Workplace Concentration of Immigrants*

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

Preliminary version:

October 20, 2008

Abstract

To what extent do immigrants work with native-born workers? How does this change as immigrants accumulate U.S. specific skills? Do other worker and firm characteristics affect the degree of workplace concentration? In this paper we explore these patterns in a sample of MSAs using a matched employer-employee database which extensively covers the establishments of selected states. We evaluate the workplace concentration of immigrants relative to native workers in detail by characteristics of employers and employees. A key finding is that immigrants are much more likely to have immigrant coworkers than are natives, and are particularly likely to work with others from the same country of origin. 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 immigrants (having controlled for recent arrival). 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 explore the extent to which this latter finding is driven by statistical aggregation issues for small establishments - we conclude that the latter is important but cannot fully account for the inverse relationship between concentration and establishment size. We also explore the mechanisms that underline the observed patterns of concentration. We find that proxies for the role of social networks as well as the importance of language skills in the production process (broadly defined) are important correlates of immigrant concentration in the workplace.

KEYWORDS: concentration, segregation, immigrant workers, assimilation.

*Disclaimer: This work is unofficial and thus has not undergone the review accorded to official Census Bureau publications. All results have been reviewed to ensure that no confidential information is disclosed. The views expressed in the paper are those of the authors and not necessarily those of the U.S. Census Bureau or the U.S. Department of the Treasury.

†U.S. Department of the Treasury; 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 primary objective in this paper is descriptive. We have unique matched employer-employee data that permit quantifying the extent of and covariates of the workplace concentration of immigrants. Our first main finding is that immigrants are much more likely to have immigrant coworkers than are natives, and are particularly likely to work with their compatriots. This 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 immigrants 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 finding that the allocation of immigrants across workplaces is far from random raises the question: what drives this workplace concentration? 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. While it is beyond the scope of this paper to fully investigation such factors, towards the end of the paper we present some exploratory analysis of whether such factors could help account for immigrant concentration. We find strong evidence that immigrants with primarily immigrant coworkers are likely to have coworkers who live in the same residential tract. While more than one underlying mechanism could generate this finding, it clearly points to some connection between workplace concentration and residential concentration. We also find evidence that immigrant concentration is lower in industries that employ immigrants with strong English language skills. This latter pattern holds even after controlling for the average education of workers in the industry.

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 exploratory 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 extends the analysis in terms of the basic 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 return 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 Corcran, 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 the importance of differences in English language skills. Immigrant assimilation into the U.S. labor market is viewed as occurring through 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 (see 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 (see Abowd et. al 2005). While more

controversial, there is some evidence that firm-level technological adoption also affects workers wages (see 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 Nicola (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 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

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

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 et. al. (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 can not 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 other studies followed that 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, Wilson and Portes (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, Wilson and Portes 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, 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 as a measure of 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), but do not directly condition on coworker characteristics. Our measure of concentration is the mean difference between immigrants and natives with the same characteristics.

We take a different approach in part because the worker characteristics that we have that vary within employer (age, gender) do not differ dramatically between immigrants and natives, and they also turn out not to have a strong correlation with immigrant concentration. Controlling for a worker's own characteristics should remove the effects of age and gender from the measured difference in coworker mean, and the estimated coefficients allow us to examine the characteristics of immigrants and natives who work in heavily immigrant workplaces.

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, for example, including controls for MSAs.

3.2 Data

We use the 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 some states only have data for a subset of those years. The workers' earnings records have also been matched to characteristics of their employer gathered in quarterly administrative 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, preliminary results based on a sample of immigrants for whom both LEHD and

Decennial 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 analysis, 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 millions of immigrants. Second, we can observe the firms in which workers are employed, 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 in future work: for example, panel data on both employers and employees that would allow us to track 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 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 primarily on the second quarter of 2000, but we take advantage of LEHD data for the 1995-2000 period because we use it 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 (listed in Table 1). 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 2 summarizes the across-MSA variation in immigrant shares for our sample of MSAs. In the average MSA in our dataset, 18.9% of workers are immigrants. In what follows, we are interested in deviations in workplace shares from the overall-average. 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.

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 (RI, defined as arriving in the last 5 years), or a more established immigrant (EI, 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.

[2]See tabulations from Andersson, Haltiwanger and Sanders (2007).

[3]Immigrants account for 27.45% of employment in single-employee businesses.

[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.

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. Here, the constant term (γ_N) represents the mean coworker share for the omitted category, which in our simplest specification consists simply of natives. Once we add controls, the omitted category is rarely of particular interest, so in place of the constant in our tables we report the predicted coworker share for natives based on the pooled distribution of the variables in x. Similarly, where we use graphs to illustrate the effects of a particular covariate, we hold other covariates constant by evaluating their effects at the pooled mean.

Coefficients γ_{EI} and γ_{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 by differences between natives and immigrants in their geographic distribution and in worker and job characteristics. In section 6, we define coworker shares for specific countries of origin and look at which immigrants are most likely to work together.

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:

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

Following our approach in presenting results from estimates without interactions, we present results from the specifications with interactions by evaluating differences between immigrants and natives based on the pooled distribution of the variables in x.

To ease computations with our 36 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 6) are estimates that are too small to be meaningfully different from zero anyway. [5]

Table 3 presents summary statistics for immigrant and native workers in our 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 also 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. Clearly, there is substantial concentration of immigrants.

In the rest of the table, the largest differences across the three groups appear in the age distribution: recent immigrants are substantially younger than natives while earlier immigrants are older. Combining

[5] So far, using statistical software to handle clustering does not seem feasible. However, we could put an upper bound on it by summarizing data at the establishment level and running regression weighted by employment.

the immigrant groups, 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. These gaps are likely driven by differences between immigrant and native women in rates of labor force participation rather than differences in the population. Comparing the distributions of employment across sectors, immigrants are more concentrated in manufacturing than are natives.[6]

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. Our restriction to employers and on-the-books employment may affect these statistics as well. Similarly, differences in distribution of employment by plant age are also small. However, immigrants are less likely than natives to work for multi-unit firms.

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 4 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 4, we report results from the base specification without any controls. This simply reproduces the differences in means one finds from the first row of 3. The subsequent rows of Table 4 show the effects of adding each of the controls one 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 significantly smaller than the overall difference. For example, if immigrants were mostly employed in a few industries, but were randomly

[6]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. See http://www.census.gov/population/cen2000/stp-159/foreignborn.pdf and http://www.census.gov/population/cen2000/stp-159/native.pdf.

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 4 shows that our measures of employer and worker characteristics account for a substantial amount of the observed concentration of immigrants in the workplace, but over half of the concentration remains unexplained. 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 19% for recent immigrants and 13% for established immigrants.

In the subsections that follow, we discuss the results of adding particular controls in Table 4 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 regression coefficients for this full model are available upon request. We do note that the R-squared for this full model with interactions is 0.501.

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 nonmetropolitan areas, only 3% of immigrants did.[7] 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 2 above.

In Table 4, in the second row we add MSA dummies to the specification in row 1, almost doubling the R-squared. Both specifications assume that the within-MSA differences between immigrants and natives are the same for each MSA, so changes in the immigrant dummy variable coefficients are readily interpreted. 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 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.

To examine which MSAs have higher workplace concentration, we add interactions between the recent and established immigrant indicators and the MSA dummies. Predictions from this model simply give us the coworker mean for each of the three groups for each of the MSAs. We plot the relationship between the actual coworker share for the immigrant groups and for natives (on the Y axis) and the overall share of immigrants in the MSA. The top panel of Figure 2 shows the patterns for recent immigrants, and the

[7] Source www.census.gov/prod/2002pubs/p23-206.pdf.

lower panel, for established immigrants. It is immediately clear for both groups that there is non-random allocation of immigrants within an MSA—that is, within an MSA, immigrants are more likely to work with immigrants than one would expect based on the share of immigrants in the population.

The solid lines plot out results from a second stage regression of the actual shares on the MSA mean share. The actual shares lie very close to the regression line for natives and established immigrants, with somewhat more variation around the regression line for recent immigrants. The vertical gap between the two lines represents our measure of immigrant concentration, which clearly rises with the overall immigrant share in an MSA. The bottom panel of Figure 2 gives the same plot for established immigrants. These figures illustrate that immigrant concentration rises with the overall immigrant share in an MSA, and that it is consistently higher for recent immigrants than for more established immigrants.

4.2 Worker Characteristics

We have limited data on the characteristics of workers—basically age and gender, in addition to knowing where a worker was born. As the third and fourth rows of Table 4 illustrate, adding these characteristics 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. The age results appear in Figure 3.

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 controls for age, gender, MSA, sector, employer age, multi-unit status (and its interaction with employer age), and employer size. The estimates for the figures are constructed using the pooled mean values of all controls except for those used in defining the categories for the bars. So for Figure 3 the pooled mean values for gender, MSA, sector, employer age, multi-unit status, and employer size 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 difference 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).

For natives, coworker share does not vary much by age. Among immigrants, coworker share rises modestly with age. 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. Still, older workers in these groups have higher concentration.

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.[8]

[8]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

4.3.1 Employer size

We measure employer size as the number of employees at the establishment. For our analysis, we are excluding establishments with only one worker since the coworker index is by construction not defined for a worker who has no coworkers. We classify employer size into the size bins as depicted in Figure 4. Adjusting for immigrant/native differences in employer size has virtually no effect on the difference in average coworker share (see Table 4). We find no effect because the distribution of employment across employer size classes is very similar for immigrants and natives; the immigrant share varies little across these classes. Because of this, as long as we constrain the size effects to be the same for immigrants and natives, the size coefficients are all quite small.

Despite this similarity in distributions, when we allow the effects to differ we find very large size effects. That is, while the share of immigrants is relatively constant across size classes, the concentration of immigrants at the workplace level falls substantially with size. This pattern is illustrated in Figure 4. 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, 45 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 31 percent. It is striking that these large effects hold even after controlling for many other factors including sector. We also find that the difference in immigrant shares between recent and more established immigrants also falls with size.

More detailed evidence on the effects of size are presented in Figure 5, which 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 there are two mechanisms driving 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 under random allocation. 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 8 shows that this is true for a particular statistical model, but we think this point holds more generally.

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 et al (2006).

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 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.[9] The lower panel of Figure 4 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 under-predict 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. As predicted by the statistical model in Appendix 8, 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 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 5 shows that except for the smallest workplaces (where concentration is by construction going to be high at the individual establishment level) we find that not much mass is concentrated at completely segregated workplaces. This suggests that even small businesses can 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, the number of open vacancies is likely to much smaller for a given small business (even if the rate of vacancies is as high or higher than larger 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 employers that face similar constraints in choosing the skill mix of their workforce. Significant variation in immigrant concentration by industry

[9] 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.

would be consistent with technological differences playing an important role in determining how employers combine employment of natives and immigrants.

