February 2014

The Urban Density Premium Across Establishments

The Urban Density Premium across Establishments †

R. Jason Faberman, Federal Reserve Bank of Chicago
Matthew Freedman, Cornell University*

February 2014

Abstract

We use longitudinal microdata to estimate the urban density premium for U.S. establishments, controlling for observed establishment characteristics and dynamic establishment behavior. Consistent with previous studies, we estimate a density premium between 6 and 10 percent, even after controlling for establishment composition, local skill mix, and the endogeneity of location choice. More importantly, we find that the estimated density premium is realized almost entirely at birth and is constant over the life of establishments. We find little evidence that the endogenous entry or exit of establishments can account for any of the estimated density premium. We interpret our results as implying that the returns to agglomeration diffuse within a city through a reallocation channel rather than through an increase in the productivity of existing firms.

Keywords: urban density premium, dynamic agglomeration economies, establishment entry and exit
JEL Codes: R12, R30, D22

† We thank Nathaniel Baum-Snow, Laurent Gobillon, and Ronni Pavan as well as seminar participants at the Federal Reserve Bank of Chicago, the Urban Economics Association Meetings, and the Conference for Urban and Regional Economics for helpful comments. The views in this paper are our own and do not reflect the views or the Federal Reserve Bank of Chicago, the Federal Reserve System, or the views of its staff members. In addition, research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the New York and Chicago Census Research Data Centers. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau. This paper has been screened to insure that no confidential data are revealed.
* Corresponding author. Department of Economics, Cornell University, 262 Ives Faculty Building, Ithaca, NY 14853. Phone: 607-254-8245. Email: freedman@cornell.edu.

1. Introduction

For years, urban economists have studied why observationally similar workers earn more in more densely populated locations. Studies have consistently found an elasticity of earnings with respect to urban density of between 3 and 10 percent. This elasticity is generally robust to controlling for a variety of factors, including the migration of skilled workers across cities, the returns to worker experience at a particular location, and labor search and matching frictions.[1] Urban economists have also examined the returns to urban density for firms.[2] Only recent studies, however, examine the relationship between urban density and firm-level outcomes at the micro level.[3] Consequently, there is little evidence on the role firm characteristics, behavior, or composition play in generating the empirical estimate of an urban density premium for employers.

In this paper, we estimate the density premium over the life cycle of U.S. establishments, controlling for a variety of observable establishment characteristics and dynamic behavior that may be endogenously related to urban density. We use microdata from the Census Bureau's Longitudinal Business Database (LBD) to relate the average earnings per worker of an establishment to the population density of its metropolitan area. Our main contributions are our ability to control for a variety of establishment-level characteristics and behavior, including entry and exit, and our ability to estimate the evolution of the density premium over time within establishments.

1 See Glaeser (1999), Glaeser and Mare (2001), Freedman (2008), Bacolod, Blum, and Strange (2009), Glaeser and Resseger (2010), Baum-Snow and Pavan (2012, 2013), and De la Roca and Puga (2012), among others.
2 These include Glaeser et al. (1992), Henderson, Kuncoro, and Turner (1995), Ciccone and Hall (1996), and Henderson (1997).
3 Examples include Henderson (2003), Moretti (2004), and Combes et al. (2012).

Consistent with previous research, we find a positive density premium; average establishment earnings increase between 6 and 10 percent with a doubling of density—depending on the sample, method, and measure used—after controlling for observable characteristics and the endogeneity of establishment location. More importantly, we find that the density premium *does not vary over an establishment's life cycle.* Specifically, we find that the density premium is constant with respect to establishment age, implying that establishments reap most of the returns to agglomeration at entry. The invariance of the density premium to establishment age is robust to a variety of controls, holds after controlling for potential endogeneity issues, and is consistent across several sub-samples of the data.

A return to agglomeration for establishments that is constant over their life cycle is important for two reasons. First, the localized accumulation and diffusion of knowledge (whether it occurs through knowledge spillovers, differential rates of "learning," or some other mechanism) are thought to be a key driver of the urban agglomeration economies that generate a density premium. The notion that the accumulation and diffusion of knowledge affects the innovation, growth, and productivity of firms is pervasive throughout the urban literature (see Audretsch and Feldman, 2004; Rosenthal and Strange, 2004; and Henderson, 2007 for reviews). Theoretical treatments of such returns in an urban context (e.g., Black and Henderson, 1999) are often based on seminal models of endogenous growth such as Romer (1986) and Lucas (1988). In these models, spillovers generate increasing returns for the firm and *accumulate over time*. Theory dictates that firms in denser areas should have faster productivity growth, all else equal, because of these accumulated returns. To the extent that average establishment earnings reflect establishment productivity, this suggests that we should instead find a density premium that rises with the age of an establishment, since accumulated spillovers should make the establishment more productive over time.

Second, several studies have found evidence of accumulated returns to agglomeration for workers (Glaeser and Mare, 2001; De la Roca and Puga, 2012; Baum-Snow and Pavan, 2012). That is, workers exhibit a greater density premium in their wages the longer they reside in a denser location. These studies attribute the resulting steeper wage-city tenure profile to faster human capital accumulation in denser areas (i.e., faster "learning"), an interpretation that is consistent with the implications of the endogenous growth models. This, too, suggests that we should find a density premium that increases with the age of an establishment, since establishments are collections of workers.

Our results do not reject the presence of spillovers, nor do they contradict the evidence on the wage-tenure profiles of workers. They do, however, suggest that the primary channel by which knowledge diffuses and accumulates within a city is a reallocation channel that relies heavily on entrepreneurs and the firm entry margin, not a spillover that directly affects the productivity of existing firms. Indeed, our findings imply that a representative firm framework, like those that postulate the existence of a "knowledge production function" (e.g., Griliches, 1979; Jaffe, 1989), cannot fully capture the dynamics of localized knowledge transmission. Our evidence is instead consistent with a story where knowledge moves across firms within a city through the entry margin. In this sense, we find a strong role for entrepreneurs as the mechanism for innovation and knowledge transmission within a metropolitan area. Similarly, our evidence is consistent with the story of spillovers and entrepreneurship put forth by Audretsch and Feldman (2004, pp. 2728-29), and the model of innovation through entrepreneurial spinoffs by Chatterjee and Rossi-Hansberg (2012). The latter model is particularly compelling because recent work by Klepper

(2010) finds a strong role for spinoffs in forming the Detroit automobile and Silicon Valley high-tech industrial clusters, while work by Wenting (2008) finds a role for localized spinoffs for innovation within the fashion industry.

Further analysis shows that the selection and sorting of establishments into or out of a metropolitan area likely do not explain our results. Establishment entry, exit, and relocation patterns reaffirm that any returns to agglomeration are realized primarily at entry, and suggest that they appear when an establishment is born and not necessarily when it moves to a new city. We find a selection effect in establishment exit patterns: low-earnings establishments are more likely to exit, and the probability of exit decreases monotonically with average establishment earnings. We do not, however, find that the probability of exit differs at any point in the earnings distribution between high-density and low-density metropolitan areas.

Similarly, an examination of the characteristics of establishments at entry and establishments that relocate across metropolitan areas suggests a limited role for sorting. We find that the highest-earnings establishments are most likely to relocate, but that all movers tend to move to *lower-density* metropolitan areas, on average, when compared to a baseline of completely random relocation. Further analysis of relocations suggests that the density premium is realized primarily at birth and not at entry into a new city. A counterfactual simulation that shuts down non-random relocation only slightly reduces our density premium estimate, suggesting that sorting on the relocation margin has little effect on our estimated density premium.

Finally, we examine the relative earnings differences of new, single-unit firms and new establishments of multi-unit firms on the premise that existing firms should be more likely to endogenously choose the location of their new establishments, and therefore have a higher estimated density premium. Our evidence does not suggest a role for sorting—entrants of multi-unit firms have relatively *lower* earnings in high-density metropolitan areas.

The next section describes the data and discusses our approach to measuring key outcomes. In Section 3, we first present our evidence on the urban density premium across establishments, which we show is robust to controlling for a variety of establishment and CBSA characteristics as well as to alternative approaches to addressing measurement and endogeneity concerns. We then go on in Section 3 to present evidence on the density premium over the establishment life-cycle. Section 4 examines to what extent establishment exit, entry, and relocation account for our results. Section 5 concludes.

2. Data and Measurement

We use establishment data from the Longitudinal Business Database (LBD) of the Census Bureau for our analysis.[4] The data include payroll and employment information for nearly every establishment in the U.S. on an annual basis, in addition to a variety of information on the establishment (e.g., industry, location, whether it is part of a multi-establishment firm). Thus, the data provide us with a rich level of detail both within and across metropolitan areas, and allow us to circumvent sampling, scope, and measurement issues that usually plague survey data. We focus on establishments in 1992 and 1997, though we use data from all available years to best identify measures such as entry, exit, and establishment age. We focus on these two years because they are Economic Census years, meaning that the U.S. Census Bureau conducted an extensive census of all businesses, so these years tend to have the most reliable measures of establishment entry and exit. These years are also the last two Census years available that have

[4] For additional details about the LBD, see the web appendix and Jarmin and Miranda (2002).

consistent measures of industry across years; the U.S. changed its classification system from the older Standard Industrial Classification (SIC) system to the North American Industrial Classification System (NAICS) in 1997. We restrict our analysis to private, non-agricultural establishments, giving us 4.9 million observations in 1992 and 5.3 million observations in 1997.

We use the Consolidated Business Statistical Area (CBSA) definition of metropolitan areas as our city-level unit of analysis, focusing only on the Metropolitan Area locations (i.e., we ignore the smaller locations classified as Micropolitan Areas under the CBSA system). This provides us with 363 CBSAs in our study.[5] Our main measure of urban density is 1990 population per square mile, which we calculate for each CBSA by aggregating population and land area data up from the county level. We use the same approach to calculate the share of the 1990 CBSA population with a college degree. We use the college share as a proxy for the average worker skill in a CBSA, but note that it is a crude proxy since it will not capture variation in other skills, both observable and unobservable, that are not related to education. We also check the robustness of our results using alternative measures of population density. These include the 1990 population level (i.e., CBSA size), CBSA employment per square mile (measured using the current year), and an area-weighted measure of population density that weights the density of subunits of a CBSA's land area by the population within that subunit. This measure has been used previously by Glaeser and Kahn (2004) and Rappaport (2008).[6]

We measure entry and exit at the annual frequency. This ensures that all exits measured in 1992 occurred during that year and all entrants measured in 1997 occurred during that year (rather than during the intervening five-year period). We define an entry (exit) as the first (last)

---

[5] These CBSAs roughly correspond to the older definitions of Metropolitan Statistical Areas and Primary Metropolitan Statistical Areas.

