describe earnings in a large number of geographic locations. Because most tracts have many close neighbors, we can accurately track the influence of agglomeration on earnings.[11]

We began the analysis with the expectation that the results would emphasize the importance of agglomeration at very short distances, as in the influential work of Rosenthal and Strange (2008) and Fu (2007). Some portions of our results do support an emphasis on the development and diffusion of ideas over very short distances. For example, the interaction between the density of employment in a particular area interacted with the proportion of college graduates in that area is a significant influence on the wages observed in a tract at short distances, such as within the tract or at a distance of 0 to 2.5 miles. In addition, a lower Duncan occupational dissimilarity index, which implies a more closely similar labor force, interacts with employment density to provide higher wages at distances of 0 to 2.5 miles.

Nevertheless, the overall results strongly indicate that employment density over a broad range of distances also has a significant effect on observed wages. Employment densities at 5 to 10

---

11 We are aware that the Census Bureau maintains the Sample Edited Detail File, which provides national data of the type used by Fu (2007). We hope eventually to address agglomeration issues within these richer data.

miles, 10 to 15 miles, and 15 to 20 miles all have substantial effects on observed wages. Employment density at greater distances has less of an effect on wages. In our data, these intermediate to long distance effects overshadow the effect of short distance agglomeration. In addition, the density of employment in a metropolitan area has a strong impact on wages, especially in large cities.

Overall, these results suggest that processes which rely on a large proportion of the geographic area of a city – such as relationships with customers and suppliers, sources of specialized inputs, or sorting related to the labor market – play a relatively greater role in urban agglomeration than processes which rely on face-to-face communication over short distances. Such patterns tend to move the wage literature away from an emphasis on short distance communication, as in Rosenthal and Strange (2008) and Fu (2007), so that the picture becomes more balanced and therefore more closely similar to the Ellison, Glaeser, and Kerr (2007; 2010) results in the coagglomeration literature.

References

Arzaghi, Mohammed and Henderson, J. Vernon (2008), "Networking Off Madison Avenue," Review of Economic Studies (October 2008), pp. 1011-1038.

Baumgarner, James R. (1988), "The Division of Labor, Local Markets, and Worker Organization," Journal of Political Economy (June 1988), pp. 509-527.

Duncan, Otis Dudley and Duncan, Beverly (1955), "A methodological analysis of segregation indexes," American Sociological Review, 20(2), pp. 210-217.

Duranton, Gilles and Puga, Diego (2004), "Micro-foundations of urban agglomeration economies," pp. 2063-2117 in Handbook of Regional and Urban Economics (J. Vernon Henderson and Jacques-Francois Thisse, Eds.), North Holland, Amsterdam, 2004.

Ellison, Glenn, Glaeser, Edward L. and Kerr, William (2007), "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," National Bureau of Economic Research Working Paper 13068, April 2007.

Ellison, Glenn, Glaeser, Edward L. and Kerr, William (2010), "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," American Economic Review (June 2010), pp. 1195-2013.

Elvery, Joel A. and Sveikauskas, Leo (2009), "A Cross-Metropolitan Analysis of the Roles of Employment Centers," presented at Regional Science International Association International meetings, Denver, November 2009.

Fu, Shihe (2007), "Smart-Café-Cities: Testing Human Capital Externalities in the Boston Metropolitan Area," Journal of Urban Economics (January 2007), pp. 86-111.

Gelman, Andrew and Hill, Jennifer (2007), Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, Cambridge, 2007.

Giuliano, Genevieve and Small, Kenneth A. (1991), "Subcenters in the Los Angeles Region," Regional Science and Urban Economics (July 1991), pp. 163-182.

Holmes, Thomas J. (1999), "Localization of Industry and Vertical Disintegration," Review of Economics and Statistics (May 1999), pp. 314-325.

Lucas, Robert E. and Rossi-Hansberg, Esteban (2002), "On the Internal Structure of Cities," Econometrica (July 2002), pp. 1445-1476.