Controlling for industry sector of employment accounts for a very modest amount of the observed immigrant concentration. 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 4). Whether we control for employer size or not has little effect on this conclusion.

Allowing immigrant concentration to differ across sectors (as illustrated in Figure 6) gives 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. The other sectors also show substantial levels of immigrant concentration, with even the least concentrated sector (finance, insurance, and real estate) having a 10 percentage point higher coworker share for established 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. For example, wholesale and retail trade rank second and third in concentration of established immigrants but third and fifth among recent immigrants.

While the findings in Figure 6 are suggestive that technological differences, broadly defined, are important for concentration, the sectoral detail in Figure 6 is quite broad. In the next section, we exploit detailed sectoral variation along with information about the characteristics of workers across industries in order to explore the factors that underlie the observed variation.

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 above in section 2, there are number of potential channels for immigrant concentration to matter for the type of technology (broadly defined), organizational structure and recruiting methods of a business. In this section, we present results of some exploratory analysis that look more directly at possible channels. In particular, we explore the role of network effects, language skills and human capital.

Data limitations imply that we explore these channels in a sequential manner. First, we explore the role of social networks at the person level by extending the regression specifications used in section 4 to include proxies for the extent of the use of social networks by businesses. Second, even after extending our analysis in this direction, we find that there is substantial residual variation in immigrant concentration across industries. Motivated by this finding, we use industry-based measures of education and language skills to explore to what extent human capital measures can explain the industry variation.

To explore social networks, we construct two measures from our data base using information on worker tract of employment and tract of residence. As one proxy for social networks, we calculate for

each worker the fraction of their coworkers that reside in the same tract that they live in.[10] This proxy, which in what follows we call a network effect, may reflect many factors. For example, as discussed in Section 2, one possibility is that referrals by current employees are an important recruitment source, and that many referrals 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 effect will, in principle, capture such effects but should more generally be viewed as capturing the extent to which residential location and employment location are correlated.

A related but distinct relationship between workplace and residence is that some businesses primarily hire workers who live near where they work. To distinguish this effect from the network measure, we construct a second variable for each employer which is the fraction of employees who live in the tract in which the business is located. In what follows, we refer to this control variable as the employee proximity index.

Summary statistics for the network index are presented in Table 5. Not surprisingly the mean of the network index is small: for the average worker the fraction of coworkers who live in the same tract is about 1.7 percent. This fraction is substantially higher for small businesses, and it falls systematically with employer size. It is also instructive to observe that while the average is small there is considerable variation across workers and it is the latter variation we are exploiting in our analysis.

Results using the network index and employee proximity index are presented in Table 6, first with only industry controls, then with industry and MSA controls, and finally with all controls. We find that the network index coefficient is large and in the predicted direction for both native born and immigrant workers. For native workers, the effect is large and negative. A native worker with a high share of coworkers who live in the same tract has a lower share of coworkers who are immigrants. For immigrant workers, the effect is given by the sum of the main and interaction terms of the network variable. In contrast to the finding for natives, the effect for immigrants is large and positive. An immigrant worker with a high share of coworkers who live in the same area has a higher share of coworkers who are immigrants. This pattern is consistent with the social network story we discussed above but may reflect any factor that links workplace concentration to residential concentration. These effects are somewhat larger with only industry and only MSA controls but remain large even with all controls.

We also find that the employee proximity variable is not important in accounting for variation in immigrant concentration in the workplace, though this may be because we chose a poor proxy for the importance of proximity. Since we have controlled for this effect, our findings on the network index are not being driven by employers who hire lots of workers that live in the same tract as their business. [11]

The network effects we find may be related to and/or interacting with the size effects discussed in section 4. To explore this, we add a three-way interaction effect between the network index, size and immigrant status to the model with all controls. The estimated interaction of size and network effects are presented in Figure 7. Figure 7 shows the finding we already emphasized—namely that holding the

[10] Census Tracts are small geographical areas with a population between 1,500 and 8,000 individuals. They 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 geographical reach of a social network, i.e. the limited distance between residents of a census tract, both in terms of geography and socio-economic factors, suggests that the within-area likelihood of interactions between members is high relative to the between areas.

[11] In future drafts, we plan on exploring additional controls to disentangle the effect we find here. In particular, we plan on constructing a measure for each worker of workers at other firms in the same workplace tract who also live in the same tract as the worker (as in Hellerstein, Neumark, and McInerney (2008a). This is a much richer measure of the propensity for workplace and residence to be connected that will control for factors such as commuting patterns and even the extent to which the connection between workplace and residence might reflect the workplace and residence location by skill. While we examine skill effects in the industry analysis below, it would be useful to control for skill effects in this micro analysis. Along these lines, we also plan on including as further controls where the worker is in the distribution of earnings as a proxy for skill.

network index constant, the immigrant concentration difference between immigrants and natives tends to fall with employer size. This pattern holds for most of the range of the network index but we also observe an interesting interaction between the network index and size. Within each of the size classes, immigrant concentration rises with the network index, but the slope of this relationship is greater for large businesses. At first glance we were puzzled by this finding since our prior was that the network effect should be even greater for smaller businesses. However, upon further reflection, it may reflect location and workplace organization decisions of larger businesses that target a specific immigrant community in terms of location for its workforce. Consider, for example, a textile plant that has chosen its location and hiring strategy to take advantage of the rapid growth of immigrants in the local area.

Having found evidence in support of network effects playing a substantial role, we now turn to exploring the role of language skills and human capital. For this analysis, we exploit the residual between industry variation in immigrant concentration after controlling for all other factors including the network effects we have just discussed. Specifically, using the specification underlying Table 6, we extract the industry effects and use these as dependent variables in this industry-level analysis. The industry effects represent the extent of immigrant concentration in the industry after accounting for all other controls. We construct measures of immigrant language skills and worker human capital by industry using the 2000 census Integrated Public Use Microdata Series (IPUMS) files and a concordance between Standard Industrial Classiciation (SIC) codes and the industry definitions on IPUMS. This concordance determines the level of industry detail, which falls between the 2 and 3 digit level (about 185 industries).

We measure English language skills among immigrants working in an industry using the fraction of *immigrant* employees who report either not speaking English, or speaking English not well. We also measure industry education levels using the fraction of workers by industry in the following education categories: high school drop-out, high school graduate (with no further education), some college attendance without receiving a bachelor's degree, and college graduates. We include only the first three measures in our regressions, so college graduates are the omitted group.

Results of regression of industry level regressions are presented in Table 7. We find evidence that concentration is greater in industries with predominantly low skilled workers. Looking at our two measures separately, immigrant concentration is higher in industries that employ more high school dropouts. Similarly, immigrant concentration is higher in industries in which immigrant employees generally have poor English language skills. When we put both measures in the equation, the language measure remains significant, while the education measures do not. These two measures of skill are highly positively correlated across industries: the fraction of employees that did not graduate from high school and the fraction of immigrant employees with poor English language skills have a correlation coefficient of 0.87. So the industries in which immigrant employees have poor English language skills are also those in which employees typically have low levels of education.

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 ask how likely it is for an immigrant from say, Mexico, to have co-workers who are Mexicans. We recognize that examining such patterns could potentially shed further light on the relative merit of various language- and cultural-based explanations for immigrant workplace concentration. It is beyond the scope of this paper to conduct a thorough analysis, but to provide perspective on the importance of this variation, we present some simple descriptive results on concentration by country of origin in this section.

We rank countries of origin by their share of employment in our sample, and present results for the top 9 countries.[12] Immigrants from the remaining countries are grouped into an ''other" category. For parsimony, we do not distinguish between recent and more established immigrants here.

Figure 8 illustrates the differences across countries of origin in coworker immigrant shares. The first bar for each country gives the raw coworker share, with the two colors giving the split between coworkers from that country of origin and immigrant coworkers from other countries. Note that the longer the bar, the less likely a group is to work with natives. So for Chinese immigrants on average more than 20% of their coworkers are other immigrants from China, while almost another 30% are immigrants from other countries. The second bar gives the same coworker shares after controlling for MSA, and the third bar presents those shares with our broadest set of controls (MSA; employer size, sector, age, and multi-unit status; worker age and gender). Including MSA dummies has only minor effects on own-country coworker shares for most country-of-origin groups. The only exception is for Cubans, who are more geographically concentrated than other immigrant groups, as illustrated in Figure 9. Controlling for MSA more substantially reduces the other-country immigration shares for most groups.

7 Concluding remarks

Using matched employer-employee data that extensively 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.

Our results here are primarily descriptive, but we also begin to examine possible underlying causes of this concentration. Based on findings in the existing literature and our own preliminary results, we see three promising avenues for exploration: the role of social networks in matching workers to employers, the importance of sharing a common language with coworkers and customers, and variation in the skill needs of employers. Our preliminary results indicate that natives who live near some of the 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 are consistent with an important role for social networks, though there are other mechanisms that could lead to a correlation between residential and employment location that we have not yet investigated. We posited that social network effects should be more important in smaller firms that are less likely to have formal human resources practices, though if anything we find the opposite: our network variable has a stronger effect in larger firms. We also find that English language skills and education are correlated with industry-level variation in concentration, though our education measures are not significant in industry-level regressions once the language measure is included.

We plan further work on the underlying causes of immigrant concentration, including developing better controls for competing explanations for ties between residential and workplace location (such as transportation infrastructure). We also plan to investigate other measures of worker skill, including measures based on wage distributions.

[12] Our list of top 9 immigrant worker source countries in 2000 includes 8 of the top 9 for the U.S. population as a whole. Our sample includes Japan, while the complete U.S. list instead includes the Dominican Republic as the 9th largest source country. The difference is likely driven by the set of states we have rather than differences in composition between the overall population and employees.

8 References

- Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky. 2005. ''*The Relation among Human Capital, Productivity and Market Value: Building Up from Micro Evidence.*" in Measuring Capital in the New Economy, in Measuring Capital in the New Economy, C. Corrado, J. Haltiwanger, and D. Sichel (eds.), (Chicago: University of Chicago Press for the NBER, 2005), Chapter 5, pp. 153-198.

- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2005. ''*The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators.*" Longitudinal Employer-Household Dynamics (LEHD) Technical paper TP-2006-01.

- Andersson, Fredrik, John Haltiwanger, and Seth Sanders. 2007. "The Role of Firms in Immigrant Assimilation and Labor Market Adjustment" Grant Application to the National Institute of Child Health and Human Development.

- Aslund, Olof and Oskar N. Skans. 2005a. ''*Measuring conditional segregation: methods and empirical examples.*" Institute for Labor Market Policy Evaluation (IFAU) working paper.

