# Workplace Concentration of Immigrants*

Fredrik Andersson, Mónica García-Pérez, John Haltiwanger, Kristin McCue, and Seth Sanders†

May 4, 2009

**Abstract**

To what extent do immigrants work with native-born workers? How does this change as immigrants accumulate U.S. specific skills? Do worker and firm characteristics explain the degree of workplace concentration? In this paper we explore these questions in a sample of MSAs using a matched employer-employee database which extensively covers employers in selected states. A key finding is that immigrants are much more likely to have immigrant coworkers than are natives. This finding is driven partly by the geographic concentration of immigrants, but the patterns hold true even within local labor markets. At the same time, most immigrants do have native coworkers: only a small share work in immigrant-only workplaces. The concentration of immigrants is higher for recent immigrants and interestingly for older arrivals among recent immigrants. We find large differences associated with establishment size – concentration is much higher in smaller establishments but is still substantial even in the largest establishments. We conclude that statistical aggregation issues for small establishments explains some of this pattern, but cannot fully account for the inverse relationship between concentration and establishment size. We also explore the mechanisms that underlie the observed patterns of concentration. We find evidence that social network effects, language skills and education help account for the sorting and concentration of immigrants in the workplace.

KEYWORDS: concentration, segregation, immigrant workers, social networks.

---

*We thank the NIH for financial support and participants at the Center for Economic Studies workshop for helpful comments on an earlier draft. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, the Comptroller of the Currency, or the U.S. Department of the Treasury. All results have been reviewed to ensure that no confidential information is disclosed.

†Credit Risk Analysis Divisions, Office of the Comptroller of the Currency; University of Maryland; University of Maryland; Census Bureau; and Duke University, respectively.

# 1 Introduction

Over the last several decades, labor markets in many U.S. cities have absorbed large inflows of new immigrants. The size of these flows has generated intense interest in their effects on the employment and wages of natives, as well as in the extent to which new immigrants have assimilated into the U.S. economy. New immigrants find employment and accumulate location-specific skills and work experience, gradually becoming integrated into local economies and potentially changing them in substantial ways. While outcomes of this process have been the subject of much research, less is known about the process itself. Which businesses hire immigrants? To what extent do immigrants work with natives? How does this change as immigrants accumulate U.S. specific skills? Do the characteristics of different immigrant groups and different geographic labor markets affect the way in which this plays out?

A lack of suitable data has limited economists' ability to address these questions. Our contribution is to bring to bear a very rich set of matched employer-employee data that allows us to identify immigrants, their coworkers, and their employers. Our unique data permit quantifying the extent of and covariates of the workplace concentration of immigrants. The paper has two broad objectives. The first is primarily descriptive. The descriptive findings show that immigrants are much more likely to have immigrant coworkers than are natives. This pattern is driven partly by the geographic concentration of immigrants, but the patterns hold true even within local labor markets. At the same time, most immigrants do have native coworkers: only a small share work in immigrant-only workplaces. The concentration of immigrants is higher for recent immigrants and, conditional on recent arrival, for older immigrants: part of the assimilation process is a movement towards more interaction with natives in the workplace, and younger arrivals are more likely to work with natives. We find large differences associated with firm size: concentration is much higher in smaller firms, but is far from zero even in the largest firms. We also find substantial variation in the extent of immigrant concentration across industries even after controlling for a detailed set of location, employer and employee characteristics.

Our second broad objective is to investigate factors that drive the observed concentration after controlling for basic employer and employee characteristics. Both the existing literature and our descriptive findings suggest that it is important to consider how businesses hire their employees and the choices that businesses make about the skill mix of their workforce. One relevant issue here is the role that language skills play in governing interactions among employees and between employees and customers. A second issue is the role of social networks in the process that matches workers and firms. A third issue is human capital - the sorting and concentration of immigrants in the workplace may reflect sorting by skills. In the second part of the paper, we explore the role of these factors. We find evidence that immigrants with primarily immigrant coworkers are likely to have coworkers who live in the same residential tract. This pattern is robust to inclusion of controls for other closely related factors such as residential segregation. We also find evidence that immigrant workers with poor English speaking ability and low levels of education are more likely to work with other immigrants, with the language variable having a particularly strong effect.

The paper proceeds as follows. Section 2 provides an overview of the relevant theoretical and empirical literature that helps guide our empirical analysis. Section 3 describes the measurement of immigrant concentration, the matched employer-employee data we use in our analysis and the methods we use to explore the correlates of immigrant concentration. In section 4 we present our main results quantifying the extent and nature of immigrant concentration across workers and businesses. Section 5 presents analysis of factors such as social networks, language skills and human capital on the patterns of immigrant concentration. Most of the analysis focuses on native born, recent immigrants and established immigrants without specific reference to country of origin. Section 6 discusses plans to extend the analysis by looking at patterns of concentration by country of origin. Concluding remarks are provided in section 7.

# 2 Background

## 2.1 Literature on earnings differences

Work examining earnings differences between whites and other groups in the U.S. has largely focused on netting out differences in skill (often captured by education and labor market experience) and geography (often using place of residence and urban residence) to assess the potential role of discrimination in labor market outcomes. This assumes that earnings differences are generated either by differing worker characteristics or differing returns to those characteristics. By extension, closing gaps in earnings requires equalizing worker characteristics and their returns across groups. Differences in returns to characteristics are assumed to reflect unobserved ways in which the wage generating process differs and is typically viewed as an upper bound on the potential for discrimination to play a role in explaining wage disparities. A huge number of papers use this approach; some classic examples that examine earnings differences relative to white men are Smith and Welch (1977) for African American men, Borjas (1982) for Hispanic men, Chiswick (1983) for Asian men, and Corcoran, Duncan, and Ponza (1983) for women.

There is also a large literature assessing the sources of earnings differences between immigrants and native born workers (for example, Chiswick (1978), or Butcher and DiNardo (2002)). These papers generally augment the basic human capital framework used in the studies above by allowing for skill differences that are specifically relevant to immigrants. These include potential differences in the value of education and work experience accumulated outside the U.S. and differences in English language skills. Immigrant assimilation into the U.S. labor market is measured as a narrowing of the earnings gap, resulting largely from increased U.S.-specific skills with time spent in the U.S. While there is debate over the speed at which the earnings gap between immigrant and native born workers closes, most studies find a substantial narrowing with time in the U.S. (see Chiswick (1978) and Borjas (1985)).

An older literature in sociology and economics stresses that earnings differences between groups may be driven by the characteristics of the firms that employ the majority and minority groups, rather than solely by human capital characteristics. Usually termed ''dual labor market theory,'' this idea gained considerable attention in the late 1960s and early 1970s (see for example Averitt (1968) or Galbraith (1971)). According to this theory, many firms (especially industrial firms) are not governed by competitive processes. Instead, these firms enjoy market power. They insulate themselves and stabilize their workforce through job training and promotional ladders (Edwards 1972). Firms that are constrained by competition do not invest in work skills and are characterized by low wages and high turnover, with low returns to human capital including job tenure.

The existence of ''good jobs'' and ''bad jobs'' by itself would not imply an earnings disadvantage to minority workers. Sociologists typically rely on a form of employer discrimination to explain why dual labor markets lead to minority disadvantage. Queuing theory suggests that good jobs always have an excess supply of applicants and firms then order workers by preferences and hire down the queue until vacancies are filled. If race or ethnicity plays a role in this ordering, a higher fraction of minority workers will be employed in the secondary market and have relatively low wages and wage growth.

While dual labor market theory per se has largely fallen out of the mainstream literature in economics and sociology, a newer literature that similarly argues that firm characteristics may be partially responsible for the level and growth in earnings of workers has gained growing acceptance. Wages appear to be positively correlated with firm productivity and firm size (Abowd, Haltiwanger, Jarmin, Lane, Lengermann, McCue, McKinney, and Sandusky 2005). While more controversial, there is some evidence that firm-level technological adoption also affects workers wages (Dunne, Foster, Haltiwanger, and Troske 2004). Lengermann (2002) finds that coworker characteristics, in addition to firm characteristics, may affect wages. Specifically, he finds that having more skilled coworkers independently raises a worker's wages. If firm characteristics play a major role in wage setting, then understanding how race and ethnicity affect the matching of workers to firms becomes important for understanding wage disparities across groups. Lengermann, McKinney, and Pedace (2004) explore the issues of sorting of immigrants across firms and find that it matters for wage differences between native born and immigrant workers.[1] We now turn to theories of worker segregation with special attention to how immigrants sort into firms.

## 2.2 Literature on segregation

Four broad overlapping theories explain segregation of workers into firms. These theories focus on sorting based on (a) productive characteristics, (b) preferences of workers or employers, (c) information available to workers or employers, or (d) cost of commuting to jobs. Some, but not all, of these theories imply that segregation results in a disadvantage for one group of workers relative to another.

There is substantial evidence of segregation by skill. For example, Kremer and Maskin (1996) look at the sorting of high and low skilled workers into firms over time and across three countries, the U.S., Britain and France. They find a high and rising correlation between worker skill levels in firms over the 1970s and 1980s. This may occur either because a firm demands a particular type of worker (for example skilled workers) or because coordination within a firm demands that workers share a common characteristic such as a common language. Cabrales, Calvo-Armengol, and Pavoni (2008) emphasize a different skill-based mechanism: if a worker's utility is a function of both absolute wages and their wages relative to those of coworkers, and if movement of workers across firms is costless, complete segregation of workers by skill is optimal. A mixed-skill workforce generates wage inequality within a firm, reducing worker utility. All workers are made better off by grouping workers with similar skills and avoiding these reference group costs. Regardless of the mechanism, segregation by skill will cause immigrant-native differences in the distribution of skill to contribute to segregation. For example, immigrants are both much more likely than natives to have an 8th grade education or less (23% vs 5.2% for natives in the 2000 census), and also more likely to have an advanced degree (10.3% vs. 8.6% for natives). Therefore, firms that specialize in hiring exclusively low-skilled or exclusively high-skilled workers will tend to have a

[1]Some of our basic findings on immigrant concentration are also found in Lengermann, McKinney, and Pedace (2004). Using the same data infrastructure that we use in this paper, they find for example differences in immigrant concentration by industry and employer size.

workforce that has a higher fraction of immigrants than the fraction in the population.

Language differences provide another productivity-based motivation for segregation. If working with someone who does not speak the same language generates transaction costs, employers may increase productivity by hiring only workers who share a common language. In this case, immigrants from non-English speaking countries may be particularly likely to be segregated, and may also be particularly likely to work with their compatriots rather than other immigrants. Lang (1986) develops a formal model of wage differences arising because of the costs to firms of having to pay a premium for bilingual workers who can bridge the language barrier. One of the results of this model is that complete segregation would occur if both capital and labor were owned by each language group. Hellerstein and Neumark (2003) find evidence that Hispanics with poor English-language skills are particularly likely to work with other Hispanics. Their data do not allow them to examine how much of this is due to Hispanic workers working for Hispanic-owned firms as in the Lang model.

Becker (1957) is the classic model of preference-based segregation. In this model, segregation of workers by race occurs as the result of discriminatory preferences on the part of co-workers. White workers would demand a premium to work with black workers. In response, firms segregate workers into separate facilities, avoiding the need to pay a wage premium to discriminating white workers. Depending on conditions including the relative size of the minority and majority group, the number of firms, and returns to scale in production, segregation may be extreme but with limited disadvantage in wages to the minority group. Dual labor market theory, described above, also generates wage differences across groups if discriminating employers put minority job candidates lower down the queue. In this case, higher wages in the primary sector ensure that a higher fraction of the majority group works in the primary sector and hence gives a wage advantage to the majority group.

Information-based theories concentrate on the mechanisms that workers use to find jobs. For example, firm use of employee referrals to fill jobs may contribute to workplace segregation. For workers, use of personal contacts to search for jobs is inexpensive and has relatively high rates of success (Holzer 1988). For employers, employee referrals provide both a low cost recruitment strategy and, on average, new hires with higher productivity and lower turnover rates (Holzer (1987), Montgomery (1991)). If workers tend to refer others who have similar characteristics, use of referrals can increase rates of segregation. Elliot (2001) finds that recent Latino immigrants are more likely than blacks or Latino natives to use personal contacts to find jobs. Weak English skills explain much of this difference. A greater reliance on referrals in small workplaces in combination with a concentration of recent immigrants in small firms also contributes to the difference.

Information flows may combine with residential segregation to contribute to workplace segregation. Neighborhoods play an important role in who you know and hence may provide important job contacts and references. Several papers have established that workers in the same firm are disproportionately from the same neighborhoods. Using data from Boston, Bayer, Ross, and Topa (2008) find that a worker is about one-third more likely to work with someone who lives in the same census block as to work with someone who lives in other blocks in their block group (typically eight or so contiguous blocks). This comparison of blocks to block groups provides important evidence that having coworkers who are neighbors does not stem from unobserved factors such as transportation routes or distance that make a place of employment a natural place to work for those living in a particular location. Many of these unobserved factors would be similar for a block group and block of residence, and so should have similar effects on the likelihood of working with more or less immediate neighbors. This paper is limited in that the exact establishment cannot be observed and sample sizes as well as the ethnic make-up of Boston restrict the authors' investigation to black-white differences.