[6] We thank Jordan Rappaport for providing us with the area-weighted data.

time an establishment appears with positive employment in the available sample of the LBD, which spans 1975 through 2005. We also measure establishment age using the full LBD sample. An establishment is assigned an initial age of zero years at entry. Since we can only identify age by observing the establishment in the LBD, we topcode age at 16 years (the maximum observed age in 1992) for both years in our sample.

We use payroll per employee as our measure of average earnings at each establishment. Doing so presents us with several measurement issues. First, payroll in the LBD covers all individuals paid during the year but employment is reported for a particular point in time (March of the year). Thus, a standard measure of payroll per employee will tend to overstate the average earnings of establishments that had high worker turnover or were rapidly contracting during the year, and tend to understate the average earnings of establishments that were rapidly expanding during the year. Second, there is the timing of the payroll and employment data. Payroll in the LBD covers all employees paid during the calendar year (January to December). However, employment is measured during the year (in March). Finally, measurement error in either payroll or employment could lead to extreme outliers in the average earnings measure, though since the data are administrative, such measurement error is limited to reporting or input errors on the part of those collecting the data.

To obtain a more accurate measure of earnings, we define the average earnings for an establishment in year t as the total annual payroll in year $t-1$ divided by the average of employment in years $t-1$ and t, or

(1) $$w_{et} = \frac{P_{e,t-1}}{\frac{1}{2}(N_{et}+N_{e,t-1})}$$

where $P_{e,t-1}$ is the total annual payroll of establishment e in year $t - 1$ and N_{et} is the reported employment of establishment e in year t. We define the average earnings of entrants as P_{et}/N_{et} and the average earnings of exits as $P_{e,t-1}/N_{e,t-1}$. We then evaluate these measures for outliers and impute an average earnings measure where necessary. We detail our evaluation and imputation algorithm in the appendix. Finally, we deflate our earnings measures using the Consumer Price Index to 1997 dollars.

Throughout our analysis, we prefer to interpret average establishment earnings as a proxy for establishment productivity. The economic significance of our findings by no means rests on this interpretation, but the interpretation allows a cleaner comparison of our findings with previous research, and related theory, that examines an explicit relationship between agglomeration and productivity. There are several issues with such an interpretation. First, earnings are a cost to the firm as much as they are a rent paid to productive labor. Second, average earnings represent an average across a distribution of workers while productivity is usually thought of as a uniform measure within an establishment.

We check the plausibility of a productivity interpretation of average establishment earnings in several ways. First, we note that our results below on the density premium across the earnings distribution parallel the findings of Combes et al. (2012), who perform a similar exercise with TFP. Namely, we find an urban density premium that increases with average earnings, while Combes et al. find a city size premium that increases with TFP. Second, our results below on establishment exit show that establishments' exit rates decline sharply with average establishment earnings. This runs counter the concern that high-earnings establishments are predominantly high-cost rather than high-productivity establishments, since we expect that the least efficient establishments would be the most likely to exit. Finally, we replicate the results of

Syverson (2004) using our average earnings measure in lieu of total factor productivity (TFP). The Syverson study is a particularly useful because it focuses on the relationship between establishment-level dynamics and local density. Syverson suggests that locations with a greater density of demand will have greater competition among local firms, and consequently, greater exit rates among their low-productivity firms. He tests and affirms the main implications of his model by looking at differences in the TFP distributions of plants in the ready-mix concrete industry across areas with different construction employment densities (the construction industry is the primary consumer of ready-mix concrete). He focuses on several key moments of the TFP distribution within each geographic area, and regresses each separately on (log) density. He finds that areas with greater demand density have a less disperse TFP distribution that exhibits greater lower truncation. These areas also have higher average TFP, larger plants, and a lower producer demand ratio. Table 1 shows our replication of the Syverson study, using average establishment earnings on a subsample that is identical to the one he uses.[7] As one can see, we find qualitatively similar results for all moments used in the Syverson study. Thus, we conclude that an interpretation of average earnings as a proxy for productivity is reasonable.

Before proceeding to the main analysis, it is worth noting how basic establishment characteristics and behavior vary with urban density, since differences in these characteristics across metropolitan areas can affect the relationship between earnings and density through a composition effect. Table 2 presents basic sample statistics as well as the simple estimated relationships between the average (log) number of employees and age of establishments, the average annual entry rate, and the average annual exit rate on (log) density. The mean of log establishment employment is 1.50 (which corresponds to mean employment, in levels, of about

[7] We provide additional details on how we conduct this analysis in the appendix.

16 workers), and the average establishment is 8 years old. About 10 percent of all establishments in each year are new entrants, and another 9 percent in each year exit. There is wide variation in these statistics across establishments, but much lower variation in their mean values across CBSAs. We estimate the relationship between these characteristics and urban density using OLS regressions that controls for the share of the CBSA population that is college educated and establishment characteristics (size, age, industry, and multi-unit firm status, excluding the characteristic used as the dependent variable).[8] Denser CBSAs tend to have smaller but older establishments, on average, though the differences in size are not statistically significant. Establishment entry rates decline with density, while exit rates are essentially unrelated with density, especially when controls for the CBSA college share and the remaining establishment characteristics are added.

3. The Density Premium over the Establishment Life Cycle

3.A. Static Estimates of the Density Premium

We preface our main results by showing that we estimate a density premium for establishments that is within the range of estimates found by previous studies. In doing so, we follow previous research by accounting for metropolitan area differences in skill mix and industrial composition, as well as additional controls for establishment composition allowed for by our data. We also control for the potential endogeneity of location choice using instrumental variable methods that are common in the urban literature. Finally, we examine the robustness of our density premium estimate under alternative measures of urban density. Our goal is to show that our baseline estimates are in line with previous research.

[8] All regressions include a dummy variable for year. In these and all subsequent establishment-level regressions in the paper, we report standard errors that are clustered at the CBSA level.

First, we note that if we regress the log of average earnings on the log of density at the CBSA level, which is what one would do if one were to use published statistics for metropolitan areas, we obtain an elasticity estimate (density premium) of 8.1 percent. Controlling for college share reduces this estimate to 7.8 percent. These estimates are roughly in line with previous estimates obtained using data on individuals rather than establishments.[9] Our main estimates come from establishment-level regressions, which allow us to control for a variety of compositional differences across metropolitan areas. For establishment *e* in CBSA *j* at year *t*, our regression model is

(2) $$\ln w_{ejt} = \alpha_t + \beta \ln D_j + \gamma C_j + \delta Z_{et} + \varepsilon_{ejt},$$

where $\ln w_{ejt}$ is the log of average establishment earnings, $\ln D_j$ is our density measure, C_j is the CBSA college share, Z_{et} is a set of establishment controls (the log of employment, fixed effects for age, fixed effects for four-digit SIC industry, and an indicator for membership in a multi-unit firm), and α_t is a year dummy. The additional establishment controls are potentially important because it is well known that earnings vary strongly with establishment characteristics like industry, size, and age.[10] To some extent, these characteristics also vary with density (Table 2). We use both years of our panel and cluster standard errors by CBSA.

Table 3 shows that, unconditionally, we find a somewhat higher density premium at the establishment level (10.2 percent) relative to using aggregate data. Controlling for the CBSA college share reduces this estimate substantially (to 8.0 percent), but additionally controlling for establishment characteristics produces a more modest reduction in the estimate (to 7.4 percent).[11] Table 4 shows the results of estimating (2) separately for different sub-groups of the data:

---

[9] See, for example, Ciccone and Hall (1996) and Sveikauskas (1975).
[10] For example, see Brown and Medoff (1989, 2003).
[11] This is consistent with recent work by Lehmer and Möeller (2010), who find an urban density premium that persists after controlling for firm size.

entering and exiting establishments, multi-unit and single-unit firms, establishments in different size categories, and establishments in different major industry groups. [12] While there is some variation within each subgroup, the estimates including all controls suggest a density premium between 5.6 and 10.1 percent.

Estimates of an urban density premium face an endogeneity issue: urban density may be a consequence rather than a cause of local productivity advantages. This is what Combes et al. (2010) refer to as the "endogenous quality of labor." Furthermore, it is not clear that population density, measured as total population per CBSA square mile, is the most accurate measure of the density of economic activity.

To deal with potential endogeneity biases, we re-estimate equation (2) using a two-stage least squares approach used by Combes et al. (2010). The approach follows the practice of using instruments that are correlated with density in such a deeply-lagged way that they are uncorrelated with productivity (or, in our case, average establishment earnings). In the case of our instruments, as well as those of Combes et al. (2010), geological and climate variables are used because, since they affect agricultural productivity so much, they are likely highly correlated with the location decisions that lead to the initial agglomeration of economic activity. At the same time, since agriculture is such a tiny share of contemporary urban employment, they are uncorrelated with current productivity. Our instruments include data on the fraction of the CBSA that is water-covered, the mean elevation and the fraction of the CBSA above 1000m, an index of terrain ruggedness, the average annual temperature and moisture, the number of potential growing days, and the fraction of the soil represented by a vector of soil types. We only

[12] We exclude age as a control for establishment characteristics in the entry regressions since, by definition, all entrants are zero years old. We also estimated the density premium across quintiles of the earnings distribution within each CBSA. Consistent with the findings of Combes et al. (2012), we find a premium that increases with establishment earnings. Our findings are in the appendix.

have such data for 283 of our 363 CBSAs, so we report both the OLS and IV results for this restricted sample.

Our results are in the first two columns of Table 5.[13] The restricted sample produces a somewhat higher estimate of the density premium (9.8 percent, compared to 7.4 percent in Table 3). Instrumenting density and college share in our 2SLS approach, however, only changes the estimated density premium from 9.8 percent to 10.4 percent, a statistically insignificant change.

Previous research has used a multiple measures of urban density. Some (e.g., Baum-Snow and Pavan, 2012) use city size, measured as total population. Others use employment density rather than population density (e.g., Ciccone and Hall, 1996). Some recent studies (Glaeser and Kahn, 2004; Rappaport, 2008) use an area-weighted population density measure on the premise that the municipal boundaries of a metropolitan area (which are used to define CBSAs) do not adequately reflect the area of economic activity.[14]

We reestimate equation (2) using OLS and four different measures of population density. The first uses population density, but weights the regression by establishment employment. This specification tests whether a worker-weighted regression produces a different result than an establishment-weighted regression. The second uses the log of the CBSA population. The third uses (log) employment density, measured using the current employment level (in 1992 or 1997) of the CBSA. The last uses the area-weighted population density from Rappaport (2008). The results are in the last four columns of Table 5. The estimated density premiums range from 5.6 to 9.5 percent, with the area-weighted measure producing the highest estimate.