McMillen, Daniel P. and Smith, Stefani (2003), "The Number of Subcenters in Large Urban Areas," Journal of Urban Economics (May 2003), pp. 321-338.

Overman, Henry G. and Puga, Diego (2010), "Labor Pooling as a Source of Agglomeration: An Empirical Investigation," pp. 133-150 in Agglomeration Economies (Edward C. Glaeser, Ed.), University of Chicago Press, Chicago, 2010.

Puga, Diego (2010), "The Magnitude and Causes of Agglomeration Economies," Journal of Regional Science (February 2010), pp. 203-219.

Rauch, James E. (1993), "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," Journal of Urban Economics, (November 1993) pages 380-400.

Redfearn, Christian L. (2007), "The Topography of Metropolitan Employment: Identifying Centers of Employment in a Polycentric Urban Area," Journal of Urban Economics (May 2007), pp. 519-541.

Rosenthal, Stuart S. and Strange, William C. (2003), "Geography, Industrial Organization, and Agglomeration," Review of Economics and Statistics (May 2003), pp. 377-393.

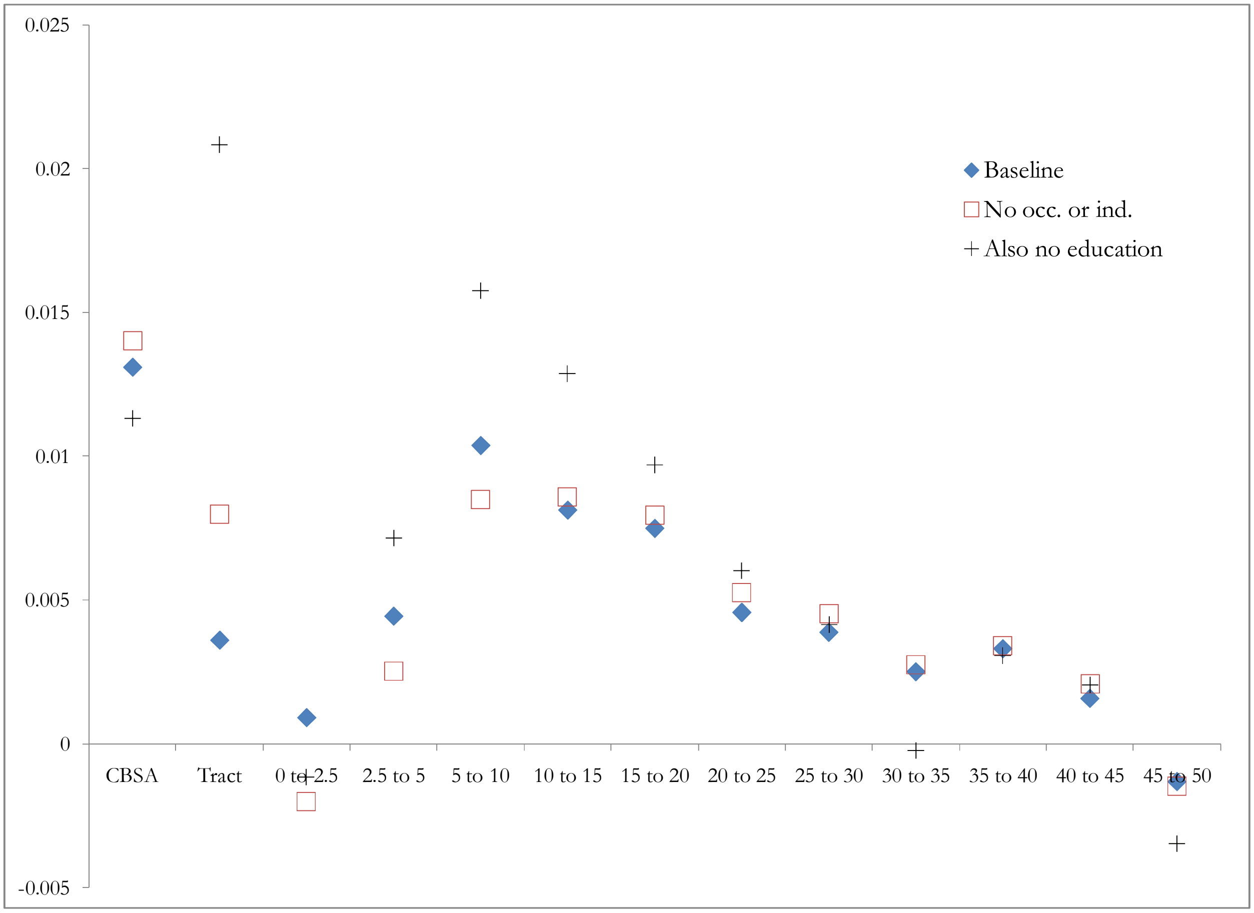
Rosenthal, Stuart S. and Strange, William C. (2004), "Evidence on the Nature and Sources of Agglomeration", pp. 2119-2172 in Handbook of Regional and Urban Economics (J. Vernon Henderson and Jacques-Francois Thisse, Eds.), North Holland, Amsterdam, 2004.

