2011

Workplace Concentration of Immigrants

Monica Garcia-Perez
St Cloud State University, migarcia-perez@stcloudstate.edu

Workplace Concentration of Immigrants*

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

October 30, 2011

Abstract

To what extent do immigrants and the native-born work in separate workplaces? Do worker and employer 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 about 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.

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

*We thank the NIH for financial support and participants at the SOLE and WEA 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.

†Office of the Comptroller of the Currency; St. Cloud State University; University of Maryland; Census Bureau; and Duke University, respectively.

1 Introduction

Over the last several decades, labor markets in many U.S. cities have absorbed large inflows of new immigrants. While the earnings and employment of these new arrivals have been the subject of much research, less is known about the kinds of employers that hire them and the factors that determine the sorting of immigrants into workplaces. 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?

The difficulty of assembling suitable data has limited researchers' 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 to which immigrants are concentrated in separate workplaces and the contribution of worker, employer and location factors to this concentration. A novel aspect of our contribution is that our data are sufficiently comprehensive to permit characterizing concentration within the workplace by country of origin.

This paper has three broad and related objectives. The first is to document the extent to which immigrants and natives work for different employers. 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. Even after controlling for location, employer and employee characteristics, we find that employees are much more likely to work with their compatriots than would happen by chance. Salvadoran and Chinese immigrants are somewhat more likely to work with immigrants from other countries than natives are, while immigrants from most other source countries are not.

Our second objective is to identify factars that can account for the observed concentration. In doing so, we rely on theories of how and why employers hire specific types of workers to identify factors of interest. Workplace concentration may reflect sorting and matching on skills and other related factors. For example, language skills may affect workplace productivity through its effects on interactions among workers and between workers and customers. Residential segregation may also play a role if businesses hire through neighborhood social networks, or if proximity or the transportation infrastructure makes access to the business's location particularly easy from a specific neighborhood. 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, generating differences in immigrant concentration. We find that our set of characteristics can account for about half of observed concentration. Of that half, 37%, 30% and 33% can be explained by worker, employer, and locational characteristics respectively. The most important worker characteristic is language proficiency; the most important employer characteristic is industry; and the most important locational factor is the residential concentration of immigrants in local neighborhoods. The importance of these factors varies by country of origin. Not speaking well is, for example, especially relevant for concentration of immigrants from Asian countries.

Our third objective is to explore hypotheses about the factors that underlie patterns of concentration. Specifically, we investigate which observable characteristics have more important effects on coworker composition for immigrants (overall and by country of origin) than for natives. We find that the impact of employer size on the share of immigrant coworkers is much more important for immigrants than natives. We also find that the role of language proficiency primarily occurs within country of origin group. That is, not speaking well is associated with a higher probability of working with compatriots, but little association with the probability of working with other immigrants. We consider the implications of these findings for alternative hypotheses about the causes of concentration.

A core advantage of our work relative to the existing literature is our very rich and dense dataset on workplace concentration, and on worker, employer and locational characteristics. Our data cover virtually all private sector workers and employers in our sample of MSAs. This allows us to accurately measure workforce concentration for even the smallest private sector businesses and to examine differences by immigrant country of origin for a number of specific source countries. It also means that we can simultaneously control for a variety of characteristics when determining the relative importance of different factors. These advantages lead to new findings on patterns of concentration among the smallest businesses, and patterns of concentration by country of origin. In addition, we find that using a more extensive set of controls leads to important quantitative differences in the share of concentration we attribute to language relative to findings in related studies.

The paper proceeds as follows. Section 2 gives an overview of the theoretical and empirical literature that helps guide our analysis. Section 3 describes our data and methodology. In section 4 we present our main results quantifying the role of observable employer and employee characteristics in accounting for patterns of immigrant concentration. Section 5 presents analysis of how employer and employee characteristics may have differential effects on immigrant concentration for immigrants and natives. Concluding remarks are provided in section 6.[1]

[1]A web appendix contains supplementary analysis and results which we refer to several times in the paper.

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) based on productive characteristics of workers, (b) based on the information available to workers and employers, (c) resulting from the residential location of workers relative to business locations, and (d) resulting from preferences of workers and employers. Because we have no direct measures of tastes, our empirical analysis focuses on factors (a) through (c).

There is substantial evidence of segregation by skill. For example, Kremer and Maskin (1996) find a high and rising correlation between coworker skill levels in firms over the 1970s and 1980s in the U.S., Britain and France. A positive correlation in skills may occur either because a firm demands workers of a particular skill level or because coordination within a firm requires 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 depends on both absolute and relative wages and movement of workers is costless, complete segregation by skill is optimal. Skill-based sorting could lead to workplace segregation of immigrants from natives because the two groups differ in their distributions of skills. Immigrants are much more likely than natives to have an 8th grade education or less (23% vs 5.2% in the 2000 census), but also more likely to have an advanced degree (10.3% vs. 8.6%). Therefore, employers that hire exclusively low-skilled or exclusively high-skilled workers will tend to have above-average immigrant employment shares.

If a shared language increases worker productivity, employers may choose workforces in which everyone speaks the same language. If so, immigrants from non-English speaking countries will be particularly likely to be segregated, and may also be particularly likely to work with immigrants who speak their language. Lang (1986) develops a formal model of wage differences that arise because employers must pay a premium for bilingual workers who can bridge the language barrier. His model implies 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) (hereafter HN) find evidence that Hispanics with poor English-language skills are particularly likely to work with other Hispanics. Portes and Wilson (1980) examine whether segregation among Cuban immigrants in Miami occurs through employment by Cuban-owned firms as in the Lang model. They 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.

Information-based theories focus on mechanisms that match workers to jobs. For example, if outside of work people interact mostly with others who have similar characteristics, employer use of employee referrals and/or employee use of personal contacts to find jobs 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. 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 combined with a concentration of recent immigrants in small firms also contribute 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] Neighbors may provide important job contacts and references. Several papers have found that those working in the same place are disproportionately from the same neighborhoods. For example, Ellis, Wright, and Parks (2007) and Wright, Ellis, and Parks (2010) find strong links between the residential concentration of immigrant groups in Los Angeles and their concentration by workplace tract and industry. 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.

Hellerstein, Neumark, and McInerney (2008) (hereafter HNM) 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 their race/ethnicity 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 find particularly large effects for Hispanics with poor English language skills and Hispanics who are immigrants. We draw on this work for ways to capture 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 HNM's goal is to establish the importance of networks for labor markets more generally, so our results are not directly comparable to theirs.

