NBER WORKING PAPER SERIES

WORKPLACE CONCENTRATION OF IMMIGRANTS

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

Working Paper 16544
http://www.nber.org/papers/w16544

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2010

We thank the NIH for financial support and participants at the SOLE meetings and workshops at the University of Chicago, the University of Kentucky, and the Center for Economic Studies for helpful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, the Comptroller of the Currency, or the U.S. Department of the Treasury. All results have been reviewed to ensure that no confidential information is disclosed. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Workplace Concentration of Immigrants
Fredrik Andersson, Mónica García-Pérez, John C. Haltiwanger, Kristin McCue, and Seth Sanders
NBER Working Paper No. 16544
November 2010
JEL No. J61,L22,R23

ABSTRACT

To what extent do immigrants and the native-born work in separate workplaces? Do worker and firm characteristics explain the degree of workplace concentration? We explore these questions using a matched employer-employee database that extensively covers employers in selected MSAs. We find that immigrants are much more likely to have immigrant coworkers than are natives, and are particularly likely to work with their compatriots. We find much higher levels of concentration for small businesses than for large ones, that concentration varies substantially across industries, and that concentration is particularly high among immigrants with limited English skills. We also find evidence that neighborhood job networks are strongly positively associated with concentration. The effects of networks and language remain strong when type is defined by country of origin rather than simply immigrant status. The importance of these factors varies by immigrant country of origin—for example, not speaking English well has a particularly strong association with concentration for immigrants from Asian countries. Controlling for differences across MSAs, we find that observable employer and employee characteristics account for almost half of the difference between immigrants and natives in the likelihood of having immigrant coworkers, with differences in industry, residential segregation and English speaking skills being the most important factors.

Fredrik Andersson
OCC, Dept of Treasury
Fredrik.Andersson@occ.treas.gov

Mónica García-Pérez
St. Cloud State University
Dept. Economics
SH354
720 4th Ave S
St Cloud, MN 56301
migarciaperez@stcloudstate.edu

John C. Haltiwanger
Department of Economics
University of Maryland
College Park, MD 20742
and NBER
haltiwan@econ.umd.edu

Kristin McCue
U.S. Census Bureau
CES/2K130E
4600 Silver Hill Rd
Washington DC 20233-6300
Kristin.McCue@census.gov

Seth Sanders
Department of Economics
and Sanford School of Public Policy
Duke University, Durham, NC 27708
seth.sanders@duke.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w16544

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 the rate at which new immigrants acquire U.S. and location-specific skills to become integrated into the local economy.[1] 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? 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. These data permit quantifying the extent of workplace concentration of immigrants and the contribution of worker, firm and location factors to this concentration.

The paper has two broad objectives. The first is descriptive: to document the extent of immigrant concentration in workplaces. We show that immigrants are much more likely to have immigrant coworkers than are natives. This is driven partly by the geographic concentration of immigrants, but holds even within local labor markets. At the same time, most immigrants do have native coworkers; only a small share work in immigrant-only workplaces. And substantial variation in immigrant concentration across firms remains after controlling for location, employer and employee characteristics.

Our second broad objective is to account for factors that drive the observed concentration. Here we address two questions. First, what fraction of immigrant concentration

[1]Borjas, Freeman, and Katz (1996) and Card (2001) are examples of studies that analyze the impact of immigrant inflows on the employment and wages of natives. Chiswick (1978), Borjas (1985), and Borjas (1994) represent classical studies on the rate of assimilation.

is accounted for by location, employer and employee characteristics? Second, do these characteristics affect the share of coworkers who are immigrants differently for immigrants and natives? These are related but distinct questions. A factor may affect the share of immigrant coworkers equally for immigrants and natives but be vitally important in explaining immigrant concentration. For example, while 99% of natives are fully proficient in English, only 80% of immigrants have complete proficiency. If English language proficiency lowers the share of coworkers who are immigrants, even if it does so equally for immigrants and natives, it likely accounts for a substantial fraction of immigrant concentration; this is simply because of the large differences in English language proficiency between immigrants and natives. On the other hand, a factor may increase the immigrant coworker share more for immigrants than natives yet be unimportant in explaining the average level of immigrant concentration. For example, if social networks are more neighborhood based for immigrants than for natives, it is likely that where a worker lives is more strongly related to the share of immigrant coworkers for immigrants than for natives; however, this might explain little of the average level of immigrant concentration if firms hire through multiple channels and neighbor referrals are a small fraction of overall hiring for both immigrants and natives.

The factors we examine are based on underlying theories of how and why firms hire specific types of workers. Workplace concentration may reflect sorting and matching on skills and other related factors. For example, language skills may play a large role in governing interactions among workers and between workers and customers, affecting firm productivity. Residential segregation may also play a role. Firms may be more likely to employ workers from a given residential location because they hire through social networks or simply because a firm's location may make access from a specific neighborhood easy, given proximity or the transportation network. In addition, differences in immigrant concentration by employer characteristics like industry and size may reflect technologies and business practices that make immigrants particularly well (or ill) suited to some types of production.

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 role of observable employer and employee characteristics in accounting for the patterns of immigrant concentration in businesses. Section 5 presents analysis of how the observable employer and employee characteristics may have differential effects on immigrant concentration for immigrants and natives. Concluding remarks are provided in section 6.

2 Background

Our work draws primarily on the literature explaining sorting of workers into firms. This literature has identified four types of sorting that may contribute to segregated workplaces: (a) sorting based on productive characteristics of workers, (b) sorting resulting from preferences of workers and employers, (c) sorting based on the information available to workers and firms, and (d) sorting resulting from the residential location of workers relative to firms.

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 positive correlation may occur either because a firm demands workers of a particular skill level or because coordination within a firm demands that workers share a common skill such as speaking a particular language. Cabrales, Calvo-Armengol, and Pavoni (2008) emphasize a different skill-based sorting 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. Skill-based sorting could lead to workplace segregation of immigrants from natives because of differences in the two groups in their distributions of skills. For example, the education distribution of immigrants is more bifurcated: immigrants are much more likely than natives to have an 8th grade education or less (23% vs 5.2% for natives in the 2000 census), but also more likely to have an advanced degree (10.3% vs. 8.6% for natives). Therefore, firms that hire exclusively low-skilled or exclusively high-skilled workers will tend to have workforces with above-average immigrant shares.

Language differences provide another productivity-based reason for segregation of immigrants. A shared language may increase worker productivity, leading firms to choose workforces in which everyone speaks the same language. If so, immigrants from non-English speaking countries may be particularly likely to be segregated, and may also be particularly likely to work with immigrants who share their language. Lang (1986) develops a formal model of wage differences that arise because firms must pay a premium for bilingual workers who can bridge the language barrier. One implication of this model is that complete segregation would occur if sufficient capital were owned by each language group. Several authors have found evidence consistent with such segregation by language. Hellerstein and Neumark (2008) find evidence that Hispanics with poor English-language skills are particularly likely to work with other Hispanics. An earlier study by Portes and Wilson (1980) examines whether segregation among Cuban immigrants in Miami occurs through employment by Cuban-owned firms as in the Lang model. The authors find that not only do Cubans work together, many work in firms owned by other Cubans. Gárcia-Pérez (2009) also finds supporting evidence that immigrant-owned small firms (mostly Hispanic or Asian-owned) are more likely to hire immigrants than are native-owned small firms.

The classic model of preference-based segregation is Becker (1957). In this model, segregation of workers by race occurs as the result of discriminatory preferences on the part of firm owners or, in another version, on the part of coworkers. If whites demand a premium to work with blacks, firms segregate workers into separate facilities to avoid paying a wage premium to discriminating whites. Such models can generate high levels of segregation but with limited disadvantage in wages to the minority group. While these models were developed to explain black-white differences in wages and labor market segregation, replacing race-based discrimination with a native aversion to working with immigrants would lead to similar implications for the segregation of immigrants.

Information-based theories concentrate on mechanisms that match workers to jobs. For example, if workers interact mostly with others who have similar characteristics, firm use of employee referrals and/or employee use of personal contacts may increase workplace segregation. Use of referrals and personal contacts may lower the costs of finding good matches, and these effects may vary across groups. For example, Holzer (1988) finds that, for workers, use of personal contacts to search for jobs is inexpensive and has relatively high rates of success. Holzer (1987) and Montgomery (1991) find that, for employers, employee referrals provide both a low cost recruitment strategy and, on average, new hires with higher productivity and lower turnover rates. 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.

These information flows may combine with residential segregation to generate workplace segregation. There is ample evidence that immigrants' places of residence are spatially concentrated.[2] Neighborhoods play an important role in who you know and hence may provide important job contacts and references. Several papers have found that those working in the same place are disproportionately from the same neighborhoods. Using data from the city of Boston, Bayer, Ross, and Topa (2008) find that a worker is about one-third more likely to work with other residents of their Census block as to work with residents of other blocks in their block group (typically eight or so contiguous blocks). This comparison to other blocks nearby provides important evidence that having coworkers who are neighbors does not stem solely from factors such as transportation routes or distance that make a location a natural place to work for those living in a particular neighborhood. Sample sizes, as well as the ethnic make-up of Boston, restrict the authors' investigation to black-white differences. It is also the case that the authors' data cannot distinguish between employees of two establishments located in the same block, so some of those working in the same location may not be coworkers.

Hellerstein, Neumark, and McInerney (2008) also present evidence of the importance of neighborhood network effects. Using matched employer-employee data, they measure the strength of social networks using the excess probability that a member of a particular race/ethnicity group works in the same establishment as neighbors from the same group. For whites they find that another worker living in the same census tract has twice the probability of working in the same establishment as what one would expect from randomness. They disaggregate their analysis for whites, blacks and Hispanics, and by education level, ability to speak English, and immigrant status within some of these groups. They find particularly large effects for Hispanics with poor English language skills and Hispanics who are immigrants.

[2]For example, 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 after some time in the U.S.