Rosenthal, Stuart S. and Strange, William C. (2008), "The Attenuation of Human Capital Spillovers," Journal of Urban Economics (September 2008), pp. 373-389.

van Soest, Daan P., Gerking, Shelby, and van Oort, Frank G. (2006), "Spatial Impacts of Agglomeration Externalities," Journal of Regional Science (December 2006), pp. 881-899.

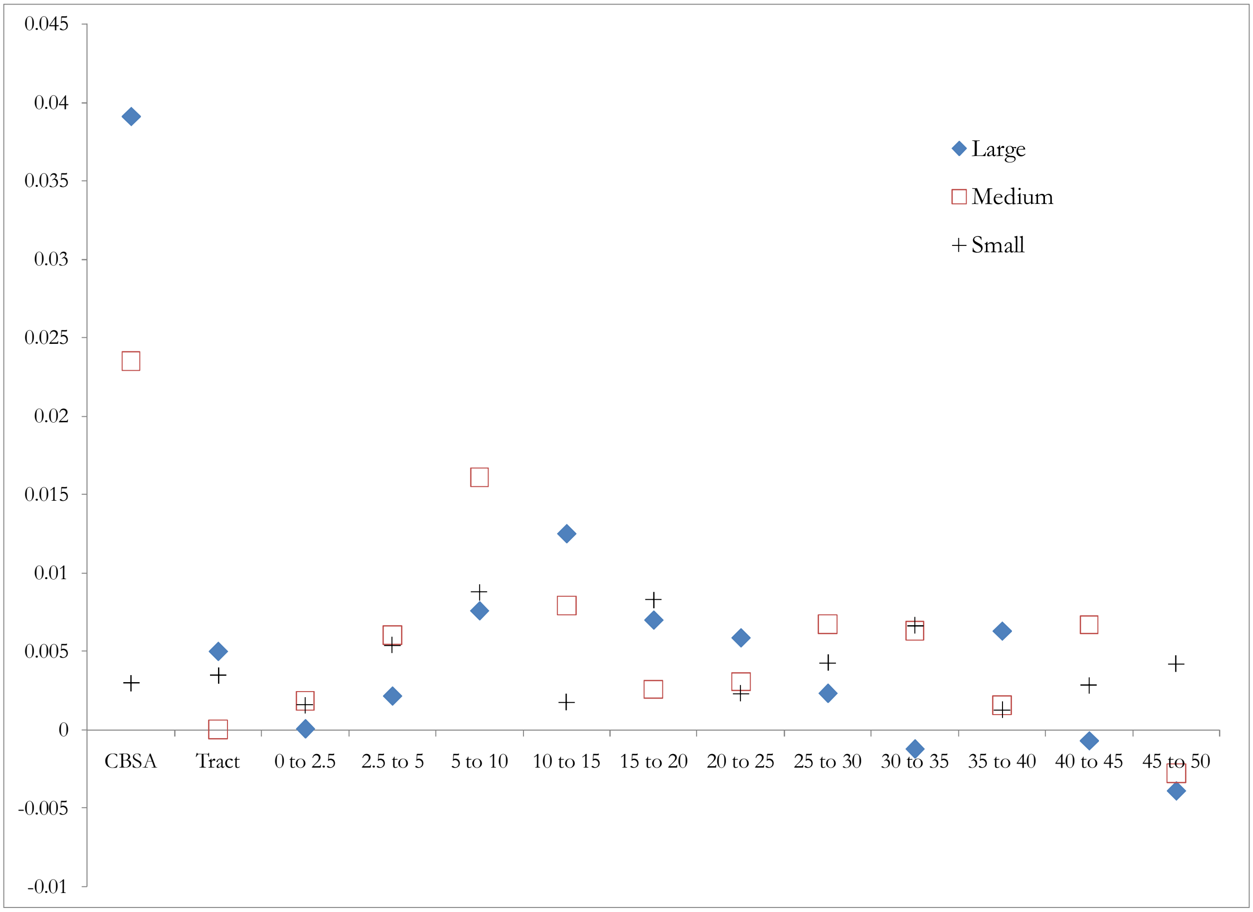
Wheaton, William C. and Lewis, Mark J. (2002), "Urban Wages and Labor Market Agglomeration," Journal of Urban Economics (May 2002), pp. 542-562