Given Hellerstein, Neumark, and McInerney's extensive work in this area and their use of some of the same data sources, it is worth clarifying how our analysis differs from

[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.

their work. The core differences stem from our more extensive data on workers. HNM construct and use the Decennial Employer-Employee Database (DEED). Starting with the 1-in-6 Decennial long form, they use sophisticated alogorithims to match workers' write-in reports on place of work to employer addresses on Census's Business Register (a list of all employer establishments). Despite missing address information and differences between businesses and their employees in how addresses are reported, HNM match 29.1% of long form workers to their work location. This implies that the DEED contains about 1-in-20 workers in the U.S.

Like HNM, we base our analysis on workers in the long form sample who match to an employer list–in our case from the Longitudinal Employer-Household Dynamics (LEHD) database. We link workers to their employers based on administrative identifiers rather than name and address matching, resulting in match rates of about 1-in-10 LEHD workers from our sample of MSAs. But our data sources give another much more important advantage: additional information from administrative records on virtually all employees covered by unemployment insurance (UI) for our sample of MSAs. For any long-form worker matching to our employer list, this gives us complete information on the share of coworkers who are immigrants as well as their countries of origin, and information on other workers who live or work nearby.

The DEED's reliance on the long-form sample for all worker information has two important effects in this context. First, employees of small firms will be under-represented in the DEED. For both HNM's work and our own, at least two workers must be observed at an establishment for either worker to be included in the analysis. With LEHD information, this has no effect on a worker's probability of inclusion; an establishment must employ more than one worker for coworker characteristics to be defined, so all relevant employers satisfy this requirement. But with only the long-form sample, this requires that each matched worker must have at least one matched coworker. In an establishment with 15 employees, imposing this restriction reduces a worker's probability of inclusion from 1-in-20 to about 1-in-40. Workers at three-employee establishments end up with only a 1-in-205 chance of inclusion, while this requirement has almost no effect on the probability of inclusion for workers at establishments with 80 or more employees.[3]

Second, the DEED will have data only on a subset of workers within most observed establishments. Having a subset of workers leads to variation in estimates of worker characteristics. With a 1-in-20 sampling rate, only 43% of 50-employee establishments included in the DEED would have at least 5 matches. To see the effects of this sampling variation, consider an establishment with 50 workers, 25 of them immigrants; if only 4 workers out of 50 are matched, the probability of observing the actual immigrant share (=0.5) is only 37.5%; in fact, with 12.5% probability, establishments would have a measured immigrant share of either 0 or 1.

HNM recognize these issues. To account for them, they use an elegant simulation approach that compares observed segregation to what one would expect to observe in their sample if employers hired randomly, drawing on statistical methods developed in Carrington and Troske (1997).[4] If observed concentration is significantly larger than expected, this is taken as evidence of non-random hiring. They also carry out these simulations allowing hiring to be random within a limited number of strata. If within strata the observed and expected concentration are the same, 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 stratification variables is small.

Because the LEHD data give us the actual immigrant share among coworkers of

[3]With either method for matching workers to their employers, probabilities of inclusion will increase with employer size if the accuracy or completeness of information provided by respondents is greater for employees of larger businesses. We find a slight bias in this direction, and reweight to off set it.

[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.

each matched long-form worker, we have no need to take within-establishment sampling variation into account in our analysis. This allows us a more flexible approach, which in turn makes it possible to examine a wider set of characteristics and to examine interactions with immigrant status. Our results will show that controlling for many characteristics simultaneously matters in this context.[5] For example, we find that adding other controls reduces by about one-third the share of concentration that we would attribute to language differences.

We do want to be clear that random assignment of workers to establishments could potentially still contaminate measures of concentration. Even in the case of no systematic variation, the expected variation in coworker shares around the true mean due to pure randomness is significant in the case of small establishments. This would be of concern with our regression approach if our control variables defined cells with observations from only a few establishment, in which case variation in coworker shares due to randomness could falsely be interpreted as systematic interpretation. The size of our data set (about 735,000 employers and 3.5 million workers drawn from relatively large MSAs), however, makes small cell sizes easy to avoid, why the law of large numbers guarantees a that random assignment of workers to establishment will have a negligible impact on the estimates.[6]

[5]See Lengermann, McKinney, and Pedace (2004) for earlier work that also took advantage of the representative coverage of employers in the Longitudinal Employer-Household Dynamics (LEHD) database to explore immigrant concentration variation across employers. This earlier work found important differences in immigrant concentration across MSAs and employer size classes, but the focus of their analysis was to use these differences to explore differences in earnings for immigrants and natives.

[6]This issue is further investigated in Web Appendix. The results from this analysis confirms that any bias on measured concentration in our sample is only modest.

3 Methodology and Data

3.1 Data

As noted above, we construct a cross-sectional sample of workers in selected MSAs by combining data from the Longitudinal Employer-Household Dynamics (LEHD) database and the 2000 Decennial Census 1-in-6 long form. Given the April 1 reference date of the census, we focus on employment in the second quarter of 2000. The LEHD database draws much of its data from complete sets of unemployment insurance (UI) earnings records for a subset of U.S. 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.[7] Basic demographic data are also available for workers, including place of birth which allows us to identify immigrants. Geocoding of addresses for both employers and places of residence allows us to examine characteristics of both locations. The LEHD data have the important advantage of allowing us to measure employer and workforce characteristics using information on all employees of all UI-covered employers in the included states. 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.

The unit of observation for businesses in the LEHD data is the establishment. An establishment is a fixed physical location where production activity takes place. Each worker's quarterly UI wage record includes a UI account number that identifies his or her firm of employment within a state in a specific quarter. Some firms have multiple locations within a state and the LEHD data identify all of the separate locations (es-

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

tablishments) within each state. Workers who are employed by a multi-establishment firm within a state are assigned to specific establishments through multiple imputations based on a rich set of information including the location of the employing firm's establishments in that state, the residential location of the worker, and the employment histories of both the worker and the establishment.[8]

While we can measure selected coworker characteristics for virtually all workers in our UI earnings sample, we match to the 1-in-6 Decennial Census long form sample to obtain two additional variables that are likely to be important in this context: education and language proficiency. We use these variables to examine whether the less educated and those with limited English are more likely to work in predominantly immigrant workplaces. The outcome of the match to the long form sample is an approximately 1-in-10 subsample of UI workers in our MSAs.[9] Matched 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.[10] 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.[11] Using these weights, matched sample

[8] A full description of the methods used for assigning workers of multi-unit firms to specific establishments can be found at Abowd, Vilhuber, McKinney, Sandusky, Stephens, Andersson, Roemer, and Woodcock (2006). We note that these methods have been developed and validated using data in which the employing establishment is identified for workers of multi-establishment firms.