Hellerstein, Neumark, and McInerney (2008a) also present evidence of neighborhood network effects. Using matched employer-employee data, they compare how likely an individual is to work in the same establishment as his neighbor, relative to the likelihood that this would result if their employer hired workers randomly from the geographic areas of residence of all individuals who work in the employer's census tract. Their dataset is large enough to disaggregate the analysis for whites, blacks and Hispanics. They find that another worker living in the same census tract has twice the probability of working in your firm than what one would expect from randomness. They do not investigate the importance of other mechanisms for sorting workers into firms.

A final theory of the sorting of workers into firms also works through residential segregation but focuses on the fact that not all jobs are equally accessible from different places of residence. Kain (1968) investigated employment patterns of blacks and whites in Chicago and Detroit. He found that blacks were unlikely to be employed in areas that were predominantly white, that blacks would have higher employment rates if housing segregation was lower, and that the movement of jobs from central cities to suburban areas depressed the employment prospects of blacks. A number of following studies compared employment differences between central city and suburban residents within an urban area. These tests often found employment prospects lower for central city residents, but controlling for unmeasured skill differences between residents of different locations remained an issue in inference. A recent study by Hellerstein, Neumark, and McInerney (2008b) questions the interpretation that a lack of jobs near where blacks live is a major source of racial employment differences. They find that the employment prospects of black residents are positively correlated with the number of nearby jobs in which blacks work, but not with the number of nearby jobs in which whites work. This indicates that even within close geographic proximity, job markets are racially segregated. They conclude that spatial mismatch has little effect on employment prospects of blacks but that what they term racial mismatch—few nearby jobs that employ blacks—has a large effect.

Clearly, residential segregation could contribute to workplace segregation of immigrants. There is ample evidence that immigrants' places of residence are spatially concentrated. Iceland (2009) describes the high level of residential segregation in the U.S. among immigrant groups but also shows that immigrants migrate to neighborhoods that are more ethnically integrated with time in the U.S. However, Portes and Wilson (1980) argue that, unlike for black Americans, residential segregation may aid immigrants—especially new immigrants–while also leading to segregation of workers in firms. Studying the post-Castro immigration from Cuba to Miami, Portes and Wilson show that not only do Cubans in the U.S. work together, many work in firms owned by other Cubans. Moreover, Cuban employees of Cuban-owned firms tended to display the same patterns of wage growth and returns to human capital as workers in firms classified as in the ''primary sector,'' that is firms with a promotion ladder, over 1000 workers, and with high average wages. While an impressive source of employment, it is not clear that this generalizes to other foreign-born groups. Capital owners specifically were forced to leave Cuba, which may have led to higher levels of capital with which to start businesses and more experience with small businesses among Cubans than among other foreign born groups. Having said this, Wilson and Portes report that much of the capital used to start these businesses was accumulated in the U.S. and not transferred from Cuban concerns.

# 3 Methodology and Data

## 3.1 Measuring immigrant concentration

We follow several recent papers that study workplace segregation (Hellerstein and Neumark (2007); Aslund and Skans (2005a), Aslund and Skans (2005b)—henceforth HN and AS) by using the share of coworkers in a particular group as a measure of exposure. That is, we exclude the worker himself when measuring the concentration of immigrants in the business he works in. For worker $i$, employed by business $j$ which has $s_j$ employees, the share of immigrants among coworkers is:

$$C_{ij} = \frac{1}{s_j - 1} \sum_{k \neq i}^{s_j} I_k \tag{3.1}$$

where $I_k$ is an indicator for whether or not worker $k$ is an immigrant. For the sake of brevity, we will refer to this simply as the coworker share. As pointed out by these authors, excluding the worker's own characteristic in calculating concentration ensures that in large samples the coworker share for both immigrants and natives should on average equal the share of immigrants in the workforce in the absence of any systematic concentration. Based on this, we use the difference between the mean coworker share for immigrants and natives to measure immigrant concentration. A positive value indicates that immigrants are more concentrated than would be expected based on random allocation. At the extreme, if immigrants worked only with immigrants and natives with natives, the difference in coworker means would equal one. A negative value for this difference would indicate that immigrants were more likely to work with natives than would be expected based on random allocation—a pattern that could arise where the two groups provide different but complementary skills.

We depart from the approach of these authors in two ways: in the way in which we condition on observable characteristics, and in choice of a normalization to gauge whether the concentration we find is large relative to some alternative. There are two types of questions that can be addressed by conditioning on observable characteristics in studying segregation: to what extent can segregation be explained by differences in the characteristics of the two groups, and which characteristics are most associated with segregation. HN and AS both focus more on the first issue, while we explore some aspects of both questions. As an example to provide some context, the immigrant and native education distributions differ, and particular employers may hire primarily from one part of the education distribution, leading to concentration of immigrants because of differences in skill. HN and AS both use the difference between measured concentration and the amount of concentration that would be generated solely by the way in which education is distributed across employers as their conditional measure of concentration. In contrast, we condition on a worker's own characteristics and on the characteristics of his or her employer (e.g. employer size and industry). Our measure of concentration is the mean difference between immigrants and natives with the same characteristics. We take this approach in part because it allows us to use simple linear regression to address both types of questions: controlling for a worker's own characteristics should remove the effects of education (for example) from the measured difference in coworker mean. The estimated regression coefficients allow us to examine the characteristics of immigrants and natives who work in heavily immigrant workplaces, and the approach also readily accommodates conditioning on continuous variables.

Both HN and AS normalize their measures of concentration, though they choose different references for the normalization. While both of their normalizations have intuitive appeal, we take a different approach. We use the immigrant-native difference in coworker shares as our measure of concentration, but in most cases also present information on the coworker share for natives as a point of reference. Our regression approach makes doing so straightforward, and also allows us to more directly illustrate patterns of concentration. For example, using the regressions to predict means for a given set of covariates allows us to illustrate the strong positive relationship between immigrant concentration and immigrant share of the workforce, when looking across groups defined by characteristics such as area of residence and employer size. In addition, the regression approach using our coworker index at the person level as the dependent variable permits us to effectively normalize our measure of concentration along a number of dimensions. For example, HN normalize to control for between MSA differences in various groups (e.g., differences in the distribution of blacks and whites across MSAs). We control for such differences directly in our regression approach by including MSAs dummies.

## 3.2 Data

We use data from the Longitudinal Employer-Household Dynamics (LEHD) database, which draws much of its data from complete sets of unemployment insurance (UI) earnings records for a subset of U.S. states. The database includes records for 1990 to 2004, though coverage in the earlier years varies across states. Workers' earnings records have been matched to characteristics of their employer gathered in quarterly administrative UI reports and through Census Bureau business censuses and surveys. Basic demographic data are also available for workers, including place of birth. For those born outside the U.S. (and its territories), we treat the year in which they first applied for a Social Security Number (SSN) as the date of their arrival. While this may not precisely date arrival, results based on a sample of immigrants for whom both LEHD and decennial population census data are available suggest that the year the individual first applied for an SSN proxies the reported year of arrival fairly well.[2] In the current anal-

[2]Here we use year of arrival only to split immigrants between those arriving very recently (within the last 5 years) and other immigrants. Comparing our classification based on date of SSN application to one based on reported year of arrival in the 2000 census, 92% of immigrants are classified in the same way according to both sources. The two measures disagree most often for Mexican immigrants: 4% report

ysis, we use data from selected metropolitan areas in 11 states. While we do not use a large number of states, our sample does include five of the six states that had immigrant populations of 1 million or more.

These data give us two unique advantages. First, we have earnings for a group large enough to include more than five million immigrants. Second, we can observe the firms in which workers are employed and identify their coworkers, allowing us to measure both employer characteristics and the characteristics of coworkers. These data have other advantages that we do not exploit here but plan to take advantage of in future work: for example, panel data on both employers and employees that would allow us to track job histories and earnings of immigrants over time in the U.S. as well as to observe contemporaneous changes for native-born workers. The main disadvantage of these data for studying immigration is that they include only on-the-books employees and so do not cover the self-employed or those working in the informal sector. Thus they likely have poor coverage of undocumented immigrants. Coverage of employment in agriculture is incomplete in the LEHD data, so we exclude all employers in that sector.

Calculating the share of coworkers who are immigrants requires at least one coworker, so we restrict our sample to businesses with at least two employees.[3] We measure concentration using a cross-section of data based on the second quarter of 2000, but we use LEHD data for the 1995-2000 period to define business age. In computing the coworker share, we use all coworkers, whether or not they hold other jobs. However, the set of observations used in our regressions includes only one job for each individual: the job where they received their highest earnings in that quarter.

We draw data from employers in 31 MSAs. We include all MSAs that have substantial foreign-born populations and are in states for which we have the required data, but we also included several smaller MSAs that experienced very rapid growth in foreign-born residents between 1990 and 2000.[4] Even in the smallest of our MSAs we have data on more than 30,000 immigrant workers, so small sample sizes are never an issue.

Table 1 summarizes the across-MSA variation in immigrant shares for our sample of MSAs. In the average MSA in our dataset, 19% of workers are immigrants. In what follows, we are interested in deviations in workplace shares from the overall-average.

arriving in the country between 1995 and 2000 while 10% applied for an SSN in that window.

[3]Immigrants account for 27% of employment in single-employee businesses, and 16% of employment in businesses with more than one employee.

[4]More precisely, we started from the list of MSAs used in Singer (2004), which included all MSAs with at least 1 million residents in 2000, and meeting at least one of the following criterion: (i) at least 200,000 foreign-born residents, (ii) a foreign-born share higher than the 2000 national average (11.1%), (iii) 1990-2000 growth rate of the foreign-born population above the national growth rate (57.4%), or (iv) above national average percentage foreign-born in 1900-1930 (''former gateways"). We drop 14 of Singer's 45 MSAs because we do not currently have access to all of the data we need from the relevant states.

Clearly the substantial variation in immigrant share across MSAs will contribute to finding immigrant concentration. The shares of both recent and established immigrants vary substantially across MSAs as well.

For roughly 10% of workers in our sample, we match in additional information on educational attainment and English language skills from the long form of the 2000 population census. Using propensity score models, we develop weights for the matched sample that allow us to closely replicate our results based on the overall sample.[5] We then use weighted estimation with the matched sample to examine the relationship between immigrant concentration and these additional measures of skill.

### 3.3 Regression specifications

Our primary empirical approach is to run a series of regressions with the coworker share as the dependent variable, and individual workers on their primary job as the unit of analysis. As a rough way to capture the way in which immigrant concentration changes with time in the U.S., we include indicators for whether an individual is a recent immigrant (defined as arriving in the last 5 years), or a more established immigrant (arriving more than 5 years ago). Since we use a cross-section of data, the differences between recent and more established immigrants confound the effects of time in the U.S. with changes in labor markets and in immigrant and native characteristics over time. We would need to exploit the panel aspect of our database to seriously address the affects of assimilation, but believe this is useful as a starting point that illustrates that assimilation effects on concentration are likely to be important.

Our initial regression specification is:

$$C_{ij} = \gamma_N + \gamma_{EI} EI_i + \gamma_{RI} RI_i + \beta x_{ij} + \epsilon_{ij} \tag{3.2}$$

where, again, $i$ denotes an individual and $j$ denotes a workplace. Subscripts EI and RI denote established and recent immigrants, respectively. Here, the constant term ($\gamma_N$) represents the mean coworker share for the omitted category, which in our simplest specification consists simply of natives. Coefficients $\gamma_{EI}$ and $\gamma_{RI}$ give us estimates of the differences between immigrants and natives in how likely they are to have immigrant coworkers. We use controls for MSA and for various worker and employer characteristics to examine the extent to which immigrant concentration can be accounted for

[5]We used the following variables in the propensity score procedure: worker age; sex; 11 country of origin groups—Mexico, China, Cuba, El Salvador, India, Korea, Japan, Vietnam, Phillipines, other countries of origin, and natives; log earnings; whether the worker was employed for each of quarters 1, 2, and 3 of 2000; four-digit industry; MSA; working population density; establishment age and size; and the number of establishments owned by the firm.

by differences between natives and immigrants in their geographic distribution and in worker and job characteristics.

Specification (3.2) assumes that the effects of covariates are the same for immigrants and natives. To examine whether this in fact holds, we use an alternative specification that includes interactions between our immigrant dummy variables and other covariates:

$$C_{ij} = \gamma_N + \gamma_{EI} EI_i + \gamma_{RI} RI_i + \beta x_{ij} + \phi_{EI} EI_i * x_{ij} + \phi_{RI} RI_i * x_{ij} + \epsilon \tag{3.3}$$

Once we add interaction terms, the intercept rarely identifies effects for a group of particular interest. To illustrate the effects of a particular covariate in specifications of form 3.3, we present predicted means for immigrants and natives for which we evaluate differences between immigrants and natives based on the pooled distribution of the variables in $x$.