[13] We present diagnostic statistics from the first stage regressions in the appendix, as well as estimates from CBSA-level regressions of (log) density on our instruments. The *F*-value on the first stage regression is 15.30 (*p*-value = 0.000), while the CBSA-level regression has an R-squared of 0.28 and a regression *F*-value of 6.90 (*p*-value = 0.000).

[14] These measures aggregate up a population density measure from subsections of the metropolitan area and weight each subsection by its population. This gives more weight to, say, downtown areas, and less weight to outlying rural areas that happen to be within the municipal boundaries of a metropolitan area.

Given the robustness of our estimated density premium across these alternative measures and specifications, and their comparability to previous findings, we conclude that the results of our subsequent analyses are not dependent on the methods, measurement, or establishment sample used, and proceed to our main analysis.

3.B. The Density Premium and Establishment Age

Productivity benefits derived from the localized accumulation and diffusion of knowledge are thought to be a key driver of an urban density premium. One example of this line of thinking is the seminal models of endogenous growth through knowledge spillovers (Romer, 1986; Lucas, 1988) where spillovers are the outcome of localized human capital accumulation. Another example involves models of spillovers and innovation, where the free flow of localized information between agents within a city fosters greater innovation (e.g., Jaffe, 1989). There is also empirical work that finds that worker wage-tenure profiles are steeper in denser cities (Glaeser and Mare, 2001; De la Roca and Puga, 2012; Baum-Snow and Pavan, 2012). This finding is attributed to a faster rate of human capital accumulation, consistent with the endogenous growth models. These models, and the related empirical work, share a common feature that the returns to agglomeration are captured through a higher rate of knowledge accumulation (through knowledge spillovers, faster "learning," or some similar process). With the exception of the evidence on wage-tenure profiles, however, this literature characterizes the productivity impact of this knowledge accumulation as a direct effect at the firm level. Therefore, we examine whether firms experience accumulated returns similar to workers, as the endogenous growth theories imply.

Specifically, we examine whether the estimated density premium for establishments increases as they age. We show below that establishment relocations are relatively rare, so an establishment's city tenure is equal to its age in most cases. Thus, a rising density premium with respect to establishment age would be evidence of a direct effect of localized knowledge spillovers on firm productivity.

We extend the regression specification in equation (2) to include an interaction between the fixed effects for establishment age and urban density,

$$\ln w_{ejt}(a) = \alpha_t^1 + \varphi(a) + \beta^1 \ln D_j + \zeta(a) \ln D_j + \gamma^1 C_j + \delta^1 \tilde{Z}_{et} + \varepsilon_{ejt}^1(a). \tag{3}$$

Above, a denotes age and $\tilde{Z}_{et}$ represents the same establishment characteristics as before except for age (industry, size, and multi-unit firm status). The coefficient of interest in this exercise is $\hat{\zeta}(a)$, since $\partial\hat{\zeta}(a)/\partial a > 0$ signifies a density premium that rises with establishment age.

In Figure 1, we present results for three specifications of equation (3), analogous to the specifications in the first, second, and fourth columns of Table 3. The figure plots $\hat{\beta}^1 + \hat{\zeta}(a)$ as a function of age for each specification. Regardless of the specification used, the density premium for establishments is essentially flat over their lifespan. In the baseline specification, the estimated elasticity varies within a relatively tight range of 9.3 to 11.1 percent. When we include all controls, the range is even tighter, between 6.9 and 7.8 percent. In comparison, each of these ranges are smaller than the reduction in the density estimate from controlling for the CBSA college share, and these differences are all well within the standard error bands for their respective specification. Figure 2 replicates the exercise for different subgroups of the data. The six panels of the figure report the coefficients $\hat{\beta}^1 + \hat{\zeta}(a)$ from the estimation of the full specification of (3) separately by (i) continuing versus exiting establishments, (ii) establishments in multi-unit and single-unit firms, (iii) establishment size class, (iv) major industry group, (v) within-CBSA earnings quintile, and (vi) using the alternative density measures from Table 4.[15] In nearly every case, the estimated density premium does not vary with age. There are even some cases where the estimates fall somewhat with age (multi-unit firms, establishments in the lowest CBSA earnings quintile), and only the construction industry shows any evidence of a notable and significant rise in the density premium with age, from 7.1 to 11.1 percent over an establishment's first 15 years. Among the alternative density measures, only the area-weighted population density measure shows a rise in the density premium, from 8.4 to 9.7 percent, but the increase is not significant.

3.C. Interpretation

We thus conclude that the premium associated with locating in a dense metropolitan area is almost entirely realized at entry. Establishments do not grow relatively more productive at denser locations through a return realized through a greater accumulation of knowledge spillovers, in contrast to models of urban growth such as Black and Henderson (1999). This does not imply that denser urban agglomeration does not lead to greater accumulated knowledge spillovers at the city level. Instead, it implies that the method by which knowledge accumulates faster in a denser city is a reallocation channel, where the returns to agglomeration are embodied in new entrants, rather than through a spillover or externality that directly affects the production of existing firms.

The evidence is consistent with a model where innovation diffuses across firms through spinoffs, as in Chatterjee and Rossi-Hansberg (2012). In their model, workers are tasked with generating innovations for a firm. Innovations vary in their quality, and those of sufficiently high

[15] We also estimated a version of (3) using the instrumental variables approach reported in Table 4. The IV estimates of the age-density interactions are almost identical to the OLS estimates. Both show that the density premium is essentially constant with respect to age.

quality induce their innovator to quit her existing firm and use the innovation to form a startup. Recent work on the Detroit automobile industry and Silicon Valley's high-tech industry by Klepper (2010), and on the fashion industry by Wenting (2008), show that a similar evolution of spinoffs played a key role in generating industrial clustering and growth within these industries. Our results thus far, however, do not distinguish between entry via new firm creation and entry through relocation. Without this distinction, models where the concentration of many firms generates a (static) benefit to density are also consistent with our results. These would include models where density itself generates an externality (Fujita and Ogawa, 1982), where firms benefit from shared access to specialized inputs (Abdel-Rahman and Fujita, 1990), and where a concentrated skilled labor pool reduces search frictions (Helsey and Strange, 1990). Later, we show that birth rather than relocation accounts for most of our result.

Before moving on, it is worth noting that there is a tension between our finding of a density premium that is independent of an establishment's age and a density premium that others have found to increase with a worker's city tenure (e.g., Glaeser and Mare, 2001; De la Roca and Puga, 2012; Baum-Snow and Pavan, 2012). It is important to keep in mind, though, that our estimates represent a combination of a fixed establishment return to agglomeration and the average return to agglomeration of its workers. In addition, the average return to workers can change over time through compositional changes in an establishment's workforce or through changes in the individual-level returns. In the extreme case where an establishment's workforce remains the same over time, earlier research suggests that our estimated density premium should rise over time. To reconcile the evidence, the establishment-specific component of the density premium would need to *decline* with age. If there were worker turnover, however, the interpretation becomes more complicated. For example, if lower-tenure workers consistently replace higher-tenure workers at a pace that keeps the average city tenure of workers roughly constant over an establishment's life, then we return to the interpretation that establishment-specific returns to urban agglomeration are unrelated to establishment age. For our results to mask an underlying increase in the establishment-specific returns over time, the average city tenure of workers would need to be decreasing over an establishment's life. How worker tenure evolves over an establishment's life cycle, however, is an open empirical question that requires access to matched employer-employee data, and therefore outside the scope of this paper. We nevertheless feel that understanding the link between worker and firm returns to density is an important part understanding urban agglomeration economies, and hope to explore this relationship further in future work.

4. The Roles of Selection and Sorting

We conclude our analysis by examining how much establishment dynamics, via entry and exit, account for the fact that we find a sizable, positive density premium for establishments that does not change over their life cycle. Recent models of urban agglomeration allow for a role for these dynamics (e.g., Behrens, Duranton, Robert-Nicoud, 2010). There are two ways that establishment turnover can affect our estimates. The first is through a selection effect. Specifically, if the least productive establishments are most likely to exit, and dense cities have greater competition that leads to higher exit rates, then we should see a stronger selection effect in denser cities. The selection effect inflates the estimated density premium through a greater lower-truncation of the establishment earnings distribution in denser cities. The second channel is a sorting effect. High-earnings firms may self-select into high-density cities, leading to an endogeneity issue in the estimation of the density premium. They can do so either at entry, which is difficult to identify, or through relocation to another city.

4.A. Selection through Exit

Recent research has tried to quantify how much of the estimated returns to agglomeration in previous studies are due to selection through firm exit. Syverson (2004) finds that higher local demand density for ready-mix concrete leads to a greater selection effect. Combes et al. (2012) examine whether selection matters more broadly using establishment TFP distributions across French metropolitan areas. They observe a strong selection effect, but find that it does not vary with city size, and therefore does not have an effect on their estimated returns to agglomeration. Their data, however, do not allow them to explicitly account for exit, which is a crucial part of models of firm selection (see Jovanovic, 1982; Hopenhayn, 1992; and Ericson and Pakes, 1995).

Our data suffer no such issues. We proceed by explicitly examining establishment exit rates as a function of their within-CBSA earnings distribution. We then compare the exit-earnings relationship in high-density versus low-density CBSAs to test for differences between the two. If there is a greater selection effect within high-density CBSAs, we should observe higher exit rates among low-earnings establishments within these CBSAs.

We group all establishments into one-percentile bins based on their ranking within their CBSA-specific earnings distribution (similar to how we grouped establishments by earnings quintile, but at a finer detail). We then pool all establishments within each percentile bin based on whether they reside in a CBSA within either the top or bottom quartile of the CBSA density distribution. Exit rates are calculated as the fraction of establishments that exit within each percentile bin. We calculate these exit rates for all establishments and for establishments aged 5 years or less. The latter group is of interest because exit rates are highest early in an establishment's life cycle. We report results that use an exit probability and average earnings measure that control for establishment characteristics (industry, size, age, multi-unit firm status). This eliminates differences in exit rates and earnings distributions that are due to observable differences in establishment composition across cities.[16]

Our estimates are in Figure 3. The top panels report exit rates a function of the earnings distribution, while the bottom panels report the difference between exit rates in high-density versus low-density CBSAs. In general, we find results consistent with the conclusions of Combes et al. (2012). There is clearly a strong selection effect for all establishments, as exit rates decline with average establishment earnings, and the highest exit rates are for establishments in the bottom 20 percent of the earnings distribution. At the same time, there is little difference in exit rates between high-density and low-density CBSAs. For all establishments, exit rates are somewhat higher in high-density CBSAs, but this difference is only marginally significant for the middle of the earnings distribution (between the 30th and 55th percentiles). For younger establishments, the differences are noisier but smaller, on average, and not significant anywhere along the earnings distribution.