Figure 1: Elasticities of median earnings with respect to density



Source and notes: See Table 1.

Figure 2: Elasticities of median earnings with respect to density for large, medium, and small CBSAs



Source and notes: See Table 3.

Table 1: Log of median earning regressions with density of CBSA, tract, and full set of rings

Log density of:	Baseline (1)		Fewer controls (2)		No human capital (3)	
CBSA	0.0131	**	0.0140	*	0.0113	~
	(0.0045)		(0.0047)		(0.0059)	
Tract	0.0036	**	0.0080	**	0.0208	**
	(0.0006)		(0.0006)		(0.0006)	
Rings						
0 to 2.5	0.0009		-0.0020	**	-0.0011	
	(0.0007)		(0.0007)		(0.0008)	
2.5 to 5	0.0044	**	0.0025	**	0.0072	**
	(0.0008)		(0.0008)		(0.0010)	
5 to 10	0.0104	**	0.0085	**	0.0158	**
	(0.0009)		(0.0010)		(0.0011)	
10 to 15	0.0081	**	0.0086	**	0.0129	**
	(0.0010)		(0.0010)		(0.0011)	
15 to 20	0.0075	**	0.0079	**	0.0097	**
	(0.0009)		(0.0010)		(0.0011)	
20 to 25	0.0046	**	0.0053	**	0.0060	**
	(0.0009)		(0.0010)		(0.0011)	
25 to 30	0.0039	**	0.0045	**	0.0041	**
	(0.0009)		(0.0009)		(0.0011)	
30 to 35	0.0025	**	0.0027	**	-0.0002	
	(0.0009)		(0.0009)		(0.0011)	
35 to 40	0.0033	**	0.0034	**	0.0031	**
	(0.0008)		(0.0009)		(0.0010)	
40 to 45	0.0016	~	0.0021	*	0.0020	*
	(0.0008)		(0.0009)		(0.0010)	
45 to 50	-0.0013		-0.0015	~	-0.0035	**
	(0.0008)		(0.0008)		(0.0010)	
Tract level control variables included						
Demographic	Yes		Yes		Yes	
Employer type	Yes		Yes		Yes	
Occupation	Yes					
Industry	Yes					
Expected education			Yes			

Source: Authors calculations from the 2000 CTPP and 2000 5% PUMS.

Notes: 41,824 tracts from 377 CBSAs with population above 150,000. Dependent variable is log of median annual earnings. Standard errors in parentheses. Significance marks: ** = sig. at 1% level or lower, ** = sig. at 5% level or higher, ~ = sig. at 10% level or higher.