Given Hellerstein, Neumark, and McInerney's extensive work in this area and their use of matched employer-employee data, it is worth clarifying how the analysis in this paper differs from their work. Both their work and our own uses information about workers from the 1-in-6 decennial long form linked to data on employers. While we have our full set of measures only for this sample, for a number of states we have some information on almost all workers from Unemployment Insurance (UI) earnings records and from the Numident file of the Social Security Administration. In particular, the Numident provides us with place of birth for almost all workers, which allows us to calculate the concentration of immigrants based on all coworkers—whether in the long-form sample or not—and in all firms. Having this additional information provides several advantages: it gives us much larger samples, but it also results in a more representative sample of employers and so gives us greater leverage to examine the effects of employer size.[3]

Hellerstein and Neumark (2008) (hereafter HN) is substantively closest to our work. Their paper focuses on measuring and accounting for worker segregation of black, Hispanic and white workers, while we consider segregation of immigrants. Because our work is aimed at a more complete accounting of sources of worker segregation, our econometric methods differ as well. HN use an elegant simulation approach, comparing observed segregation to what would be expected if employers hired randomly, drawing on statistical methods developed in Carrington and Troske (1997).[4] If observed concentration is larger than expected, this is taken as evidence of non-random hiring. They carry out these simulations allowing hiring to be random within a limited number of

[3]See Lengermann, McKinney, and Pedace (2004) for earlier work using the LEHD infrastructure files that also took advantage of the representative coverage of employers to explore immigrant concentration variation across employers. This earlier work found important differences in immigrant concentration across MSAs and employer size classes, although the focus of the analysis was to use these differences to explore differences in earnings for immigrants and natives.

[4]Carrington and Troske (1998) also used matched Census employer-employee data to measure segregation across businesses, using simulations of the effects of random matching to distinguish between random and systematic segregation by gender in manufacturing.

strata. If within strata the observed and expected concentration are the same, then they take this as evidence that these strata explain the unconditional level of worker concentration. This method works well as long as the number of strata is small.

We employ more traditional regression-based methods and decompositions. This makes it easier to control for many factors at the same time and to assess the marginal contributions of specific factors. This approach also easily accommodates examining which characteristics have a particularly strong relationship to concentration by simply adding interaction terms.

HN do point out that their simulation approach mitigates a potentially serious problem: while in expectation random hiring would equalize the fraction of coworkers who are immigrants across natives and immigrants, with small samples in a particular strata, a finding of unequal coworker shares is not necessarily inconsistent with random hiring. Empirically, the authors establish that while potentially an issue, this bias was small in their sample and did not qualitatively affect results. This small sample problem would be of concern with our regression approach if our control variables defined cells with small samples. But small sample sizes are of less concern given our data set: our measures of concentration are based on all employees (34.3 million) of about 735,000 employers, and our sample consists of only relatively large MSAs, so cells based on small samples are uncommon.

We also have drawn on findings from Hellerstein, Neumark, and McInerney (2008) on the importance of network effects in determining the distribution of workers. But our aim is to identify the importance of these effects in accounting for immigrant concentration, while Hellerstein, Neumark, and McInerney (2008)'s goal is to establish the importance of networks for labor markets more generally.

3 Methodology and Data

3.1 Data

We construct a cross-sectional sample of workers in selected MSAs for the second quarter of 2000 by combining data from the Longitudinal Employer-Household Dynamics (LEHD) database and the 2000 Decennial Census 1-in-6 long form. The LEHD database 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 2008, though coverage in the earlier years varies across states. Workers' earnings records have been matched to characteristics of their employer gathered in quarterly administrative UI reports and through Census Bureau business censuses and surveys.[5] Basic demographic data are also available for workers, including place of birth which allows us to identify immigrants. The LEHD data have the unique advantage of allowing us to measure employer and workforce characteristics using information on all employees of all UI-covered employers in the included states. Thus, we can identify basic characteristics, including immigrant status, of all coworkers. Their main disadvantage for studying immigration is that they include only on-the-books employees, leaving out the self-employed and those working in the informal sector. Thus they likely have poor coverage of undocumented immigrants. Coverage of employment in agriculture is incomplete, so we exclude that sector.

While we can measure selected coworker characteristics for virtually all workers in our UI earnings sample, we need to match to the 1-in-6 Decennial Census long form sample to obtain measures of education and language proficiency. The outcome of that match is an approximately 1-in-10 subsample of the UI earnings sample.[6] Matched

[5]A full description of the LEHD data infrastructure can be found at Abowd, Vilhuber, McKinney, Sandusky, Stephens, Andersson, Roemer, and Woodcock (2006).

[6]From the full 1-in-6 long form sample, we exclude those who do not report any employment at the

workers have a slightly lower immigrant coworker share than workers in the full sample, and there seems to be a tendency for older, longer-tenure workers at large establishments and in older, multi-unit firms to be overrepresented in the matched sample.[7] But generally these differences are small. We estimate a propensity score model and use it to create weights for the matched sample that adjust for selection on observables.[8] Using these weights, matched sample results without education and language measures closely replicate regression results based on the complete UI earnings sample.[9]

We limit the matched sample to workers employed in 31 selected metropolitan areas in 11 states, with our choice of areas being based on the presence of substantial immigrant populations and data availability. While we use a small number of states, they include five of the six states in which the 2000 foreign-born population exceeded 1 million. In addition to cities with large immigrant populations, we also include several MSAs with smaller immigrant populations but with very rapid growth in foreign-born residents between 1990 and 2000.[10] We include all matched employees of non-agricultural businesses located in a sample MSA, whether or not they live in the MSA. This gives us a sample of 3.5 million workers. Even the smallest of our MSAs has more than 3,000

time of the census. In addition, not all long form respondents can be matched to the UI data, either because the information needed for matching is missing or because no match can be found.

[7]Unweighted sample means for the full and matched sample are in web appendix Table W-2 for immigrants and in Table W-3 for natives. The web appendix for this paper can be found at `http://econweb.umd.edu/~haltiwan/workplace_conc_oct29_2010_web_appendix.pdf`. Tables and figures with the prefix "W-" all appear in this web appendix.

[8]We use the following variables to estimate propensity scores: worker age and sex; 11 country of origin groups—Mexico, China, Cuba, El Salvador, India, Korea, Japan, Vietnam, Philippines, other countries of origin, and natives; log earnings; whether the worker was employed for each of quarters 1, 2, and 3 of 2000; three-digit industry; MSA; working population density; establishment age and size; and the number of establishments owned by the firm. Industry categories are based on the 1990 Industrial Classification System used in household surveys. This classification is based on SIC codes, but categories are somewhat more aggregate than 3-digit SIC categories.

[9]Compare Tables W-4 and W-5 in the web appendix.

[10]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 share foreign-born in 1900-1930 (''former gateways"). We drop 14 of Singer's 45 MSAs because we do not have the data we need for those areas.

immigrant workers.

The average immigrant workforce share across our 31 MSAs is 19% but immigrants are less than 11% of the work force in eight MSAs, while three MSAs have workforces that are more than 35% immigrant.[11] Even with random assignment to jobs within a local labor market, these substantial differences across areas would make immigrants more likely to work together than to work with natives, simply because immigrants are disproportionately in the MSAs with high immigrant shares. Since our interest is in how workers are matched with employers within a local labor market, we include MSA dummy variables in all of our specifications so that estimates are based on within-MSA variation.

We follow HN, Aslund and Skans (2005a), and Aslund and Skans (2005b) 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 the absence of any systematic concentration, in large samples the mean coworker share for both immigrants and natives should equal the share of immigrants in the workforce. Based on this property, we use the difference between the mean coworker share for immigrants and natives to measure immigrant concentration. A positive value indicates that immigrants are more concentrated than would be expected based on random

[11] See web appendix Table W-1 for more detail.

allocation.[12] 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.

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.[13] In computing the coworker share, we equally weight 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 (primary job).

Figure 1 plots the cumulative distribution of immigrant coworker share for natives and for immigrants as of the second quarter of 2000.[14] In our sample of immigrant-rich MSAs, 10% of natives work in native-only workplaces (coworker share=0), while the share of immigrants working for immigrant-only businesses is considerably smaller (2.8%). About 10% of the median native's coworkers are immigrants, while for the median immigrant, the share is about 32%. For reference purposes, we include a third line giving the cumulative distribution that would apply if immigrants and natives were randomly assigned to employers in a manner that preserves the size distribution of employment. This simulated distribution depends only on the overall immigrant share (18.7% in our sample, on a weighted basis) and the size distribution of employment. By assumption, the random assignment distribution is identical for immigrants and natives.

Because the likelihood of extreme values occuring randomly is quite low in large

[12] With the caveat of potential small sample bias discussed in HN.

[13] In our sample of MSAs, immigrants account for 27% of employment in single-employee businesses, and 16% of employment in businesses with more than one employee.

[14] Figure W-1 in the web appendix shows that the full sample exhibits very similar patterns.

samples, and because large employers account for a substantial share of employment, about 60% of workers would have between 17% and 20% immigrant coworkers if workers were grouped randomly. If all employers had only two employees, the random assignment graph would look quite different: 81.3% of workers would have a coworker share of 0, while the other 18.7% would have a coworker share of 1. In contrast, if all employers had 100 employees, 30% of workers would have coworker share in the 17%-20% range. Over half of our sample works for businesses with 100 or more employees, and more than 90% works for businesses with at least 10 employees, which is why very few workers would be expected to have no immigrant coworkers under random assignment.

The observed distributions of coworker shares for both natives and immigrants differ substantially from what would obtain under random assignment. Under random assignment, we would expect the share with only native coworkers to be well below the 10% observed for natives (but only a bit above the 2.2% observed for immigrants), while the share of employees working only with immigrants would be close to zero. Our analysis focuses on the mean difference in coworker shares between immigrants and natives, which is close to the median difference illustrated here by the horizontal gap between the distribution functions where they cross 0.5 on the vertical axis. The median coworker share for both immigrants and natives is quite difference from the random assignment value of .187, but note that the immigrant/native difference is substantial even at the 10th percentile.