Table 6 presents further evidence against a role for selection. In this exercise, we explicitly follow cohorts of entrants within the top and bottom quartile of the CBSA density distribution for their first 5 years and compare the relative difference in the evolution of earnings distributions among the surviving establishments. The estimates again show a clear selection effect within each CBSA group—mean earnings are higher and there is much less dispersion for surviving establishments, with much of the reduction in dispersion occurring in the bottom half of the earnings distribution. The difference in the selection effect between the high-density and

[16] We also performed the exercise using the raw earnings and exit probabilities and obtain very similar results.

low-density CBSAs, however, is small. If anything, high-density CBSAs have a relatively smaller increase in mean earnings and less truncation of their earnings distribution. Thus, we conclude that firm selection does not account for the density premiums estimated in the previous sections.

4.B. Sorting through Relocation

The estimated density premium for establishments can also be affected by sorting. That is, the premium may be overstated because productive establishments sort into dense areas. Studies that focus on the density premium for workers can control for such sorting by estimating the within-worker density premium for individuals who migrate across cities. Examining sorting for establishments is more complex because establishments can sort along two margins: relocation or entry. The longitudinal information on the establishment's location allows us to examine establishment relocations, but it can suffer from miscoding issues. We have an accurate measure of entry, but one cannot separately identify whether the location at entry was endogenously chosen or whether it happened to be the residence of the entrepreneur. Because of these limitations, little empirical research has been done on the sorting of firms into cities (an exception is Duranton and Puga, 2001, using French data). We conclude our analysis by examining whether high-earnings establishments sort into high-density CBSAs, either by relocation or by entry.

We start with relocating establishments. Because of the potential for the miscoding of location, we identify relocations in the data as establishments whose county code changes no more than once during our sample period.[17] We focus primarily on continuing establishments

[17] Geographic miscoding generally occurs because of a reporting error in a particular year. When they are corrected, it appears as if the establishment relocated then relocated back to their original location. We have no reason to believe that the likelihood of miscoding is correlated with either establishment earnings or CBSA density.

over the 1991-92 and 1996-97 periods. Restricting the analysis to moves between CBSAs identifies just over 81,000 moves, which represents 1.03 percent of all continuous establishments.[18] By comparison, entrants constitute 10.3 percent of all establishments.

Table 7 reports the differences in the basic characteristics of continuous establishments who move versus those that remain in the same CBSA. Relocating establishments tend to be younger, slightly larger, and less likely to be part of a multi-unit firm, on average. They also have earnings that are 10.7 percent higher than stayers, on average, implying that higher-earnings establishments are in fact more likely to sort. Relocating establishments come from CBSAs that are slightly denser (1.1 percent) and have a somewhat higher share of individuals with at least a college degree (0.4 percentage points). At the same time, they tend to move to CBSAs that are considerably less dense (21.7 percent) and have lower shares of their population with a college degree (0.7 percentage points). Thus, while establishments with higher earnings are more likely to move, the average relocation goes in the opposite direction of what a sorting story would suggest.

Figure 4 examines the behavior of relocating establishments across the full earnings distribution to see whether the simple means in Table 7 mask a richer pattern of relocation across metropolitan areas. As with the selection exercise of the previous section, we allocate all continuous establishments into one-percentile bins based on the earnings distribution of their CBSA of origin.[19] We then calculate the fraction of establishments that relocate within each

---

[18] Inter-CBSA moves represent 35 percent of all inter-county moves. In addition, 79 percent of all moves out of a CBSA are to another CBSA (the remainder are to non-metropolitan areas). We reject about 26 percent of potential relocations because of multiple changes in an establishment's county code.

[19] As before, we use a residual measure of earnings the probability of moving that controls for establishment size, age, industry, and multi-unit firm status. Results using the unconditional earnings and move measures are very similar to those reported in the figure. In the appendix, we present additional results using the earnings distribution measured across the full sample of establishments (i.e., one that is not CBSA-specific). Again, the results are qualitatively similar to those in the figure.

percentile bin. These estimates are presented in the first panel of the figure. Next, we calculate the change in (log) CBSA density for the subset of establishments that relocate. We plot the average change within each percentile in the bottom panel of the figure.

The thick lines in each panel represent the estimates derived directly from the data. The top panel shows that relocations are disproportionately concentrated in the top 20 percent of the within-CBSA earnings distribution. Relocation rates are essentially constant at just under 1 percent for establishments in the bottom 60 percent of the earnings distribution. Those in the top 5 percent of the distribution are 64 percent more likely to move than those in the bottom 60 percent. The bottom panel shows that establishments move to lower-density CBSAs, on average, across almost the entire earnings distribution. Establishments in the bottom 60 percent of the distribution move to a CBSA with a population density that is about 21 percent lower than the density of their original CBSA, on average. The decline is smaller as one moves further up the earnings distribution. Only the top 5 percent of establishments experience a small but statistically insignificant increase in CBSA density, on average.

CBSAs vary widely in their size (i.e., number of establishments) and in their earnings distributions. Consequently, the case where moves are purely random, in the sense that moves do not depend on establishment earnings and are not directed towards a particular location, may generate spurious relations between earnings, move probabilities, and changes in CBSA density. To examine how much our estimates differ from a baseline of random relocations, we generate a simulated panel of establishments that inhabit the empirically observed distribution of CBSAs and allow them to move with equal probability to a randomly selected CBSA. We then calculate the probability of moving and the average change in density by the percentile of their origin CBSA's earnings distribution (as we do with the LBD data). We provide the details of this simulation in the appendix. Our generated baseline is very much in the spirit of Ellison and Glaeser (1997), who compare how much industrial concentration differs from the case where industries cluster completely randomly through a "dartboard" approach.

Our results are also in Figure 4. The thick dashed lines present the difference between the estimates obtained from the data and our simulated estimates. The probability of moving is qualitatively unchanged by construction, since, in the simulated data, all establishments have a probability of moving equal to the sample mean. The change in CBSA density, relative to the simulated estimates, is actually more negative than what we observe in the data. Our simulation predicts that density should rise 39 percent, on average, when establishments move, regardless of where an establishment lies within its origin-CBSA's earnings distribution. Thus, relative to the case of random relocations, we find that relocating establishments in the bottom 60 percent of the distribution experience a 60 percent decline in the population density of their CBSA, while those in the top 5 percent experience a 32 percent decline, on average. Establishments at the high end of the earnings distribution tend to move to denser CBSAs in relative terms, but in absolute terms, all establishments move to less dense CBSAs.

We consider the effects of these relocations on our density premium estimates to be small, however, for several reasons. First, while high-earnings establishments are the most likely to move, the change in the density of their CBSAs relative to other movers (23 percent) is small relative to the standard deviation in CBSA density measured across establishments (105 percent) or CBSAs (94 percent). Second, relocation rates are fairly low on average. The highest-earnings establishments, which are the most mobile, only have a 1.6 percent chance of moving in a given year. Most establishments have less than a one percent chance. Finally, the density estimate derived from a regression using our simulated establishment data is only slightly different from the in the estimate using the LBD data. Specifically, if we estimate equation (2) on the sample of continuous LBD establishments, we obtain a density premium estimate of 0.072 (standard error of 0.008). If we estimate the same equation on our simulated establishments, we obtain an estimate of 0.067 (standard error of 0.003). This rough comparison suggests that establishment sorting accounts for only 7 percent of our estimated density premium.

We conclude this subsection by showing that relocations also have little effect on our finding that establishments realize a density premium almost entirely upon entry. One might worry that our earlier results do not distinguish between entry into a CBSA through birth and entry into a CBSA through relocation. The distinction has theoretical implications because a density premium realized primarily at birth is consistent with a model of spillovers through spinoffs, while a premium realized primarily at entry into a new city (regardless of age) would be consistent with models where firms gain returns to agglomeration through local comparative advantage, localized access to shared inputs, or other factors that benefit all firms at a given location.

We use our data on relocations between 1991 and 1997 to distinguish between establishment age and city tenure. We focus on the single cross-section of establishments active in 1997, which allows us calculate city tenure with accuracy up to its first 6 years. We then run a version of equation (3) that also allows the density premium to vary with city tenure, conditional on the establishment having moved at least once since 1991. The estimated equation is

$$\text{(4)} \qquad \ln w_{ejt}(a,\tau) = \alpha^2 + \varphi^2(a) + \beta^2 \ln D_{jt} + \zeta^2(a) \ln D_{jt} + \gamma^2 C_{jt} + \delta^2 \tilde{Z}_{et} + \theta(\tau) + \eta I(m_t) + \mu(\tau) \ln D_{jt} + \nu(\tau)[\ln D_{jt} \cdot I(m_t)] + \varepsilon^2_{ejt}(a,\tau).$$

The regression is the same as before, except that it now includes dummies for whether the establishment has moved to its current CBSA since 1991, $I(m_t)$, city tenure, $\theta(\tau)$, with tenure τ

top-coded at 6 years, and interaction of the city tenure dummies with (log) density, both unconditionally, and conditional on relocating from somewhere else. Figure 5 presents the estimates of the evolution of the density premium with respect to age ($\hat{\beta}^2 + \hat{\zeta}(a)$) and with respect to city tenure ($\hat{\beta}^2 + \hat{\mu}(\tau) + \hat{\nu}(\tau)$). We estimate that the density premium is about 6.0 percent at birth, independent of city tenure, and is again roughly constant over the life of the establishment. Conditional on age, we find that entry into a new city increases the density premium by 0.9 percent the first year, but the increase is not significant. The additional effect of density on relocating establishments varies between -0.6 percent and 1.6 percent in subsequent years, but these differences are also not statistically significant. As with age, we find no significant rise in the estimated premium with city tenure. Thus, birth, rather than entry at any point in the life-cycle, is the time when establishments reap their returns to density.