- Aslund, Olof and Oskar N. Skans. 2005b. ''*Will I see you at work? Ethnic workplace segregation in Sweden 1985-2002.*" Institute for Labor Market Policy Evaluation (IFAU) working paper.

- Averitt, Robert T. 1968. *The Dual Economy: The Dynamics of American Industrial Structure.* New York: Norton.

- Bayer, Patrick, Ross S. and Topa G. (2008)''*Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes,*" Journal of Political Economy, forthcoming.

- Becker Gary S. 1957. *The Economics of Discrimination.* Chicago, University of Chicago Press.

- Borjas, George J. 1982. ''*The Earnings of Male Hispanic Immigrants in the United States,*" Industrial and Labor Relations Review, Vol. 35, No. 3, pp. 343-353.

- Borjas, George J. 1985. ''*Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants,*" Journal of Labor Economics, Vol. 3, No. 4 (October), pp. 463-489.

- Butcher, Kristin F., John Dinardo, 2002. ''*The Immigrant and Native-Born Wage Distributions: Evidence from United States Censuses,*" Industrial and Labor Relations Review, Vol. 56, No. 1 (October), pp. 97-121.

- Cabrales, Antonio , Calvó-Armengol, A, and Pavoni, N. 2008. ''*Social Preferences, Skill Segregation, and Wage Dynamics,*" Review of Economic Studies, Vol. 75(1) pp. 65-98.

- Chiswick Barry R. 1978. ''*The Effect of Americanization on the Earnings of Foreign-born Men,*" The Journal of Political Economy, Vol. 86, No. 5 (October), pp. 897-921.

- Chiswick Barry R. 1983. ''*An Analysis of the Earnings and Employment of Asian-American Men,*" Journal of Labor Economics, Vol. 1, No. 2 (April), pp. 197-214 .

- Corcoran, Mary, Greg J. Duncan, and Michael Ponza. 1983. ''*A Longitudinal Analysis of White Women's Wages.*" Journal of Human Resources, Vol. 18, No. 4 (Autumn), pp. 497-520.

- Dunne, T., Foster, L., Haltiwanger, J, and Troske, K.R. 2004. ''*Wage and Productivity Dispersion In US Manufacturing: The Role of Computer Investments*", Journal of Labor Economics, 22 (2), 397-429.

- Edwards, Richard C. 1972. ''*The Social Relations In the Firm and Labor Market Structure.*" Pp. 3-26 in Labor Segmentation, edited by R.C. Edwards, M. Reich and D. M Gordon. Lexington, Mass: Heath.
- Elliott, James R. 2001. ''*Referral Hiring and Ethnically Homogeneous Jobs: How Prevalent is the Connection and for Whom?.*" Social Science Research 30: 401-425.
- Galbraith, John K. 1971. *The New Industrial State.* New York: Mentor.
- Hellerstein, Judith K. and Neumark, David. 2003. ''*Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set,*" Annales d'Economie et Statistique, no. 71-72, pp. 19-78.
- Hellerstein, Judith K. and Neumark, David, and McInerney Melissa. 2008a. ''*Measuring the Importance of Labor Market Networks,*" NBER Working Paper.
- Hellerstein, Judith K. and Neumark, David, and McInerney Melissa. 2008b. ''*Spatial Mismatch or Racial Mismatch,*" Journal of Urban Economics, Vol 64, No.2, pp. 464-479.
- Hellerstein, Judith and David Neumark. 2007. ''*Workplace segregation in the United States: Race, Ethnicity, and Skill.*" Center for Economic Studies discussion paper 07-02.
- Holzer, Harry J. 1987. ''*Hiring procedures in the firm: Their economic determinants and outcomes.*" NBER working paper 2185.
- Holzer, Harry J. 1988. ''*Search method use by unemployed youth.*" Journal of Labor Economics 6(1): 1-20.
- Iceland, John. 2009. *Where we Live Now: Immigration and Race in the United States.* University of California Press.
- Kain, John. 1968. ''*Housing Segregation, Negro Employment and Metropolitan Decentralization,*" Quarterly Journal of Economics, Vol 82, No 2, May. pp. 175-97.
- Kremer, Michael and Maskin, Eric S. 1996. *''Wage Inequality and Segregation by Skill.''* NBER Working Paper No. W5718.
- Lang, Kevin. 1986. ''*A Language Theory of Discrimination.* " Quarterly Journal of Economics 101(2):363-381.
- Lengermann, Paul. 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 2002-10.
- Lengermann, Paul, Kevin McKinney, and Robert Pedace. 2004. ''*New Evidence on Immigration and Labor Market Outcomes Using Linked Longitudinal Employer-Employee Data.*" Working paper.
- Montgomery, James D. 1991. ''*Social networks and labor market outcomes: Toward an economic analysis.*" American Economic Review 81(5): 1408-1418.
- Portes, Alejandro and Wilson, Kenneth L. 1980, ''*Immigrant Enclaves: An Analysis of the labor market experiences of Cubans in Miami.*" The American Journal of Sociology, 86(2):295-319.
- Singer, Audrey. 2004. ''*The Rise of New Immigrant Gateways.*" Washington: Brookings Institution.
- Smith, James P., and Finis R. Welch. 1977. ''*Black-White Male Wage Ratios: 1960-70,*" The American Economic Review, Vol. 67, No. 3 (June), pp. 323-338.

Appendix: 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 clearly not consistent with random allocation, and that concentration is particularly high in small businesses. This raises the question of whether we should expect some 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 some segregation, then uses the model to simulate how employer size might affect differences in coworker immigrant share.

Suppose that an employer of given size s draws its workforce 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.

Assume that the outcome of each draw can be described using the binomial probability mass function:

$$b(i,s) = \binom{i}{s} p_D^i (1-p_D)^{s-i} \tag{A-1}$$

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 the composition of the workforce, and a parameter θ which 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 $d = 0$—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 A-1 illustrates the shape of $d()$ for various values of θ.

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 more likely than natives to be 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. (e.g. with $s = 2$, immigrants are always half of the workers in discarded first round draws.)

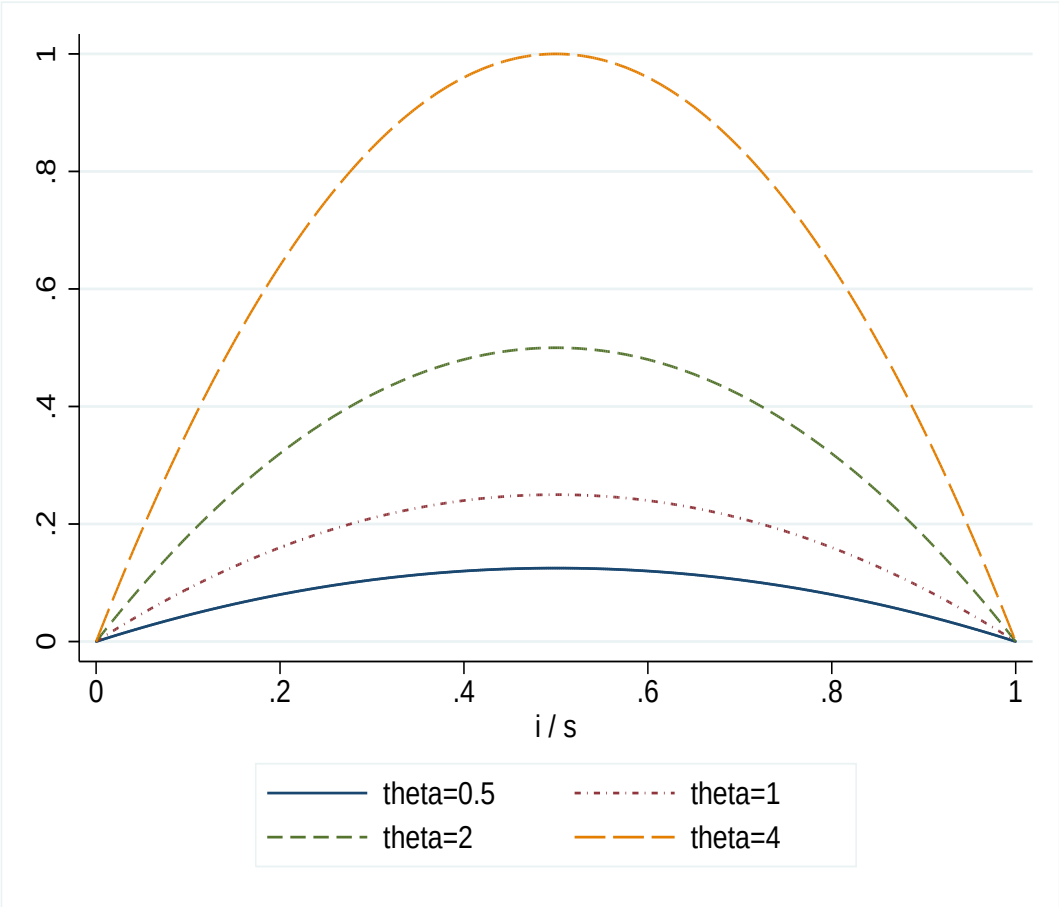
Thus the second draw is also binomial, but the immigrant share is given by:

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

and:

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

Figure A-1: **Shape of function d**



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 A-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 θ 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:

$$E(cw_I|s) = \sum_{i=0}^{s} \left(Pr(i, s|p_0, \theta) * i * \frac{i-1}{s-1} \right) \tag{A-5}$$