3.2 Regression specifications and decompositions

Our empirical approach is based on a series of regressions with the coworker share as the dependent variable, and individual workers on their primary job as the unit of analysis. To ease computation with over 3 million workers, we use linear regression models rather than adopting an approach that accounts for the limited range of the dependent variable. As Figure 1 illustrated, most of the mass of the distribution is not at either 1 or 0, which mitigates some of the problems inherent in the linear model. There is a strong positive correlation in the coworker share among employees of the same business that generates a downward bias in conventionally estimated standard errors in all worker-level regressions. To avoid this, we use the Huber-White variance estimator, allowing for arbitrary correlation of errors among employees of the same establishment.

Our initial regression specification is:

$$C_{ij} = \gamma_N^{base} + \gamma_I^{base} I_i + \theta^{base} msa_{ij} + \epsilon_{ij}^{base} \tag{3.2}$$

where i denotes an individual and j denotes a workplace. I and N denote immigrants and natives, respectively. In (3.2), the constant term γ_N^{base} represents the mean coworker share for the omitted category, which in this simplest specification consists of natives in the omitted MSA. Coefficient γ_I^{base} gives us the mean within-MSA difference between immigrants and natives in how likely they are to have immigrant coworkers, and thus represents our base measure of immigrant concentration.[15]

We compare results from (3.2) to an augmented regression to address our first question: Which characteristics of workers and employers are important in accounting for immigrant concentration? Therefore, we add a vector of worker and employer characteristics x_{ij} to obtain:

[15]In an earlier draft of this paper, we explored differences in concentration between recent immigrants—those arriving between 1995 and 2000—and more established immigrants. Web appendix Tables W-4 and W-6 give information on differences in means for the two groups. Some of the results presented here are broken down for the two groups in web appendix Tables W-8, W-9, and W-10. Recent immigrants have higher coworker shares than more established immigrants. These differences confound the effects of time in the U.S. with changes in immigrant characteristics across entering cohorts. We would need to exploit the panel aspect of our database to seriously address the effects of assimilation, but this difference suggests that assimilation effects on concentration are likely to be important.

$$C_{ij} = \gamma_N^{main} + \gamma_I^{main} I_i + \theta^{main} msa_{ij} + \beta^{main} x_{ij} + \epsilon_{ij}^{main} \tag{3.3}$$

To the extent that $\gamma_I^{main} < \gamma_I^{base}$, the vector of characteristics in x partially account for the raw immigrant concentration.

We quantify the contributions of various sets of characteristics using a decomposition developed by Gelbach (2009). Let $\delta = (\gamma_I^{base} - \gamma_I^{main})$ represent the amount of immigrant concentration explained by the characteristics included in x. Gelbach notes that the formula for omitted variable bias gives a natural way to decompose δ. If x has K components then δ can be decomposed into K additive terms with the contribution of the k^{th} variable given by $\delta^k = \beta^{k,main} * \alpha_I^k$, where the α_I^k are coefficients estimated from the K auxiliary regressions:

$$x_{ij}^k = \alpha_N^k + \alpha_I^k I_i + \eta_{ij} \tag{3.4}$$

This decomposition makes clear that two things must occur for a factor to account for a substantial share of immigrant concentration: (i) the factor must be strongly correlated with immigrant concentration even when conditioning on other controls ($\beta^{k,main}$ is large); and (ii) there must be a large average difference between immigrants and natives in x^k (α_I^k is large).

We then turn to our second question: In which kinds of jobs and among which sorts of workers is immigrant concentration highest? For example, is concentration higher in small firms, or for certain education groups? (3.3) assumes that immigrant concentration is the same within cells defined by the covariates, so the reduction in concentration between (3.2) and (3.3) is driven by differences in the distribution of immigrants and natives across cells. To address this second question, we add interactions between the immigrant dummy variable and other covariates to (3.3) to allow the covariate coefficients to differ for immigrants and natives, obtaining:

$$C_{ij} = \gamma_N^{int} + \gamma_I^{int} I_i + \theta^{int} msa_{ij} + \lambda^{int} I_i * msa_{ij} + \beta^{int} x_{ij} + \phi_I^{int} I_i * x_{ij} + \epsilon_{ij}^{int} \tag{3.5}$$

Finally, we apply a Oaxaca-like version of Gelbach's decomposition to equation (3.5). This allows us to parcel out the contributions of differential effects of covariates on immigrant concentration, in addition to the effects of differences in means. In this case, $\delta^* = (\gamma_I^{base} - \gamma_I^{int})$ is the amount of immigrant concentration explained by x and its interactions.

The decomposition simplifies to:

$$\delta^* = \sum_{i=1}^{K} [\beta^{k,int} * (\bar{X}_I^k - \bar{X}_N^k) + \phi_I^{k,int} * \bar{X}_I^k] \tag{3.6}$$

The first term then quantifies the contribution that a particular variable makes to immigrant concentration through the immigrant-native difference in means—evaluated using the main-effect coefficient (i.e. the effect for natives)—while the second gives the contribution from differences in how the covariate affects coworker shares for the two groups (evaluated at the immigrant mean).

One issue in applying this to (3.5) is that there are many possible decompositions. The interaction terms allow the extent of immigrant concentration to vary with the value of x, so any decomposition conditions on a particular value of the vector x. In $\delta^* = \gamma_I^{base} - \gamma_I^{int}$, γ_I^{int} represents the level of concentration for immigrants when $x = 0$. So, for example, if we simply switch which industry we are omitting, the amount being decomposed and the amount attributed to industry differences changes. With the variables as originally scaled, $x = 0$ is not a particularly interesting value, so we instead present decompositions at the mean value of vector x for natives.

3.3 Descriptive statistics

Table 1 presents summary statistics for immigrant and native workers in our matched sample. The first row gives mean coworker shares. For the average native, about 14% of coworkers are immigrants, while 37% of the coworkers of immigrants are immigrants. The immigrant-native difference in coworker means—our measure of concentration—is .229, indicating substantial concentration.[16]

The following rows give demographic information for each group. Immigrants are slightly older than natives in our sample. Men substantially outnumber women among working immigrants, while among working natives men are more narrowly in the majority. Differences between immigrant and native women in rates of labor force participation likely contribute to these gaps. Most immigrant workers arrived in the U.S. as children or young adults. Immigrants are much more likely to be high school drop-outs than are natives, but immigrants are also overrepresented among those with advanced degrees.

The category "Does not speak English well" consists of those who speak a language other than English at home, and report that they speak English "not well" or "not at all". Unsurprisingly, immigrants are more likely than natives to fall into this category, but note that it does include some natives. Mean log earnings on the primary (highest earnings) job are very similar for immigrants and natives, and immigrants are more likely than natives to work for their 2000-Q2 employer in at least one of the surrounding quarters. Differences in job tenure likely contribute to the slightly higher earnings of immigrants, as transitory jobs are likely to have particularly low quarterly earnings because most will involve less than three full months of work. They may also be associated with relatively low wage rates and part-time work.

[16]This measure of concentration does not control for MSA differences in the immigrant population. In all of regressions below, we control for MSA effects so our focus is on within MSA concentration. Controlling for MSA effects alone reduces concentration to 0.171

We find only minor differences between immigrants and natives in employer characteristics. Immigrants are more likely to work in the smallest establishments, and less likely to work in the largest, but overall the differences are small. Differences in the distribution of employment by establishment age are also small. However, immigrants are less likely than natives to work for multi-unit firms. Immigrants are more concentrated in manufacturing than are natives, but otherwise the differences by sector are not particularly large.[17]

The last three rows give means for three additional measures that we construct to explore the relationship between workplace concentration and neighborhood networks. Each of these is based on information on worker tract of employment and/or tract of residence.[18] Because we only have data on those who work, we base these variables on workers residing in a particular tract rather than all tract residents.

The first measure is simply the share of immigrants in a worker's tract of residence, which we use to capture residential segregation. As can be seen in Table 1, immigrants in our sample of MSAs are more likely to live in tracts with high immigrant shares than are natives, but even so the majority of their neighbors are natives.

We construct a second variable for each worker by calculating the share of employees at other businesses located close to his employer (defined as other employers in the same tract) who also live in the worker's residential tract. The denominator is the number of

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

[18]Census tracts are small geographic 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 geographic reach of a social network: the limited distance between residents of a census tract—both in terms of geography and socio-economic factors—suggests that the likelihood of interactions among residents of the same tract is high relative to the likelihood of interactions between residents of different tracts.

employees working for other employers in a worker's tract of employment. The numerator is the number among that group who live in the worker's residential tract.[19] Proximity or convenient transportation links may make residents of certain neighborhoods likely to work at a particular location, resulting in a relationship between workplace and residence. This measure of the general propensity for workplace and residence locations to be connected will control for commuting patterns but will also reflect other connections between workplace and residence such as sorting across locations by skill. We refer to this as a shared commute index. For the average worker, there is not a strong association between place of employment and particular tracts of residence: the mean for this variable is only 0.3% for immigrants, and 0.5%for natives.

Our third measure is intended as a proxy for the presence of neighborhood-based job networks. For each worker we calculate the fraction of their coworkers who also reside in the worker's tract of residence. So, for example, if a business hired three residents each from four different residential tracts, each worker would have a neighborhood network index of 2/11, as two of their 11 coworkers would be from the their neighborhood. The mean of the network index is small: for both immigrants and natives, 1.9% of coworkers live in the same tract. While the averages are small there is considerable variation across workers. For instance, the mean is substantially higher for workers employed in small businesses and it falls systematically with employer size. We have in mind that referrals by current employees may be an important recruitment source, and many referrals may come about through contacts with neighbors (as discussed in Section 2). If so, where neighborhood referrals are important we would expect to find people who work together also living close together. This network variable should capture such effects but may more generally capture the extent to which residential location

[19] In our sample, there are on average 49 employers per tract (excluding tracts that are strictly residential). Seven percent of tracts with employment have only one employer, and for those tracts, the variable is zero. Only 9% of workers in our sample work in single-employer tracts.

and employment location are correlated. To control for this, we include the residential segregation and shared commute variables as controls when we explore the role of the network variable.