4.C. Sorting through Entry

As a final exercise, we document differences in the earnings distribution of entrants between high-density and low-density CBSAs. Direct identification of establishment sorting through entry is virtually impossible because one cannot distinguish whether a birth occurred in a particular metropolitan area because the location was endogenously chosen over other locations or because the entrepreneur lives there. Figueiredo, Guimarães, and Woodward (2002) and Michelacci and Silva (2007) both find a strong "home bias" in the startup location choices of entrepreneurs. Using data from Italy, Michelacci and Silva find that the fraction of entrepreneurs operating where they were born was significantly larger than the fraction of dependent workers working where they were born. Since we cannot identify this distinction in our data, we only examine whether there is at least suggestive evidence of sorting in the earnings differences among entrants.

We examine differences in the earnings distribution across high-density and low-density metropolitan areas for entrants in absolute terms and relative to incumbent establishments. We focus on the earnings differences of entrants that are new firms relative to the differences of new establishments of multi-unit firms. The working hypothesis is that a multi-unit firm is more likely to endogenously choose the location of its new establishment, whereas a new, single-establishment firm is more likely to start up where its entrepreneur is resides. Therefore, a sorting mechanism should be stronger for the entrants of multi-unit firms.

Table 8 presents the statistics for the earnings distribution of entrants into CBSAs of the highest quartile and lowest quartile of the CBSA density distribution, using an earnings measure that conditions out establishment characteristics. Entrants in high-density CBSAs have earnings that are 21.2 percent higher, on average, than the earnings of entrants in low-density CBSAs. The difference in median earnings is nearly as large (18.0 percent).[20] Relative to the earnings of incumbent establishments in their respective CBSA categories, entrants in high-density CBSAs have earnings that are only 1.3 percent higher, on average, and relative median earnings are essentially equal. Entrants of multi-unit firms exhibit similar earnings differences between high-density and low-density CBSAs. In absolute terms, their mean earnings are 16.3 percent higher and their median earnings are 13.9 percent higher in high-density CBSAs. Relative to incumbents, multi-unit firm entrants in high-density CBSAs have mean earnings that are 3.6 percent *lower*, and median earnings that are 4.1 percent lower, in high-density CBSAs. More importantly, relative to all entrants, multi-unit entrants have mean earnings that are 4.9 percent lower, and median earnings that are 4.1 percent lower, in high-density CBSAs. These results are

[20] Without controlling for establishment characteristics, the difference in mean earnings is 26.0 percent and the difference in median earnings is 24.2 percent.

not a clear rejection of a role for sorting, but do not present any first-order evidence for the sorting of high-earnings entrants into high-density CBSAs.

5. Conclusions

The density premium is a key feature of urban agglomeration. A large body of research has examined the premium afforded to individual workers, but few studies have explored the premium afforded to individual establishments. Using longitudinal microdata on U.S. establishments, we estimate a density premium of about 7.4 percent after controlling for the college share of a metropolitan area and observable establishment characteristics. The estimate varies little across various subgroups of the data and is robust to instrumenting for the simultaneous determination of density and productivity as well as the use of alternative measures of density.

Our main finding is that the urban density premium does not rise with establishment age (or city tenure), implying that any potential returns to agglomeration are realized at establishment birth and fixed over time. The result is in contrast to previous research on workers that find a steeper wage-city tenure profile in larger cities. We interpret the result as evidence that knowledge accumulation occurs within a city through a reallocation channel rather than through a spillover that directly affects the production of incumbent firms. Our findings do not, however, preclude a role for knowledge spillovers as a source of urban growth. Models of innovation through firm spinoffs (e.g., Chatterjee and Rossi-Hansberg, 2012) are consistent with our interpretation, as are any models that stress a prominent role for firm entry in the transmission of returns to urban agglomeration. The key insight is that, to the extent that returns to urban agglomeration occur via the diffusion of knowledge (whether it be through learning, spillovers, or something else), such knowledge diffuses through the creation of new businesses and not necessarily through the exchange of knowledge between existing businesses.

We also examine how much firm selection (through exit) and firm sorting (through relocation or entry) drive our results. While we find evidence of a selection effect in general, exit rates follow a similar pattern across the earnings distribution for both high-density and low-density metropolitan areas, suggesting that selection is not a main driver of observed urban density premia. We find weak evidence of sorting among establishments that relocate—high-earnings establishments are the most likely to relocate, and they move to relatively denser metropolitan areas, on average. At the same time, we find that all relocating establishments tend to move to *lower-density* metropolitan areas in absolute terms. We hypothesize that new establishments of multi-unit firms should be the most likely to have their location endogenously chosen, yet we find that they have weaker evidence of sorting, relative to new single-unit firms.

We conclude by identifying three avenues for future research. The first is advancing urban theory to include a more prominent role for firm entry, for the reasons we note above. The second is a deeper examination of the location choices and characteristics of new businesses. Our data allow us to provide only cursory evidence on firm sorting, though we provide robust evidence that much of the returns to urban agglomeration occur at entry. Understanding firm behavior at entry is therefore critical to understanding the nature of agglomeration economies. The third is an exploration of the relationship between a firm's return to city tenure and a worker's return to city tenure. There is a tension between our results and earlier research on workers. Geographic differences in the composition and turnover of an establishment's workforce no doubt play a role. Documenting and quantifying this role would provide a better understanding of how urban agglomeration affects both workers and firms.

References

Abdel-Rahman, Hesham, and Masahisa Fujita, 1982. "Product Variety, Marshallian Externalities, and City Sizes," *Journal of Regional Science*, 30(2): 165-183.

Audretsch, David B., and Maryann P. Feldman, 2004. "Knowledge Spillovers and the Geography of Innovation," in *Handbook of Urban and Regional Economics*, Volume 4, J. Vernon Henderson and Jacques-François Thisse (eds.), North Holland: Amsterdam.

Bacolod, Marigee, Bernardo Blum, and William Strange, 2009. "Skills and the City," *Journal of Urban Economics,* 65(2): 127-135.

Baum-Snow, Nathaniel, and Ronni Pavan, 2012. "Understanding the City Size Wage Gap," *Review of Economic Studies*, 79(1): 88-127.

______, 2013. "Inequality and City Size," *Review of Economics and Statistics*, forthcoming.

Behrens, Kristian, Gilles Duranton, and Frederic Robert-Nicoud, 2010. "Productive Cities: Sorting, Selection, and Agglomeration," CEPR Discussion Paper No. 7922.

Black, Duncan, and J. Vernon Henderson, 1999. "A Theory of Urban Growth," *Journal of Political Economy*, 107(2): 252-284.

Brown, Charles, and James Medoff, 1989. "The Employer-Size Wage Effect," *Journal of Political Economy,* 97(5): 1027-59.

______, 2003. "Firm Age and Wages," *Journal of Labor Economics*, 21(3): 677-97.

Chatterjee, Satyajit, and Esteban Rossi-Hansberg, 2012. "Spinoffs and the Market for Ideas," *International Economic Review*, 53(1): 53-93.

Ciccone, Antonio, and Robert E. Hall, 1996. "Productivity and the Density of Economic Activity," *American Economic Review,* 86(1): 54-70.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux, 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 80(6): 2543-2594.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux, 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Effects," in *Agglomeration Economics*, Edward L. Glaeser (ed.), University of Chicago Press: Chicago, IL.

De la Roca, Jorge, and Diego Puga, 2012. "Learning by Working in Big Cities," CEPR Discussion Paper 9243.

Duranton, Gilles, and Diego Puga, 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review*, 91(5): 1454-77.

Ellison, Glenn, and Edward L. Glaeser, 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105(5): 889-927.

Ericson, Richard, and Ariel Pakes, 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies,* 62(1): 53-82.

Figueiredo, Octávio, Paulo Guimaraes, and Douglas Woodward, 2002. "Home-Field Advantage: Location Decisions of Portuguese Entrepreneurs," *Journal of Urban Economics*, 52(2): 341-361.

Freedman, Matthew, 2008. "Job Hopping, Earnings Dynamics, and Industrial Agglomeration in the Software Publishing Industry," *Journal of Urban Economics*, 64(3): 590-600.

Fujita, Masahisa, and Hideaki Ogawa, 1982. "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations," *Regional Science and Urban Economics*, 12(2): 161-196.
Glaeser, Edward L., 1999. "Learning in Cities," *Journal of Urban Economics,* 46(2): 254-77.
Glaeser, Edward L., and Matthew Kahn, 2004. " Sprawl and Urban Growth," in *Handbook of Urban and Regional Economics*, Volume 4, J. Vernon Henderson and Jacques-François Thisse (eds.), North Holland: Amsterdam.
Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer, 1992. "Growth in Cities," *Journal of Political Economy*, 100(6): 1126-1152.
Glaeser, Edward L., and David C. Mare, 2001. "Cities and Skills," *Journal of Labor Economics*, 19(2): 316-342.
Glaeser, Edward L., and Matthew G. Resseger, 2010. "The Complementarity between Cities and Skills," *Journal of Regional Science,* 50(1): 221-244.
Griliches, Zvi, 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *Bell Journal of Economics*, 10(1): 92-116.
Helsey, Robert W., and William C. Strange, 1990. "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics,* 20(2): 189-212.
Henderson, J. Vernon, 1997. "Externalities and Industrial Development," *Journal of Urban Economics*, 42(3): 449-470.
______, 2003. "Marshall's Scale Economies," *Journal of Urban Economics*, 53(1): 1-28.
______, 2007. "Understanding Knowledge Spillovers," *Regional Science and Urban Economics*, 37(4): 497-508.
Henderson, J. Vernon, Ari Kuncoro, and Matthew Turner, 1995. "Industrial Development in Cities," *Journal of Political Economy*, 103(5): 1067-1090.
Hopenhayn, Hugo, 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica,* 60(5): 1127-50.
Klepper, Steven, 2010. "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics*, 67(1): 15-32.
Jaffe, Adam, 1989. "Real Effects of Academic Research," *American Economic Review*, 79(5): 957-70.
Jarmin, Ron, and Javier Miranda,2002. "The Longitudinal Business Database," Center for Economic Studies Working Paper no. 02-17, U.S. Census Bureau.
Jovanovic, Boyan, 1982. "Selection and the Evolution of Industry," *Econometrica*, 50(3): 649-70.
Lehmer, Florian, and Joachim Möeller, 2010. "Interrelations between the Urban Wage Premium and Firm-Size Wage Differentials: A Microdata Cohort Analysis for Germany," *Annals of Regional Science*, 45(1): 31-53.
Lucas, Robert Jr., 1988. "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1): 3-42.
Michelacci, Claudio, and Olmo Silva, 2007. "Why So Many Local Entrepreneurs?" *Review of Economics and Statistics*, 89(4): 615-633.
Moretti, Enrico, 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions," *American Economic Review*, 94(3): 656-90.
Rappaport, Jordan, 2008. "Consumption Amenities and City Population Density," *Regional Science and Urban Economics*, 38(6): 533-552.