[9] Not all long form respondents can be matched to administrative data, either because the information needed for matching is missing or because no match can be found. We match about 70% of long-form respondents who live in our sample of MSAs and report working for a non-agricultural private sector employer or a state or local government.

[10] Appendix Tables A.1 and A.2 give comparisons of the full and matched sample characteristics.

[11] 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.

results closely replicate regression results based on the complete UI earnings sample.[12]

We limit the matched sample to workers employed in 31 selected metropolitan areas in 11 states, with our choice of areas 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.[13] 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. We have more than 3,000 immigrant workers who match to long form census data even in the smallest of our MSAs.

The average immigrant workforce share across our 31 MSAs is 19% but immigrants are less than 11% of the workforce in eight MSAs, while they are are more than 35% of the workforce in three MSAs. 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

[12]Comparing the first column in Tables A.1 and A.2 to Table 1 illustrates the close correspondence between means for the full sample and those for the weighted matched sample.

[13]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.

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 significant positive value indicates that immigrants are more concentrated than would be expected based on random allocation. At the extreme, if immigrants worked only with immigrants and natives with natives, the difference in coworker means would equal one. A significant 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.[14] 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. In our sample of immigrant-rich

[14] In our sample of MSAs, immigrants account for 27% of employment in single-employee businesses.

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%. Our analysis focuses on the mean difference in coworker shares between immigrants and natives, which is close to the median difference illustrated here. 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.

Clearly the observed distributions are inconsistent with random assignment. Because the likelihood of extreme values occurring randomly is quite low in large 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. The share with only native coworkers would 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.

3.2 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 in our set of MSAs, 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.

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. 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 ″Speaks English very well″ consists of those who report that they speak English ″very well″ along with those who speak only English at home (in which case they were not asked how well they speak it). Unsurprisingly, immigrants are more likely than natives to fall into categories other than ″very well″, but note that even the category ″Not at all″ includes some natives.[15] 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. These jobs may also be associated with relatively low wage rates and part-time work.

We find only minor differences between immigrants and natives in broadly defined employer characteristics. Immigrants are more likely to work in the smallest establishments, and less likely to work in the largest, but overall the differences by employer size are small, as are differences by establishment age. 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 broad sector are not particularly large.[16]

[15]Almost 90% of natives in the poorly or not at all categories speak Spanish at home. Roughly 40% report Puerto Rican origin, another 30% report Mexican American origin, 20% are other Spanish speakers, and the remainder report other languages (more often European than Asian).

[16]Our estimates of employment distributions across sectors differ from published estimates from the

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.[17] Because we only have data on those who work, we base these variables on workers residing in a particular tract rather than all residents of the tract.

The first measure is simply the share of immigrants in a worker's tract of residence, which we use to control for residential segregation. Neighbors act as contacts and references for job opportunities so concentration of immigrants in the neighborhood can contribute to immigrant concentration in the workplace. As can be seen in Table 1, immigrants in our sample of MSAs are substantially 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 who also live in the worker's residential tract. The denominator is the number of 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.[18] Proximity or convenient transportation links may make residents of certain neighborhoods likely to work at a particular location, resulting

2000 population census for several reasons: we include only a subset of MSAs; we exclude agriculture and use sectors 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 comparison 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).

[17]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.

[18]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.

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 that influence concentration. We refer to this as our 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 a specific type of neighborhood-based job network. Neighborhood contacts and references may make neighbors being more likely to be coworkers. 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 overall, the mean is substantially higher for workers employed in small businesses and falls systematically with employer size.

3.3 Regression specifications

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 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.

As a basis for comparison, our simplest 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.

We next add a vector of worker and employer characteristics x_{ij}:

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

Comparing results from (3.3) to (3.2) allows us to address our first question: Which characteristics of workers and employers are important in accounting for immigrant concentration? 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:

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

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) within MSA, there must be a large average difference between immigrants and natives in x^k (α_I^k is large).

4 Accounting for immigrant concentration

Here we quantify the extent to which observable employer and employee characteristics can account for patterns of concentration, looking first at concentration of all immigrants and then turning to concentration of immigrants from specific countries of origin.

4.1 Basic results

Using (3.2) as our starting point, average within-MSA concentration (γ_N^{base}) is 0.171; that is, the average share of immigrant coworkers is 17.1 percentage points more for immigrants than for natives working in the same MSA, as compared to the overall concentration in our sample of 22.9 percentage points reported in Table 1 not accounting for MSA effects.

Table 2 presents coefficient estimates for specification (3.3). Controlling for observable employee and employer characteristics reduces estimated concentration from 0.171 (γ_N^{base}) to 0.083 (γ_I^{main})—a roughly 50% reduction. The top panel of Table 3 reports those estimates, while the bottom panel breaks out the share of that reduction accounted for by particular types of characteristics. Three factors stand out as important: English language skills, industry of employment, and the share of a worker's neighbors who are immigrants. Together these account for 94% of explained concentration (and 48% of total concentration), with the next runners up (education and the interaction of firm age with multi-unit status) contributing less than 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 associated with reduced English proficiency even when controlling for numerous other factors. Given the large share of U.S. immigrants of Hispanic origin, it is worth comparing our findings to HN's findings on the importance of language for Hispanic/white concentration. Using the same language grouping (but no other controls), HN find that about one-third of all Hispanic/white within-MSA concentration is attributable to segregation by language. In our sample, if we include only language (and MSA) controls, language explains 28% of overall immigrant concentration. But using the broader set of controls given in Table 2, we attribute about 18% of overall concentration to language.[19] While in both cases we find that language is important, looking at language separately from other factors produces results that overstate its importance.[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

[19]Table 3 shows that language accounts for 35% of *explained* concentration but this in turn implies language accounts for 18% of *overall* concentration.

[20]We note that HN in other parts of their study explore factors such as education and employer size but not at the same time as exploring the role of language. Their methodological approach is not well suited to controlling for many factors simultaneously.

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 one-third contribution of residential segregation across Census tracts within MSAs. This points to a very strong relationship between living with immigrants and working with them. Note that neither of the other tract-level variables (the network index and the shared commute index) accounts for much of the concentration. The coefficient in Table 2 indicates that network effects have a positive and statistically significant effect, which is consistent with the hypothesis that network effects increase the likelihood of working with immigrants. However, there is not much difference between immigrants and natives in the mean value ot the network index, as seen by Table 1. Because of this, the network variable cannot account for much immigrant concentration.