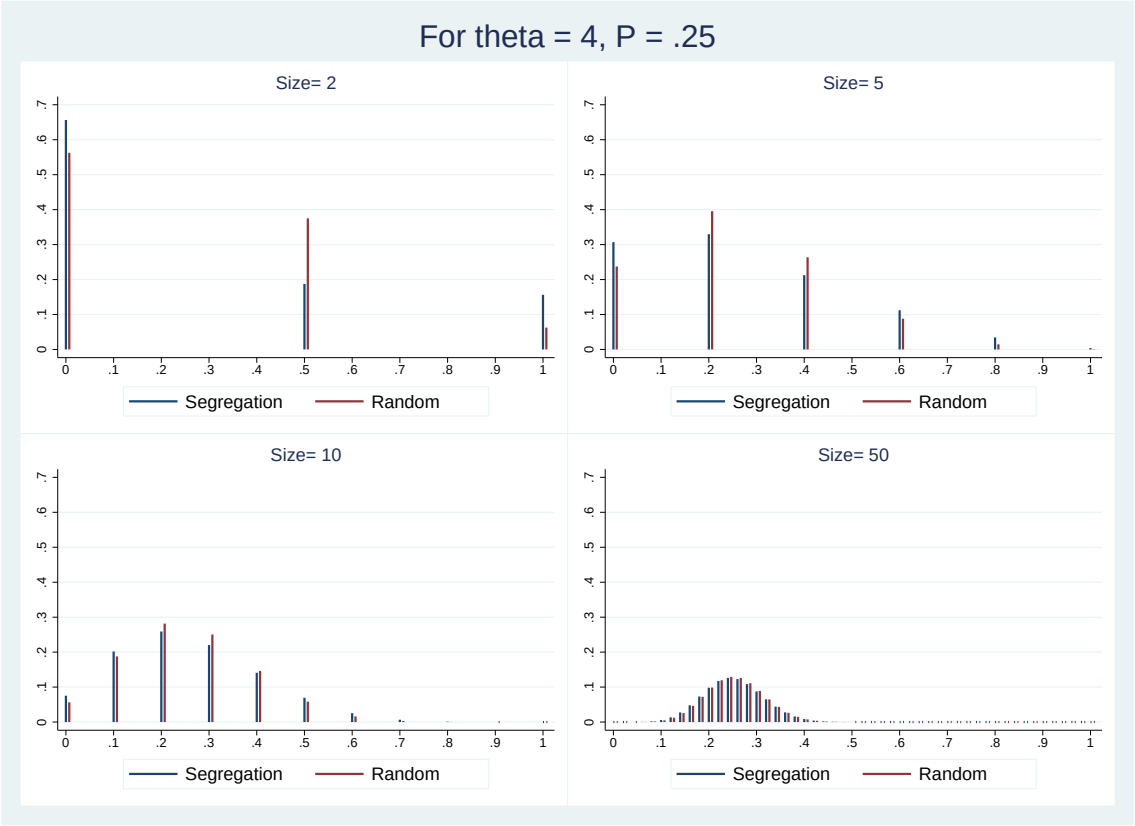
and for natives,

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

The difference is then:

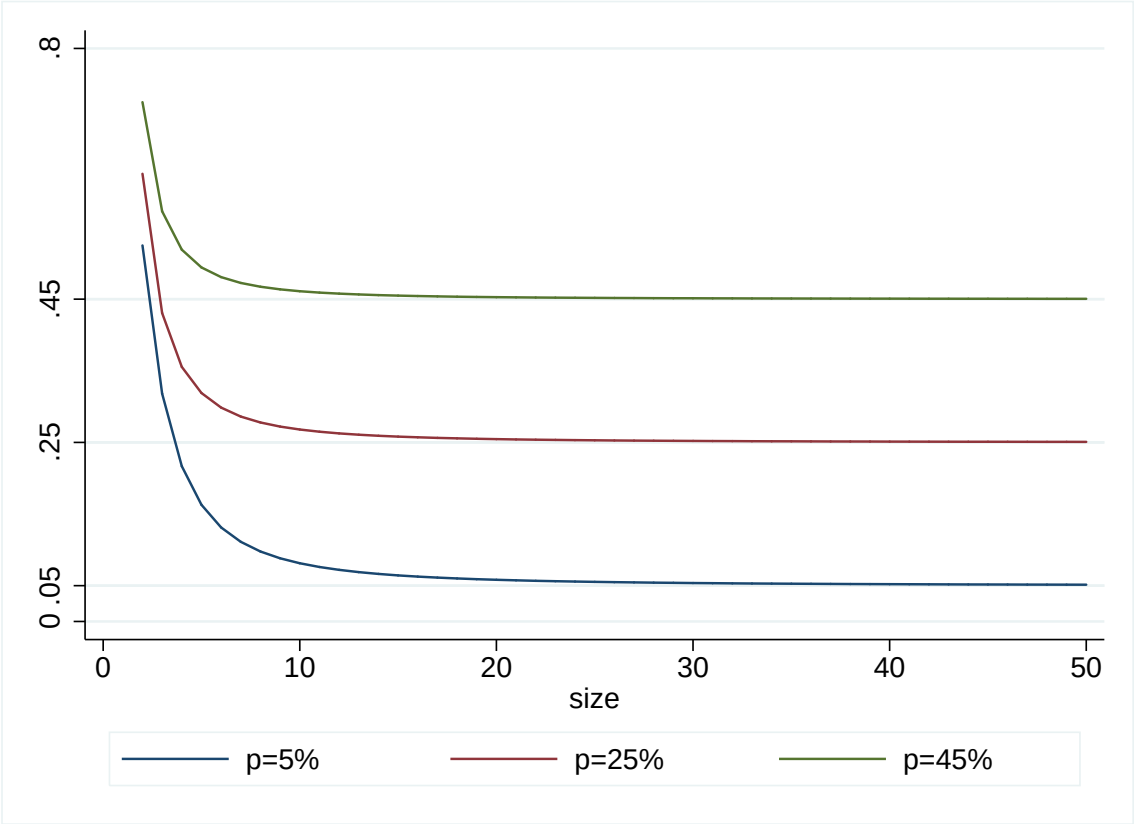
$$E(cw_N - cw_I|s) = \sum_{i=0}^{s} \left(Pr(i, s|p_0, \theta) * \frac{i(i-1) - (s-1)i}{s-1} \right) \tag{A-7}$$

Figure A-2: Immigrant share distribution with and without segregation



Figures A-3 to A-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 A-3 graph gives the mean by firm size for immigrants, figure A-4 is for natives, and figure A-5 gives the difference between them. Figure A-6 repeats figure A-5, except that it is parameterized to represent a lower level of segregation ($\theta = 1$).

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



Examination of these figures makes a couple of patterns clear:

- For very small employers (< 10 employees), the model can generate a large difference in coworker means, even with a relatively mild tendency to segregate.
- Even for large theta, this model generates essentially no segregation in large firms.

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

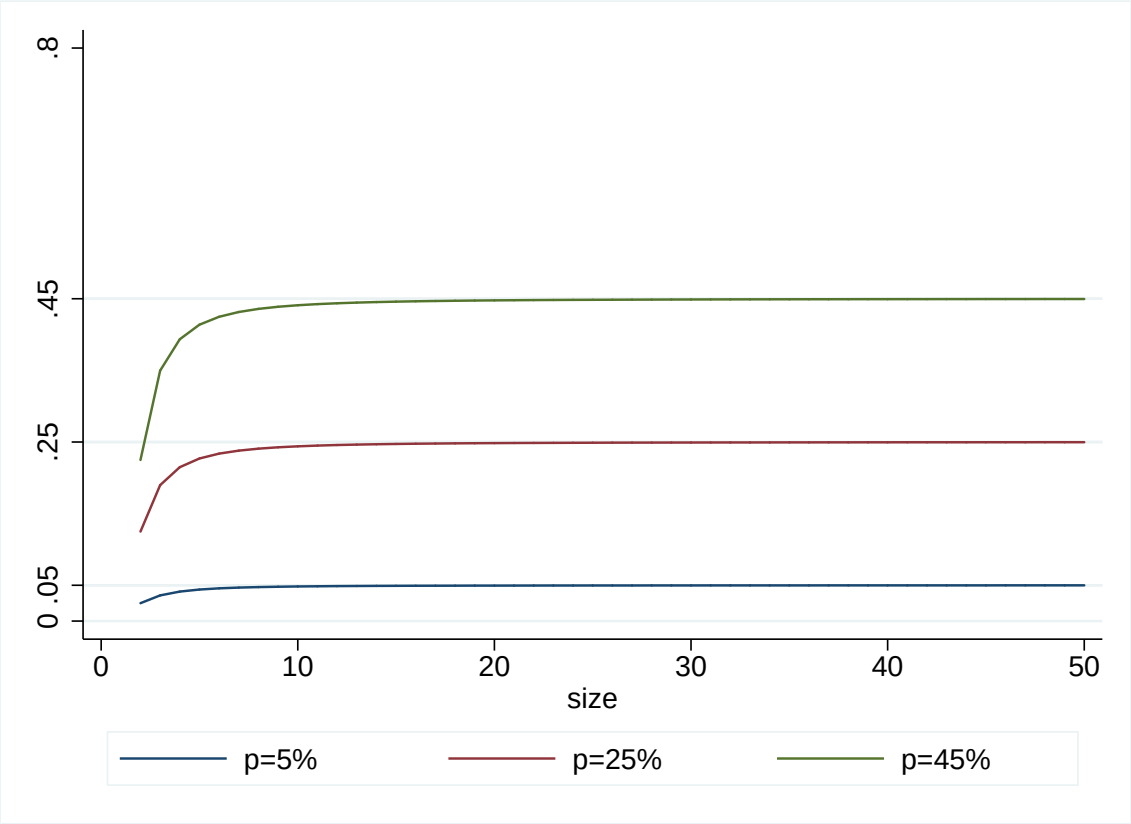


Figure A-5: Immigrant-native difference in coworker mean and employer size ($\theta = 4$)