Romer, Paul, 1986. "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94(5): 1002-1037.

Rosenthal, Stuart, and William Strange, 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Urban and Regional Economics*, Volume 4, J. Vernon Henderson and Jacques-François Thisse (eds.), North Holland: Amsterdam.

Sveikauskas, Leo, 1975. "The Productivity of Cities," *Quarterly Journal of Economics,* 89(3): 319-413.

Syverson, Chad, 2004. "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(6): 1181-1222.

Wenting, Rik, 2008. "Spinoff Dynamics and the Spatial Evolution of the Fashion Industry, 1858-2005," *Journal of Economic Geography*, 8(5): 593-614.

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

Table 1. Comparison of Results for Firm Selection in the Concrete Industry

Moment	Estimate from Syverson (2004), using TFP for y_{et}	Estimate from the LBD, using avg. earnings for y_{et}
Interquartile range of distribution of ln y_{et}	-0.015 (0.004)	-0.028 (0.013)
Median value of ln y_{et}	0.018 (0.003)	0.095 (0.015)
Size-weighted mean of ln y_{et} [1]	0.024 (0.004)	0.081 (0.015)
Tenth percentile of distribution of ln y_{et}	0.056 (0.010)	0.080 (0.027)
Mean plant size[1]	0.211 (0.012)	0.065 (0.016)
Producer-demand ratio[2]	-0.363 (0.015)	-0.680 (0.033)
Number of Observations	665	410

Notes: Table reports the estimates from the regression of the listed moment on a measure of demand density (the log of construction employment) and a year dummy across geographic locations (BEA Census Economic Areas for Syverson, and our sample of CBSAs for the LBD.) Estimates in the first column come from Syverson (2004, "Model 2" on p. 1206.), and estimates in the second column are authors' estimates from the LBD. See text for more details. Standard errors are in parentheses.

1. Size is measured as the log of total sales in Syverson (2004) and as the log of employment in the LBD.
2. The producer-demand ratio is the number of plants per 1,000 construction employees.

Table 2. Basic Statistics on Relationships between CBSA Establishment Characteristics and Density

	ln *Size* (employees)	Age (years)	Entry Rate (share of estabs.)	Exit Rate (share of estabs.)
Sample Mean	1.50	8.01	0.103	0.092
Std. Deviation across Establishments	1.37	6.80	0.304	0.289
Std. Deviation across CBSAs	0.10	0.86	0.009	0.015
OLS regression on ln(Density) and College Share				
$\ln D_j$	-0.007	0.189	-0.004	0.000
	(0.012)	(0.052)	(0.001)	(0.002)
College Share, C_j	-0.186	-2.139	0.034	0.020
	(0.090)	(0.672)	(0.015)	(0.012)
Establishment Controls	Yes	Yes	Yes	Yes
R^2	0.332	0.195	0.087	0.131
Number of Observations		10,256,604		

Notes: Table reports summary statistics for the listed variables in each column, as well as the results of regressions of the listed variables on the log of 1990 population density and the share of the 1990 population with a college degree. All regressions include a year dummy. Establishment characteristics, where listed, include the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC. Standard errors, clustered by CBSA, are in parentheses.

Table 3. Establishment-Level Relations between Earnings and Density

	All Establishments			
	(1)	(2)	(3)	(4)
$\ln D_j$	0.102	0.080	0.092	0.074
	(0.007)	(0.010)	(0.010)	(0.010)
College Share, C_j		1.024		0.883
		(0.102)		(0.093)
Year effects?	Yes	Yes	Yes	Yes
Controls for establishment characteristics?	No	No	Yes	Yes
R^2	0.014	0.017	0.310	0.313
Number of Observations	10,256,604			

Notes: Table reports estimates from the regression of the log of average establishment earnings on the listed variables for our sample of establishment-year observations from the LBD. Establishment characteristics include the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC. Standard errors, clustered by CBSA, are in parentheses.

Table 4. Establishment-Level Relations between Earnings and Density by Sub-Group

	Entrants and Exits		*Multi- & Single-Unit Firms*	
	Entrants	Exits	Single-Unit	Multi-Unit
$\ln D_j$	0.076 (0.011)	0.079 (0.013)	0.080 (0.010)	0.058 (0.009)
College Share, C_j	1.135 (0.129)	0.815 (0.100)	0.945 (0.108)	0.700 (0.088)
Year effects?	Yes	Yes	Yes	Yes
Controls for establishment characteristics?	Yes	Yes	Yes	Yes
R^2	0.257	0.271	0.279	0.460
Number of Observations	1,063,789	950,456	7,587,861	2,668,379

	By Establishment Size				
	1 to 9 Employees	10 to 99 Employees	100 to 249 Employees	250 to 999 Employees	1,000+ Employees
$\ln D_j$	0.079 (0.010)	0.064 (0.010)	0.067 (0.009)	0.075 (0.012)	0.071 (0.013)
College Share, C_j	0.926 (0.104)	0.769 (0.084)	0.862 (0.111)	0.749 (0.113)	0.703 (0.421)
Year effects?	Yes	Yes	Yes	Yes	Yes
Controls for establishment characteristics?	Yes	Yes	Yes	Yes	Yes
R^2	0.270	0.521	0.539	0.517	0.521
Number of Observations	7,578,426	2,437,528	171,787	58,707	10,156

	By Major Industry Group				
	Construction	Manu-facturing	Retail Trade	Finance & Prof. Services	Local Services
$\ln D_j$	0.084 (0.019)	0.072 (0.016)	0.064 (0.016)	0.101 (0.012)	0.056 (0.005)
College Share, C_j	0.882 (0.214)	0.911 (0.135)	0.770 (0.099)	1.134 (0.124)	0.803 (0.099)
Year effects?	Yes	Yes	Yes	Yes	Yes
Controls for establishment characteristics?	Yes	Yes	Yes	Yes	Yes
R^2	0.154	0.279	0.254	0.219	0.280
Number of Observations	982,179	635,839	2,554,622	1,795,447	2,730,177

Notes: Table reports estimates from the regression of the log of average establishment earnings on the listed variables for our sample of establishment-year observations from the LBD. Establishment characteristics include the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC. Standard errors, clustered by CBSA, are in parentheses.

Table 5. Establishment-Level Relations between Earnings and Density, Alternative Specifications

Density Measure	Population Density	Population Density	Population Density	Population (Level)	Employment Density	Area-Wtd. Population Density
Estimation Method	OLS, Restricted Sample	IV, Restricted Sample	OLS, Emp.-Weighted	OLS	OLS	OLS
$\ln D_j$	0.098 (0.007)	0.104 (0.009)	0.067 (0.010)	0.056 (0.006)	0.073 (0.010)	0.095 (0.006)
College Share, C_j	0.898 (0.099)	1.067 (0.237)	0.767 (0.096)	0.798 (0.106)	0.799 (0.091)	0.840 (0.108)
Year effects?	Yes	Yes	Yes	Yes	Yes	Yes
Controls for establishment characteristics?	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.317	0.315	0.490	0.313	0.313	0.313
Number of Observations	7,761,264		10,256,604			

Notes: Table reports estimates from the regression of the log of average establishment earnings on the listed variables for our sample of establishment-year observations from the LBD. All regressions includes controls for year and observable establishment characteristics (the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC). The first two columns report the OLS and 2SLS estimates on a subsample of 283 CBSAs where we have the data on climate and geology that we use as instruments. These instruments include the fraction of the CBSA that is water-covered, the fraction above 1000m elevation, an index of the ruggedness of the land, the average annual temperature and moisture, the number of growing days, and the fraction of the land containing a set of 8 different soil types. The last four columns report OLS estimates where we use alternative specifications of urban density: the log of population density using an employment-weighted regression, the log of population, the log of employment density, and the log of population density generated by weighting each subunit of a metropolitan area by its total population. Standard errors, clustered by CBSA, are in parentheses.

Table 6. Statistics on the Earnings Distribution of Surviving Entrants in High- and Low-Density CBSAs

	Change in Distributional Statistic: Five-Year Survivors - Entrants		
Statistic	Low-Density CBSAs	High-Density CBSAs	*Difference-in-Difference: High-Density – Low-Density*
Mean Earnings	0.268	0.252	-0.016
Median Earnings	0.208	0.200	-0.008
Standard Deviation	-0.224	-0.204	0.020
Interquartile Range	-0.269	-0.222	0.047
90-10 Ratio	-0.613	-0.545	0.068
50-10 Ratio	-0.435	-0.384	0.051
Entrant Survival Rate	0.485	0.479	
Observations	38,820	332,018	

Notes: Table reports distributional statistics of earnings of entering establishments that survived to their fifth year, pooled across the top quarter (high density) or bottom quarter (low density) of CBSAs, ranked by 1990 population density. Statistics are based on an estimate of average establishment earnings that controls for establishment characteristics (the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC).

Table 7. Summary Statistics of Relocating Establishments

	Non-Relocating Establishments	Relocating Establishments	Difference
(log) Earnings	9.830	9.937	0.107
	(0.812)	(0.903)	[0.003]
Size (employees)	19.08	19.67	0.59
	(110.68)	(85.22)	[0.30]
Age (years)	8.59	6.34	-2.25
	(5.57)	(5.16)	[0.02]
Percent in Multi-Unit Firms	27.3	23.5	-3.8
	(44.6)	(42.4)	[0.1]
(log) Density at Origin	5.936	5.947	0.011
	(1.048)	(1.068)	[0.002]
(log) Density at Destination	---	5.730	-0.217[1]
		(1.049)	[0.005]
College Share at Origin	22.39	22.78	0.39
(Percent)	(5.46)	(6.09)	[0.02]
College Share at Destination	---	22.07	-0.71[1]
(Percent)		(6.18)	[0.03]
Observations	7,799,688	81,766	

Notes: Table reports summary statistics for establishments who relocated from one CBSA to another versus establishments that remained in place. Standard deviations of the statistics are in parentheses. Standard errors for the difference between statistics are in brackets.