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 concentration by country of origin. That is, we can estimate how likely it is for an immigrant from Mexico (for example) 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]

[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

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. In place of the residential segregation measure in Table 2, we use 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 immigrants, 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 both 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 exception is Salvadorans, who, relative to natives, are roughly twice as likely to work with immigrants from other countries as with other Salvadorans. Based on results that we do not present here, 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,

we have rather than differences in composition between the overall population and employees.

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 countries of origin these factors fully explain the excess tendency to work with immigrants from other countries. 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; even for these two countries, covariates explain more than two-thirds of the excess non-compatriot concentration.

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.[23] However, the importance of these factors differs for own- versus other-country concentration, and varies across country groups. Residential segregation accounts for virtually all (92%) of the explained variation in own-concentration for Cubans, and the majority of explained variation for all countries except for Mexico, India and the Philippines.

[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 a shared language may share other characteristics as well. Note that we find no such tendency to work together for Cubans and Mexicans, or for Cubans and Salvadorans, despite a shared language.

[23]Appendix Table A.4 gives details.

The industry distribution of employment accounts for more than half of the explained concentration for immigrants from India and the Philippines, while residential segregation and the industry distribution of employment each count for about one-third of explained variation for immigrants from Mexico.

Differences in English language skill make important but smaller contributions than residential segregation and industry to explaining own-country concentration. Not speaking well is relatively more important for own-country concentration in Asian countries including China, Japan, Korea, and Vietnam. English proficiency is the most important factor in accounting for other-country concentration for all countries except India and the Philippines (source countries from which about 95% of immigrants speak English well or very well).

4.3 Taking stock

The results thus far point to four main findings. First, there is substantial concentration of immigrants in workplaces. Second, the covariates most strongly associated with concentration are industry, language skills and residential segregation. Other studies have found a role for language skills and residential segregation on related outcomes – we find that the impact of these covariates, while still important, is diminished by taking into account other factors, including employer characteristics. Third, a substantial share of this concentration takes the form of immigrants working with their compatriots. Fourth, even after accounting for many employer and worker characteristics, including employer location, industry and size, substantial concentration remains within groups defined by these characteristics. These results are based on specifications that assume that the effect of covariates on the probability of working with immigrants are the same for immigrants and natives. We relax this assumption in the next section.

5 Digging deeper: Do factors affect immigrants and natives differently?

Our main effects specification (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. We now consider specifications that add interactions between immigrant status and other covariates to (3.3) to allow the covariate coefficients to differ across groups, obtaining:

$$C_{ij} = \gamma_N^{int} + \gamma_I^{int} I_i + \theta^{int} msa_{ij} + \lambda^{int} I_i * msa_{ij} + \beta^{int} x_{ij} + \phi_I^{int} I_i * x_{ij} + \epsilon_{ij}^{int} \tag{5.1}$$

The interaction terms allow us to identify characteristics associated with particularly high or low levels of concentration. We can then look more closely at the mechanisms underlying concentration discussed in section 2.

We focus here on identifying immigrant/native differences in how specific characteristics affect the probability of working with immigrants. We have computed the Gelbach decomposition using this more flexible specification and (in unreported results) find that the findings from the main-effects analysis continue to hold. Part of the reason is that this more flexible specification yields only a modest increase in overall explanatory power: it increases the R^2 from 0.52 to 0.55. Evaluating concentration at the pooled sample means of all variables, unexplained concentration is 0.067 in the fully interacted model, compared to 0.083 with the main effects model. Thus the value of this exercise is not large gains in explanatory power but rather identifying characteristics associated with particularly high or low levels of concentration.

5.1 Overall patterns

The full specification (5.1) yields too many interaction coefficients to usefully present the full set in a table. We use graphs to illustrate the key categorical interactions, and then present predicted levels of concentration for the important continuous variables in Table 5. We do not present results for variables such as education and age where the estimated interaction effects are negligible. In both the table and figures, we evaluate concentration at the pooled mean of all other variables.

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. In Figure 2 we see that, while coworker share rises for both groups as English proficiency falls, it generally rises more for immigrants than natives. The gap is highest among those who speak English poorly rather than those who do not speak it at all. The finding that not speaking English well matters more for immigrants than natives suggests that it is more than simply language skills that are at work here.

Employer size effects are of particular interest because they potentially reflect a number of factors that influence concentration. Size may matter if production processes vary across establishments of different sizes, leading to differing demand for particular sets of skills. 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. The hiring process is also 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.

Figure 3 shows that concentration (the vertical distance between immigrant and native bars) falls substantially with employer size. Natives are somewhat less likely to work with immigrants in small establishments than in large ones, while immigrants are much more likely to work with other immigrants in the smallest establishments. For the smallest firms, much of the concentration comes from segregated workplaces—those with only immigrant or only native employees.[24] About three-quarters 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. For example, less than 20% of immigrants working at establishments with 5-9 workers have only immigrant coworkers, and all-immigrant workplaces account for a negligible share of employment in size class 20-49 and larger. All-native workplaces account for a negligible share of employment for establishments with 100 or more employees.

Taken at face value, the results by employer size provide support for the hypothesis that employee interactions with coworkers and customers are especially important at small businesses. But caution should be used in making this interpretation, as such a pattern could also arise 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 an increase in 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, becoming small for employers with 10-20 employees and negligible for larger employers.[25] Because concentration continues to decline across the top four size classes in Figure 3, we conclude that at least part of the observed decline in concentration with size reflects the economic factors discussed above.

[24]Figure A.1 in the web appendix shows these patterns.
[25]See section A.3 of the appendix.

We know from Table 3 that in 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 substantial variation across industries in the extent of concentration. Since it is impractical to illustrate differences across all detailed industries, Figure 4 gives predicted coworker shares for immigrants and natives (and the difference between them) for each detailed industry that accounts for at least one percent of employment in our sample. The industries are ordered from highest to lowest levels of concentration (i.e. the difference in coworker shares). Concentration tends to be highest in industries with large overall immigrant shares, but that is not always the case. For example, nursing facilities and hotels have similar immigrant shares overall, but concentration is much higher in nursing facilities than in hotels. This finding is consistent with a more critical role for interaction and coordination between the staff and customers (patients) in nursing facilities than in hotels.

In Table 5 we report concentration patterns for continuous variables of particular interest. As we saw in Table 2, those who live with immigrants are also more likely to work with them. While this pattern holds for both natives and immigrants, a somewhat larger effect among immigrants makes concentration particularly high for those living in predominantly immigrant areas. This can be seen by the increase in concentration (the difference in coworker shares) in the far right column in moving from the 10th to the 90th percentile of the residential segregation distribution. We also find that our neighborhood network index is positively associated with concentration and this effect is slightly larger for immigrants. Concentration is especially high at the 90th percentile of the network index.