To ease computations with over 30 million records, we use linear regression models rather than adopting an approach that accounts for the limited range of the dependent variable. In this draft, we also ignore the effect of clustering within employer in estimating the standard errors. For most of our specifications, the dependent variable mean is not close to either 0 or 1, which mitigates some of the problems inherent in the linear model. The strong positive correlation in the coworker share among employees of the same business will lead to a downward bias in our estimated standard errors in all worker-level regressions. Given the huge size of our sample, the results we present would generally remain significant at standard levels even if the corrected standard errors were 100 times larger. The few exceptions (in Table 7) are estimates that are too small to be meaningfully different from zero anyway. [6]

## 3.4 Descriptive statistics

Table 2 presents summary statistics for immigrant and native workers in our full sample. The first row gives coworker shares for the three groups. For the average native, about 15% of coworkers are immigrants, while 42% of the coworkers of recent immigrants are fellow immigrants, and 36% of the coworkers of established immigrants are immigrants. The immigrant-native difference in coworker means—our measure of concentration—is .272 for recent immigrants and .214 for more established immigrants, indicating sub-

[6] So far, using statistical software to handle clustering does not seem computationally feasible for our full sample. However, we could put an upper bound on the effect on standard errors by summarizing data at the establishment level for immigrants and natives, and then running our regressions weighted by employment and clustering on establishment. This reduces the number of records to less than two times our number of establishments and cluster size to at most 2.

stantial concentration.

The following rows give demographic information for each group. Recent immigrants are substantially younger than natives while earlier immigrants are older. Combining the two, immigrants are slightly older than natives in our sample. Men substantially outnumber women among both recent and established immigrants, while among natives men are more narrowly in the majority. Differences between immigrant and native women in rates of labor force participation likely contribute to these gaps. Overall, more established immigrants are also more likely to have entered the country before reaching prime working age. This is in part by definition, as immigrants who are old enough to work and arrived in the country within the last five years could not have arrived when they were young children.

Comparing employer characteristics, we generally find only minor differences between immigrants and natives, though if our sample included the self-employed and off-the-books employment we might see a somewhat different picture. Immigrants are more concentrated in manufacturing than are natives, but generally the differences by sector are not particularly large.[7] The size distributions look surprisingly similar. Immigrants are more likely to work in the smallest firms, and less likely to work in the largest, but overall the differences are small. Similarly, differences in distribution of employment by establishment age are also small. However, immigrants are less likely than natives to work for multi-unit firms.

The statistics on earnings show rather large differences, with recent immigrants having lower earnings than natives, while established immigrants have higher earnings. Differences in job tenure likely explain some of this pattern, as recent immigrants are most likely to be in a job that lasts only one quarter, while established immigrants are least likely to be in such a transitory job. Transitory jobs are likely to have particularly low quarterly earnings because most will involve less than three full months of work. They may also be associated with relatively low wage rates and part-time work.

We construct three additional measures from our database using information on worker tract of employment and tract of residence—measures that we use in section 5 to explore the relationship between workplace concentration and social networks. The first of these is simply the share of immigrant workers living in a worker's tract of residence which we use to capture the degree of residential segregation.[8] Because we only

[7] Comparing our estimates to published 2000 population census estimates is inexact for several reasons: our analysis includes only a subset of MSAs; our sectors are defined based on SIC codes while the 2000 industry codes are NAICS based; and we exclude the self-employed and those working off the books, both of which may be included in household estimates of employment. But for reference purposes, in the 2000 decennial census 17% of immigrants and 14% of natives worked in manufacturing, while 8% of immigrants worked in construction compared to 7% of natives (Census Bureau 2005).

[8] Census tracts are small geographic areas with a population between 1,500 and 8,000 individuals. They

have data on those who work, we use the share of immigrants among workers residing in a particular tract rather than the share among all residents. As can be seen in Table 2, immigrants are more likely to live with other immigrants than are natives, but there is little difference in residential segregation between recent and earlier immigrants. On average, both groups live in tracts that are majority non-immigrant.

As a proxy for social networks, we calculate for each worker the fraction of their coworkers who also reside in the worker's tract of residence. This proxy, which we refer to as a neighborhood network index, may reflect many factors. For example, as discussed in Section 2, referrals by current employees may be an important recruitment source, and many referrals may come about through contacts with neighbors. If so, where neighborhood referrals are important we would expect to find people who work together also living close together. Our network variable will, in principle, capture such effects but may more generally capture the extent to which residential location and employment location are correlated. For this reason, we include the residential segregation variable (described above) along with a shared commute variable (described below) as controls when we explore the role of the network variable. Not surprisingly the mean of the network index is small: for the average worker, 1.7% of coworkers live in the same tract. The mean is substantially higher for small businesses and it falls systematically with employer size. While the average is small there is considerable variation across workers and it is the latter variation we exploit in our analysis.

Proximity or convenient transportation links may make residents of certain neighborhoods likely to work at a particular location, which would also result in a relationship between workplace and residence. To control for this effect when using the network measure, we construct two additional variables for each worker based on the share of employees at other businesses located close to his employer (defined as other employers in the same tract) who also live in the worker's residential tract (as in Hellerstein, Neumark, and McInerney 2008a). These measures of the propensity for workplace and residence locations to be connected will control for commuting patterns but will also reflect other connections between workplace and residence such as sorting across locations by skill. We refer to this as a shared commute index which we split between immigrant co-commuters and native co-commuters. The denominator for both components of the shared-commute index is the number of employees working for other employers in a worker's tract of employment.[9] The numerator for the immigrant co-commuter vari-

are designed to be relatively homogeneous with respect to socio-economic characteristics. As such, they are arguably well-suited to serve as a proxy for the geographic reach of a social network: the limited distance between residents of a census tract—both in terms of geography and socio-economic factors—suggests that the likelihood of interactions among residents of the same tract is high relative to the likelihood of interactions between residents of different tracts.

[9] In our sample, there are on average 49 employers per tract (excluding tracts that are strictly residen-

able is the number among that group who are immigrants and who live in the worker's residential tract; the native co-commuter variable is defined analogously. These shares are quite small, but differ between immigrants and natives.

Figure 1 provides some basic information on the distribution of our dependent variable. The three lines plot the cumulative distribution of immigrant coworker share for natives and for recent and more established immigrants. About 13% of natives work in native-only workplaces (having coworker immigrant share=0) in our sample of immigrant-rich MSAs, but the share of immigrant employment in immigrant-only businesses is surprisingly small (2.8% of immigrants). In this set of MSAs, about 10% of the median native's coworkers are immigrants, while for established immigrants the share at the median is about 34%, and for recent immigrants, the share is about 41%.

# 4 Accounting for immigrant concentration

We carry out two sets of exercises to examine the degree and nature of immigrant concentration. First, we address the extent to which observable factors can account for immigrant concentration using a series of regressions with the coworker share as the dependent variable based on specification (3.2). Second we apply specification (3.3) in which we add interactions between the immigrant dummy variables and our explanatory variables. In doing so, we let the difference between coworker shares for immigrants and natives vary with observable characteristics which allows us to determine in what sort of workplaces and for which kinds of workers we see the most concentration.

Table 3 presents estimates of the key parameters from the first set of regressions. The first two columns present estimates of the coefficients on the dummy variables identifying our two immigrant groups—recent immigrants, defined as those who arrived between 1995 and 2000, and more established immigrants who entered before 1995.

In the first row of Table 3, we report results from the base specification without any controls. This simply reproduces the differences in means one finds from the first row of 2. The subsequent rows of Table 3 show the effects of adding each of the sets of controls. We include MSA dummies in all but the first row, but add the other controls one set at a time. Note that in doing this we are allowing the immigrant share of employment to vary with the controls, but assuming that within-cell immigrant concentration is the same for all control categories. Our intent here is to determine whether any of the employer or worker characteristics available to us identify cells with a large share of immigrant employment, but within which the immigrant-native differences are sig-

tial). 7% of tracts with employment have only one employer, and for those tracts, the shared commute variables are zero. Only 9% of workers in our sample work in single-employer tracts.

nificantly smaller than the overall difference. For example, if immigrants were mostly employed in a few industries, but were randomly distributed across workplaces within industry, industry controls would reduce the concentration coefficients to zero because there is no concentration within industry. If that were the case, then explaining immigrant concentration would boil down to explaining why immigrants worked in different industries than natives.

In broad terms, Table 3 shows that our measures of employer and worker characteristics account for a substantial amount of the observed concentration of immigrants in the workplace, but about half of the concentration remains unexplained. Differences in the immigrant share of employment across MSAs and detailed industries account for roughly one-third of total concentration. The other control with substantial explanatory power is the share of immigrants in a worker's residential tract. Living in an immigrant-rich residential tract is positively correlated with the share of coworkers who are immigrants. Thus, the difference between immigrants and natives in the likelihood of working with immigrants is substantially smaller for those living in neighborhoods with similar immigrant shares than it is for immigrants and natives overall. In the last row, we include all of our controls but still find that, compared to natives, the difference in share of coworkers who are immigrant is 14% for recent immigrants and 9% for established immigrants.

In the subsections that follow, we discuss the results of adding particular controls in Table 3 along with results from our second exercise based on (3.3) in which we add interactions with the immigrant dummy variables. Because the patterns identified by the interaction terms are easier to grasp visually, we present the findings from this exercise through graphs of predicted coworker shares. The R-squared for the full model with interactions is 0.567.

## 4.1 Location

Covariates have the greatest potential to account for differences in coworker means when their distribution differs substantially between immigrants and natives. Geography is one dimension along which there are substantial differences. In this section we look only at differences across metropolitan areas, but in section 5 we explore how differences in location within MSAs may also contribute to immigrant concentration.

Immigrants are much more likely than natives to live in the largest metropolitan areas in the U.S. For example, in 2000 55% of immigrants lived in the 9 metro areas having populations of at least 5 million, compared to 27% of natives. While 21% of natives lived in non metropolitan areas, only 3% of immigrants did (Schmidley 2001). Even if immigrants were randomly sorted into jobs within their local labor markets, the fact that many natives live in areas with few immigrants would lead us to find substantial concentration in coworker means for the nation as a whole. By restricting our sample to urban areas that have many immigrants, we increase the overall share of immigrant coworkers above the national average. At the same time, we reduce the difference between immigrants and natives by excluding areas where natives work with few immigrants. But substantial variation in the immigrant share of employment remains across our sample MSAs, as illustrated in Table 1 above.

In Table 3, including MSA dummies almost doubles the R-squared. The reduction in the recent immigrant coefficient between row (1) and row (2) indicates that roughly one-fifth of the overall difference between recent immigrants and natives simply reflects differences in their geographic distribution: unsurprisingly, the metropolitan areas in which immigrants work have higher immigrant shares than the areas in which the average native works. Similarly, about one-quarter of the native/immigrant differential for established immigrants is due to differences in which cities they live in. When we allow immigrant concentration to differ across MSAs (results not reported), we find that immigrant concentration rises very consistently with the overall immigrant share in an MSA, and that concentration is consistently higher for recent immigrants than for more established immigrants.

## 4.2 Worker Demographics

We have limited data on the demographic characteristics of workers in our full sample—basically age and gender, in addition to knowing the country in which a worker was born. As the third and fourth rows of Table 3 illustrate, adding age and gender to the specification with MSA dummy variables has essentially no effect. Allowing the effects to differ between natives and immigrants shows that age does have a weak association with immigrant concentration, though gender does not. As Figure 2 illustrates, older immigrants are somewhat more likely to work with other immigrants than are younger immigrants, but there is little difference by age for natives. Note that because recent immigrants have by definition arrived within the preceding 5 years, age and age at arrival are necessarily highly correlated for that group so what we observe are the combined effects. We would need to move beyond the cross section we are using here to disentangle their effects for recent immigrants.

We use similar methods for most of the following bar charts, so it is useful to clarify how the estimates were constructed for the first chart. The coworker shares here are based on regression estimates from specifications that include all of the sets of variables listed in Table 3. We constructed the estimates presented in the figures using the pooled mean values of all controls except for those used in defining the categories for the bars. So for Figure 2 pooled mean values for all variables except age are used to get predicted values for each age and immigrant status group. The age dummy values are set according to the labels on each of the three clusters of bars. The differences between bars for a given age group are determined by the coefficients on the immigrant group dummy variables and by the product of the interaction effects for the group with the pooled mean the other controls. The age interaction terms determine how much the bars vary across age categories for a given group (i.e. natives, recent, or established immigrants) while the pooled means and coefficients for other variables determine the average level of the bars for a group.

## 4.3 Employer characteristics

Fortunately, we have a rich set of employer characteristics in our data. Most of the measures we use are defined for an establishment (or business location). The measures include establishment size (measured by employment), detailed industry and detailed location.[10]

### 4.3.1 Employer size

We classify employer size into the size bins depicted in Figure 3. Recall that we are excluding establishments with only one worker since the coworker index is by construction not defined for a worker who has no coworkers, so the bins begin with size 2. When we constrain the size effects to be the same for immigrants and natives, adjusting for immigrant/native differences in employer size has virtually no effect on the difference in average coworker share (see Table 3). This is unsurprising given that the distribution of employment across employer size classes (given in Table 2) is similar for immigrants and natives; the immigrant share varies little across these classes.