4 Accounting for immigrant concentration

In this section, we quantify the extent to which observable employer and employee characteristics can account for patterns of concentration. We first estimate a regression with the coworker share as the dependent variable based on specification (3.2) and then apply the Gelbach decomposition to this specification. Table 2 presents the coefficient estimates from the regression. Table 3 gives the results of the decomposition.

4.1 Basic results

We use the average within-MSA concentration as our starting point. This average is 0.171, indicating that on average the share of immigrant coworkers is 17.1 percentage points more for immigrants than for natives working in the same MSA. Table 3 shows that controlling for observable employee and employer characteristics reduces estimated concentration from 0.171 to 0.096, a reduction of about 45%. The bottom part of the table gives the share of that reduction accounted for by particular types of characteristics. Three factors stand out as accounting for significant shares: difficulty speaking English, industry of employment, and the share of a worker's neighbors who are immigrants. Together these account for 92% of explained concentration (and 40% of total concentration), with education and the interaction of firm age with firm size the next runners up with about 3% each.

Language skills make a large contribution to explaining concentration both because most of those who do not speak English well are immigrants, and because of the substantial increase in coworker share (7.7%) associated with poor English skills even when controlling for numerous other factors. Given the large share of U.S. immigrants of Hispanic origin, it is worth pointing out that our results on language are generally consistent with HN's findings for Hispanic/white concentration. Using the same language grouping, HN find that about 15% of all Hispanic/white concentration is attributable to segregation by language, while we attribute about 9% of overall immigrant/native concentration to language.[20]

The substantial contribution of industry comes about because the distribution of employment across detailed industries is quite different for immigrants and natives. This seems somewhat surprising given that in Table 1 the distribution across sectors shows only modest differences. To try to bring out where these differences are important, we split the contribution into differences in immigrant employment by sector and then into the contributions of within-sector detail. This split is somewhat sensitive to how the detail is specified, but using the modal 3-digit industry within each sector as omitted categories (as we do here), differences across broad sectors (particularly the high share of immigrants in manufacturing) and then differences across detailed industries within services appear to be the most important contributors.

The other striking result is the almost 40% contribution of residential segregation by Census tract within MSA. While we do not think of residential segregation as exogenous, the large contribution it makes in this accounting framework does point to a very strong relationship between living with immigrants and working with them.

[20]I.e. 20.4% of .171-.096 is .015 or about 9% of .17. Using a more detailed breakdown of language skills, HN attribute almost one-third of Hispanic/white concentration to language.

4.2 Country of origin differences

In the analysis above, we simply distinguish between natives and immigrants, but our data also permit exploring patterns of immigrant concentration by country of origin. For example, we can estimate how likely it is for an immigrant from Mexico to have coworkers who are Mexican. These patterns are useful in considering the extent to which overall levels of immigrant concentration reflect concentration by country of origin rather than a more general phenomenon of non-natives working together. To make this manageable, we rank countries of origin by their share of employment in our sample, and carry out the analysis separately for immigrants from the top nine countries.[21]

Table 4 presents estimates of the extent of concentration by country of origin for these nine countries. Columns 1 and 3 give coefficients on the country-specific dummy variable from regressions using the share of coworkers from that country as the dependent variable. Columns 2 and 4 give coefficients on the country-specific dummy variable from regressions using the share of coworkers from other countries of origin as the dependent variable (e.g. non-Cuban immigrants in the first row). The first two columns are from regressions that include only country and MSA dummies as controls, while the third and fourth columns add the other variables used in Table 2, except that the residential segregation measure is split into nine country-specific shares in a worker's residential tract and the remainder, which gives the share of immigrants from other countries.

The first entry indicates that for the average Cuban immigrant the share of coworkers who are Cuban is 16.7 percentage points higher than the share for the average native within the same MSA. The entry in the second column shows that for Cuban immi-

[21] Our list of the top nine immigrant worker source countries in 2000 includes eight of the top nine for the U.S. population as a whole. Our top-nine list includes Japan, while the top nine based on overall U.S. population instead includes the Dominican Republic. The difference is likely driven by the set of MSAs we have rather than differences in composition between the overall population and employees.

grants, the share of coworkers who are immigrants from other countries is only 6.6 percentage points higher than the share of non-Cuban immigrant coworkers for natives. For each of the other countries as well, immigrants are significantly more likely than natives to work with their compatriots and with other immigrants as well. For most countries of origin, immigrants are much more likely to work with their compatriots than with other immigrants. The one exception is Salvadorans, who, relative to natives, are roughly twice as likely to work with immigrants from other countries as with other Salvadorans. While the results do not appear in the table, this largely reflects a propensity for Salvadorans and Mexicans to work together. Given such a propensity, the large Salvadoran other-immigrant effect likely reflects the fact that Mexican immigrants greatly outnumber Salvadoran immigrants in our sample of MSAs. In general, Asian immigrants are slightly less likely than natives to work with Mexican immigrants, and in most cases with Salvadoran immigrants as well.[22]

Columns 3 and 4 of Table 4 report estimates of the same coefficients when we include our full set of covariates. A comparison of Columns 1 and 3 shows how much the added controls contribute to accounting for concentration measures by country of origin. For Cuba, adding covariates reduces the Cuban concentration by close to half, from 0.167 to 0.094—roughly similar to the magnitudes we observed in Table 3 for all immigrants. We find a similarly large reduction in concentration for Mexicans, and roughly a 30% reduction for Salvadorans. For Asian immigrant groups—particularly Korean and Japanese immigrants—we find more modest reductions in concentration from adding covariates. While observable factors only partially explain compatriot concentration, for most country groups these factors fully explain the excess tendency to work with

[22]While the finding that Mexicans and Salvadorans are much more likely to work with each other than with other immigrants suggests the importance of a shared language, countries with shared a shared language may share other characteristics as well. Note that we find no such tendancy to work together for Cubans and Mexicans, or for Cubans and Salvadorans, despite a shared language.

immigrants from other countries.[23] Applying the Gelbach decomposition, we find that the same three factors account for most of the explained variation in country-level concentration as for overall concentration: residential segregation, English language skills, and industry of employment.[24] However, the importance of these factors differs for own- versus other-country concentration, and varies across country groups. Virtually all (93%) of the reduction in own-concentration for Cubans is accounted for by the strong relationship between living with Cuban immigrants and working with them, combined with Cubans' greater tendency to live in tracts with large shares of Cuban residents. For the other countries, residential segregation by own country of origin is important, but differences in the distribution of employment across industries contribute as well. The industry distribution of employment accounts for more than half of the explained concentration for immigrants from India and the Philippines, with residential segregation accounting for the majority of the reduction for all other countries except Mexico. Differences in English language skill make little contribution to explaining own-country concentration, but account for at least 24% of the reduction in other-country concentration for all countries except India and the Philippines.[25] Immigrants from India and the Phillipines have quite low rates of difficulty with English (5% and 4% respectively).

4.3 Taking stock

The results thus far point to three main findings. First, there is substantial concentration of immigrants in workplaces. Second, the covariates most strongly associated with con-

[23]With the full set of controls, only immigrants from El Salvador and China appear substantially more likely than natives to work with immigrants from other countries; and in these cases, covariates explain more than half of the excess non-compatriot concentration.

[24]See web appendix Table W-14 for details.

[25]Note that the decompositions here are based on models in which the effect of English language skills on the coworker share for any particular country of origin is assumed to be the same no matter what the country of origin. Below, we examine the effects when we allow the coefficient on English in the Cuban coworker share regression (for example) to differ between natives, Cubans, and non-Cubans.

centration are industry, language skills and residential segregation. Third, a substantial share of this concentration takes the form of immigrants working with their compatriots. Even after accounting for many employer and worker characteristics, including employer location, industry and size, concentration remains substantial within employer and worker characteristic groups. These results are based on specifications that assume that the effect of covariates on concentration are the same for immigrants and natives. We relax this requirement in the next section.

5 Where is Concentration Greatest?

In this section, we turn to exploring in which kinds of jobs and for which kinds of workers do we observe the highest levels of concentration. We do so by estimating models that allow for covariates to have differential effects on the coworker characteristics of immigrants and natives using specification (3.3). The coefficients on interaction terms allow us to identify characteristics associated with particularly high or low levels of concentration. We can then look more closely at whether the observed patterns are consistent with the alternative mechanisms underlying concentration discussed in section 2. We also examine whether the interaction effects account for an important share of the variation in concentration.

5.1 Overall patterns

Estimation of the full specification (3.3) yields too many interaction coefficients to usefully present the full set in a table. For brevity, we present graphs illustrating the more interesting categorical interactions, and then present coefficient estimates for selected continuous variables in Table 5. We note that adding the full set of interactions increases the R^2 to 0.55 as compared to 0.51 without the interactions.

We begin by illustrating the differential effects of our human capital variables: English language skills and education. As is the case in each of the following figures, the coworker shares in Figure 2 are predictions based on regression estimates from specifications that include all of the controls listed in Table 3 along with a full set of interactions with the immigrant dummy variable. We calculate predictions using mean values of all controls for the sample as a whole (pooling immigrants and natives), except that we assign values to the categorical variables that define the bars.

We know (from Table 2) that there are substantial differences in concentration associated with the ability to speak English and (from Table 3) that differences between immigrants and natives in the ability to speak English explain a significant share of overall concentration. Here we allow differential effects of language skills and find (as Figure 2 illustrates) that concentration is greater among those who do not speak English well. This is consistent with HN's findings that Hispanics with poor language skills were particularly likely to work with other Hispanics. We return to the role of language below using country-of-origin groups.

Figure 3 illustrates differences in concentration across education categories. While the differences across education levels are not large, concentration is lowest among high school dropouts—both immigrant and native high school dropouts are quite likely to work with immigrants, but the difference between them is relatively small. Note that, in our sample, 31% of workers who did not graduate from high school are immigrants, but both native and immigrant coworker shares for these workers are well below .31 because most workplaces include workers from more than one education group. Concentration is highest among those with advanced degrees.