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. Concentration is 15% lower at the 90th percentile of the earnings distribution than at the 10th percentile, holding other variables constant. This pattern suggests that concentration falls with worker skill and in particular with components of skill not captured by education or language, since we control for those variables in making the comparison.

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

To help us further understand the role of network and language effects, 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 specification is an extension of that reported in Table 4 where we now include interaction effects of the observable worker, employer and locational characteristics with worker country of origin. The ”main” effect in the first column gives the effect of the network index for natives: a higher value of the network index is associated with a slightly reduced probability of working with immigrants from each of these countries. The ”own” effect gives the difference in the network effect between workers from the designated country and natives, while the “other” column gives differences in the network effect between immigrants from other countries and natives. The own effects are consistently large while the other effects are consistently small. For immigrants, working with neighbors is highly correlated with working with compatriots, but not with having coworkers from other countries. That is, network effects appear to work primarily through networks of compatriots.

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 other countries either for natives or other immigrants.[26] 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 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 large main effect. Note that, combining main and interaction effects, the implied effect of not speaking English well for immigrants from Mexico (0.019+0.016=.035) 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.[27]

[26] 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.

[27] In further analysis, we also examined whether the network and language effects are stronger within immigrant groups that speak Spanish. We reran 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.

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.

Overall, our results are consistent with the importance of several different mechanisms emphasized in the literature as contributing to segregated workplaces, including sorting with respect to productive characteristics of workers, information available to workers and employers, and the impact of residential segregation.

We find that roughly half of immigrant concentration can be explained by our set of observable worker, employer and location characteristics. Of the half that can be explained, 37%, 30% and 33% can be explained by worker, employer and locational characteristics respectively.

In particular, immigrants who live near coworkers are more likely to work with others who are immigrants. The effect for natives—they are more likely to work with natives if they live near coworkers— is similar, but much smaller. These findings hold even when controlling for a variety of other factors that could lead to a correlation between residential and employment location (e.g., residential segregation and commuting patterns).

Workers who do not speak English well are more likely to have immigrant coworkers, and that concentration increases as we move down the earnings distribution. These effects are of interest in their own right because they suggest that 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.

Furthermore, taking into account of detailed country of origin information, we find that immigrants who work together are particularly likely to be compatriots; this is especially true for immigrants who have poor English language skills.

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.

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

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`.

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.

ELLIS, M., R. WRIGHT, AND V. PARKS (2007): "Geography and the Immigrant Division of Labor," *Economic Geography*, 83(3).

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.

WRIGHT, R., M. ELLIS, AND V. PARKS (2010): "Immigrant Niches and the Intrametropolitan Spatial Division of Labour," *Journal of Ethnic and Migration Studies*.

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
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
Speaks English (**)	Very well	55.4	97.6
	Well	23.6	1.5
	Poorly	16.0	0.7
	Not at all	5.1	0.2
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: Unit of observation is a worker. N=2,965,225 natives, 600,761 immigrants. Figures represent percentages, except for log earnings. Estimates are weighted using propensity score weights. (**) Very well category includes those who speak only English at home, and those that speak another language at home and speak English very well.

Table 2: Coworker share regression, main effects model

Covariate		Coefficient	Std Error
Immigrant		0.0830	0.0007
Education	High school drop-out	0.0114	0.0004
	High school graduate	0.0015	0.0002
	Bachelor's degree	0.0049	0.0003
	Advanced degree	0.0103	0.0006
Speaks English	Well	0.0557	0.0007
	Poorly	0.0921	0.0010
	Not at all	0.1007	0.0017
Continuity on 2000-Q2 job	Worked Q1	0.0024	0.0006
	Worked Q3	0.0019	0.0005
	Worked Q1 and Q3	0.0019	0.0006
Log quarterly earnings on primary job		0.0018	0.0002
Worker age	Age<30	-0.0034	0.0003
	30<= Age <40	-0.0016	0.0002
Female		0.0017	0.0002
Employer size	2-4 employees	0.0222	0.0018
	5-9	0.0055	0.0016
	10-19	-0.0060	0.0015
	20-49	-0.0077	0.0015
	50-99	-0.0037	0.0015
	100-499	0.0042	0.0015
Firm has more than 1 establishment		-0.0301	0.0012
Establishment age	<= 1 year	0.0009	0.0017
	2-4 years	0.0026	0.0014
Firm has >1 estab * Estab age	<=1 year	-0.0027	0.0024
	2-4 years	0.0005	0.0018
Immigrant share of workers in residential tract		0.1807	0.0021
Neighborhood network index		0.0439	0.0036
Shared commute index		-0.4214	0.0105

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	0.083
Contribution to reduction in coefficient	**Percents**
Individual characteristics (total)	37.2
Log earnings	0.2
Quarters of work	0.0
Age and sex	0.2
Language	34.8
Education	2.0
Employer characteristics (total)	30.1
Firm size	0.3
Firm age and multi-unit status (interacted)	2.8
Industry	27.0
Sector	11.9
Sum of within sector detail	15.2
Manufacturing detail (73 3-digit industries)	3.1
Transportation, communications, utilities (14 inds)	1.0
Wholesale (18 industries)	0.7
Retail (33 industries)	0.5
FIRE (4 industries)	0.8
Services (51 industries)	9.1
Neighborhood characteristics (total)	32.7
Immigrant share of workers living in residential tract	32.1
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. Within-sector, the omitted category for the detailed industry dummies is the modal 3-digit industry for that sector.

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	0.093	-0.012
	(0.008)	(0.003)	(0.005)	(0.002)
El Salvador	0.063	0.148	0.044	0.032
	(0.001)	(0.002)	(0.001)	(0.001)
Mexico	0.157	0.021	0.086	-0.025
	(0.001)	(0.001)	(0.001)	(0.001)
China	0.200	0.139	0.164	0.044
	(0.005)	(0.003)	(0.004)	(0.003)
India	0.155	0.054	0.135	0.015
	(0.005)	(0.002)	(0.004)	(0.002)
Japan	0.140	0.026	0.136	-0.009
	(0.005)	(0.003)	(0.005)	(0.003)
Korea	0.188	0.047	0.178	-0.022
	(0.005)	(0.003)	(0.005)	(0.003)
Philippines	0.095	0.050	0.075	0.012
	(0.002)	(0.001)	(0.001)	(0.001)
Vietnam	0.181	0.086	0.154	-0.005
	(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: Predicted coworker shares and concentration for selected values of covariates