[10]There are some technical issues in assigning workers to multi-unit establishments in the LEHD data. The UI wage records at the person-level include state-specific employer identifiers which identify the firm that a worker is employed by. The UI wage records link to ES-202 records filed by the firm that provide employment, payroll, industry, and location information for each of the firm's establishments in that state. LEHD has developed algorithms for assigning workers to multi-unit establishments which multiply impute an establishment identifier to affected workers based on the worker's place of residence; the locations, sizes, and ages of the employing firm's establishments; and the timing of the worker's employment. Once a worker is assigned to a specific establishment while working for a given employer, the worker remains with that establishment as long as the worker remains employed with that employer. We weight each implicate based on the estimated probability of being employed at that establishment. More details are available in Abowd, Vilhuber, McKinney, Sandusky, Stephens, Andersson, Roemer, and Woodcock (2006).

Despite this similarity in distributions, when we allow the effects to differ between immigrants and natives we find large size effects. That is, while the share of immigrants is relatively constant across size classes, the concentration of immigrants in the workplace falls substantially with size, as illustrated in Figure 3. Natives are slightly more likely to work with immigrants in larger firms than in smaller firms, while immigrants are much less likely to work with other immigrants in larger firms. For example, 37 percent of coworkers are immigrants for recent immigrants who work at establishments with 10-19 workers, while for recent immigrants at establishments with 500 or more employees, that figure is 26 percent. It is striking that these large effects hold even after controlling for many other factors including detailed industry. We also find that the difference in immigrant shares between recent and more established immigrants falls with size.

To illustrate more concretely how these differences arise, Figure 4 gives cumulative distributions of employment across coworker shares for different employer size classes. The size of the gap between the native and immigrant cumulative distributions represents the size of the differences in means, or the amount of concentration. For the smallest firms, much of the concentration comes from segregated workplaces—those with only immigrant or only native employees. About three-quarters of natives in this size class work only with other natives, while roughly half of recent immigrants and two out of five established immigrants work only with other immigrants. Looking across the different size classes, the share of employment accounted for by all-immigrant and all-native workplaces falls quickly as firm size increases.

We think two mechanisms drive this pattern. One is a size effect we find interesting—a greater tendency for immigrants to work with natives in larger firms. The second is a statistical artifact that arises from the fact that the variance across employers in the coworker share falls with employer size. Given some size-neutral tendency to group like workers together, the difference in mean coworker share will tend to fall as the variance of the mean falls—that is, with employer size.

To see this, consider 2-employee firms. The only possible outcomes are complete segregation (2 natives or 2 immigrants), or integration (1 native, 1 immigrant). If workers are randomly allocated to employers, the expected values of mean coworker shares for immigrants and natives will both equal the overall immigrant share of the (employer size=2) workforce—a difference of 0. But given some tendency to group like workers together, moving some of the weight of the distribution towards segregated workplaces has a relatively large effect on the mean difference because it moves immigrants towards workforces with coworker share=1, and natives towards coworker shares of 0. As employer size increases, extreme values become less likely. If we think of some process shifting weight away from integrated workforces to those with more segregation, with larger firms this has a smaller effect on the mean difference because less of the weight ends up at extreme values. Appendix A shows that this is true for a particular statistical model, but we think this point holds more generally.

This statistical effect is not particularly interesting but we need some way to gauge how much of the size effect it accounts for. Because the change in variance with sample size falls off quite quickly as size increases, we think, consistent with the statistical model in Appendix A, that the statistical effect is unlikely to account for size effects among firms with more than 20 employees. Thus it might be reasonable to think of size effects based on the portion of our sample with at least 20 employees as representing the economic size relationship, while in smaller firms the size effect combines the economic and statistical relationships. Based on this assumption, we fit a flexible functional form to the size effect for the portion of our sample with at least 20 employees, and then use the fitted model to predict the size effect for smaller firms.[11] The lower panel of Figure 3 superimposes this estimated/extrapolated relationship on the actual size-specific means.

For each of our three groups, we separately fit the relationship between mean coworker share and firm size over the range of firm size above 20. The points marked on each line represent the mean predicted coworker share for that employer size grouping. For example, in the lower graph, the 23% marked on the established immigrant line for the 500+ size group is the mean predicted value for established immigrants in this size range—a bit lower than the actual 27% share which is labeled in the upper graph. For groups 2-4, 5-9, and 10-19, the actual coworker share does not influence the fit of the model. The model projection fits the native means closely, which is unsurprising given that the native mean varies little with size. For immigrants, the projections underpredict the coworker means, with a particularly large gap for recent immigrants in the smallest firm size classes. If we take the projection as tracing out the real size effect, the evidence is consistent with a modest underlying size effect. Given that interpretation, the gap between the actual and projected mean then represents the purely statistical effect of size. Consistent with the statistical model in Appendix A, this effect is large for very small firms, but rapidly decreases with size.

We think that the size effects, especially after controlling for the statistical aggregation effects, potentially reflect a number of factors that influence concentration as described in section 2. One reason that size may matter is that the production process (even within industrial sectors) varies across establishments of different sizes. Job

[11] We use linear, quadratic and cubic functional forms to predict the size effect for smaller firms. The quadratic and cubic specifications gave very similar results. We show the quadratic results here.

tasks and division of labor are likely less formal in small establishments, with all workers more likely to interact with coworkers and customers. As such, more concentrated workplaces permit immigrant workers working alongside with immigrants to potentially overcome language and related barriers. Still, as Figure 4 shows, we find that not much mass is concentrated at completely segregated workplaces, except in the smallest workplaces where establishment-level concentration will be high even under random allocation. This suggests that even small businesses find ways to organize their production activities to permit native and immigrant workers working side-by-side.

A related argument is that the hiring process is likely to be more informal for small businesses. Moreover, vacancies are likely to occur less often in small businesses, even if vacancy rates are as high or higher than in medium to large businesses. Both of these effects might make social networks more important in the hiring process for small businesses. At this point of the analysis we cannot distinguish between these or alternative channels for our findings on the role of employer size. For now, we highlight the importance of employer size but we also explore some of these channels in our analysis below.

#### 4.3.2 Industry

Industry differences in immigrant concentration are of particular interest to us because, with the data sources we use here, industry provides the best way of grouping together employers that face similar constraints in choosing the skill mix of their workforce. Significant variation in immigrant concentration by industry would be consistent with technological differences playing an important role in determining how employers combine employment of natives and immigrants.

Controlling for detailed industry reduces our measure of concentration by about 13% for recent immigrants and 15% for established immigrants, while substantially increasing the explanatory power of the regression (as illustrated in Table 3). Whether we control for employer size or not has little effect on this conclusion. It is impractical to illustrate differences across 185 detailed industries, but Figure 5 illustrates differences by broad sector to give a sense of where immigrants are most concentrated. The figure orders sectors according to coworker shares for recent immigrants. Manufacturing is the most immigrant intensive sector in our data: even among natives, immigrants account for more than one out of five coworkers. The concentration of immigrants is also highest in manufacturing: despite the large coworker share for natives, the share for immigrants is about double the native share. Other sectors also show substantial levels of immigrant concentration at least for recent immigrants, with even the least concentrated sector (finance, insurance, and real estate) having an 18% percentage point higher coworker share for recent immigrants than for natives. Note that using the coworker share for established immigrants (or for natives) to order the sectors would change the ranking of sectors—there is less consistency across groups in ranks by sector than we found when looking at variables such as size.

# 5 Exploring social networks, language skills, and human capital as possible explanations for concentration

Section 4 has three main findings. First, there is substantial concentration of immigrants at the workplace. Second, even after accounting for many employer and worker characteristics including employer location, industry and size, concentration remains substantial within employer and worker characteristic groups. Third, the differences in coworker means between immigrant and native workers vary substantially with employer and worker characteristics. The most interesting interaction effects we find are by employer size and industry. These effects are especially intriguing because they arguably reflect differences in how businesses organize their workplaces. As discussed in section 2, a number of potential channels for the type of technology (broadly defined), organizational structure, and recruiting methods of a business to generate immigrant concentration. In this section, we present results of analysis that looks more directly at these possibilities.

To do so, we add variables intended to capture the effects of social networks, language skills and human capital to the regression specifications in section 4. As a proxy for social networks, we use the neighborhood network variable defined in section 3.4. As discussed there, this network measure is likely to be correlated with a variety of factors that connect the workplace to the place of residence. Accordingly, we include the residential segregation and shared commute variables as controls. We also include controls for education and ability to speak English. These variables are interesting in their own right since the concentration of immigrants may reflect sorting by skills captured by these characteristics. But we also include them because social network effects may be related to language skills and level of education.

In this analysis, we use our matched sample since it is for this sample we have education and English language measures. Table 4 shows unweighted summary statistics for our matched sample. Comparing Table 4 to Table 2 illustrates that most of the differences between the matched and full samples are modest. Matched natives have a somewhat lower coworker share than in the full sample, but the shares for immigrants are little changed. There seems to be a general tendency for workers at multi-unit firms to be overrepresented in the matched sample, but the difference is large only for established immigrants. Workers with very transitory jobs also tend to be somewhat overrepresented. The most dramatic differences are in the mean values for the shared commute and neighborhood network index which are much smaller in the matched sample for all three groups. Table A-1 gives weighted statistics for the matched sample to illustrate that the weights we construct bring us reasonably close to replicating the observable characteristics of the full sample.

Table 5 presents summary statistics for the additional variables that we can construct using the matched data. These estimates are weighted using our propensity score weights to make the sample representative of workers in our set of MSAs. Immigrants are much more likely to be high school drop-outs than are natives, particularly very recent immigrants. But immigrants are also overrepresented among those with advanced degrees. The English language measure we use is based on a sequence of questions asked on the census long form questionnaire. All respondents are asked whether they speak English or another language at home. Those who speak another language at home are then asked to categorize how well they speak English–not at all, not well, well, or very well. The group we describe as not speaking English well includes all those responding either "not at all" or "not well" to this second question. Unsurprisingly, recent immigrants are more likely than others to report not speaking English well. Note that this group includes a small fraction of natives, presumably primarily second generation immigrants.

Our analysis of the contribution of these additional variables is in two steps. First, we estimate the model without any interactions to examine the role of these variables in accounting for immigrant concentration. Table 6 is the extension of Table 3 with these additional variables. The first two rows correspond to the first two rows in Table 3, but estimated on the matched sample using our weights. The third row corresponds to the "All of the above" row in Table 3. While estimated concentration is somewhat lower for established immigrants using the matched sample, the estimates using all of the controls from the full sample match Table 3's results very closely. Adding the additional controls without interactions modestly increases the explanatory power of the model, with the English language measure having a more substantial effect than the education measure.

In Table 7 we show results for the specification that includes a full set of interactions so that the effect of the variables can differ between immigrants and natives. To simplify the model, we categorize all immigrants together rather than distinguishing between recent and more established immigrants. In Table 7 we only report the interaction coefficients for the new variables of interest but note that the same type of patterns by employer and employee characteristics discussed in section 4 are present. It is also of interest to observe that the fully interacted model now accounts for a substantially larger share of variation – the R-squared in Table 7 is 0.605.

The results in Table 7 support the hypothesis that social networks play an important role in workplace concentration. The network index variable is positively associated with concentration: natives who work with their neighbors have fewer immigrant coworkers, while immigrants who work with their neighbors have more. It is important to emphasize that this pattern holds controlling for a rich set of employer and employee characteristics and controls for shared commute patterns, residential location, education and language skills.

Turning to the other effects of interest, the coefficients on education controls and the English language variable indicate that high school drop-outs and those with poor English language skills are more likely to work with immigrants. While significant, even the largest education effect—the difference between high school drop-outs and those with some college—is quite small. The effect of having limited ability to speak English is substantial: immigrants who do not speak English well have about 25% more immigrant coworkers.

The control variables are also of interest. The residential segregation index has the expected pattern. A higher share of immigrants in the residential tract is associated with a higher share of coworkers who are immigrants for both natives and immigrants. The shared commute pattern has a somewhat anomalous pattern. We note however that we have found that the pattern of these coefficients is highly sensitive to the inclusion of the residential segregation index. Thus, it appears that these two controls are capturing related effects that may be difficult to identify separately. We note that the main effects of interest are robust to including only one of these controls (i.e., either residential segregation or shared commute).

# 6 Country of origin differences

In the analysis above, we distinguish between natives, recent immigrants and established immigrants. Our data also permit exploring how the patterns of immigrant concentration vary by country of origin. That is, instead of only asking how likely it is for immigrants to have co-workers that are immigrants, we can ask how likely it is for an immigrant from say, Mexico, to have co-workers who are Mexicans. Examining such patterns could potentially shed further light on the relative merit of various language- and culture-based explanations for immigrant workplace concentration. In particular, we would expect social networks to have much stronger effects within country-of-origin groups, while language-based explanations would imply similar effects for immigrants from different countries that share the same language. We plan to investigate these issues in the next draft of this paper.