1. Estimate represents the difference in log density (or college share) between the origin and destination CBSA.

Table 8. Statistics on the Earnings Distribution of Entrants in High- and Low-Density CBSAs

All Entrants

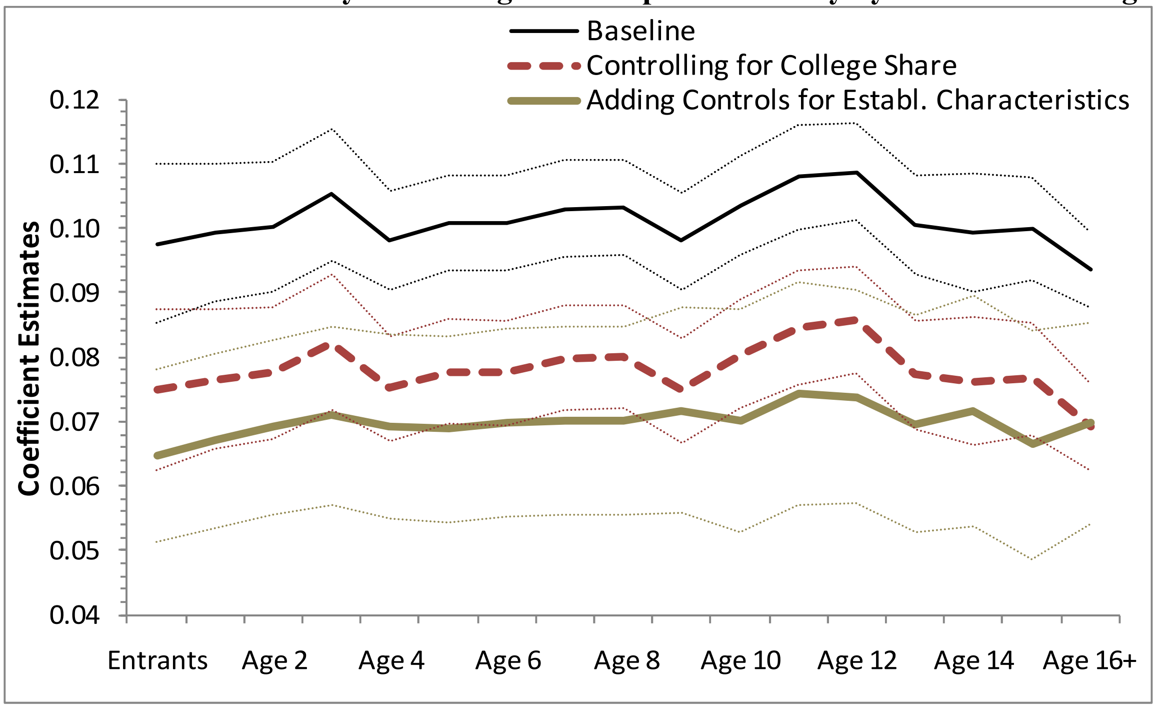
			High – Low Density Difference	
Statistic	Low-Density CBSAs	High-Density CBSAs	*Absolute Difference*	*Relative to Incumbents*
Mean Earnings	9.640	9.852	0.212	0.013
Median Earnings	9.756	9.936	0.180	0.000
Standard Deviation	0.883	0.910	0.026	-0.014
Observations	80,092	693,139		

Entrants of Multi-Unit Firms

			High – Low Density Difference		
Statistic	Low-Density CBSAs	High-Density CBSAs	*Absolute Difference*	*Relative to Incumbents*	*Relative to All Entrants*
Mean Earnings	9.860	10.023	0.163	-0.036	-0.049
Median Earnings	9.906	10.045	0.139	-0.041	-0.041
Standard Deviation	0.677	0.737	0.060	0.019	0.033
Observations	18,530	165,691			

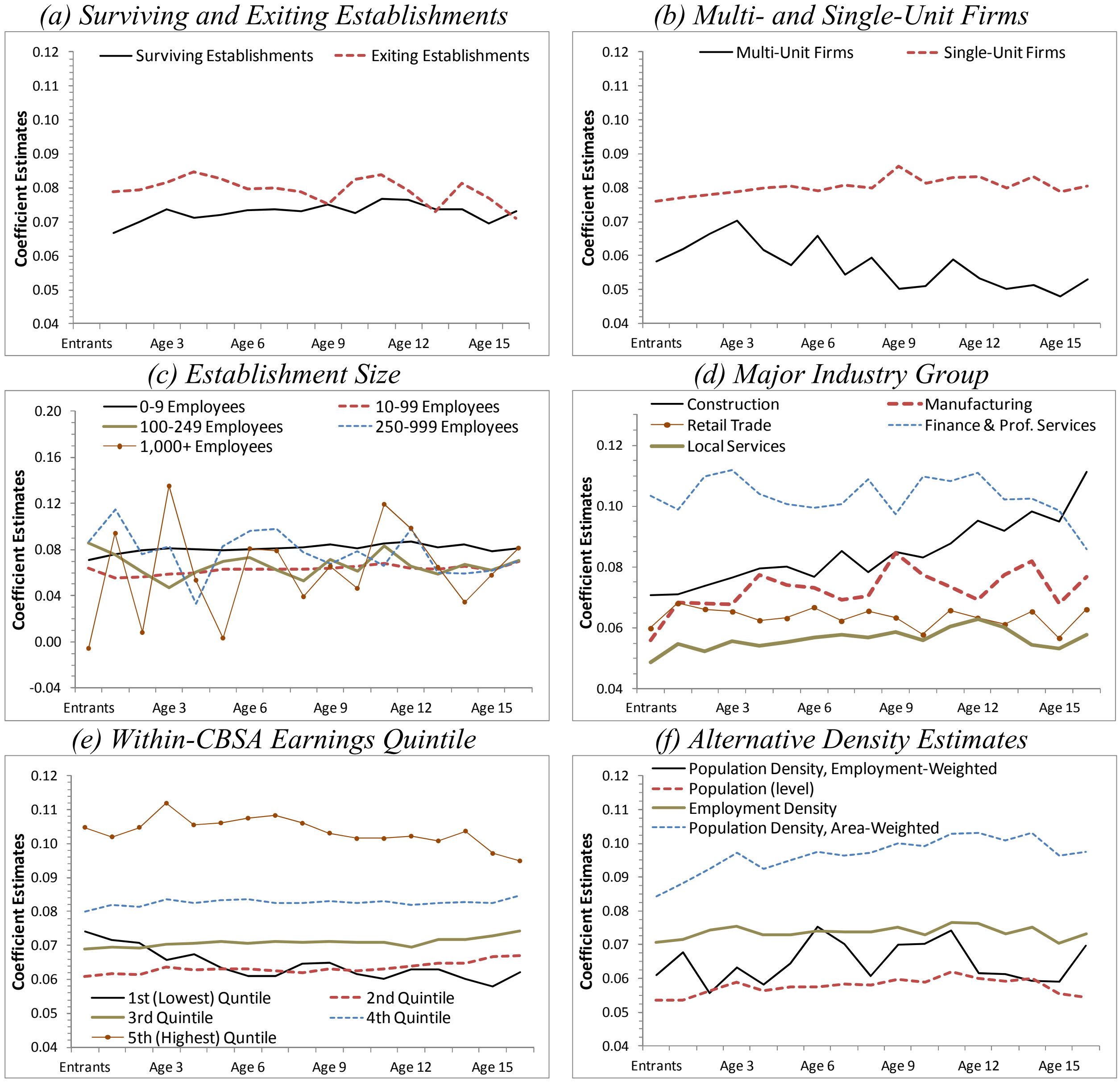
Notes: Table reports distributional statistics of earnings of entering establishments pooled across the top quarter (high density) or bottom quarter (low density) of CBSAs, ranked by 1990 population density. Statistics based on an estimate of average establishment earnings that controls for establishment characteristics (the log of establishment employment, a dummy for whether the establishment is part of a multi-unit firm, fixed effects for age, and fixed effects for four-digit SIC).

Figure 1. Estimated Elasticity of Earnings with respect to Density by Establishment Age



Notes: The figure plots the predicted elasticity of earnings with respect to density as a function of age. Estimates come from equation (3) in the text. See text for details. Thin dashed lines represent standard error bands, with standard errors clustered by CBSA.

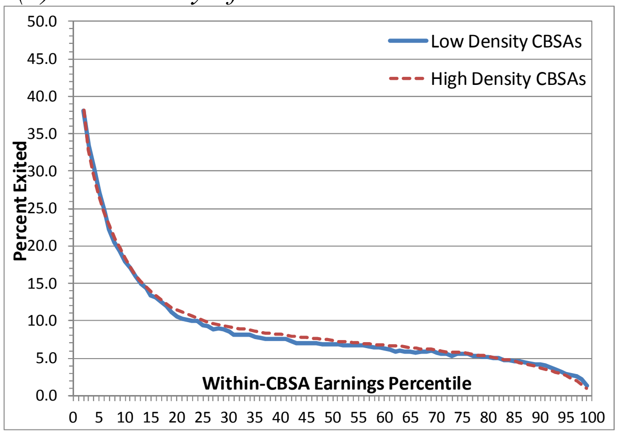
Figure 2. Elasticity of Earnings with respect to Density by Establishment Age and Sub-Group



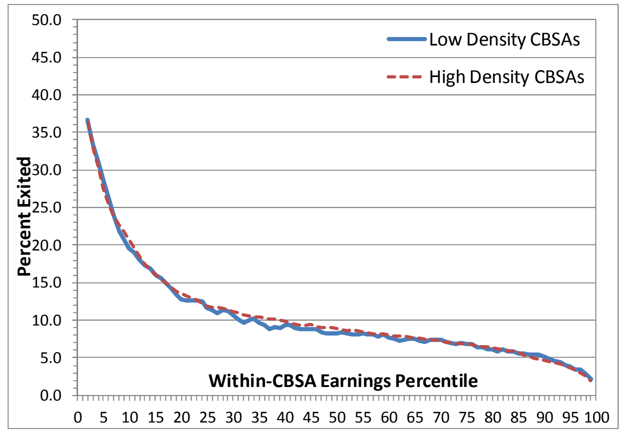
Notes: Each panel of the figure plots the predicted elasticity of earnings with respect to density as a function of age within each reported category. Estimates come from equation (3) in the text. See text for details. Thin dashed lines represent standard error bands, with standard errors clustered by CBSA.