	Immigrants	Natives	Difference
Predicted coworker share at mean for all variables	0.233	0.165	0.067
Holding other Xs at mean, prediction for:			
Residential segregation			
10th percentile	0.199	0.139	0.060
90th percentile	0.290	0.211	0.079
Network index			
Index=0	0.226	0.167	0.059
90th percentile	0.239	0.164	0.075
Log earnings			
10th percentile	0.239	0.159	0.080
90th percentile	0.228	0.170	0.058

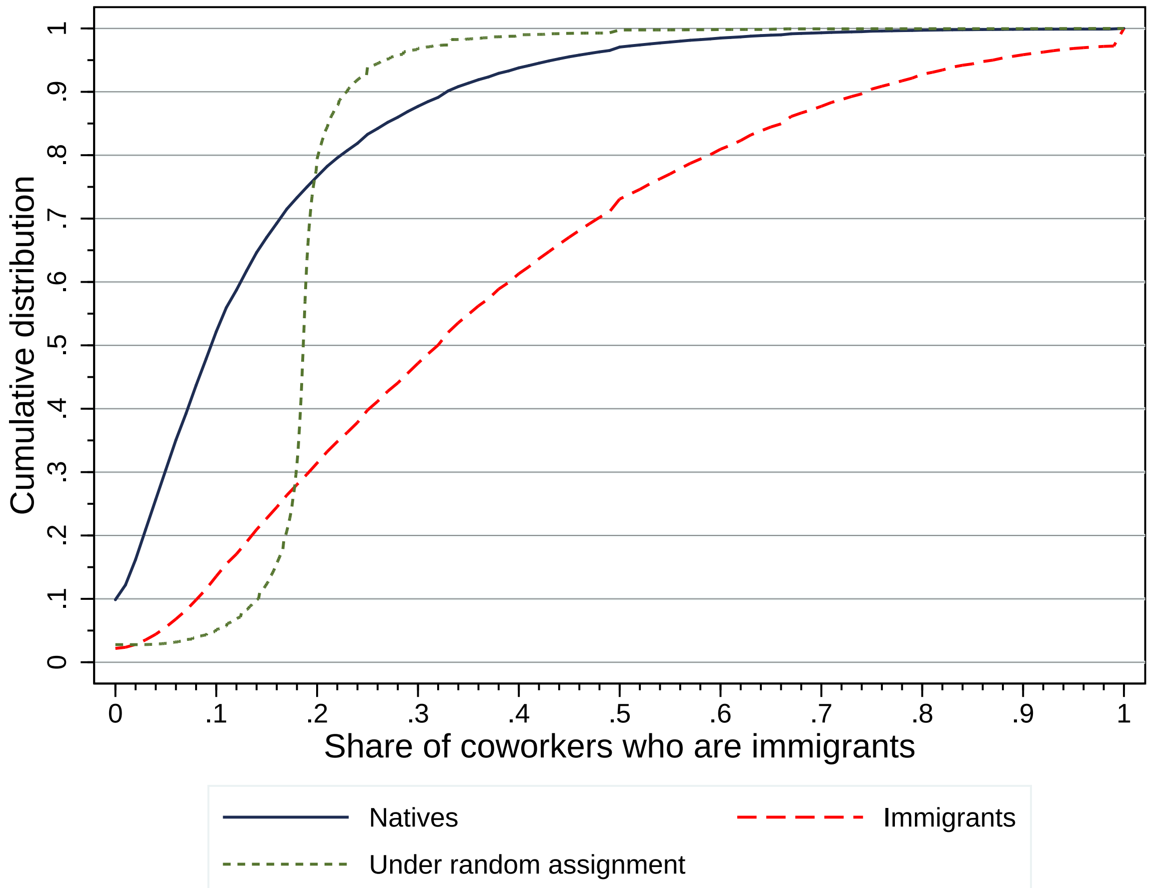
Note: Estimates give predicted immigrant/native difference in coworker mean based on fully interacted model. Predictions based on pooled means for all variables except the variable in the left column and immigrant status. Predicted concentration (the 3rd column) equals the prediction with all Xs at the mean (1st row) plus interaction coefficient(s) times difference between the indicated point in the X distribution and the mean.

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

	Neighborhood index			Speaks English poorly		
	Main	**Own**	**Other**	**Main**	**Own**	**Other**
Cuba	-0.001	0.364	-0.004	0.002	0.058	0.001
	(0.001)	(0.033)	(0.002)	(0.000)	(0.005)	(0.000)
El Salvador	-0.004	0.486	-0.006	0.002	0.014	0.000
	(0.000)	(0.045)	(0.001)	(0.000)	(0.002)	(0.000)
Mexico	-0.020	0.465	-0.009	0.016	0.019	-0.012
	(0.001)	(0.017)	(0.003)	(0.001)	(0.001)	(0.001)
China	-0.002	0.366	0.008	0.000	0.064	0.005
	(0.000)	(0.033)	(0.002)	(0.000)	(0.006)	(0.000)
India	-0.002	0.581	0.003	0.000	0.074	-0.001
	(0.000)	(0.031)	(0.002)	(0.000)	(0.008)	(0.000)
Japan	-0.001	0.315	-0.002	0.001	0.146	-0.000
	(0.000)	(0.071)	(0.001)	(0.000)	(0.011)	(0.000)
Korea	-0.002	0.218	-0.006	0.001	0.115	0.000
	(0.000)	(0.044)	(0.001)	(0.000)	(0.009)	(0.000)
Philippines	-0.003	0.616	-0.002	0.000	0.033	-0.001
	(0.001)	(0.042)	(0.001)	(0.000)	(0.005)	(0.000)
Vietnam	-0.003	0.555	-0.001	0.002	0.076	0.003
	(0.000)	(0.032)	(0.002)	(0.000)	(0.004)	(0.000)

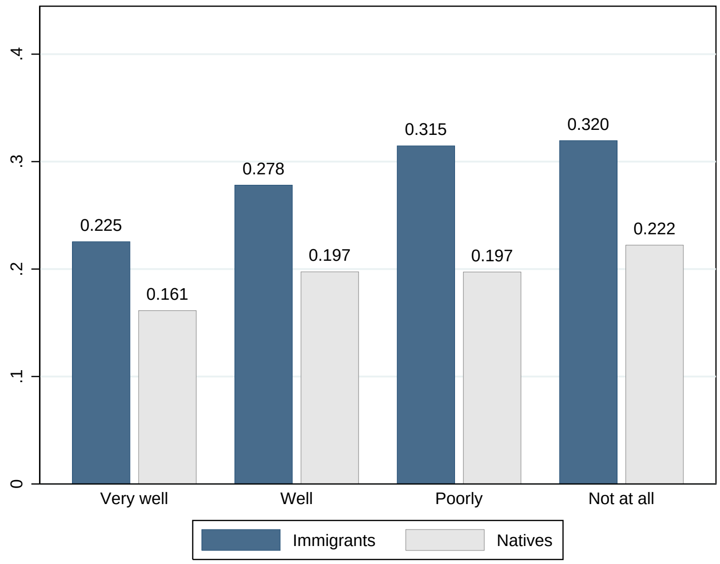
Notes: Standard errors appear directly below coefficient estimates. Each line in the table presents estimates from a separate regression with the share of coworkers from the indicated country as the dependent variable. The specification also includes main effects and own/other interactions for two other language categories (speaks English well, and does not speak English at all), 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.