# 7 Concluding remarks

Using matched employer-employee data that comprehensively cover employment in our sample of MSAs, we find that immigrants are much more likely to work with each other—and hence less likely to work with natives—than would be expected given random allocation of workers. This is in part driven by the distribution of immigrants across MSAs, but within an MSA, substantial concentration remains. We find evidence that suggests that immigrant assimilation into the U.S. workforce includes a tendency to have more native coworkers with more time in the U.S. We also document that immigrant concentration is greatest in small firms, and varies substantially across industries.

After presenting descriptive results, we examine possible underlying causes of this concentration. We find evidence that supports the hypothesis that social networks, language skills and education are important factors in accounting for workplace concentration. Our results indicate that natives who live near coworkers are more likely to work with others who are native born. The effect for immigrants is similar—they are more likely to work with immigrants if they live near coworkers—but larger. These findings hold even when controlling for a variety of other factors (e.g., residential segregation and commuting patterns) that could lead to a correlation between residential and employment location. We also find that workers who do not speak English well and less educated workers are more likely to have immigrant coworkers. These effects are of interest in their own right since they suggest some of the workplace concentration we observe is associated with sorting by skill and language but including these controls also demonstrates the robustness of our findings on social network effects.

# References

ABOWD, J. M., J. C. HALTIWANGER, R. S. JARMIN, J. I. LANE, P. LENGERMANN, K. MCCUE, K. MCKINNEY, AND K. SANDUSKY (2005): "The Relation among Human Capital, Productivity and Market Value: Building Up from Micro Evidence," in *Measuring Capital in the New Economy*, chap. 5, pp. 153–198. University of Chicago Press for the NBER, Chicago.

ABOWD, J. M., L. VILHUBER, K. MCKINNEY, K. SANDUSKY, B. STEPHENS, F. ANDERSSON, M. ROEMER, AND S. WOODCOCK (2006): "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," LEHD technical paper, U.S. Census Bureau.

ASLUND, O., AND O. SKANS (2005a): "Measuring Conditional Segregation: Methods and Empirical Examples," Working Paper, Institute for Labor Marker Policy Evaluation (IFAU).

——— (2005b): "Will I See you at Work? Ethnic Workplace Segregation in Sweden 1985-2002," Working Paper, Institute for Labor Marker Policy Evaluation (IFAU).

AVERITT, R. T. (1968): *The Dual Economy: The Dynamics of American Industrial Structure*. Norton, New York.

BAYER, P., S. ROSS, AND G. TOPA (2008): "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economics*, 116(6), 1150–1196.

BECKER, G. S. (1957): *The Economics of Discrimination*. University of Chicago Press, Chicago.

BORJAS, G. J. (1982): "The Earnings of Male Hispanic Immigrants in the United States," *Industrial and Labor Relations Review*, 35(3), 343–353.

——— (1985): "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3(4), 463–489.

BUTCHER, K. F., AND J. E. DINARDO (2002): "The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses," *Industrial and Labor Relations Review*, 56(1), 97–121.

CABRALES, A., A. CALVO-ARMENGOL, AND N. PAVONI (2008): "Social Preferences, Skill Segregation, and Wage Dynamics," *Review of Economic Studies*, 75(1), 65–98.

CENSUS BUREAU, U. S. (2005): "Census 2000 STP-159)," Special Tabulations, U.S. Census Bureau, See `http://www.census.gov/population/cen2000/stp-159/foreignborn.pdf` and `http://www.census.gov/population/cen2000/stp-159/native.pdf`.

CHISWICK, B. R. (1978): "The Effect of Americanization on the Earnings of Foreign-born Men," *Journal of Political Economics*, 86(5), 897–921.

——— (1983): "An Analysis of the Earnings and Employment of Asian-American Men," *Journal of Labor Economics*, 1(2), 197–214.

CORCORAN, M., G. DUNCAN, AND M. PONZA (1983): "A Longitudinal Analysis of White Women's Wages," *Journal of Human Resources*, 16(4), 497–520.

DUNNE, T., L. FOSTER, J. C. HALTIWANGER, AND K. TROSKE (2004): "Wage and Productivity Dispersion in U.S. Manufacturing: The Role of Computer Investments," *Journal of Labor Economics*, 22(2), 397–429.

EDWARDS, R. C. (1972): "The Social Relations In the Firm and Labor Market Structure," in *Labor Segmentation*, ed. by M. R. R.C. Edwards, and D. M. Gordon, pp. 3–26. Mass: Heath, Lexington.

ELLIOT, J. R. (2001): "Referral Hiring and Ethnically Homogeneous Jobs: How Prevalent Is the Connections and for Whom?," *Social Science Research*, 30(1), 401–425.

GALBRAITH, J. K. (1971): *The New Industrial State*. Mentor, New York.

HELLERSTEIN, J. K., AND D. NEUMARK (2003): "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set," *Annales D'Économie et de Statistique*, pp. 19–78.

——— (2007): "Workplace Segregation in the United States: Race, Ethnicity and Skill," CES 07-02, Center of Economic Studies Discussion Paper.

HELLERSTEIN, J. K., D. NEUMARK, AND M. MCINERNEY (2008a): "Neighbors and Co-workers: Measuring the Importance of Labor Market Networks," Working Paper 14201, NBER.

——— (2008b): "Spatial Mismatch or Racial Mismatch," *Journal of Urban Economics*, 64(2), 464–479.

HOLZER, H. (1987): "Hiring Procedures in the Firm:Their Economic Determinants and Outcomes," Working Paper 2185, NBER.

——— (1988): "Search Method Use by Unemployed Youth," *Journal of Economic Literature*, 6(1), 1–20.

ICELAND, J. (2009): *Where we Live Now: Immigration and Race in the United States*. University of California Press.

KAIN, J. (1968): "Housing Segregation, Negro Employment and Metropolitan Decentralization," *Quarterly Journal of Economics*, 82(2), 175–95.

KREMER, M., AND E. MASKIN (1996): "Wage Inequality and Segregation by Skill," Working Paper 5718, NBER.

LANG, K. (1986): "A Language Theory of Discrimination," *Quarterly Journal of Economics*, 2(101), 363–381.

LENGERMANN, P. (2002): "Is It Who You Are, Where You Work, or With Whom You Work? Reassessing the Relationship Between Skill Segregation and Wage Inequality," LEHD technical paper, U.S. Census Bureau.

LENGERMANN, P., K. MCKINNEY, AND R. PEDACE (2004): "New Evidence on Immigration and Labor Market Outcomes Using Linked Longitudinal Employer-Employee Data," mimeo, U.S. Census Bureau, LEHD.

MONTGOMERY, J. (1991): "Social Networks and Labor Market Outcomes: Toward an Economic Analysis," *American Economic Review*, 81(5), 1408–18.

PORTES, A., AND K. L. WILSON (1980): "Immigrant Enclaves: An Analysis of the Labor Market Experiences of Cubans in Miami," *The American Journal of Sociology*, 86(2), 295–319.

SCHMIDLEY, A. D. (2001): "Profile of the Foreign-Born Population in the United States: 2000," Current Population Reports, Series P23-206, U.S. Census Bureau, `http://www.census.gov/prod/2002pubs/p23-206.pdf`.

SINGER, A. (2004): *The Rise of New Immigrant Gateways*. Brookings Institution, Washington.

SMITH, J. P., AND F. WELCH (1977): "Black-White Male Wage Ratios: 1960-70," *American Economic Review*, 67(3), 323–338.

# A Simulations of employer size effects in a statistical model with segregation

If immigrants and natives are randomly allocated to jobs in proportion to their presence in the working population, the expected difference between immigrants and natives in the share of coworkers who are immigrant is zero regardless of employer size. However, we find that the distribution of immigrants across workplaces is inconsistent with random allocation, and that concentration is particularly high in small businesses. This raises the question of whether we should expect a general tendency to segregate to have the same effects on measured concentration in small and large businesses. The following sets up a statistical model that incorporates a tendency to segregate. The model is then used to simulate concentration by employer size. Under this model, the tendency to segregate has a much larger effect on concentration for very small employers than for those of modest or large size.

Suppose that employers of size $s$ draw their workforces randomly from the population, but that some fraction of initial draws that involve an integrated workforce (i.e. some natives and some immigrants) are rejected and replaced with a new draw. For simplicity, we treat these draws as with replacement and assume that all employers are the same size, rather than dealing with a distribution of employer sizes. Assume that the outcome of each draw can be described using the binomial probability mass function:

$$b(i,s) = \begin{pmatrix} i \\ s \end{pmatrix} p_D^i (1-p_D)^{s-i} \tag{A.1}$$

where $i$ represents the number of immigrants in the workforce draw, $s$ represents employer size, and $p_D$ represents the fraction of workers who are immigrants in the group being sampled in draw $D$. For the initial draw, the parameter $p_0$ will equal the overall share of immigrants in the workforce.

Suppose that employers discard a draw with probability $d$ which depends on workforce composition and a parameter $\theta$ that indexes the tendency to segregate ($0 \leq \theta \leq 4$).

$$d(i;s,\theta) = \frac{i}{s}\left(\frac{s-i}{s}\right)\theta \tag{A.2}$$

If an employer draws only immigrants or only natives, then $d = 0$ and the original draw is kept. If there are some of both types of employees, then the workforce is redrawn with probability $d$. This shifts some of the probability mass from more integrated towards more segregated types of employee mixes. Figure 1 illustrates the shape of $d()$ for various values of $\theta$.

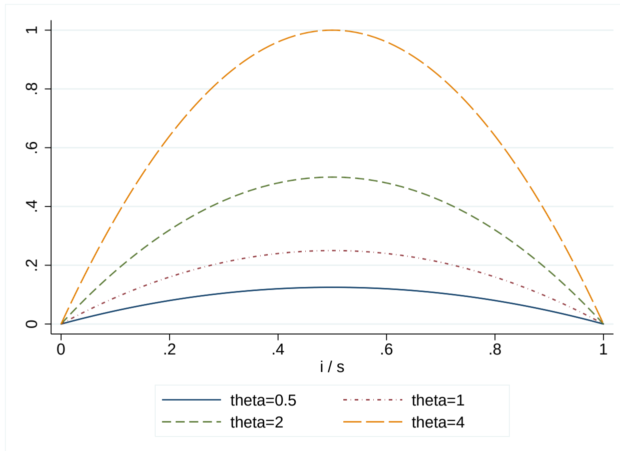
For $\theta = 4$, all draws with immigrants making up exactly half the workforce ($i/s = .5$) are discarded in the first round. However, even with $s = 2$, the final distribution includes some workforces with i/s = .5 because 1 immigrant and 1 native can be drawn in the second round.

If immigrants account for a small share of the population, they are disproportionately included in integrated workforces in the first draw. Because of this, the population that the second draw is taken from has a somewhat higher share of immigrants than the initial population. For example, with $s = 2$ immigrants are always half of the workers in discarded first round draws, no matter what $p_0$ is.

Thus while we assume that the final draw is also binomial, the relevant immigrant share is given by:

$$p_1 = \frac{\sum_{j=1}^{s} b(j,s;p_0) * d(j;s,\theta) * j}{\sum_{j=1}^{s} b(j,s|p_0) * d(j;s,\theta) * s} \tag{A.3}$$

## Figure 1: Shape of function d




and

$$Pr(i;s,p_0,\theta) = b(i,s|p_0) * (1 - d(i;s,\theta)) + b(i,s|p_1) * \left( \sum_{j=0}^{s} b(j,s|p_0) * d(j;s,\theta) \right) \quad \text{(A.4)}$$

where the first term represents the probability that the initial draw has $i$ immigrants and is not discarded, and the second term represents the probability that the final draw has $i$ immigrants and that an initial draw was discarded.

For the simple case $s = 2$ and $\theta = 4$ (so $d = 1$ for the only integrated workforces—those with 1 immigrant, 1 native), $p_1 = .5$, and the probability of observing a workforce with 1 immigrant and 1 native in the final distribution simplifies to $p_0(1-p_0)$ (half the binomial probability). Figure 2 illustrates the difference between the distribution of the coworker mean with segregation and without for employers of varying size. It uses parameter values $\theta = 4$ and $p_0 = .25$. Smaller values of $\theta$ would reduce the shift in the distribution, while smaller values of $p_0$ shift the weight of both distributions to the left.