Table 2: Log of median earning regressions with less detailed rings for comparison

Log density of:	Tract and first ring combined		Tract and first ring separated		Fewer controls		No human capital	
	(1)		(2)		(3)		(4)	
CBSA	0.0172	**	0.0174	**	0.0178	**	0.0141	**
	(0.0037)		(0.0037)		(0.0039)		(0.0049)	
Tract			0.0030	**	0.0056	**	0.0174	**
			(0.0005)		(0.0005)		(0.0005)	
Rings								
0 to 5	0.0098	**	0.0073	**	0.0031	**	0.0100	**
	(0.0006)		(0.0006)		(0.0006)		(0.0007)	
5 to 25	0.0310	**	0.0313	**	0.0318	**	0.0460	**
	(0.0010)		(0.0010)		(0.0010)		(0.0012)	
25 to 50	0.0053	**	0.0053	**	0.0073	**	-0.0015	
	(0.0012)		(0.0012)		(0.0013)		(0.0014)	
Tract level control variables included								
Demographic	Yes		Yes		Yes		Yes	
Employer type	Yes		Yes		Yes		Yes	
Occupation	Yes		Yes					
Industry	Yes		Yes					
Expected education					Yes			

Source: Authors' calculations from the 2000 CTPP and 2000 5% PUMS.

Notes: 41,824 tracts from 377 CBSAs with population above 150,000. Dependent variable is log of median annual earnings. In column (1), own tract employment and area is included when calculating the employment density of the 0 to 2.5 mile ring. In the remaining columns, the tract density is calculated separately from the 0 to 2.5 ring. Standard errors in parentheses. Significance marks: ** = sig. at 1% level or lower, ** = sig. at 5% level or higher, ~ = sig. at 10% level or higher.

Table 3: Log of median earning regressions with CBSA size interaction

Area	Log density		Log density interacted with medium CBSA indicator		Log density interacted with small CBSA indicator	
CBSA	0.0391	*	-0.0156		-0.0361	~
	(0.0188)		(0.0266)		(0.0199)	
Tract	0.0050	**	-0.0050	**	-0.0015	
	(0.0008)		(0.0013)		(0.0011)	
Rings						
0 to 2.5	0.0001		0.0018		0.0015	
	(0.0010)		(0.0018)		(0.0016)	
2.5 to 5	0.0022	~	0.0039	~	0.0032	~
	(0.0012)		(0.0022)		(0.0018)	
5 to 10	0.0076	**	0.0085	**	0.0012	
	(0.0014)		(0.0026)		(0.0022)	
10 to 15	0.0125	**	-0.0046		-0.0108	**
	(0.0015)		(0.0029)		(0.0023)	
15 to 20	0.0070	**	-0.0044		0.0013	
	(0.0015)		(0.0027)		(0.0022)	
20 to 25	0.0059	**	-0.0028		-0.0036	
	(0.0014)		(0.0025)		(0.0022)	
25 to 30	0.0023	~	0.0044	~	0.0019	
	(0.0013)		(0.0024)		(0.0021)	
30 to 35	-0.0012		0.0075	**	0.0079	**
	(0.0013)		(0.0025)		(0.0020)	
35 to 40	0.0063	**	-0.0048	*	-0.0050	*
	(0.0012)		(0.0023)		(0.0020)	
40 to 45	-0.0007		0.0074	**	0.0035	~
	(0.0011)		(0.0023)		(0.0019)	
45 to 50	-0.0039	**	0.0011		0.0081	**
	(0.0011)		(0.0022)		(0.0018)	

Source: Authors calculations from the 2000 CTPP.

Notes: 41,824 tracts from 377 CBSAs with population above 150,000. Dependent variable is log of median annual earnings. Each column has a set of coefficients from the same regression. Standard errors in parentheses. Significance marks: ** = sig. at 1% level or lower, ** = sig. at 5% level or higher, ~ = sig. at 10% level or higher.