Figure 3. Establishment Exit Probabilities by CBSA Earnings Percentile

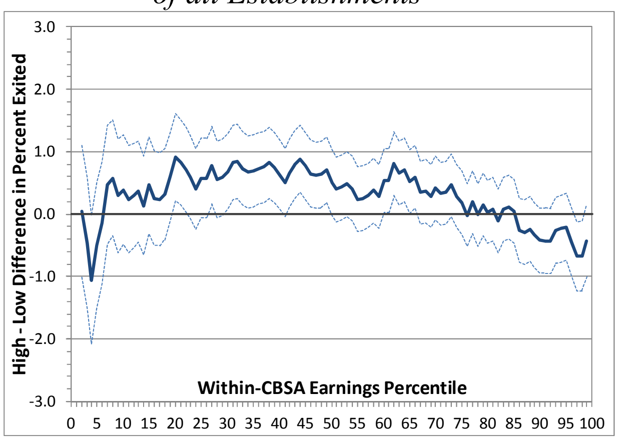
(a) Probability of Exit: All Establishments



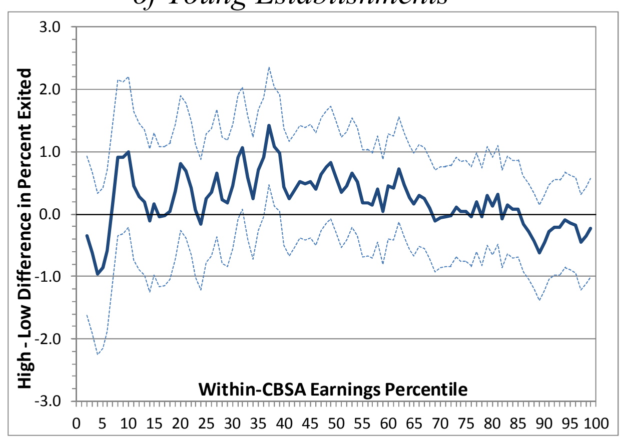
(b) Probability of Exit: Young Establishments



(c) High – Low Difference: Exit Probability of all Establishments



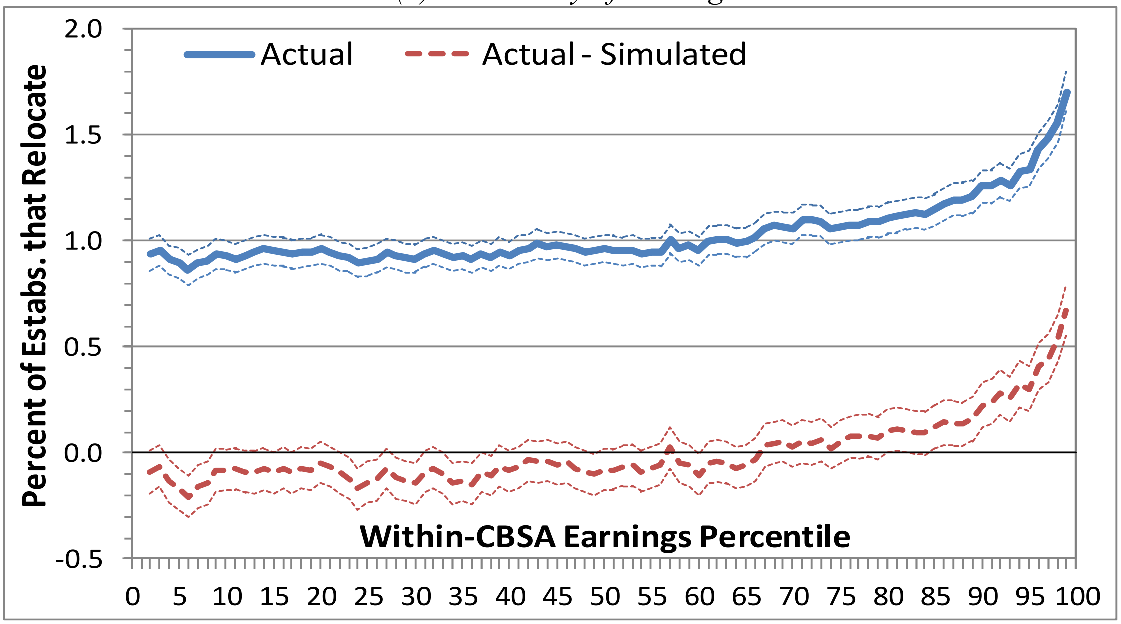
(d) High – Low Difference: Exit Probability of Young Establishments



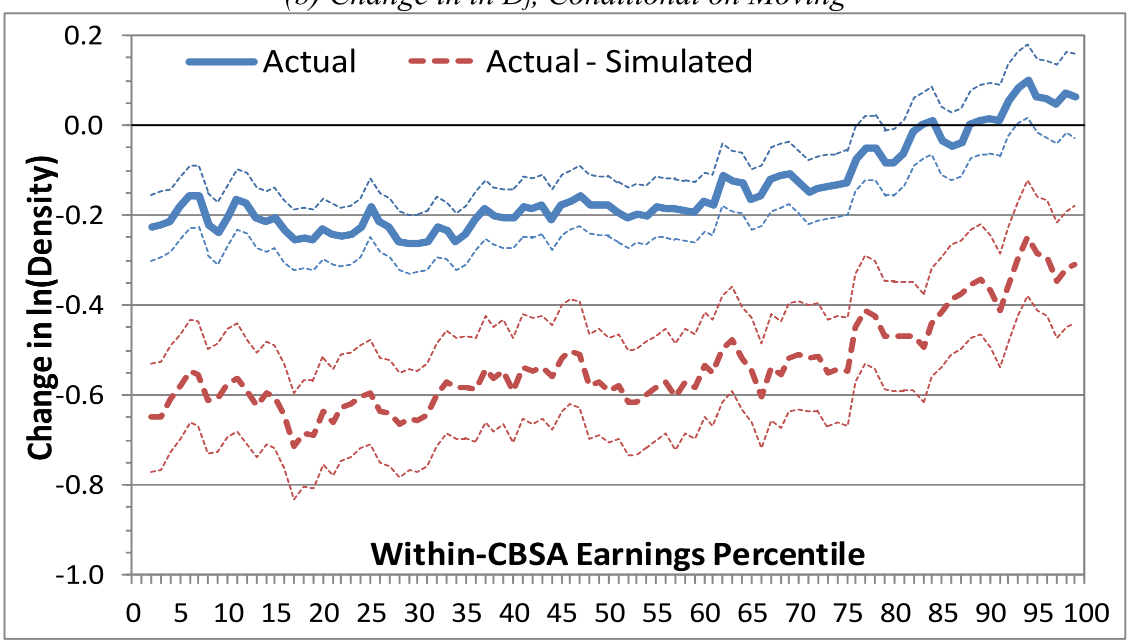
Notes: Top panels report the probability of an establishment exiting by percentiles of the within-CBSA earnings distribution for CBSAs grouped into the highest and lowest quartiles of the CBSA density distribution. Bottom panels report the difference in exit probabilities between high-density and low-density CBSAs. The left panels are for all establishments while the right panels are for establishments aged 5 years or less. All probabilities and earnings are conditional on establishment characteristics (size, age, industry, multi-unit firm status). All panels report 3-percentile, centered averages to smooth the estimates.

Figure 4. Establishment Relocation Probabilities and Density Changes by CBSA Earnings Percentile

(a) Probability of Moving

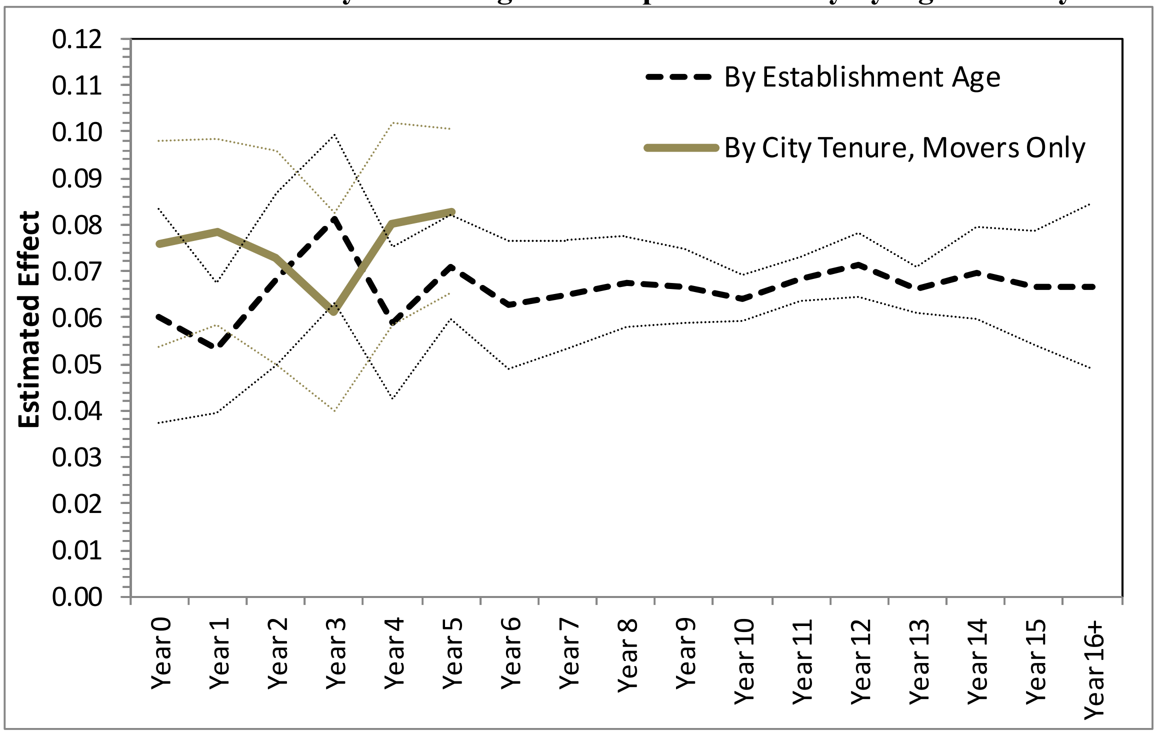


(b) Change in ln D_j, Conditional on Moving



Notes: Top panel reports the probability of an establishment relocating from one CBSA to another by percentiles of the earnings distribution of the origin CBSA. Bottom panel reports the change in an establishment's CBSA (log) density, conditional on relocating, by the same percentile measure. Thick solid blue lines represent outcomes from the data using an earnings and move probability measure that conditions out establishment characteristics (size, age, industry, multi-unit firm status). Thick dashed red lines represent estimates from the data less the outcome predicted from a simulation where all establishments have an equal probability of a move to a random CBSA. Thin dashed lines represent 95 percent confidence intervals. All panels report 3-percentile, centered averages to smooth the estimates.

Figure 5. Estimated Elasticity of Earnings with respect to Density by Age and City Tenure



Notes: The figure plots the predicted elasticity of earnings with respect to density as a function of age and as a function of CBSA tenure, independent of age and conditional on moving to the CBSA within the previous 6 years. Estimates come from equation (4) in the text. See text for details. Thin dashed lines represent standard error bands, with standard errors clustered by CBSA.