Employer size effects are of particular interest because they potentially reflect a number of factors that influence concentration as described in section 2. One reason that size may matter is that production processes vary 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. If this is the case, more concentrated workplaces may permit immigrant workers to overcome language and related barriers. A related argument is that the hiring process is likely to be less formal for small businesses. Moreover, vacancies are likely to occur less often in small businesses, even if vacancy rates are as high or higher than in medium to large businesses. Both of these effects might increase the importance of social networks in the hiring process for small businesses.

We find a strong negative association between size and concentration. While the size coefficients in Table 2 indicate that the employment share of immigrants is relatively constant across size classes, Figure 4 shows that concentration falls substantially with size. Natives are somewhat less likely to work with immigrants in smaller establishments than in large establishments, while immigrants are much more likely to work with other immigrants in the smallest establishments. For example, 33% of coworkers are immigrants for immigrants working at establishments with 5-9 workers, while at establishments with 500 or more employees, the coworker share for immigrants is 22%. It is striking that these large effects hold even after controlling for many other factors, including detailed industry. For the smallest firms, much of the concentration comes from segregated workplaces—those with only immigrant or only native employees.[26] About two-thirds of natives in the 2-4 employee size class work only with other natives, while roughly 40% of immigrants work only with other immigrants. But the share of employment accounted for by all-immigrant and all-native workplaces falls quickly as employer size increases.

We think two mechanisms drive this pattern. One is a size effect resulting from employer and employee behavior—a greater tendency for immigrants to work with immi-

[26] See Figures W-8 and W-9 in the web appendix for additional detail on coworker shares by establishment size.

grants in smaller firms. The second is a mechanical effect 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. As we illustrate in a simple statistical model in the web appendix, the contribution of this statistical artifact should fall quickly with size.[27] If this statistical artifact is generally proportional to the variance of the mean, declines in concentration with establishment size for employers with 20 or more employees will be primarily the result of the behavioral rather than the mechanical effect. Based on the modest decline in concentration over the 20-500+ range of size, we conclude that size has a small negative effect on concentration.

In Figure 5 we illustrate concentration patterns for single- and multi-unit firms by firm age. There is little difference in concentration by firm age for single units, and looking at younger firms, little difference between multi-units and single-units. What stands out is the category of multi-units that have been in existence for at least five years. The immigrant share of employment is somewhat lower for these long-established multi-units, and the level of immigrant concentration is only half what it is for younger firms and older single units.

We know from Table 3 that even with the main-effects specification, detailed industry had a relatively large role in explaining concentration, reflecting substantial immigrant/native differences in the kinds of businesses they work for. When we include interactions between industry and immigrant status, we also find systematic differences across industries in within-industry concentration. Since it is impractical to illustrate differences across the full set of detailed industries, Figure 6 gives coworker shares and concentration rates for each of the detailed industries that account for at least one per-

[27]See section C of the web appendix.

cent of employment in our sample. The industries are ordered by the level of concentration, with electrical equipment and machinery (not elsewhere classified) having the greatest concentration and also the greatest coworker share for immigrants.[28] While industries with high coworker share for immigrants often also have high levels of concentration, that is not always the case. For example, nursing facilities and hotels have similar predicted immigrant coworker shares, but quite difference levels of concentration. Immigrants and natives are much more likely to work together in the same hotels than they are to work in the same nursing facilities.

In Table 5 we report main and interaction coefficients for continuous variables of particular interest. The residential segregation index has the expected pattern: those who live with immigrants are also more likely to work with them. For both natives and immigrants, a higher share of immigrants in a worker's residential tract is associated with a higher share of coworkers who are immigrants, but the effect is somewhat larger for immigrants. For earnings, we find that concentration falls as we move up the distribution: high-earnings natives are more likely to work with immigrants than lower earnings natives, while high-earnings immigrants are less likely to work with other immigrants. We also find that our neighborhood network index is positively associated with concentration: natives who work with their neighbors have fewer immigrant coworkers, while immigrants who work with their neighbors have more. Thus workers who seem more likely to have found their jobs through neighborhood ties are also more likely to work with others of their own type. This pattern holds even with controls for employer, employee, and residential characteristics—in particular, even with controls for the share of immigrants living in their residential tract.

[28]Elementary and secondary schools actually have negative concentration—natives and immigrants with mean characteristics are more likely to work together than would be expected with random allocation. Our employer identifiers for public schools generally identify school districts rather than individual schools, so school workplaces may have higher levels of concentration than this indicates.

5.2 Using country of origin to further explore the roles of language and social networks

To help us understand the possible mechanisms underlying the network and language results, Table 6 gives coefficients on network and English language skill measures for our top nine immigrant countries of origin. Each row in the table presents estimates from a separate regression with the share of coworkers from the indicated country as the dependent variable. The own effects here are again consistently large: for immigrants, working with neighbors is highly correlated with working with compatriots even after controlling for the share of neighbors who are from the same country. The "main" effect given in the first column gives the effect for natives of the network index on share of coworkers from the designated country, while the "other" column gives differences between the effect for natives and the effect for immigrants from other countries. These coefficients are consistently small, indicating that working with neighbors has little relationship with the probability of working with immigrants from countries other than their own country.

With the exception of the results for Mexico, the effects of language skills also occur primarily within country-of-origin group. Not speaking English well is associated with a higher probability of working with compatriots, but little association with the probability of working with immigrants from most countries for natives or other immigrants.[29] The own-country effects of language are largest for immigrants from Asian countries in our sample, particularly Japan and Korea.

In the results for Mexico, the main effect for not speaking English well is large relative to main effects for other countries. In this specification, the main effect gives the effect for natives who do not speak English well, and speak a language other than English at

[29]Note that these estimates condition on the share of neighborhood residents from these nine countries of origin and the share coming from all other non-U.S. countries.

home. Most members of this group speak Spanish at home and almost one-third report they are of Mexican-American origin, both factors that might account for the effect we find here.[30] Note that, combining main and interaction effects, the implied effect of not speaking English well for immigrants from Mexico (.021+.009=.03) is within the range of implied effects for the own group in regressions for other countries. Similarly, while the "other" interaction has a relatively large negative coefficient in the row for Mexico, it is offset by the main effect.[31]

5.3 Contribution of within-cell differences to concentration

To help quantify the significance of the differences in effects discussed here, we return to the specifications for all immigrants and pursue two additional exercises. First, we apply the Oaxaca-like Gelbach decomposition given in (3.6) which identifies the contribution of differences in coefficients as well as differences in means. Second, since the decomposition only provides insights about the contribution of covariates and coefficients evaluated at the specified mean, we also illustrate how variables of interest contribute to the variation across workers by presenting predicted values for particular points in the distributions of key characteristics.

[30] Almost 90% of these natives speak Spanish at home. Roughly 40% report they are of Puerto Rican origin, another 30% report they are of Mexican American, 20% are other Spanish speakers, and the remainder report other languages (more often European than Asian).

[31] In results which we do not include here, we also examined whether the network and language effects are stronger within immigrant groups that speak Spanish. We did this by rerunning the two regressions with share of coworkers from Mexico and from El Salvador as the dependent variable. In constructing controls we split up the other immigrant group into immigrants from Spanish speaking countries (including countries with primarily Spanish speaking populations that are not in this table) and those from countries speaking other languages. The results gave little support to the hypothesis that network effects are stronger within groups defined by a shared language. The language effects for immigrants from Spanish-speaking countries were only slightly larger than the effects for natives, but recall that the natives who do not speak English well primarily speak Spanish. When we break up other-immigrant effects into country-specific effects for each of our nine countries, we find that immigrants from Vietnam who do not speak English well are more likely to work with immigrants from China than with natives or immigrants from other countries (with a similar cross-effect for Chinese immigrants). This appears primarily due to the 12% of immigrants from Vietnam who speak Chinese as their first language, who have a relatively high probability of working with immigrants from China.

Table 7 presents the results from the extended Gelbach decomposition. As mentioned earlier, with interaction terms the decomposition varies depending on where in the distribution of characteristics the effects are evaluated: the results with interaction terms yield residual concentration of 0.116 evaluated at the immigrant mean values, while the same set of controls and interactions evaluated at the native means yields a residual concentration measure of 0.069. The split between contributions from differences in means and differences in coefficients also depends on where in the distribution these effects are measured. Evaluated at immigrant means, the contribution of coefficients is by definition zero leaving all the contribution due to characteristics. Evaluated at native means, the contribution of coefficients is much larger.[32]

Given this sensitivity, the more robust statements from this decomposition are about the combined contribution of characteristics and coefficients since the sum of the contribution of both is reasonably similar regardless of where the terms are evaluated. Viewed from that perspective, we see that characteristics and coefficients like language, industry and residential segregation make the greatest contributions. In that respect, the results for Table 7 largely reinforce the insights from Table 3. But the contribution of language in accounting for concentration is somewhat lower based on the fully interacted model. Residential segregation accounts for a slightly greater decrease in the fully interacted model, though its share falls somewhat. Industry's contribution also rises in level while maintaining its share. Firm age and multi-unit status contribute substantially more in this version of the decomposition, because the interaction terms allow for the differences in concentration between well-established multi-units and other firms seen in Figure 5.

When we do the decomposition using the mean for natives, the decomposition in

[32]The contribution ϕ_I^k is multiplied by α_I^k, the difference in the mean of x^k between immigrants and natives. If natives have the same mean as immigrants for x^k, the contribution of the coefficients is zero as the coefficients are multiplied by zero. The web appendix presents some additional estimates to illustrate that, while there are several ways to carry out the decomposition, our conclusions from this decomposition are robust to the various alternatives. See Tables W-11 and W-12.

equation (3.6) becomes:

$$\delta_I^* = \sum_{i=1}^{K} \beta^k * (\bar{X}_I^k - \bar{X}_N^k) + \phi_I^k * (\bar{X}_I^k - \bar{X}_N^k) \tag{5.1}$$

Some of the patterns in Table 7 might seem surprising given results reported earlier in this section. For example, Figure 4 shows that the effect of employer size on immigrant concentration is quite different for immigrants and natives. Likewise, Table 5 shows that the effects of the network index are much more positive for immigrants than for natives. Given the large differences in coefficients, it might seem surprising that these effects contribute little in Table 7. But in the decomposition the effects attributed to differences in coefficients depend critically on where these effects are evaluated – and for both of these variables there is not much difference between native and immigrant means. And neighbors are a small fraction of coworkers; for example, fewer than 1 out of 10 workers has 4% or more of their coworkers living in the same Census tract. Even though the effect of social networks on immigrant concentration is estimated to be large, the low level implied by our measure leaves little room for this factor to account for much of the variation in coworker shares.