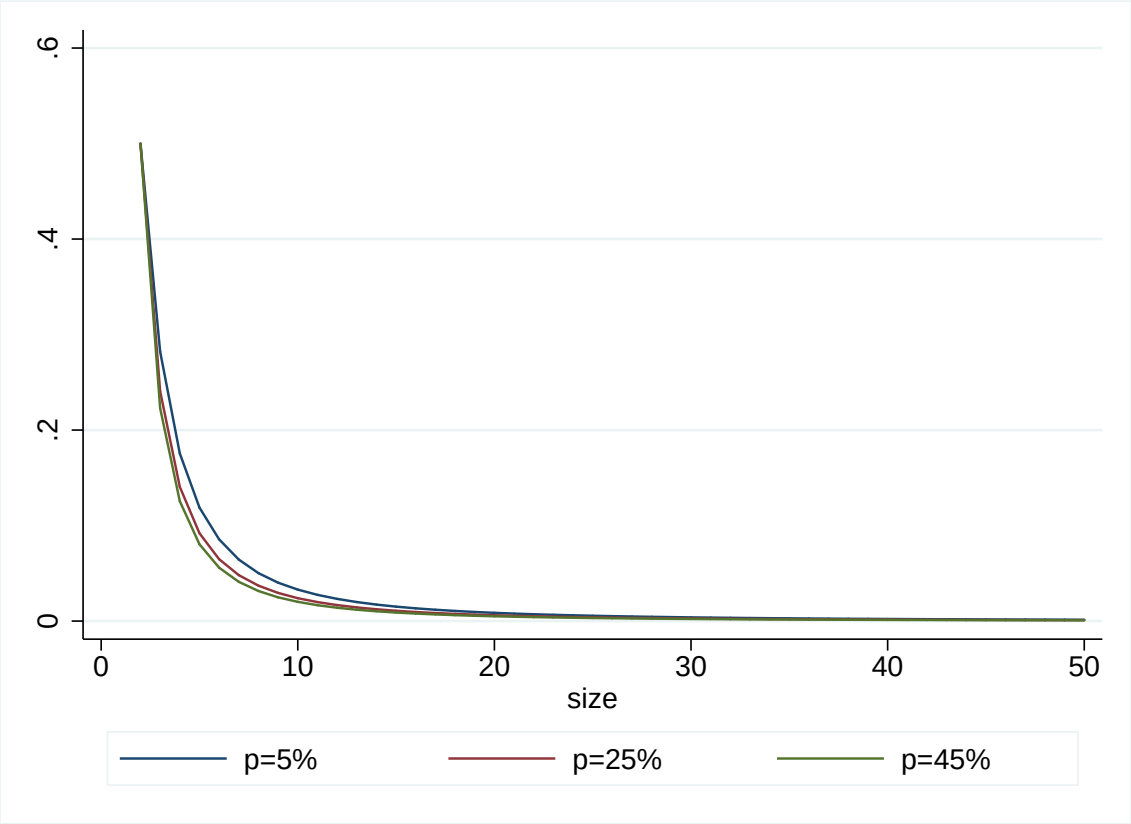
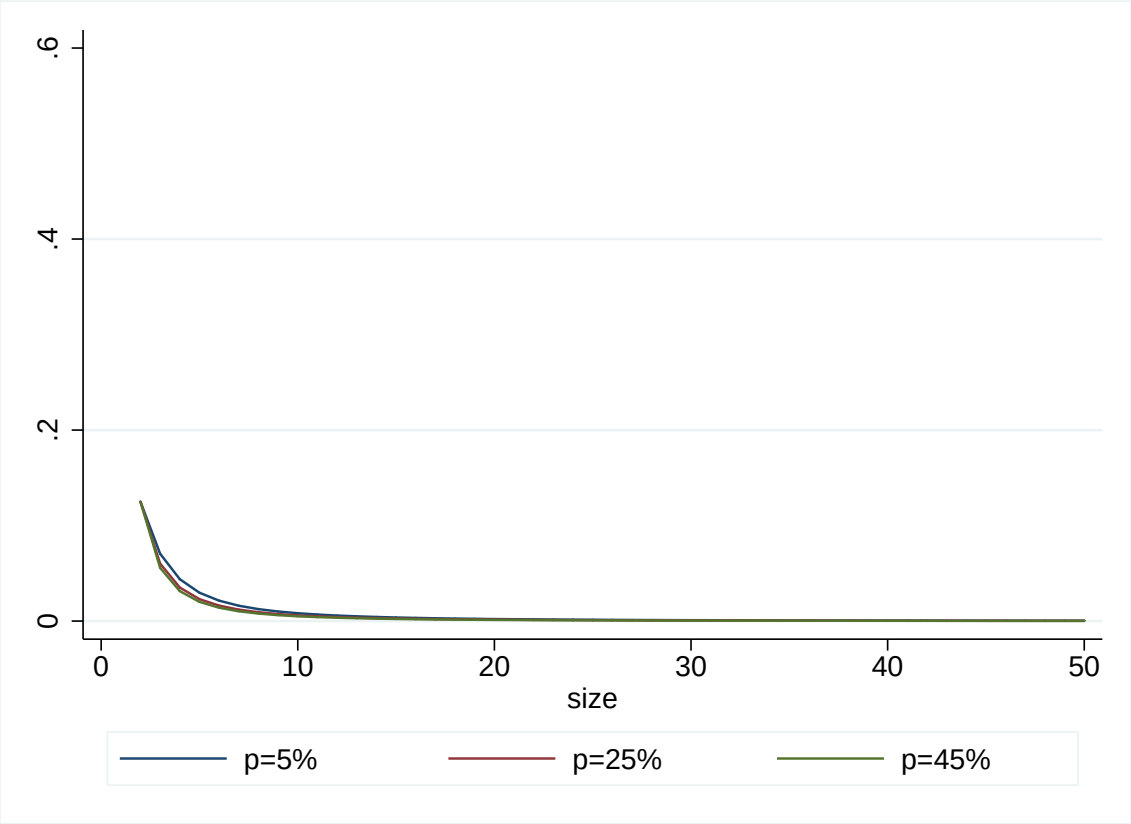


Figure A-6: Immigrant-native difference in coworker mean and employer size ($\theta = 1$)



TABLES

Table 1: **List of Sample MSAs**

	Name
1	Austin, TX
2	Baltimore, MD
3	Bergen-Passaic, NJ
4	Charlotte, NC
5	Chicago, IL
6	Dallas, TX
7	Denver, CO
8	Fort Lauderdale, FL
9	Fort Worth, TX
10	Greensboro, NC
11	Houston, TX
12	Jersey City, NJ
13	Los Angeles, CA
14	Miami, FL
15	Middlesex-Somerset, NJ
16	Minneapolis, MN
17	Newark, NJ
18	Oakland, CA
19	Orange County, CA
20	Orlando, FL
21	Philadelphia, PA
22	Pittsburgh, PA
23	Portland, OR
24	Raleigh-Durham-Chapel Hill, NC
25	Riverside-San Bernardino, CA
26	Sacramento, CA
27	San Diego, CA
28	San Francisco, CA
29	San Jose, CA
30	Tampa, FL
31	West Palm Beach, FL

Table 2: **Variation in Immigrant Shares of Workforce across Sample MSAs**

	% Immigrant	% Recent Immigrant	% Established Immigrant
Mean	18.86	3.40	15.46
Std.Dev	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 3: **Characteristics of Immigrant and Native Workers**

	Recent immigrant	Established immigrant	Native
Coworker Share	*42.1*	*36.3*	*14.9*
Age			
Age<30	*43.64*	*19.72*	*29.27*
30<Age<40	*35.61*	*33.24*	*29.98*
Age>40	*20.76*	*47.04*	*40.74*
Sex			
Female	*43.18*	*43.57*	*48.3*
Male	*56.82*	*56.43*	*51.67*
Age at arrival (*)			
0<12	*1.13*	*14.68*	.
12-25	*36.2*	*49.63*	.
26-35	*36.97*	*24.81*	.
35+	*25.70*	*10.89*	.
Sector			
Construction	*4.66*	*5.49*	*5.88*
Manufacturing	*18.92*	*21.32*	*12.84*
Transportation & Utilities	*3.72*	*5.20*	*6.50*
Wholesale	*6.80*	*7.02*	*6.46*
Retail	*22.72*	*18.29*	*21.41*
FIRE	*3.17*	*5.57*	*7.18*
Services	*40.01*	*37.11*	*39.72*
Establishment Size			
2-4	*2.64*	*2.95*	*2.34*
5-9	*5.84*	*6.06*	*5.67*
10-19	*8.33*	*8.26*	*8.58*
20-49	*15.29*	*14.37*	*14.94*
50-99	*14.37*	*13.33*	*13.60*
100-499	*31.38*	*30.86*	*29.55*
500 or more	*22.15*	*24.18*	*25.31*
Establishment age			
0-1	*12.30*	*10.72*	*10.79*
2-4	*25.55*	*22.01*	*23.55*
Age 5 or more	*62.15*	*67.27*	*65.66*
Firm Unit			
Single unit (one establishment)	*68.42*	*65.68*	*58.62*
Multi unit (Multiple establishments)	*31.58*	*34.32*	*41.38*

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 35,966,450 workers in total for our 31 MSAs.

Table 4: **Contribution of Covariates to Immigrant Concentration**

Covariates included	**Recent immigrant**	**Established immigrant**	**R-square**
None	0.272	0.214	0.198
MSA	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
All of the above	0.193	0.130	0.467

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 3). 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. The categories for worker age, employer size and employer age are the same as in Table 3. There are 185 detailed industry categories. All standard errors are less than 0.0001.

Table 5: **Distribution of Network Index by Employer Size and Worker Type**

	All workers		Immigrants		Natives	
	Mean	**Stdev**	**Mean**	**Stdev**	**Mean**	**Stdev**
Neighborhood Network Index	0.017	*0.042*	0.018	*0.048*	0.017	*0.040*
By Size categories						
2-4	0.102	*0.178*	0.131	*0.221*	0.093	*0.167*
5-9	0.054	*0.092*	0.052	*0.107*	0.052	*0.088*
10-19	0.031	*0.052*	0.031	*0.052*	0.031	*0.052*
20-49	0.019	*0.032*	0.017	*0.029*	0.019	*0.033*
50-99	0.013	*0.022*	0.011	*0.020*	0.013	*0.023*
100-499	0.009	*0.017*	0.008	*0.015*	0.009	*0.017*
500 +	0.007	*0.014*	0.006	*0.012*	0.007	*0.015*

Note:The unit of observation is a worker. N=31,572,732. These are the workers with full information on their residential location.

Table 6: **Network and Proximity Index Effects from Coworker Share Regressions**

	Industry Controls Only	**MSA/Industry Controls**	**All Controls**
Neighborhood Network Index	-0.169	-0.106	-0.095
Employee Proximity Index	-0.014	-0.007	-0.006
Network Index * Immigrant	0.680	0.654	0.450
Proximity Index * Immigrant	0.000	0.002	-0.011
R-Square	0.326	0.504	0.521

Note: All standard errors are below 0.001. The model with ''Industry Controls Only" does include industry dummies and interactions with immigrant dummy. The unit of observation is a worker. N=31,572,732. These are the workers with full information on their residential location.