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



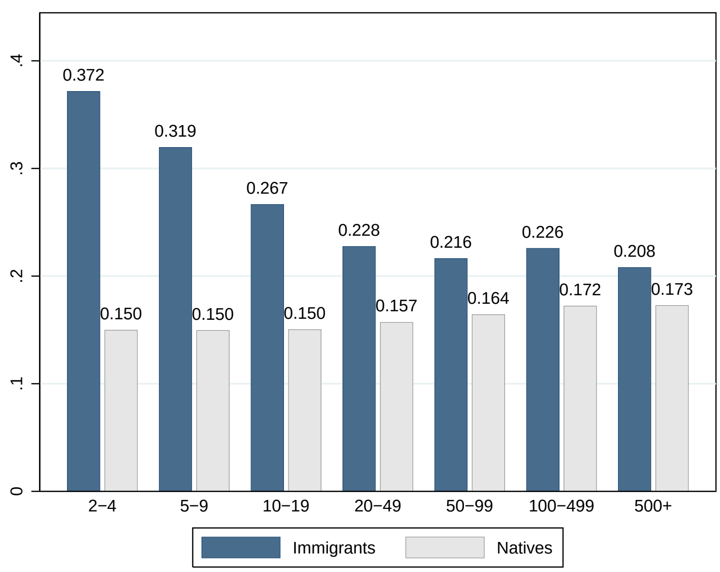
Note: The CDF under random assignment is constructed by first simulating the distribution of coworker shares conditional on employer size S by drawing 4,000 binomial random variates for S trials with p=.187 (share immigrant in our sample), and then using the number of immigrants (=number of successes 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 how well a worker speaks English



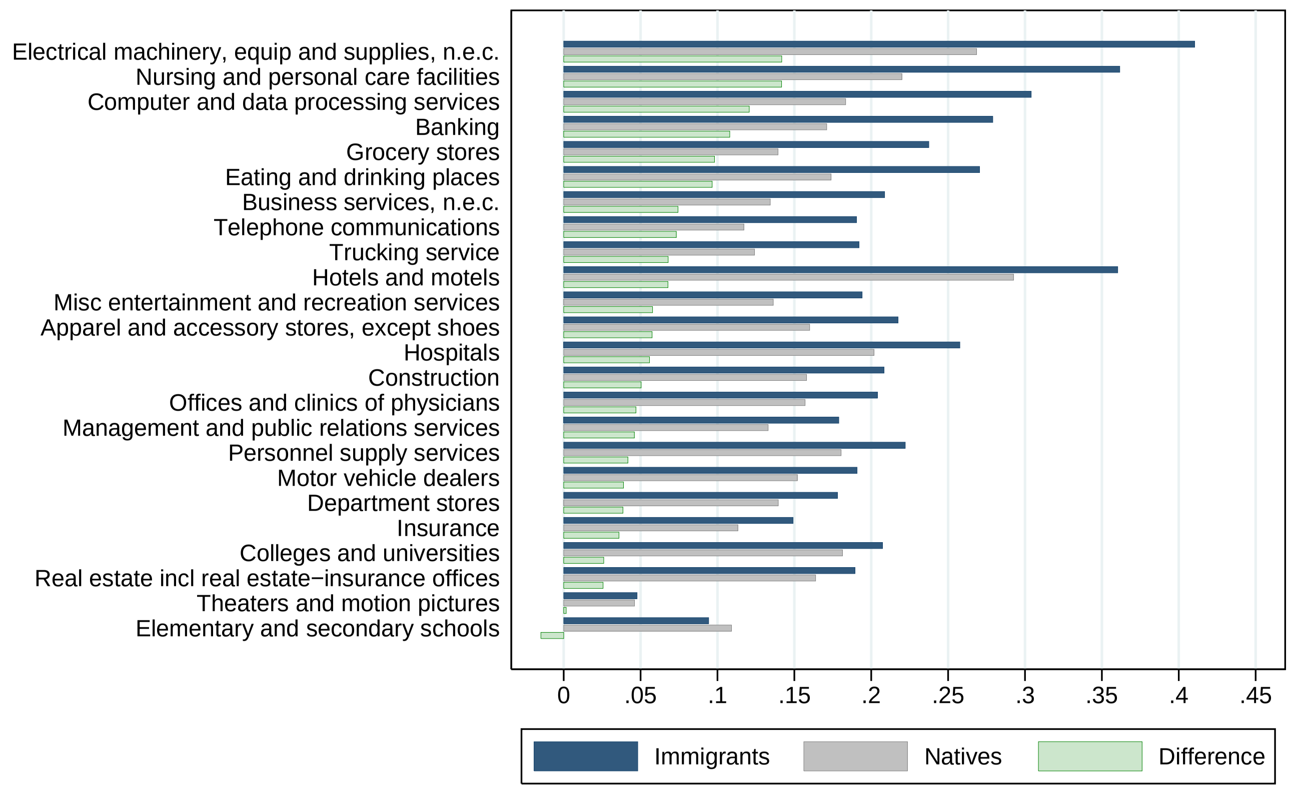
Note: Based on predictions in which all variables except language and immigrant status are set to pooled mean values. Model used includes interactions between the immigrant dummy variable and all other covariates. Those who speak only English at home are categorized as speaking it very well.

Figure 3: 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 4: 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.

NOT INTENDED FOR PUBLICATION

Workplace Concentration of Immigrants: Appendix

This appendix includes supplementary tables, figures and analysis. It is organized into three sections. Section A.1 contains tables of summary statistics that compare the full sample (all workers from the UI wage records in our sample of MSAs) to the matched sample (subset of workers with records matched to the Decennial long form data). Section A.2 has some supplementary tables for the country of origin analyses. Section A.3 includes an analysis of a statistical artifact that arises in examining concentration by employer size for very small firms (A.3.1), and a figure with additional detail on differences by firm size (A.1).