While many of these variables do not contribute to differences at the mean, several are important in explaining variation in immigrant concentration over the distribution of these characteristics. To illustrate this, Table 8 shows predicted concentration changes across the values of key covariates. Mean log earnings are quite similar for immigrants and natives, and in both decompositions earnings made essentially zero contribution to explaining differences in mean concentration. But Table 8 shows that concentration is 20 to 30% lower at the 90th percentile of the earnings distribution than at the 10th percentile, holding other variables constant. Our network variable also makes essentially no contribution to explaining differences at the mean, but concentration is substantially lower for those who do not work with anyone who lives in their tract (the majority of the sample), than for those at the 90th percentile of the network index.

6 Concluding Remarks

Using matched employer-employee data that comprehensively cover employment in our sample of MSAs, we find that immigrants are much more likely to work with each other—and hence less likely to work with natives—than would be expected given random allocation of workers. This is in part driven by the distribution of immigrants across MSAs, but within MSAs substantial concentration remains. We document that immigrant concentration is greatest in small firms, and varies substantially across industries. We find evidence that immigrant social networks, poor English language skills and living in neighborhoods with many other immigrants are each significantly associated with greater workplace concentration of immigrants. Immigrants who work together are quite likely to be compatriots; this is particularly true for immigrants who have poor English language skills.

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

We find that roughly half of immigrant concentration cannot be explained by our set of observable worker, employer and location characteristics. Of the half that can be explained, 20%, 40% and 40% can be explained by worker, employer and locational characteristics respectively. The most important worker characteristic associated with immigrant concentration is language proficiency; the most important employer characteristic is detailed industry; and the most important locational factor is the residential concentration of immigrants in census tracts. Our measure of social networks is highly correlated with immigrant concentration - an immigrant who works in a firm that hires many workers from the same Census tract has many more immigrant coworkers. But this measure of social networks does not account for much of the observed immigrant concentration. Even among workers who have neighbors as coworkers, those neighbors are a small fraction of their coworkers. While the estimated effect of social networks on immigrant concentration is large, the low level implied by our measure leaves little room for this factor to account for much of the average difference in immigrant coworker shares between immigrants and natives.

References

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

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

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

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

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

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

——— (1994): "The Economics of Immigration," *Journal of Labor Economics*, 32(4), 1667–1717.

BORJAS, G. J., R. B. FREEMAN, AND L. F. KATZ (1996): "Searching for the Effect of Immigration on the Labor Market," *American Economic Review*, 86, 246–251.

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

CARD, D. (2001): "Immigrant Inflows, Native Outflows and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19, 22–64.

CARRINGTON, W. J., AND K. TROSKE (1997): "On Measuring Segregation in Samples With Small Units," *Journal of Business and Economic Statistics*, 15(4), 402–409.

CARRINGTON, W. J., AND K. R. TROSKE (1998): "Sex Segregation in U.S. Manufacturing," *Industrial and Labor Relations Review*, 51(3), 445–464.

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

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

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

GÁRCIA-PÉREZ, M. (2009): "Does It Matter Who You Work for and Who You Work With? The Impact of Owners and Coworkers on Hiring and Wages," mimeo, St. Cloud State University.

GELBACH, J. B. (2009): "When Do Covariates Matter? And Which Ones, and How Much?," Working Paper 09-07, University of Arizona Department of Economics.

HELLERSTEIN, J. K., AND D. NEUMARK (2008): "Workplace Segregation in the United States: Race, Ethnicity and Skill," *Review of Economics and Statistics*, 90(3), 459–477.

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

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

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

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

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

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

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

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

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

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

Table 1: Sample characteristics

		Immigrants	Natives
Coworker share		37.2	14.3
Worker age	Age<30	24.0	33.0
	30< Age <40	34.0	26.1
	Age>40	42.1	40.9
Male		55.6	51.2
Age at arrival (*)	<= 12	12.5	
	13-25	47.4	
	26-35	27.2	
	36+	12.9	
Education	High school drop-out	32.6	17.0
	High school graduate	18.6	25.3
	Some college	17.1	25.8
	Bachelor's degree	21.7	24.1
	Advanced degree	10.0	7.8
Does not speak English well		21.1	0.9
Log quarterly earnings on primary job		8.4	8.3
Employed by Q2 employer in	Q1 and Q3	68.4	64.4
	Q1 or Q3	24.9	27.1
	Neither Q1 nor Q3	6.7	8.4
Establishment size	2-9 employees	8.9	7.8
	10-49	22.5	23.2
	50-99	13.2	13.4
	100-499	30.6	29.5
	500 or more	24.8	26.1
Firm has multiple establishments		34.6	43.3
Establishment age	<=1 years	11.8	11.5
	2-4 years	22.9	24.4
	5 or more years	65.3	64.1
Sector	Construction	5.5	6.0
	Manufacturing	20.2	12.4
	Transportation/utilities	3.5	4.9
	Wholesale	6.6	6.1
	Retail	20.0	23.1
	FIRE	4.7	6.5
	Services	39.5	41.1
Immigrant share of workers in residence tract		36.7	14.8
Shared commute index		0.3	0.5
Neighborhood network index		1.9	1.9

Notes: The unit of observation is a worker. N=2,965,225 natives and 600,761 immigrants. All figures except log earnings represent percentages. The estimates are weighted using our propensity score weights based on the probability of matching. (*) Year of application for a Social Security Number is used as a proxy for time of arrival in the U.S.

Table 2: Regression model with main effects

Covariate		Coefficient	Std Error
Immigrant		0.0960	0.0007
Education	High school drop-out	0.0142	0.0004
	High school graduate	0.0021	0.0002
	Bachelor's degree	0.0045	0.0003
	Advanced degree	0.0094	0.0006
Does not speak English well		0.0774	0.0010
Continuity on 2000-Q2 job	Worked Q1	0.0026	0.0006
	Worked Q3	0.0021	0.0005
	Worked Q1 and Q3	0.0024	0.0006
Log quarterly earnings on primary job		0.0017	0.0002
Worker age	Age<30	-0.0045	0.0003
	30<= Age <40	-0.0016	0.0002
Female		0.0014	0.0002
Employer size	2-4 employees	0.0226	0.0018
	5-9	0.0056	0.0016
	10-19	-0.0060	0.0015
	20-49	-0.0077	0.0015
	50-99	-0.0037	0.0015
	100-499	0.0043	0.0015
Firm has more than 1 establishment		-0.0304	0.0012
Establishment age	<= 1 year	0.0008	0.0017
	2-4 years	0.0025	0.0014
Firm has >1 estab * Estab age	<=1 year	-0.0025	0.0024
	2-4 years	0.0007	0.0018
Immigrant share of workers in residential tract		0.1902	0.0021
Neighborhood network index		0.0447	0.0036
Shared commute index		-0.4248	0.0106

Controls include MSA and detailed industry in addition to the variables listed in the table. The unit of observation is a worker. N=3,549,111. Estimation of standard errors accounts for correlation between error terms for workers employed at the same establishment.

Table 3: Contribution of Covariates to Immigrant Concentration

Mean immigrant-native difference in model with:	
1. MSA dummies only	0.171
2. Full set of controls	0.096
Contribution to reduction in coefficient	**Percents**
Individual characteristics (total)	23.9
Log earnings	0.2
Quarters of work	0.1
Age and sex	0.4
Language	20.4
Education	2.8
Employer characteristics (total)	35.8
Firm size	0.4
Firm age and multi-unit status (interacted)	3.3
Industry	32.1
Sector	14.1
Sum of within sector detail	17.9
Manufacturing detail (73 3-digit industries)	3.7
Transportation, communications, utilities (14 inds)	1.2
Wholesale (18 industries)	0.8
Retail (33 industries)	0.5
FIRE (4 industries)	1.0
Services (51 industries)	10.8
Neighborhood characteristics (total)	40.3
Immigrant share of workers living in residential tract	39.7
Neighborhood network index	0.2
Shared commute index	0.4

Notes: Figures in the first two rows give the predicted difference in mean coworker share between immigrants and natives. The rows in the bottom panel of the table give the percentage of the difference in coefficients between rows 1 and 2 accounted for by that particular set of controls.

Table 4: Concentration by Country-of-Birth

	MSA + country dummies		Full specification	
	Own country	**Other country**	**Own country**	**Other country**
Cuba	0.167	0.066	0.094	-0.000
	(0.008)	(0.003)	(0.005)	(0.002)
El Salvador	0.063	0.148	0.045	0.048
	(0.001)	(0.002)	(0.001)	(0.001)
Mexico	0.157	0.021	0.090	-0.013
	(0.001)	(0.001)	(0.001)	(0.001)
China	0.200	0.139	0.165	0.063
	(0.005)	(0.003)	(0.004)	(0.003)
India	0.155	0.054	0.135	0.024
	(0.005)	(0.002)	(0.004)	(0.002)
Japan	0.140	0.026	0.136	0.002
	(0.005)	(0.003)	(0.005)	(0.002)
Korea	0.188	0.047	0.178	-0.007
	(0.005)	(0.003)	(0.005)	(0.003)
Philippines	0.095	0.050	0.076	0.022
	(0.002)	(0.001)	(0.001)	(0.001)
Vietnam	0.181	0.086	0.155	0.015
	(0.003)	(0.002)	(0.003)	(0.002)

Notes: Standard errors appear directly below coefficient estimates. The own-country effects are estimates of the coefficient on the relevant country dummy from regressions with dependent variable = country-specific coworker share variable. It gives the excess probability, relative to natives, of working with compatriots. The other-immigrant estimates are estimates of the coefficient on that country's dummy from regressions with dependent variable = immigrant coworker share excluding that country of origin. It gives the excess probability, relative to natives, of working with immigrants who are not compatriots. All regressions include MSA dummies, dummy variables for these 9 countries of origin, plus an additional dummy for all other countries of origin excluding the U.S. The full specification additionally includes controls for industry, establishment size, firm age and multi-unit status, worker age, sex, log earnings, quarters of work, neighbor network index, shared commute index for natives and immigrants, education, English language skill, and the immigrant shares in a worker's residential tract accounted for by immigrants from each of these 9 countries plus the share for all other foreign countries of origin.