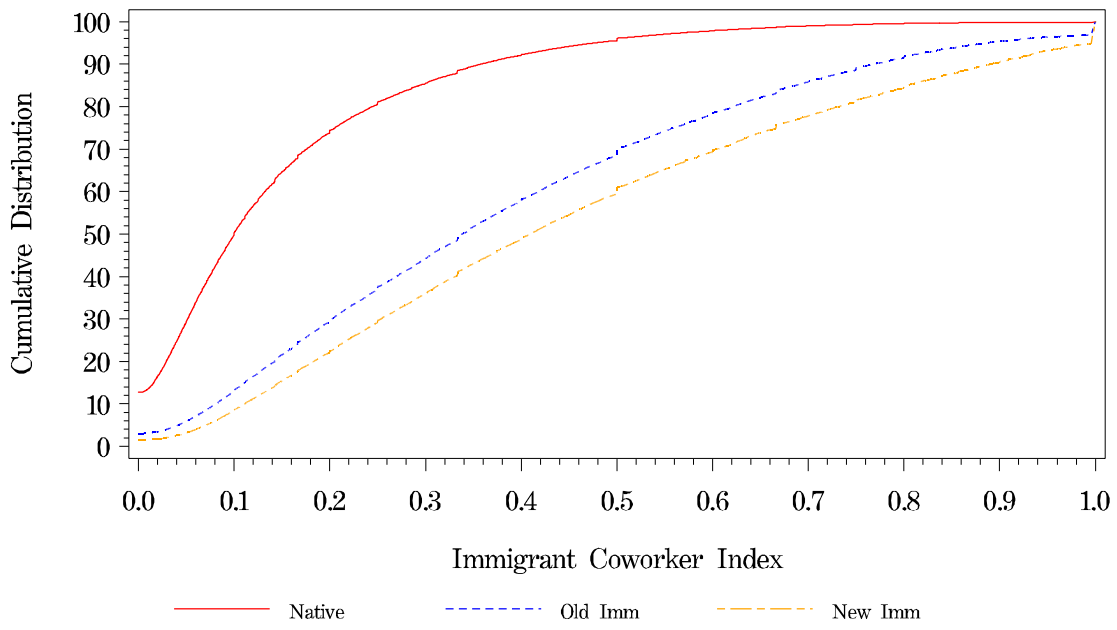
Table 7: **Immigrant Concentration and Industry Skill Measures from Industry Level Regressions**

Covariates	**(1)**		**(2)**		**(3)**	
	Coeff.	**SE**	**Coeff.**	**SE**	**Coeff.**	**SE**
Constant	0.08***	*0.03*	0.01	*0.01*	0.04	*0.03*
High School Dropout	0.17 **	*0.07*			-0.10	*0.09*
High School Graduate	0.15***	*0.05*			0.06	*0.05*
Some College	-0.24***	*0.08*			-0.12	*0.08*
Low English Ability			0.30***	*0.03*	0.29***	*0.06*
R-square	0.353		0.411		0.422	

Note: Regression at the industry level. Number of observations 185. SE denotes standard errors. Results are from 2nd-stage industry-level regressions of mean residual immigrant concentration by industry on industry measures of employee skills constructed from IPUMS 2000 census files. The residuals are from a first-stage worker-level regression of coworker share on MSA dummies; employee age and gender; employer size, age, and multi-unit status interacted with age (All Controls in Table 6). Number of workers in the first stage 31,572,732. * 10%, ** 5%, ***1% significant level.

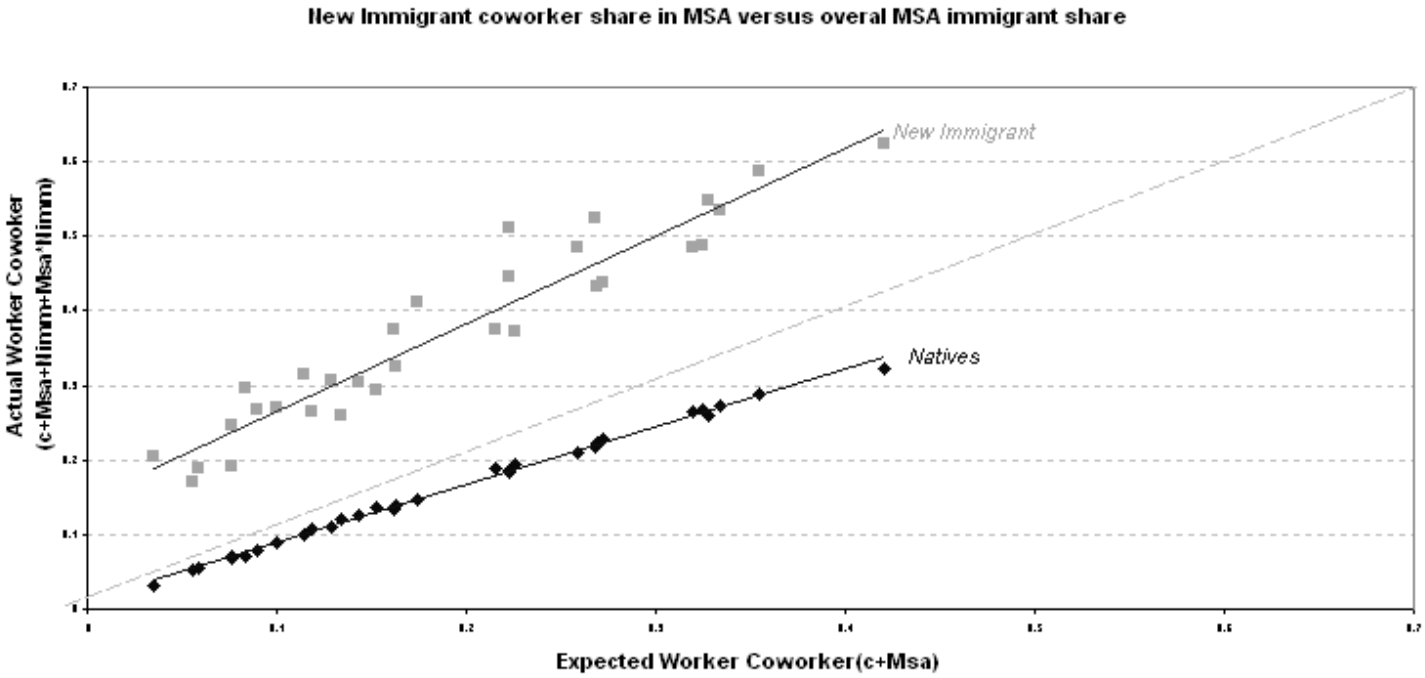
FIGURES

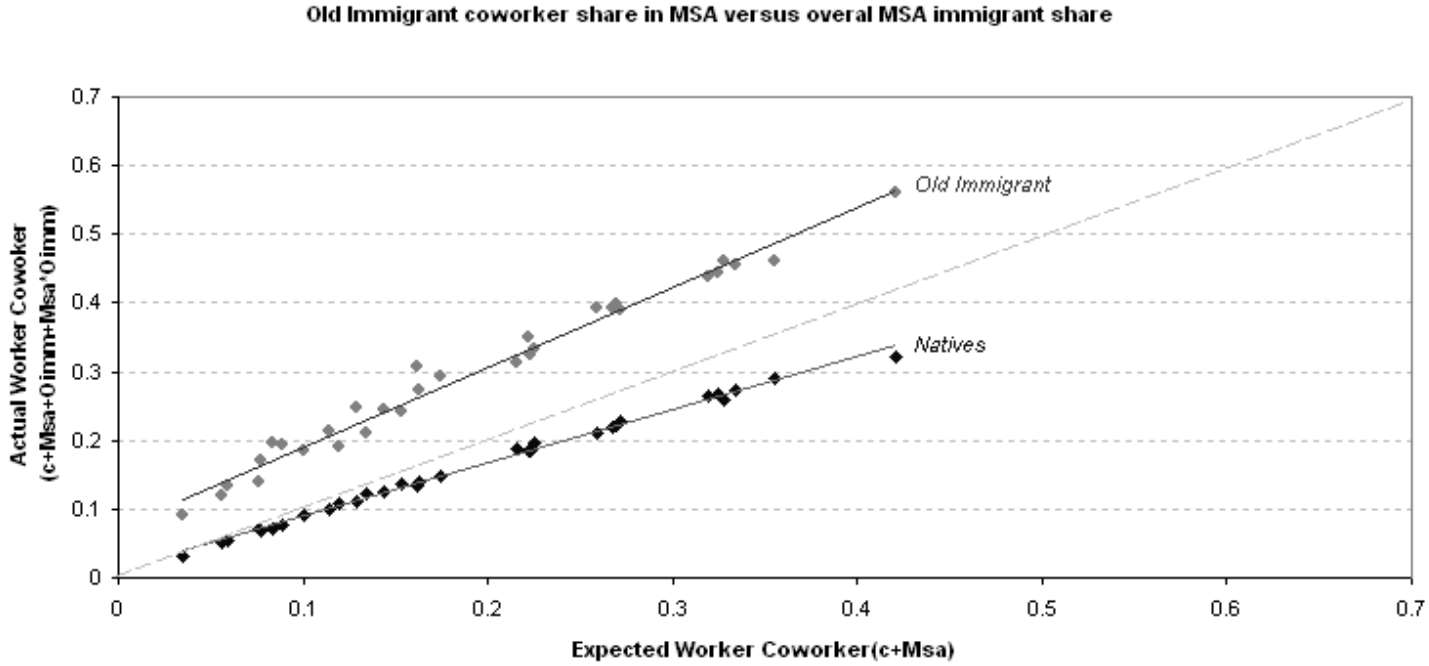
Figure 1: Cumulative Distribution of Coworker Share by Worker Type



Source= LEHD database year 2000 second quarter

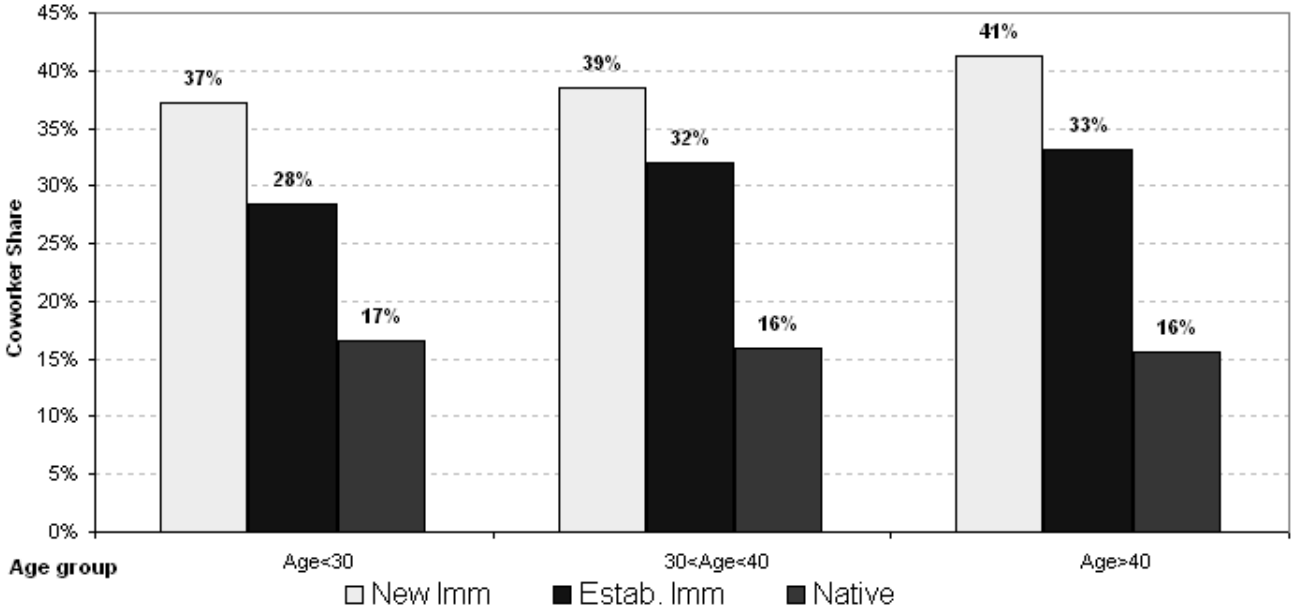
Figure 2: Relationship between Coworker Share and Immigrant Workforce Share Across MSAs