For immigrants, mean share of coworkers who are immigrant for employer size $s$ is:

$$E(cw_I|s) = \sum_{i=0}^{s} \left( Pr(i|I_j = 1; s, p_0, \theta) * \frac{i-1}{s-1} \right) = \sum_{i=0}^{s} \left( Pr(i;s,p_0,\theta) * \frac{i}{sp_0} * \frac{i-1}{s-1} \right) \quad \text{(A.5)}$$

and for natives,

$$E(cw_N|s) = \sum_{i=0}^{s} \left( Pr(i;s,p_0,\theta) * \frac{(s-i)}{s(1-p_0)} * \frac{i}{s-1} \right) \quad \text{(A.6)}$$

The difference is then:

$$E(cw_N - cw_I|s) = \sum_{i=0}^{s} \left( Pr(i;s,p_0,\theta) * \frac{i[p_0(s-i) - (i-1)(1-p_0)]}{s(s-1)p_0(1-p_0)} \right) \quad \text{(A.7)}$$

Figures 3 to 5 plot out the relationship between employer size and coworker means for various values of the immigrant share of the overall workforce $p$(different colored lines in each graph), using segregation parameter $\theta = 4$. Figure 3 graph gives the mean by firm size for immigrants,

**Figure 2:** Immigrant share distribution with and without segregation

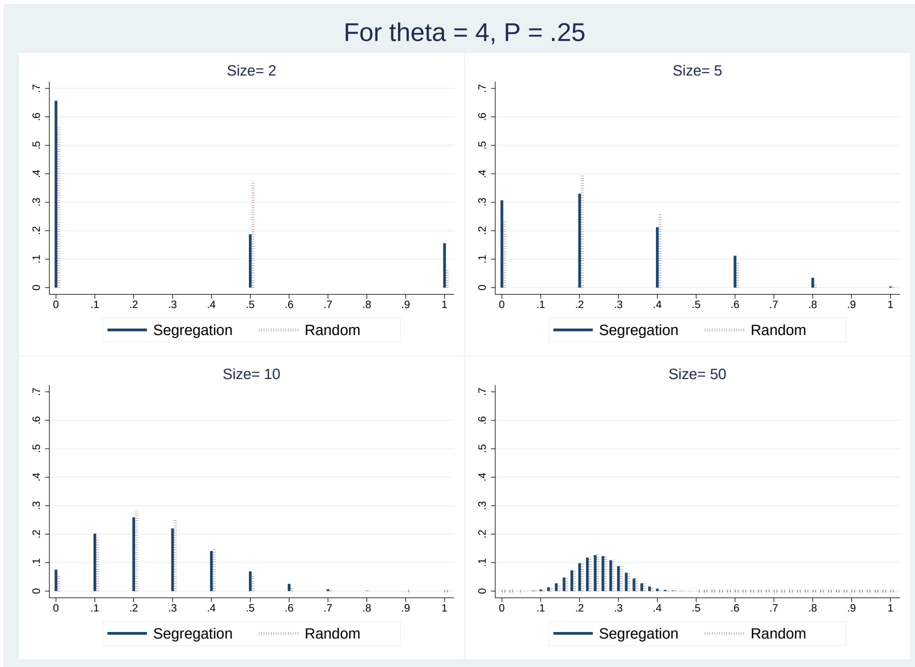



Figure 4 is for natives, and Figure 5 gives the difference between them. Figure 6 repeats Figure 5, except that it is parameterized to represent a lower level of segregation ($\theta = 1$). Examination of these figures makes a couple of patterns clear: (i) For very small employers ($< 10$ employees), the model can generate a large difference in coworker means, even with a relatively mild tendency to segregate. (ii) Even for large theta, this model generates essentially no segregation in large firms.

**Figure 3:** Immigrant coworker mean and employer size ($\theta = 4$)

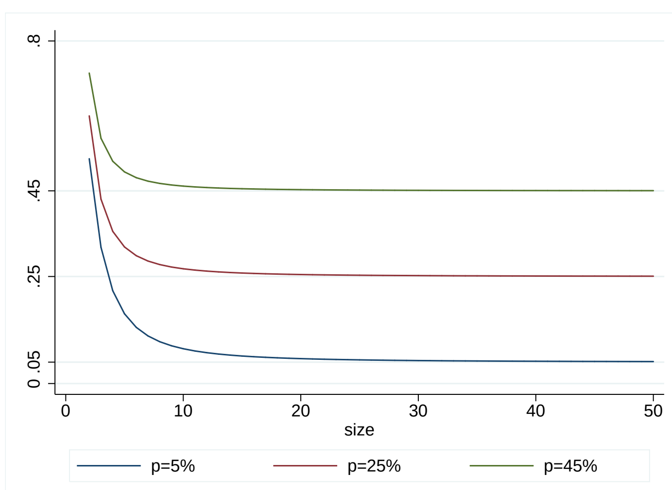


**Figure 4:** Native coworker mean and employer size ($\theta = 4$)

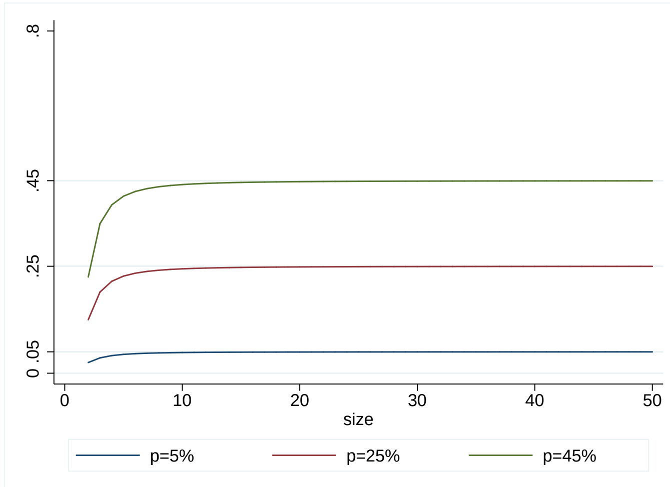



**Figure 5:** Immigrant-native difference in coworker mean and employer size ($\theta = 4$)

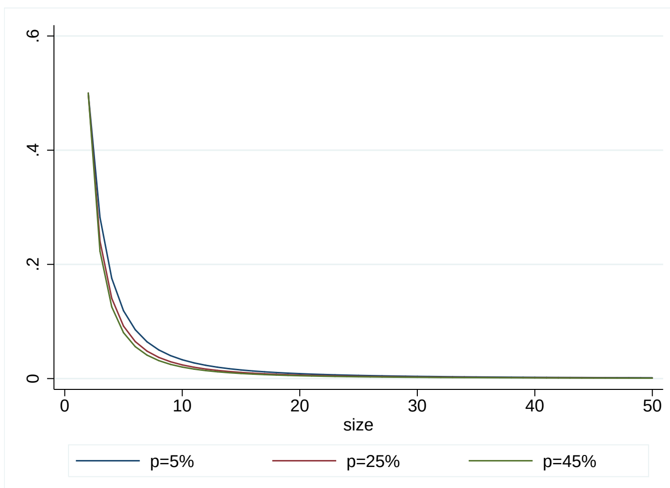
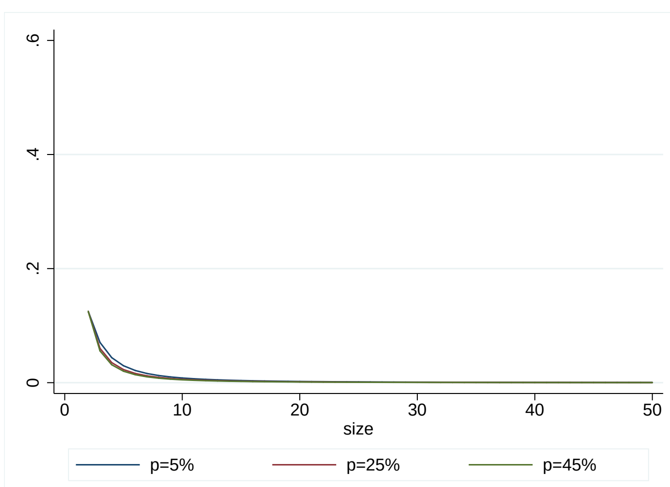



**Figure 6:** Immigrant-native difference in coworker mean and employer size ($\theta = 1$)



## Table A-1: Characteristics of Weighted Matched Sample

| Weighted | | Immigrants Recent | Immigrants Established | Native |
|---|---|---|---|---|
| **Coworker Share** | | 42.0 | 36.1 | 14.3 |
| **Age** | | | | |
| | Age<30 | 44.6 | 19.8 | 30.9 |
| | 30<Age<40 | 35.7 | 34.3 | 28.6 |
| | Age>40 | 19.7 | 45.0 | 40.0 |
| **Male** | | 56.9 | 62.4 | 51.9 |
| **Age at arrival (*)** | | | | |
| | <12 | 0.9 | 12.3 | . |
| | 12-25 | 36.3 | 51.5 | . |
| | 26-35 | 38.4 | 25.4 | . |
| | 35+ | 24.5 | 10.8 | . |
| **Establishment size** | | | | |
| | 2-9 | 8.4 | 9.2 | 7.8 |
| | 10-49 | 23.6 | 23.2 | 22.9 |
| | 50-99 | 13.4 | 13.4 | 13.4 |
| | 100-499 | 29.1 | 30.6 | 29.1 |
| | 500 or more | 25.0 | 23.1 | 26.8 |
| **Firm has multiple establishments** | | 34.6 | 36.3 | 42.6 |
| **Establishment age** | | | | |
| | 0-1 years | 13.0 | 11.3 | 11.9 |
| | 2-4 years | 26.2 | 22.2 | 24.6 |
| | Age 5 or more | 60.8 | 66.6 | 63.6 |
| **Sector** | | | | |
| | Construction | 4.85 | 5.47 | 5.71 |
| | Manufacturing | 18.75 | 21.70 | 13.43 |
| | Transportation & utilities | 2.60 | 3.77 | 5.26 |
| | Wholesale | 6.46 | 6.73 | 6.18 |
| | Retail | 23.99 | 18.86 | 22.46 |
| | FIRE | 3.01 | 5.14 | 6.71 |
| | Services | 40.34 | 38.34 | 40.26 |
| **Log quarterly earnings on primary job** | | 8.15 | 8.54 | 8.34 |
| **Consecutive quarters on 2000-Q2 job:** | | | | |
| | Quarter before AND after | 57.5 | 70.4 | 64.2 |
| | Quarter before OR after (not both) | 34.7 | 24.0 | 28.0 |
| | Neither quarter before NOR after | 7.8 | 5.6 | 7.8 |
| **Immigrant share of workers in residence tract** | | 36.5 | 36.3 | 14.2 |
| **Neighborhood network index** | | 2.00 | 1.50 | 1.78 |
| **Shared commute index:** | | | | |
| | Immigrant co-commuters | 0.13 | 0.09 | 0.04 |
| | Native co-commuters | 0.21 | 0.17 | 0.47 |

Source: LEHD database and author calculations.
(*) Year of application for a SSN is used as a proxy for time of arrival in the U.S.
Note: The unit of observation is a worker. All figures represent percentages. There are 3,549,111 matched workers in total for our group of MSAs.

**Table A-2: Regression model without interactions**

| Covariates | Coefficient |
|---|---|
| **Recent immigrant** | 0.1283 |
| **Established immigrant** | 0.0784 |
| **Neighborhood network index** | 0.0410 |
| **Shared commute index** | |
| Immigrant co-commuters | 1.9923 |
| Native co-commuters | -0.5932 |
| **Log quarterly earnings on primary job** | 0.0020 |
| **Consecutive quarters on 2000-Q2 job** | |
| Worked for employer in quarters 1 and 2 of 2000 | 0.0036 |
| Worked for employer in quarters 2 and 3 of 2000 | 0.0016 |
| Worked for employer in quarters 1, 2, and 3 of 2000 | 0.0016 |
| **Education categories** | |
| College graduate | 0.0013 |
| Advanced degree holder | 0.0075 |
| High school drop-out | 0.0188 |
| High school graduate | 0.0024 |
| **Does not speak English well** | 0.0759 |
| **Immigrant share of workers in residential tract** | 0.1859 |
| **Establishment age** | |
| 0-1 years | -0.0023 |
| 2-4 years | 0.0039 |
| **Firm has more than 1 establishment** | -0.0315 |
| **Firms has >1 estab * Establishment age** | |
| Establishment age 0-1 years | 0.0009 |
| Establishment age 2-4 years | -0.0001 |
| **Worker age** | |
| Age<30 | -0.0088 |
| 30–40 | -0.0019 |
| **Female** | 0.0023 |
| **Establishment size** | |
| 2-4 | 0.0234 |
| 5-9 | 0.0054 |
| 10-19 | -0.0035 |
| 20-49 | -0.0068 |
| 50-99 | -0.0032 |
| 100-499 | 0.0045 |

Note: All standard errors are below 0.001. Controls in all columns include MSA and detailed industry in addition to the variables listed in the table. The unit of observation is a worker. N=3,549,111.

# TABLES

**Table 1: Variation in Immigrant Share of Workforce across Sample MSAs**

| | Percent Immigrant | | |
|---|---|---|---|
| | **Total** | **Recent** | **Established** |
| **Mean** | 18.86 | 3.40 | 15.46 |
| **Standard Deviation** | 10.27 | 1.85 | 8.57 |
| **P25** | 10.57 | 1.94 | 8.52 |
| **Median** | 16.26 | 2.92 | 13.54 |
| **P75** | 26.60 | 4.37 | 22.82 |
| **P90** | 32.58 | 6.03 | 27.23 |

Source: Authors calculations based on LEHD UI-ES202 database.
Note: Unit of observation is an MSA. Immigrant shares are measured as of the second quarter of 2000, and recent immigrants are those arriving between 1995 and 2000. The table presents fuzzed percentiles values.