Table 4: Log of median earning regressions with expected education interaction

Area	Log density (1)		% BA in area (2)		Log density interacted with % BA (3)	
CBSA	-0.0433		-0.0083		0.0023	
	(0.0368)		(0.0063)		(0.0014)	
Tract	-0.0081	**	0.0057		0.0004	**
	(0.0016)		(0.0049)		(0.0001)	
Rings						
0 to 2.5	-0.0099	**	-0.0002		0.0003	**
	(0.0024)		(0.0006)		(0.0001)	
2.5 to 5	0.0010		0.0025	**	0.0000	
	(0.0029)		(0.0007)		(0.0001)	
5 to 10	0.0046		0.0021	*	0.0001	
	(0.0037)		(0.0009)		(0.0001)	
10 to 15	-0.0068	~	-0.0013		0.0005	**
	(0.0038)		(0.0008)		(0.0001)	
15 to 20	-0.0069	~	-0.0014	~	0.0005	**
	(0.0036)		(0.0007)		(0.0001)	
20 to 25	-0.0096	**	-0.0014	*	0.0005	**
	(0.0036)		(0.0007)		(0.0001)	
25 to 30	0.0047		0.0004		0.0000	
	(0.0033)		(0.0007)		(0.0001)	
30 to 35	0.0098	**	0.0007		-0.0003	~
	(0.0038)		(0.0007)		(0.0001)	
35 to 40	0.0000		-0.0013	*	0.0002	
	(0.0035)		(0.0007)		(0.0001)	
40 to 45	0.0007		0.0002		0.0000	
	(0.0037)		(0.0007)		(0.0001)	
45 to 50	0.0006		0.0004		-0.0001	
	(0.0033)		(0.0005)		(0.0001)	

Source: Authors calculations from the 2000 CTPP and 2000 5% PUMS.

Notes: 41,824 tracts from 377 CBSAs with population above 150,000. Dependent variable is log of median annual earnings. Each column has a set of coefficients from the same regression. "% BA" refers to the % BA expected based on the occupation mix in the tract. See text for details. Standard errors in parentheses. Significance marks: ** = sig. at 1% level or lower, ** = sig. at 5% level or higher, ~ = sig. at 10% level or higher.

Table 5: Log of median earning regressions with interactions of occupational dissimilarity index

Area	Log density (1)		Occupational dissimilarity index (2)		Log density interacted with index (3)	
CBSA	0.0158	**	0.0008		-0.0001	
	(0.0054)		(0.0006)		(0.0001)	
Tract	0.0024	**				
	(0.0006)					
Rings						
0 to 2.5	0.0058	**	0.0006	*	-0.0002	**
	(0.0015)		(0.0003)		(0.0000)	
2.5 to 5	0.0065	**	-0.0002		-0.0001	
	(0.0019)		(0.0003)		(0.0000)	
5 to 10	0.0097	**	-0.0002		0.0000	
	(0.0020)		(0.0003)		(0.0001)	
10 to 15	0.0019		-0.0007	*	0.0002	**
	(0.0021)		(0.0003)		(0.0001)	
15 to 20	0.0080	**	0.0000		0.0000	
	(0.0021)		(0.0003)		(0.0001)	
20 to 25	0.0085	**	0.0007	*	-0.0001	*
	(0.0021)		(0.0003)		(0.0001)	
25 to 30	0.0044	*	-0.0002		0.0000	
	(0.0020)		(0.0002)		(0.0001)	
30 to 35	0.0008		-0.0004		0.0001	
	(0.0021)		(0.0003)		(0.0001)	
35 to 40	-0.0016		-0.0003		0.0002	**
	(0.0020)		(0.0002)		(0.0001)	
40 to 45	0.0013		-0.0001		0.0000	
	(0.0020)		(0.0002)		(0.0001)	
45 to 50	-0.0029		-0.0001		0.0001	
	(0.0018)		(0.0002)		(0.0000)	

Source: Authors calculations from the 2000 CTPP.

Notes: 41,824 tracts from 377 CBSAs with population above 150,000. Dependent variable is log of median annual earnings. Each column has a set of coefficients from the same regression. The occupational dissimilarity index is higher the more dissimilar occupation mix is from the tract's occupation mix. See text for details. Standard errors in parentheses. Significance marks: ** = sig. at 1% level or lower, ** = sig. at 5% level or higher, ~ = sig. at 10% level or higher.