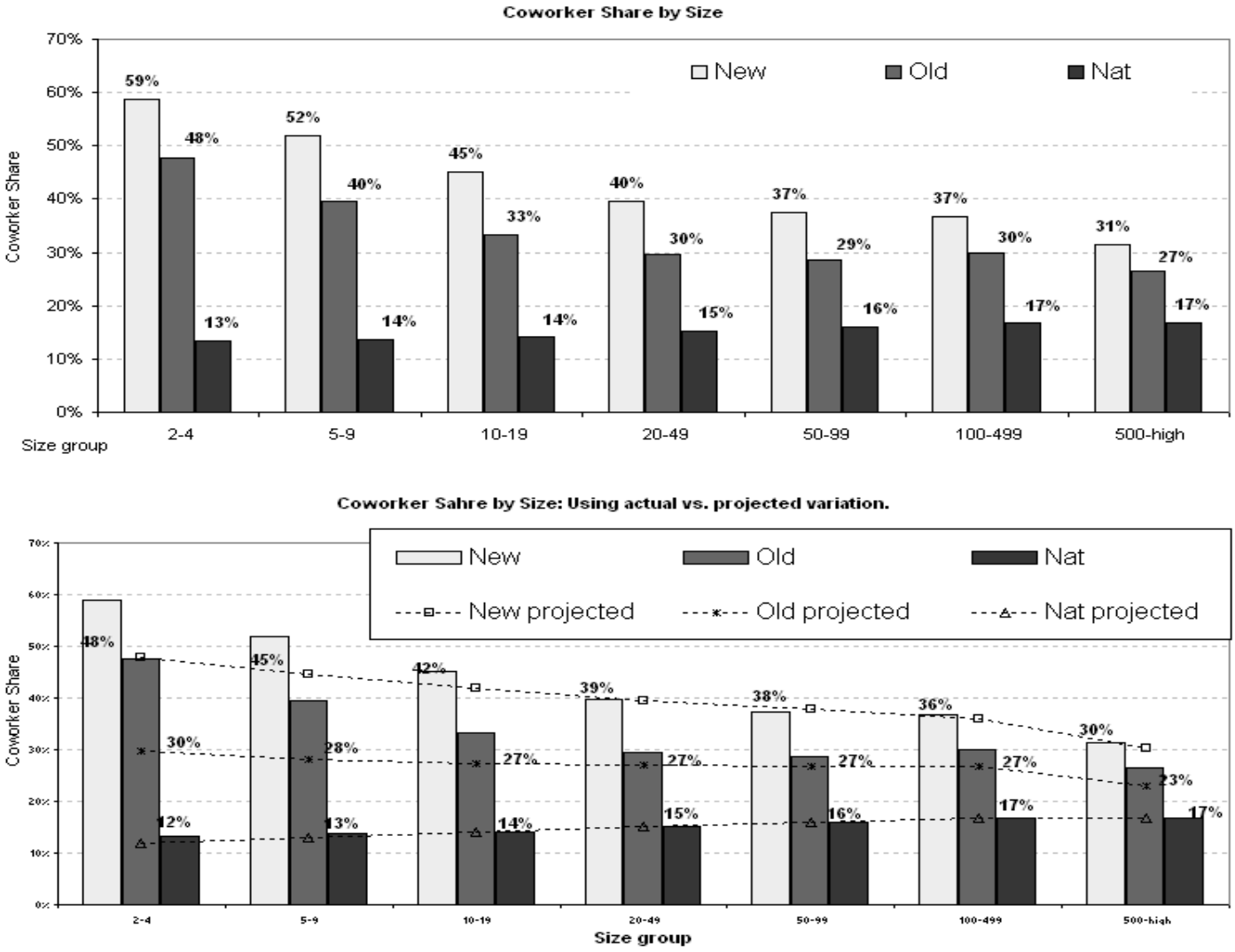
Source:Authors calculation based on LEHD database 2000 quarter 2.

Figure 3: Coworker share by age of employee



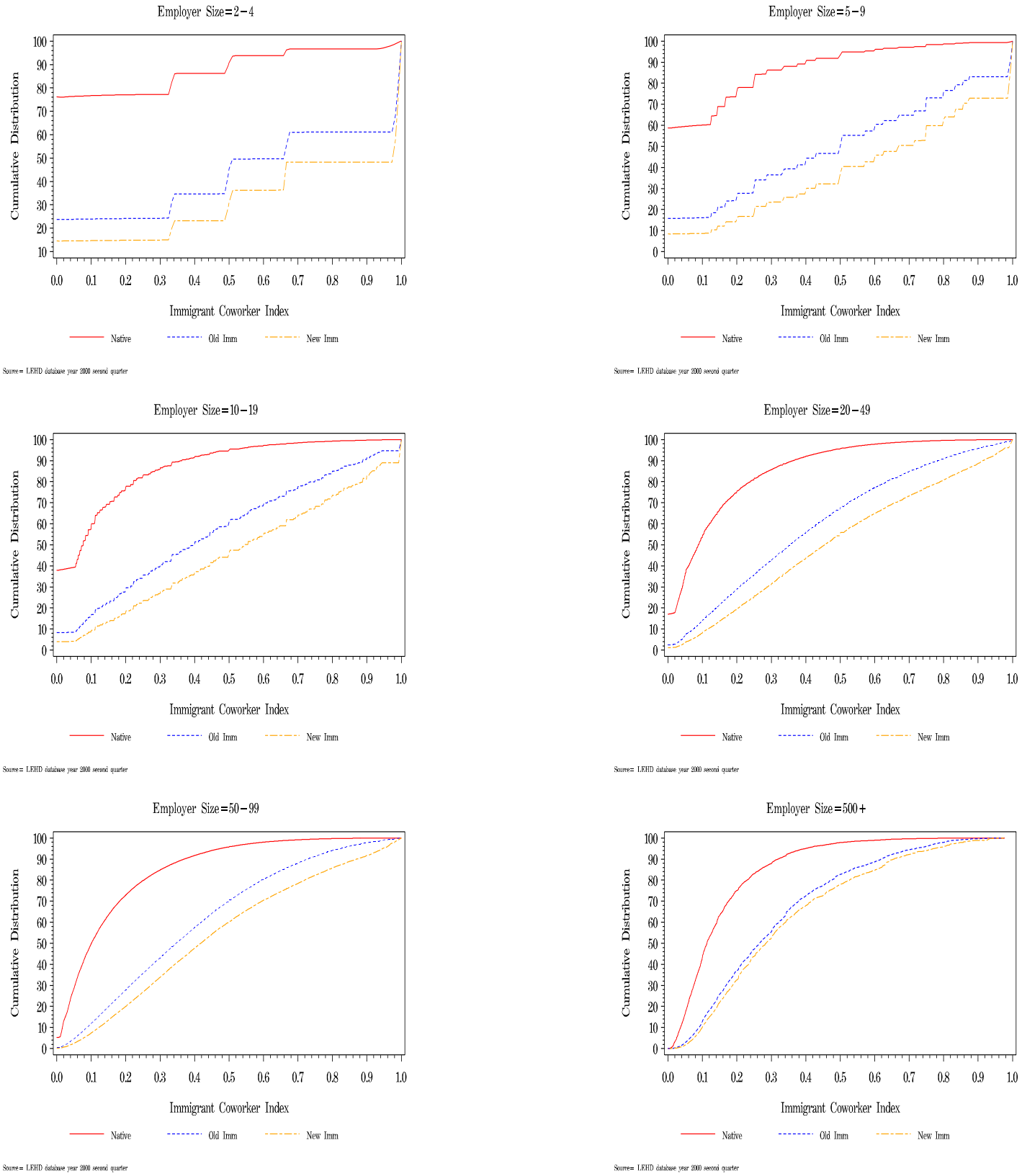
Note: Size, Sector, plant age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

Figure 4: Coworker share by employer size



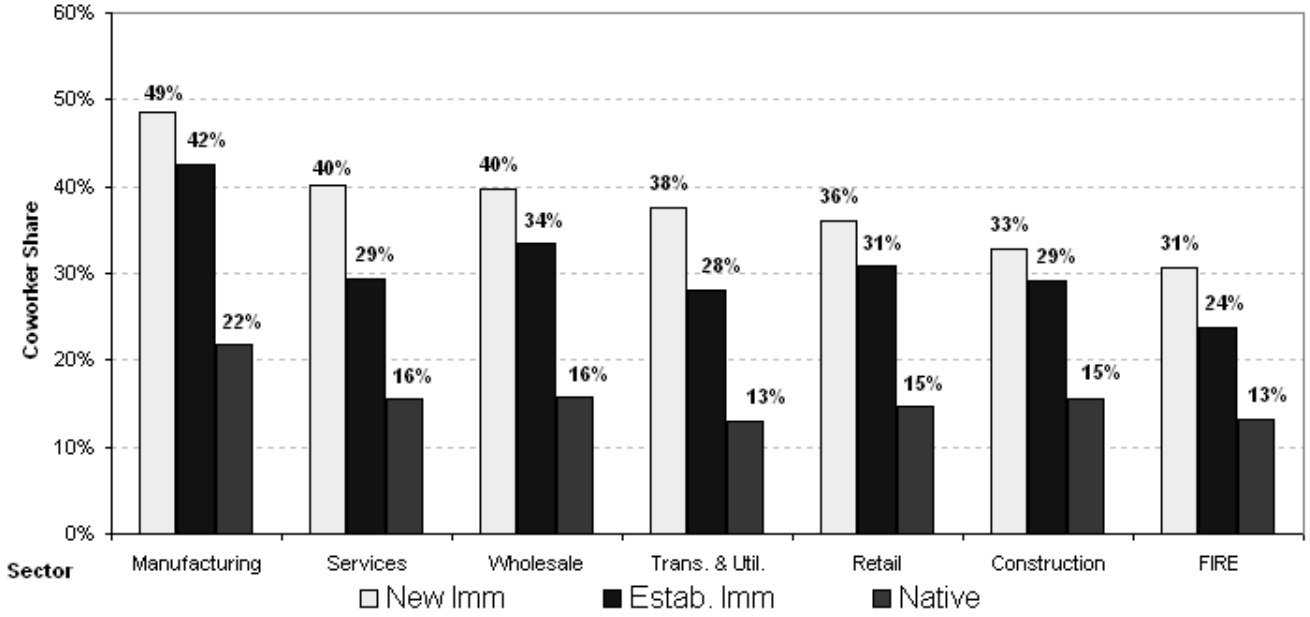
Notes: Evaluated at pooled mean for other control variables–MSA, sector, immigrant demographics, establishment age interacted with multi-unit status. Sector, individual's age, plant age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