Table 2: Characteristics of Immigrant and Native Workers, Full Sample

| | Immigrants | | Natives |
|---|---|---|---|
| | Recent | Established | |
| **Coworker share** | 42.1 | 36.3 | 14.9 |
| **Age** | | | |
| Age<30 | 43.6 | 19.7 | 29.3 |
| 30<Age<40 | 35.6 | 33.2 | 30.0 |
| Age>40 | 20.8 | 47.0 | 40.7 |
| **Male** | 56.8 | 56.4 | 51.7 |
| **Age at arrival (*)** | | | |
| <12 | 1.1 | 14.7 | . |
| 12-25 | 36.2 | 49.6 | . |
| 26-35 | 37.0 | 24.8 | . |
| 35+ | 25.7 | 10.9 | . |
| **Establishment size** | | | |
| 2-9 | 8.5 | 9.0 | 8.0 |
| 10-49 | 23.6 | 22.6 | 23.5 |
| 50-99 | 14.4 | 13.3 | 13.6 |
| 100-499 | 31.4 | 30.9 | 29.6 |
| 500 or more | 22.2 | 24.2 | 25.3 |
| **Firm has multiple establishments** | 31.6 | 34.3 | 41.4 |
| **Establishment age** | | | |
| 0-1 | 12.3 | 10.7 | 10.8 |
| 2-4 | 25.6 | 22.0 | 23.6 |
| Age 5 or more | 62.2 | 67.3 | 65.7 |
| **Sector** | | | |
| Construction | 4.7 | 5.5 | 5.9 |
| Manufacturing | 18.9 | 21.3 | 12.8 |
| Transportation & utilities | 3.7 | 5.2 | 6.5 |
| Wholesale | 6.8 | 7.0 | 6.5 |
| Retail | 22.7 | 18.3 | 21.4 |
| FIRE | 3.2 | 5.6 | 7.2 |
| Services | 40.0 | 37.1 | 39.7 |
| **Log quarterly earnings on primary job** | 8.20 | 8.52 | 8.39 |
| **Consecutive quarters on 2000-Q2 job** | | | |
| Quarter before AND after | 58.7 | 70.9 | 65.6 |
| Quarter before OR after (not both) | 32.9 | 23.4 | 26.6 |
| Neither quarter before NOR after | 8.5 | 5.7 | 7.8 |
| **Immigrant share of workers in residence tract** | 36.5 | 35.2 | 15.2 |
| **Neighborhood network index** | 2.10 | 1.78 | 1.70 |
| **Shared commute index:** | | | |
| Immigrant co-commuters | 0.14 | 0.11 | 0.05 |
| Native co-commuters | 0.24 | 0.23 | 0.48 |

Source: LEHD database and author calculations.
(*) Year of application for a SSN is used as a proxy for time of arrival in the U.S.
Note: The unit of observation is a worker. Employer characteristics and earnings are for the first quarter 2000 job with the highest earnings. All figures except for log earnings represent percentages. There are 35,966,450 workers in total for our 31 MSAs.

**Table 3: Contribution of Covariates to Immigrant Concentration (Full Sample)**

| Covariates | Recent immigrant | Established immigrant | R-square |
|---|---|---|---|
| **Full sample** | | | |
| No covariates | 0.272 | 0.214 | 0.198 |
| MSA dummies | 0.224 | 0.156 | 0.379 |
| MSA+: | | | |
| Worker age | 0.225 | 0.155 | 0.379 |
| Worker sex | 0.224 | 0.156 | 0.379 |
| Employer size | 0.224 | 0.156 | 0.380 |
| Employer age | 0.225 | 0.156 | 0.379 |
| Employer age * Multi-unit | 0.221 | 0.154 | 0.387 |
| Industry detail | 0.195 | 0.133 | 0.460 |
| Size and industry | 0.195 | 0.135 | 0.461 |
| Log earnings and full-quarter controls | 0.223 | 0.115 | 0.379 |
| Neighborhood network index | 0.221 | 0.155 | 0.381 |
| Shared commute index variables | 0.222 | 0.156 | 0.379 |
| Immigrant share in residential tract | 0.181 | 0.132 | 0.471 |
| All of the above | 0.143 | 0.089 | 0.495 |

Notes: Figures in the first two columns give the predicted difference in mean coworker share between the immigrant group and natives. As a point of reference, the mean coworker share for natives in the first line is .149 (as in Table 2). It is also .149 for all other specifications if evaluated at the native mean for all included covariates, but somewhat higher if evaluated at the pooled sample mean. The unit of observation is a worker. N=35,966,450 for the full sample. The variables are as described in Table 2, except that we use 185 detailed industry categories in place of sector, and use more detailed size categories for establishments with fewer than 50 employees. All standard errors are less than 0.0001.

**Table 4: Characteristics of Matched Sample Workers (Unweighted)**

| | Immigrants | | |
|---|---|---|---|
| | Recent | Established | Native |
| **Coworker share** | 41.3 | 36.1 | 13.6 |
| **Age** | | | |
| Age<30 | 43.6 | 19.0 | 31.0 |
| 30<Age<40 | 35.8 | 32.7 | 26.0 |
| Age>40 | 20.6 | 48.3 | 43.0 |
| **Male** | 55.9 | 55.2 | 50.8 |
| **Age at arrival (*)** | | | |
| <12 | 0.8 | 11.0 | . |
| 12-25 | 35.4 | 58.4 | . |
| 26-35 | 38.5 | 21.5 | . |
| 35+ | 25.3 | 9.2 | . |
| **Establishment Size** | | | |
| 2-9 | 7.8 | 8.21 | 9.2 |
| 10-49 | 22.7 | 24.2 | 22.3 |
| 50-99 | 13.4 | 14.2 | 13.2 |
| 100-499 | 29.5 | 31.8 | 30.81 |
| 500 or more | 26.7 | 21.6 | 24.6 |
| **Firm has multiple establishments** | 33.6 | 44.6 | 44.8 |
| **Establishment age** | | | |
| 0-1 | 12.5 | 10.0 | 11.4 |
| 2-4 | 25.9 | 19.2 | 14.3 |
| Age 5 or more | 61.6 | 70.8 | 74.3 |
| **Sector** | | | |
| Construction | 5.2 | 4.9 | 5.9 |
| Manufacturing | 17.9 | 17.7 | 12.4 |
| Transportation & Utilities | 2.6 | 17.8 | 6.3 |
| Wholesale | 6.3 | 5.7 | 6.0 |
| Retail | 24.5 | 16.5 | 22.9 |
| FIRE | 2.9 | 4.3 | 6.4 |
| Services | 40.6 | 33.1 | 40.2 |
| **Log quarterly earnings on primary job** | 8.21 | 8.54 | 8.34 |
| **Consecutive quarters on 2000-Q2 job:** | | | |
| Quarter before AND after | 56.0 | 70.1 | 64.1 |
| Quarter before OR after (not both) | 32.9 | 21.6 | 27.3 |
| Neither quarter before NOR after | 11.1 | 8.3 | 8.6 |
| **Immigrant share of workers in residence tract** | 35.8 | 34.6 | 13.5 |
| **Neighborhood network index** | 0.97 | 0.88 | 0.95 |
| **Shared commute index:** | | | |
| Immigrant co-commuters | 0.07 | 0.06 | 0.03 |
| Native co-commuters | 0.13 | 0.14 | 0.32 |

Source: LEHD database and author calculations.
(*) Year of application for a SSN is used as a proxy for time of arrival in the U.S.
Note: The unit of observation is a worker. Employer characteristics and earnings are for the first quarter 2000 job with the highest earnings. All figures except for log earnings represent percentages. There are 3,549,111 matched workers in total for our group of MSAs.

Table 5: Characteristics of Immigrant and Native Workers, Matched Sample

| Weighted | Immigrants Recent | Immigrants Established | Native |
|---|---|---|---|
| **Education categories** | | | |
| High school drop-out | 34.9 | 28.0 | 17.1 |
| High school graduate | 19.4 | 15.8 | 25.4 |
| Some college | 13.5 | 15.0 | 26.0 |
| Bachelor's degree | 20.8 | 33.0 | 23.9 |
| Advanced degree | 11.5 | 8.2 | 7.6 |
| **Does not speak English well** | 31.4 | 16.9 | 1.1 |

Source: LEHD database and author calculations.
Note: The unit of observation is a worker. All figures represent percentages. There are 3,549,111 matched workers in total for our group of MSAs.

Table 6: Contribution of Covariates to Immigrant Concentration (Matched Sample)

| Covariates | Recent immigrant | Established immigrant | R-square |
|---|---|---|---|
| **Matched sample** | | | |
| No covariates | 0.269 | 0.186 | 0.175 |
| MSA dummies | 0.218 | 0.139 | 0.369 |
| Full sample specification | 0.148 | 0.090 | 0.498 |
| Full sample specification +: | | | |
| Education controls | 0.146 | 0.087 | 0.500 |
| English language controls | 0.129 | 0.080 | 0.504 |
| Education and English controls | 0.128 | 0.078 | 0.506 |

Notes: Figures in the first two columns give the predicted difference in mean coworker share between the immigrant group and natives. As a point of reference, the mean coworker share for natives in the first line is .143 (as in Table A-1). It is also .14 for all other specifications if evaluated at the native mean for all included covariates, but somewhat higher if evaluated at the pooled sample mean. The unit of observation is a worker and N=3,549,111. The variables are as described in Table A-1, except that we use 185 detailed industry categories in place of sector, and use more detailed size categories for establishments with fewer than 50 employees. All standard errors are less than 0.001.

Table 7: Network Effects from Coworker Share Regressions

| Covariates | (1) |
|---|---|
| Neighborhood network index | -0.081 |
| Network index * Immigrant | 0.443 |
| Native shared commute index | 0.033 |
| Native shared commute * Immigrant | -0.754 |
| Immigrant shared commute index | -0.774 |
| Immigrant shared commute * Immigrant | 0.351 |
| Immigrant share in residential tract | 0.075 |
| Immigrant residential share * Immigrant | 0.053 |
| High school drop-out | 0.016 |
| High school graduate | 0.002 |
| College graduate | 0.002 |
| Graduate degree | 0.004 |
| High school drop-out * Immigrant | -0.004 |
| High school graduate * Immigrant | 0.009 |
| College graduate * Immigrant | 0.009 |
| Graduate school degree * Immigrant | 0.009 |
| Does not speak English well | 0.216 |
| Does not speak English well* Immigrant | 0.035 |
| R-Square | 0.605 |

Note: All standard errors are below 0.001. Controls in all columns include MSA, detailed industry, employer age and size, worker age and sex, along with interactions with immigrant for each, in addition to the variables listed in the table. The unit of observation is a worker. N=3,549,111.