Table 5: Selected Coefficients from Fully Interacted Model

Covariates	Coefficients	Standard Errors
Immigrant share in residential tract	0.161	0.0018
Immigrant residential share * Immigrant	0.052	0.0028
Log earnings	0.003	0.0002
Log earnings * Immigrant	-0.008	0.0004
Neighborhood network index	-0.072	0.0026
Network index * Immigrant	0.436	0.0088

Note: In addition to the variables listed in the table, controls include main effects and interactions with the immigrant dummy for MSA, detailed industry, establishment size, firm age and multi-unit status, worker age, sex, log earnings, quarters of work, shared commute index for natives and immigrants, education, and the immigrant share in a worker's residential tract. The unit of observation is a worker. N=3,565,986. Robust standard errors allow for correlated errors among employees of the same establishment.

Table 6: Network and Language Effects on Concentration by Country-of-Birth

	Neighborhood index			Does not speak English well		
	Main	**Own**	**Other**	**Main**	**Own**	**Other**
Cuba	-0.001	0.364	-0.004	0.002	0.055	-0.000
	(0.001)	(0.033)	(0.002)	(0.000)	(0.004)	(0.000)
El Salvador	-0.004	0.486	-0.005	0.002	0.010	-0.000
	(0.000)	(0.045)	(0.001)	(0.000)	(0.002)	(0.000)
Mexico	-0.020	0.466	-0.008	0.018	0.013	-0.015
	(0.001)	(0.017)	(0.003)	(0.001)	(0.001)	(0.001)
China	-0.002	0.366	0.008	0.000	0.065	0.004
	(0.000)	(0.033)	(0.002)	(0.000)	(0.006)	(0.000)
India	-0.002	0.584	0.003	0.000	0.065	-0.001
	(0.000)	(0.031)	(0.002)	(0.000)	(0.008)	(0.000)
Japan	-0.001	0.325	-0.002	0.000	0.110	-0.001
	(0.000)	(0.073)	(0.001)	(0.000)	(0.011)	(0.000)
Korea	-0.002	0.223	-0.006	0.001	0.080	0.000
	(0.000)	(0.044)	(0.001)	(0.000)	(0.008)	(0.000)
Philippines	-0.003	0.617	-0.002	0.000	0.025	-0.002
	(0.001)	(0.042)	(0.001)	(0.000)	(0.004)	(0.000)
Vietnam	-0.003	0.556	-0.001	0.001	0.066	0.002
	(0.000)	(0.032)	(0.002)	(0.000)	(0.004)	(0.000)

Notes: Standard errors appear directly below coefficient estimates. Each line in the table presents estimates from a separate regression with the share of coworkers from the indicated country as the dependent variable. The specification also includes main effects and own/other interactions for MSA, detailed industry, establishment size, firm age and multi-unit status, worker age, sex, log earnings, quarters of work, shared commute index for natives and immigrants, education, and the immigrant shares in a worker's residential tract accounted for by immigrants from each of these 9 countries plus the share for all other foreign countries of origin. The Main column gives the cofficient on the indicated variable. The Own column gives the coefficient on that variable interacted with a dummy for that row's country of origin, and the Other column gives the coefficient on that variable interacted with a dummy for immigrants from other countries of origin.

Table 7: Decomposition of predicted raw and conditional concentration measures

Mean immigrant-native difference, evaluated at native means, in model with:		
1. MSA dummies only	0.171	
2. Full set of controls with interactions	0.069	
Contribution to reduction in difference between rows 1. and 2. (percents)	Differences in Xs	Differences in Coefficients
MSA (main effects in base, but interactions not)		8.0
Individual characteristics (total)	9.8	5.9
Log earnings	0.3	-0.8
Quarters of work	-0.1	0.9
Age and sex	-0.3	1.7
Language	7.6	5.4
Education	2.4	-1.2
Employer characteristics (total)	22.8	25.3
Firm size	0.1	1.0
Firm age and multi-unit status (interacted)	1.4	5.4
Industry	21.3	10.9
Sector	9.8	6.3
Sum of within sector detail	11.5	4.6
Manufacturing detail (73 3-digit industries)	0.9	-2.8
Transportation, communications, utilities (14 inds)	0.1	0.2
Wholesale (18 industries)	0.5	0.4
Retail (33 industries)	2.0	2.2
FIRE (4 industries)	-0.1	-0.0
Services (51 industries)	8.1	4.7
Neighborhood characteristics (total)	24.6	11.5
Immigrant share of workers living in residential tract	24.7	10.5
Neighborhood network index	-0.2	0.1
Shared commute index	0.2	0.8
Total column share	57.3	42.7

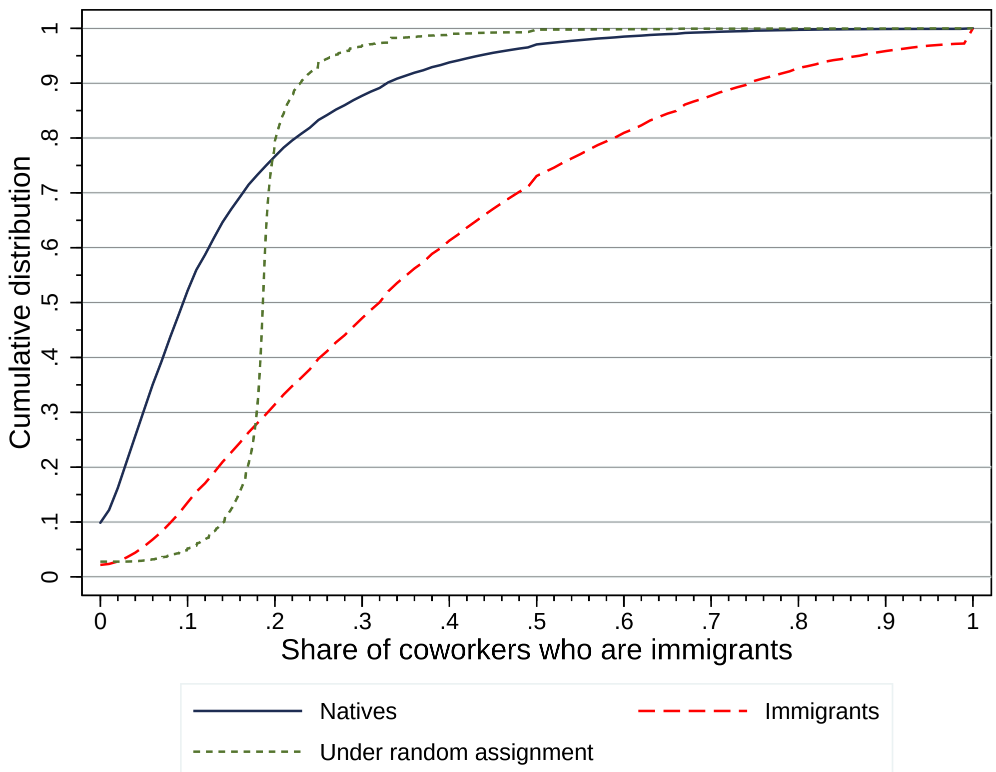
Notes: Regressions use controls deviated from the native mean. The differences in Xs are evaluated using the coefficients for natives, while differences in coefficients are evaluated using immigrant Xs deviated from the native mean.

Table 8: Predicted concentration for selected values of covariates

	Pooled mean	Immigrant mean	Native mean
Predicted concentration at mean for all variables	0.078	0.116	0.069
Holding other Xs at mean, prediction for:			
Log earnings			
10th percentile	0.092	0.131	0.083
90th percentile	0.067	0.106	0.058
English			
Speaks well	0.077	0.110	0.069
Does not speak well	0.104	0.137	0.096
Residential segregation			
10th percentile	0.069	0.098	0.069
90th percentile	0.093	0.122	0.087
Network index			
Index=0	0.070	0.108	0.061
90th percentile	0.086	0.123	0.077
Firm size			
Small estab (2-4 employees)	0.234	0.270	0.225
Small estab (10-19 employees)	0.127	0.164	0.119
Large estab (>=500 employees)	0.045	0.082	0.037

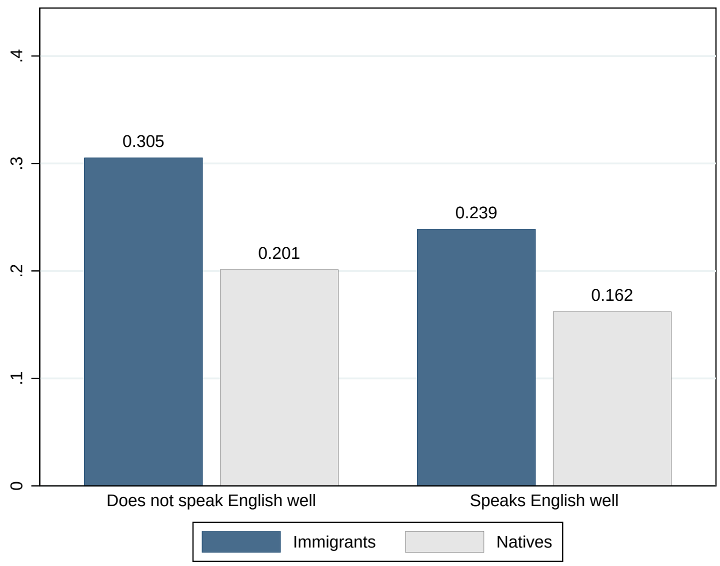
Note: Estimates give predicted immigrant/native difference in coworker mean based on fully interacted model. Predictions based on mean value specified in column heading, except setting the variable in the left column to the indicated value. Predicted concentration equals the prediction with all Xs at the mean (1st row) plus interaction coefficient(s) times difference between the indicated point in the X distribution and the mean.