Figure 5: Cumulative Distribution of Coworker Share by Worker Type and Employer Size



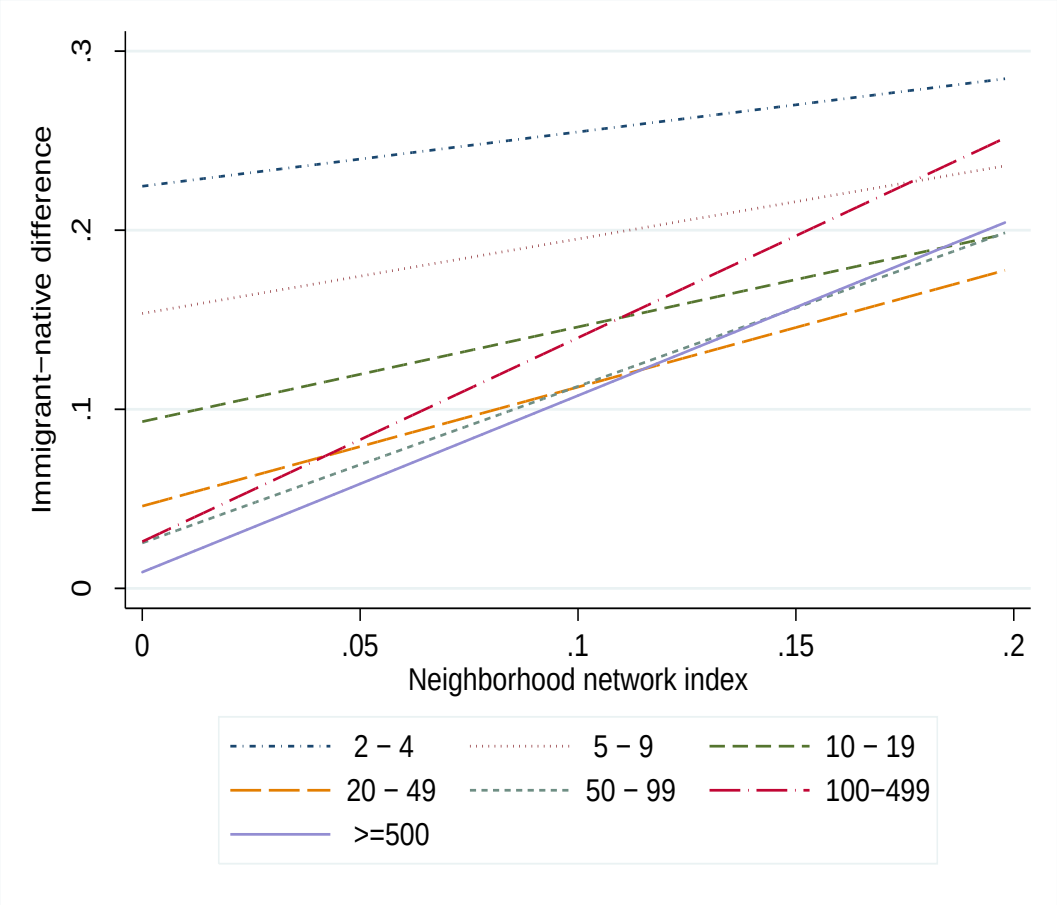
Source=LEHD database. Year 2000 second quarter.

Figure 6: Coworker share by employer sector



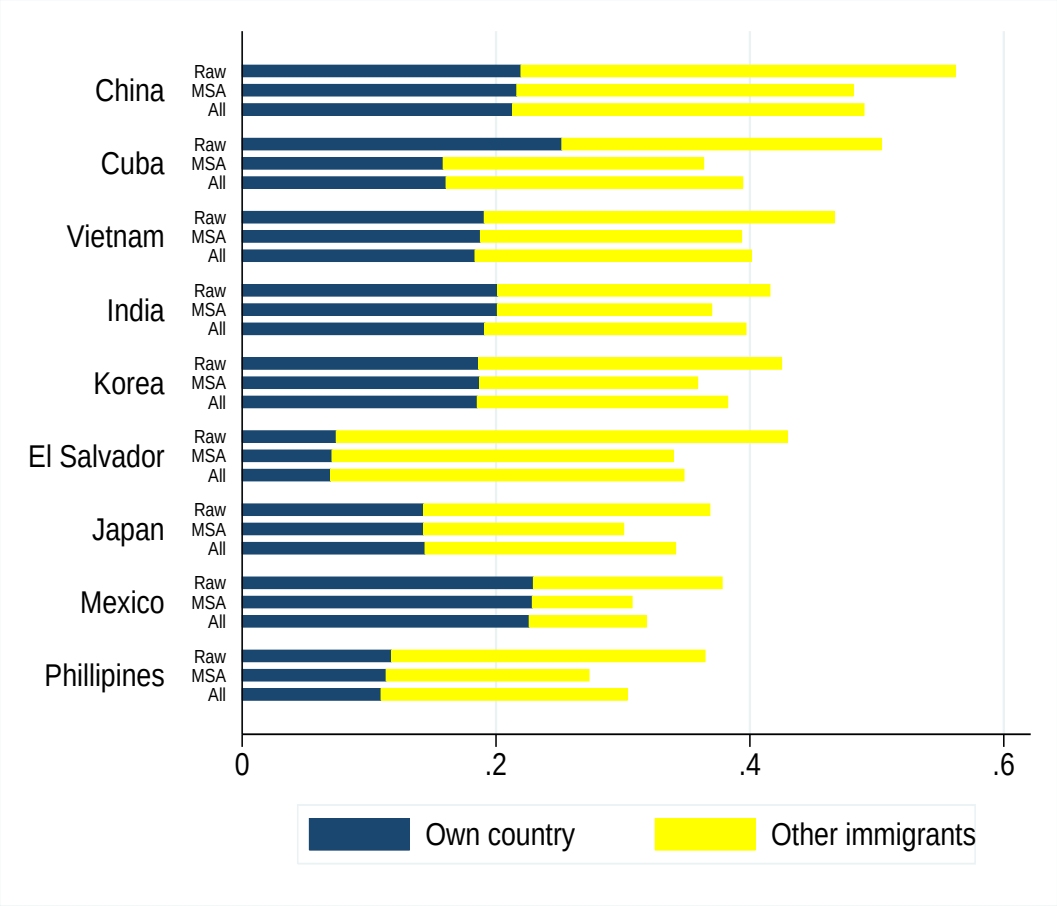
Note: Size, individual's age, plant age, sex, units and MSA groups use total population distribution. Using full two-way interactions with individual status.

Figure 7: Concentration as a function of network index for different employer-size groups



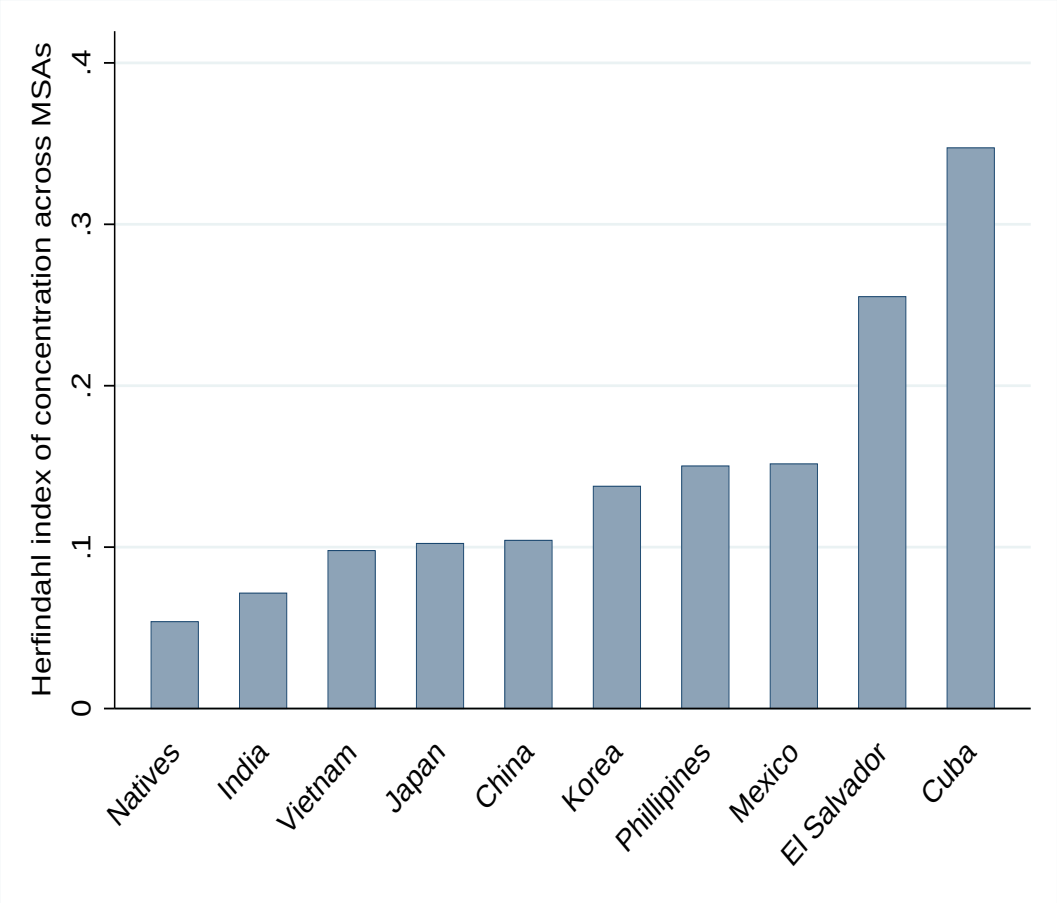
Notes:For this graph we use estimates from full regression. The full regression includes all of our standard controls, including interactions between controls and an immigrant dummy variable. The unit of observation is a worker. N=31,572,732. The regression also includes our proximity index and its interaction with an immigrant dummy variable.

Figure 8: Coworker shares by country of origin



Notes: Raw denotes coworker country-of-origin share estimated with no controls. MSA denotes estimates that include MSA dummy variables as the only control. All denotes estimates that include controls for MSA, person and establishment age, sex, sector, establishment size, and multi-unit status.
Source:Authors calculation based on LEHD database 2000 quarter 2.

Figure 9: Geographic concentration of immigrants by country of origin



Notes: For each country of origin, the index is calculated as the sum of squared population shares for each of our 31 MSAs.
Source:Authors calculation based on LEHD database 2000 quarter 2.