# FIGURES

**Figure 1:** Cumulative Distribution of Coworker Share by Worker Type

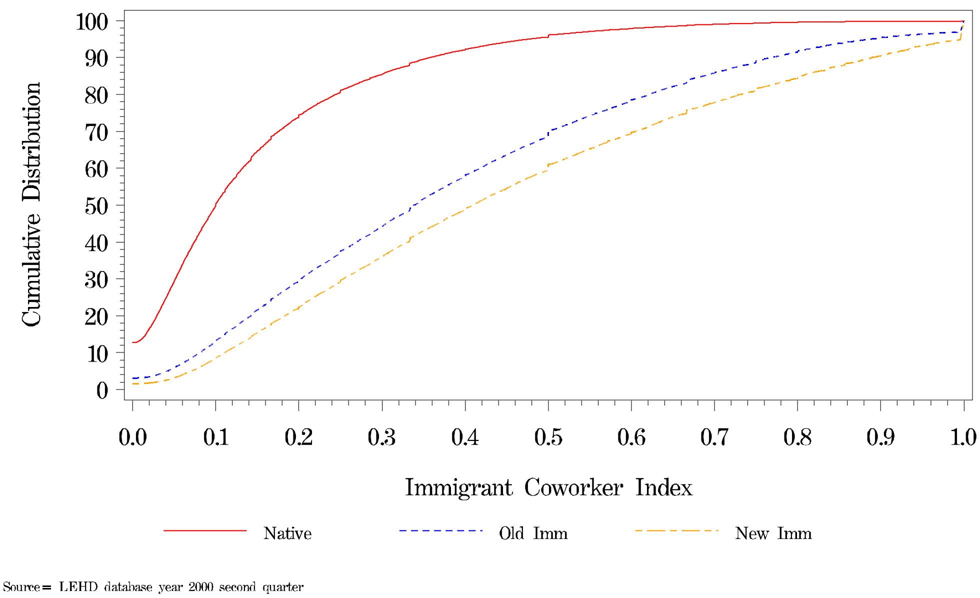
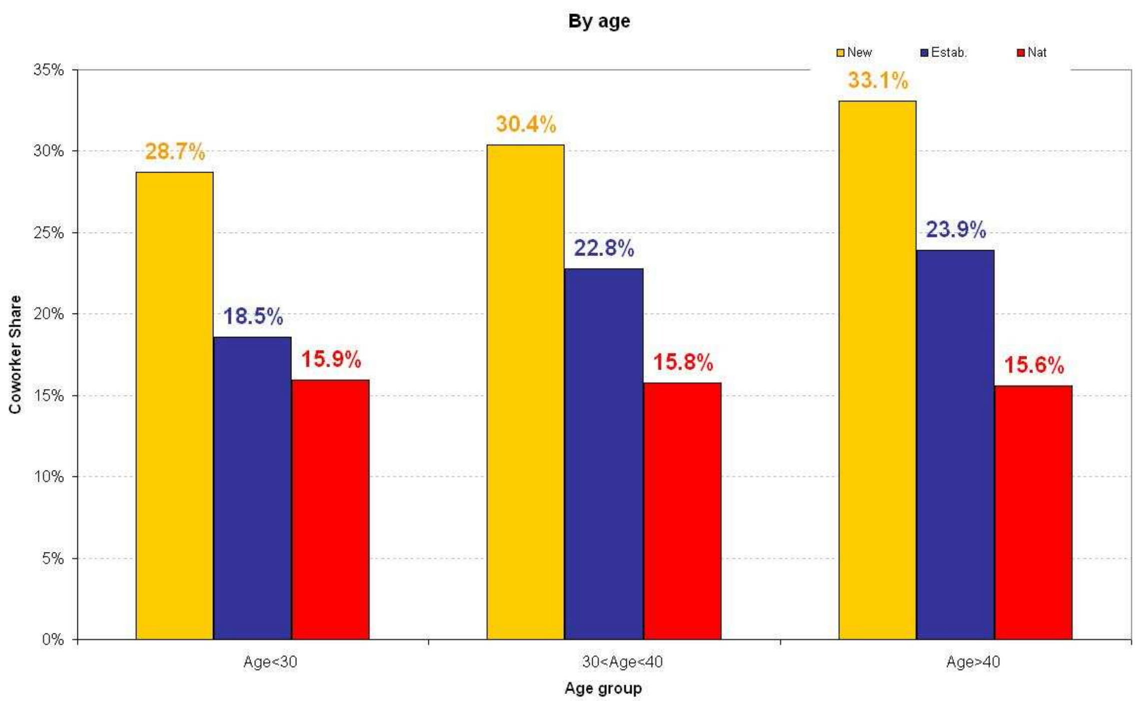
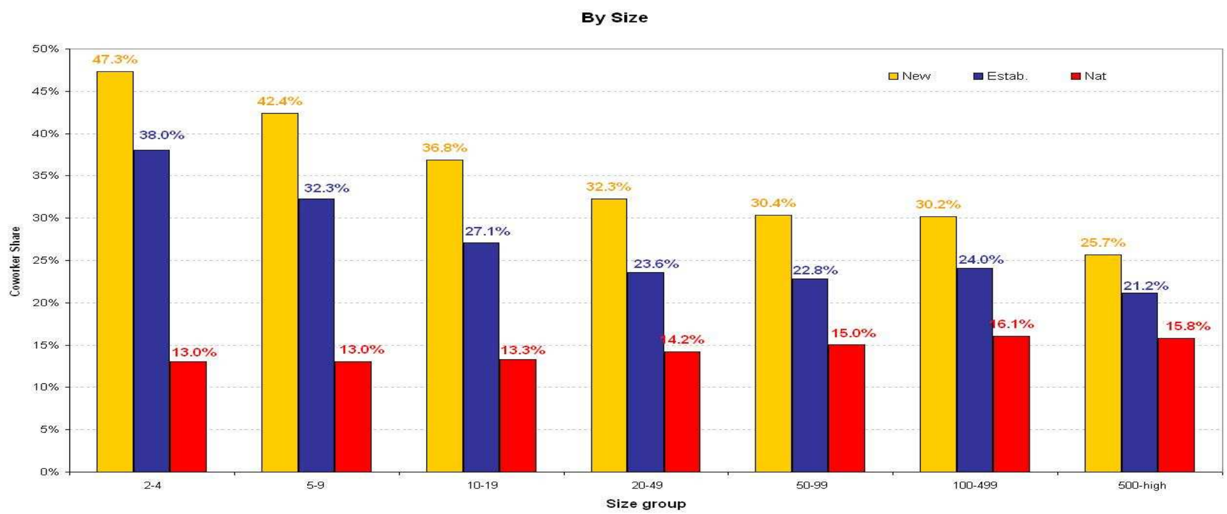


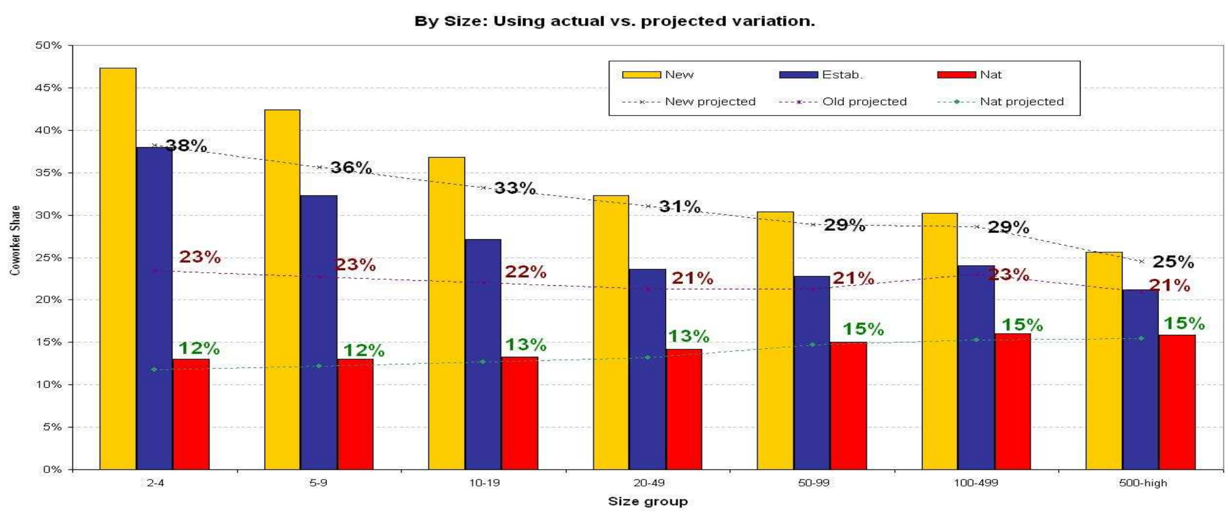
**Figure 2:** Coworker share by age of employee




Note: Size, Sector, establishment age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

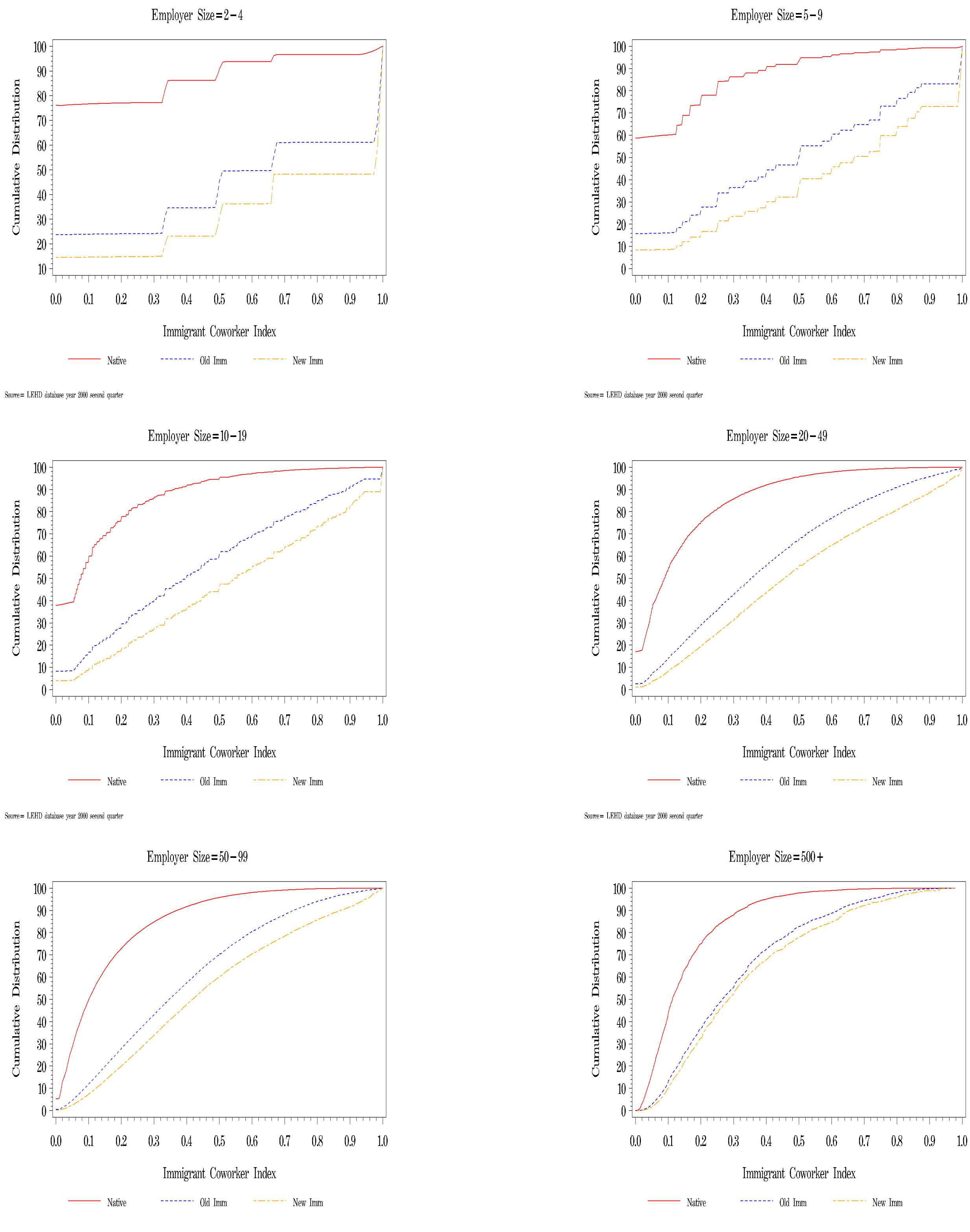
**Figure 3:** Coworker share by employer size



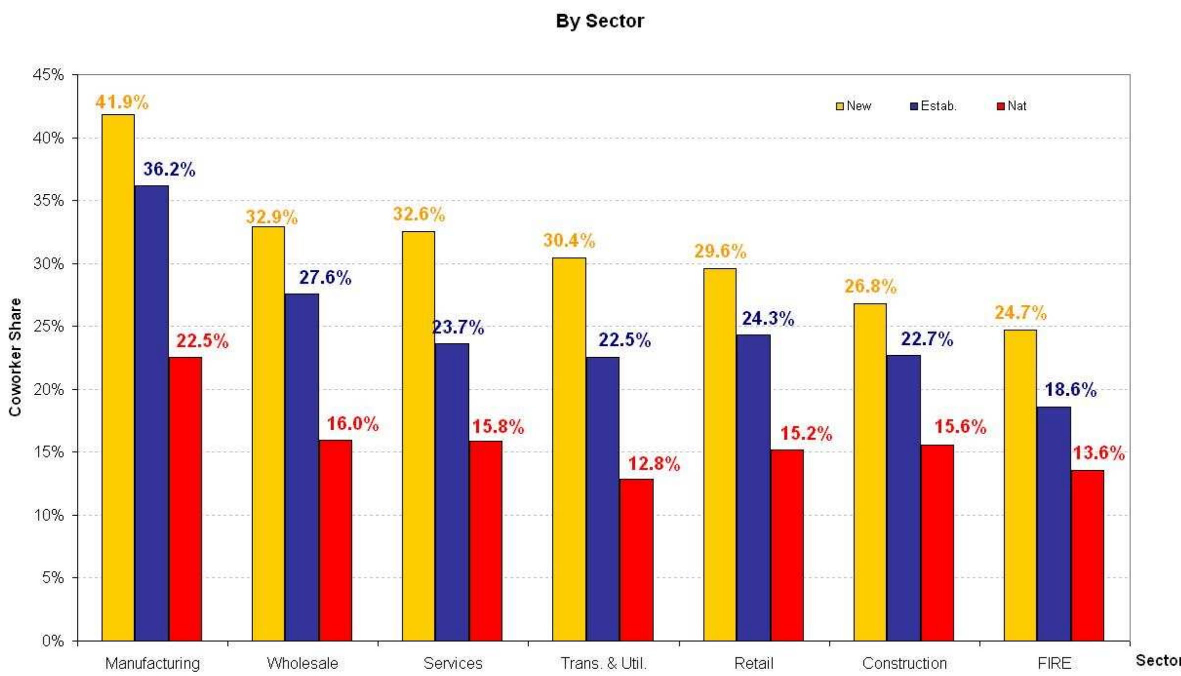

Notes: Evaluated at pooled mean for other control variables–MSA, sector, immigrant demographics, establishment age interacted with multi-unit status. Sector, individual's age, establishment age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

**Figure 4:** Cumulative Distribution of Coworker Share by Worker Type and Employer Size




Source=LEHD database. Year 2000 second quarter.

**Figure 5:** Coworker share by employer sector




Note: Size, individual's age, establishment age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.