Figure 1: Cumulative Distribution of Coworker Share for Natives and Immigrants



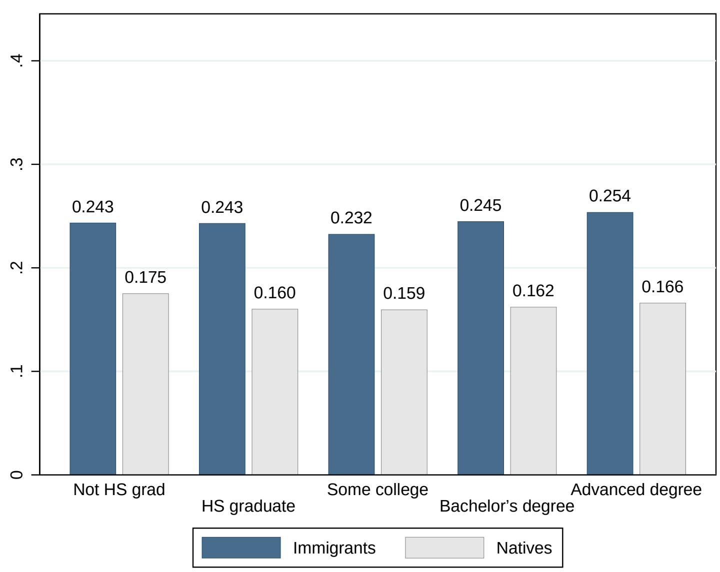
Note: The CDF under random assignment is constructed by first simulating the distribution of coworker shares conditional on employer size S by drawing 4,000 binomial random variates for S trials with p=.187 (share immigrant in our sample), and then using the number of immigrants (=number of successes in S trials) to calculate coworker shares. We simulate the distribution for each value of employer size from S=2 to 2,000. The distribution of employers becomes thinner as S increases, while the distribution of coworker shares changes little as S increases for large S. So for employer sizes above 2,000, we group employers into size ranges–using intervals of 200 for employer sizes 2,000-8,000, 1,000 for employer sizes 9,000-20,000, and 10,000 for employer sizes above that level. We then sum up the conditional probabilities for each coworker share across values of S using the empirical distribution of employer size as weights.

Figure 2: Coworker share by whether worker speaks English well



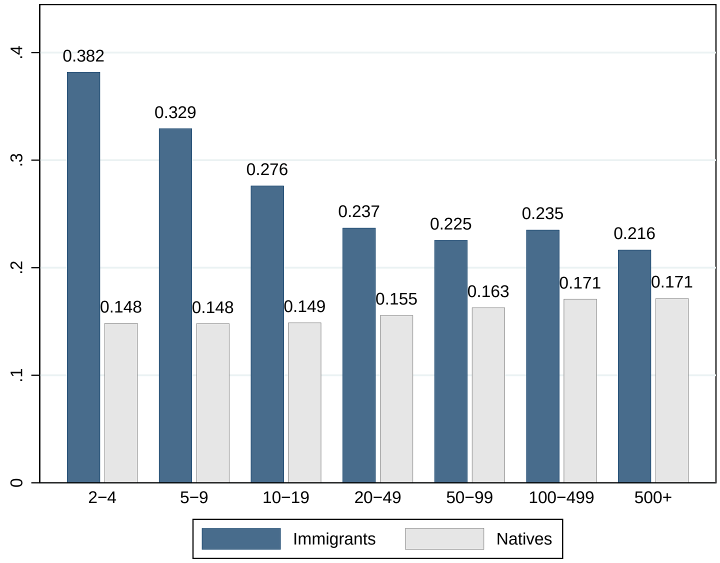
Note: Based on predictions in which all variables except language variable and immigrant status are set to pooled mean values. Model used for prediction includes interactions between the immigrant dummy variable and all other covariates.

Figure 3: Coworker share by education level



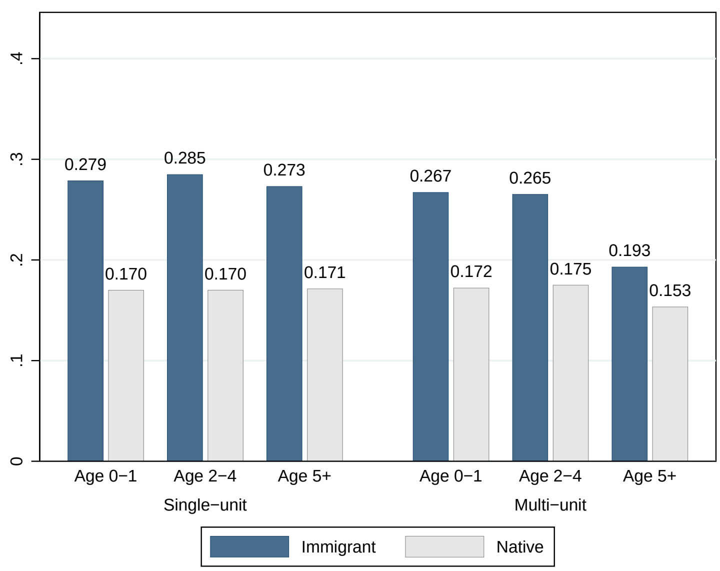
Note: Based on predictions in which all variables except education and immigrant status are set to pooled mean values. Model used for prediction includes interactions between the immigrant dummy variable and all other covariates.

Figure 4: Coworker share by employer size



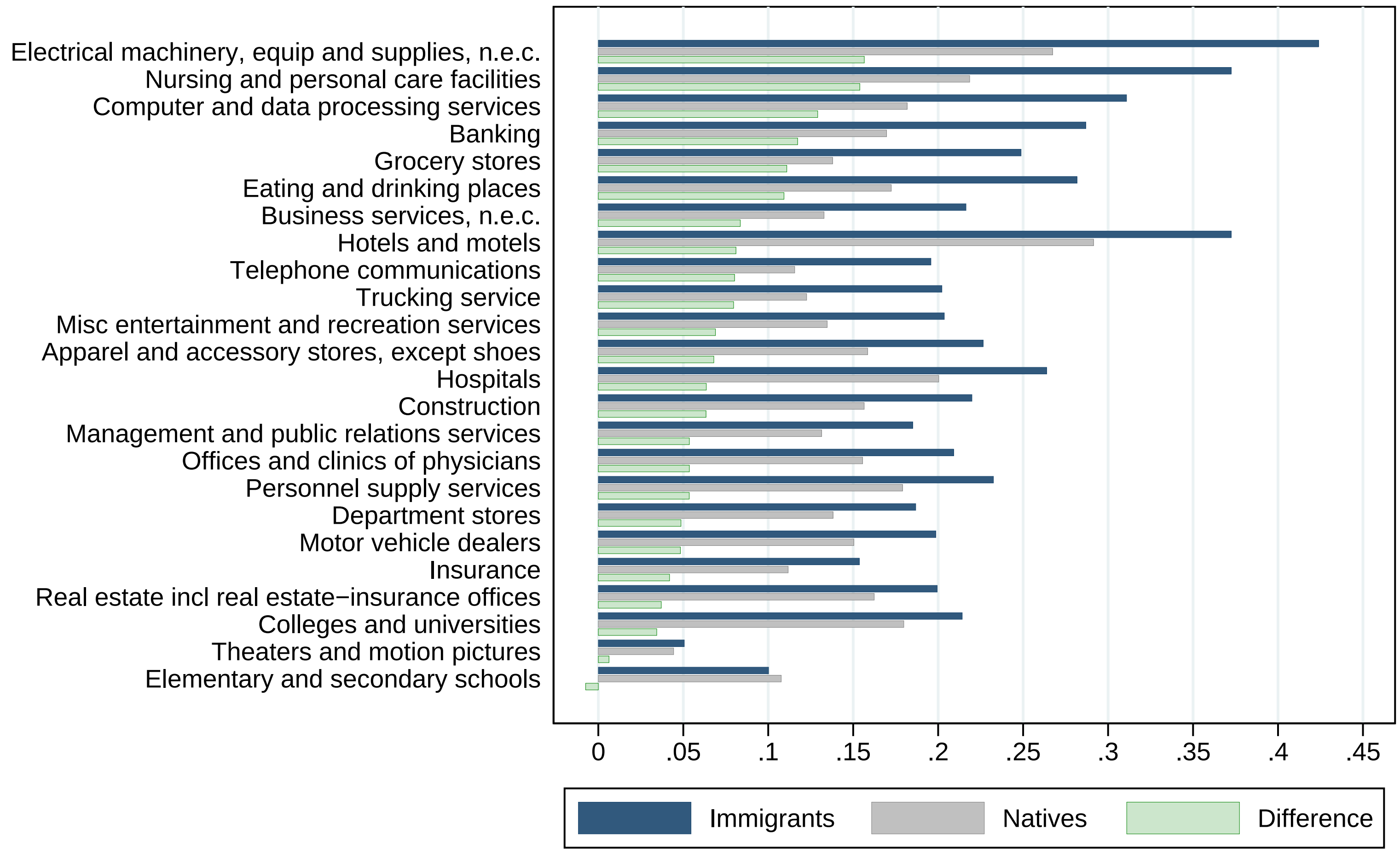
Note: Based on predictions in which all variables except employer size and immigrant status are set to pooled mean values. Model used for prediction includes interactions between the immigrant dummy variable and all other covariates.

Figure 5: Coworker share by firm type and age



Note: Based on predictions in which all variables except firm type, firm age, and immigrant status are set to pooled mean values. Model used for prediction includes interactions between the immigrant dummy variable and all other covariates.

Figure 6: Coworker shares for largest detailed industries



Note: Figure includes all 3-digit industries that accounted for at least 1% of sample employment. Based on predictions in which all variables except detailed industry and immigrant status are set to pooled mean values. Model used for prediction includes interactions between the immigrant dummy variable and all other covariates.