A.1 Summary Statistics

The tables here provide the following information:

- Comparisons between the full and matched samples on an unweighted basis for all immigrants are in Table A.1 and for all natives are in Table A.2.
- A comparison of Table 1 in the main text to the first column in Tables A.1 and Table A.2 illustrates how closely the weighted matched sample lines up with the full sample.

Table A.1: Characteristics of Immigrants in Full and Matched Samples (Unweighted)

		Full	Matched
Coworker share		37.7	36.3
Worker age	Age<30	23.3	22.3
	30< Age <40	33.7	33.3
	Age>40	42.9	44.4
Male		56.1	55.0
Age at arrival	<= 12	12.4	12.5
	13-25	47.5	47.6
	26-35	26.9	27.1
	36+	13.2	12.8
Education	High school drop-out		31.8
	High school graduate		18.4
	Some college		17.1
	Bachelor's degree		22.2
	Advanced degree		10.5
Does not speak English well			20.4
Log quarterly earnings on primary job		8.5	8.5
Continuity of 2000-Q2 job	Q1 and Q3	69.3	71.1
	Q1 or Q3	24.6	23.3
	Neither Q1 nor Q3	6.0	5.6
Establishment size	2-9 employees	9.3	8.7
	10-49	23.0	21.9
	50-99	13.3	12.9
	100-499	31.0	30.8
	500 or more	23.4	25.6
Firm has multiple establishments		33.5	35.8
Establishment age	<=1 year	12.4	11.4
	2-4 years	23.5	22.6
	5+ years	64.1	66.1
Sector	Construction	5.2	5.2
	Manufacturing	20.8	21.3
	Transportation/utilities	4.0	3.6
	Wholesale	7.2	6.6
	Retail	19.2	19.3
	FIRE	5.0	4.9
	Services	38.5	39.1
Immigrant share of workers in residence tract		36.7	35.9
Neighborhood network index		1.8	1.9
Shared commute index		0.3	0.3

Notes: The unit of observation is a worker. N=600,761 for the matched sample and N=6.2 million for the full sample. All figures except log earnings represent percentages.

Table A.2: Characteristics of Natives in Full and Matched Samples (Unweighted)

		Full	Matched
Coworker share		14.5	13.6
Worker age	Age<30	32.3	30.9
	30< Age <40	26.8	25.9
	Age>40	40.8	43.1
Male		51.3	50.5
Education	High school drop-out		15.8
	High school graduate		25.5
	Some college		25.7
	Bachelor's degree		24.7
	Advanced degree		8.2
Does not speak English well			0.8
Log quarterly earnings on primary job		8.4	8.4
Continuity of 2000-Q2 job	Q1 and Q3	65.6	67.3
	Q1 or Q3	26.6	25.6
	Neither Q1 nor Q3	7.8	7.2
Establishment size	2-9 employees	8.3	7.8
	10-49	23.8	22.7
	50-99	13.5	13.1
	100-499	29.6	29.6
	500 or more	24.9	26.8
Firm has multiple establishments		42.2	44.7
Establishment age	<=1 year	12.0	11.1
	2-4 years	24.8	23.9
	5+ years	63.3	65.1
Sector	Construction	5.7	5.7
	Manufacturing	12.1	13.3
	Transportation/utilities	5.4	5.0
	Wholesale	6.6	6.2
	Retail	22.0	22.3
	FIRE	7.1	6.7
	Services	41.2	40.8
Immigrant share of workers in residence tract		15.7	14.0
Neighborhood network index		1.7	1.9
Shared commute index		0.5	0.5

Notes: The unit of observation is a worker. N=3.0 million for the matched sample and N=26.4 million for the full sample. All figures except log earnings represent percentages

A.2 Supplemental Tables for Country of Origin Analyses

Table A.3 presents some summary statistics by country of origin. Table A.4 presents the Gelbach decomposition by country of origin for the main effects model. Table A.5 presents the language cross-effects by country of origin for the interacted model.

Table A.3: Selected sample characteristics by country of origin

	Share of	Share speaking English:			Share of neighborhood workers who are:			
	workers	Well	Poorly	Not at all	Same group	Mexican	Other immigrants	Natives
Mexico	4.9	27.5	26.1	10.1	26	.	15	59
El Salvador	0.7	28.7	25.7	9.7	8	19	22	51
Cuba	0.7	19.9	21.0	13.0	31	2	25	42
India	0.8	17.4	4.3	0.5	6	3	22	69
Philippines	1.3	22.8	3.4	0.2	9	8	21	62
China	0.5	31.1	24.8	8.6	10	4	29	57
Japan	0.2	22.0	9.3	0.3	1	4	20	75
Korea	0.4	27.2	17.8	1.5	4	4	25	67
Vietnam	0.8	35.6	26.8	3.3	12	9	21	59
Other	8.4	20.1	10.3	2.6	17	4	9	69

Notes: The figures in the table are percentages. Those in the first column would add to 100 if natives were included. Those in the right panel add to 100 for each row. In the "Other" row, the same group share includes all neighborhood immigrant workers who are not from one of the listed countries, while "Other immigrants" includes all neighborhood immigrant workers who are from the listed countries aside from Mexico.

Table A.4: Main-effects decomposition by country of origin

	Individual char			Employer char		Neighborhood char		
	Language	Education	Other	Industry	Other	Own res seg	Other res seg	Other
Own								
Mexico	18	7	-0	36	2	36	0	-0
El Salvador	11	3	0	21	1	64	-0	0
Cuba	4	0	0	2	1	92	2	0
India	1	3	1	51	1	43	-2	1
Philippines	3	-0	1	57	2	36	2	0
China	12	-0	0	13	2	76	-3	0
Japan	17	3	1	24	3	51	1	0
Korea	15	1	-0	12	8	65	-2	0
Vietnam	16	0	0	32	1	50	0	0
Other								
Mexico	72	2	0	15	3	3	4	0
El Salvador	38	4	-0	28	3	-1	27	0
Cuba	48	2	1	16	6	5	21	1
India	30	-1	1	34	5	-11	39	2
Philippines	34	-2	2	31	-1	-2	36	1
China	44	2	1	32	4	-6	22	1
Japan	51	-2	4	20	6	-1	21	2
Korea	42	-1	1	27	8	0	22	2
Vietnam	45	2	0	33	4	-4	19	1

Notes: The figures in the table are contributions based on Gelbach decompositions for the regressions used in Table 4. The “Own” panel gives shares of the difference in coefficients between columns (1) and (3), while the “Other” panel gives shares of the difference in coefficients between columns (2) and (4).

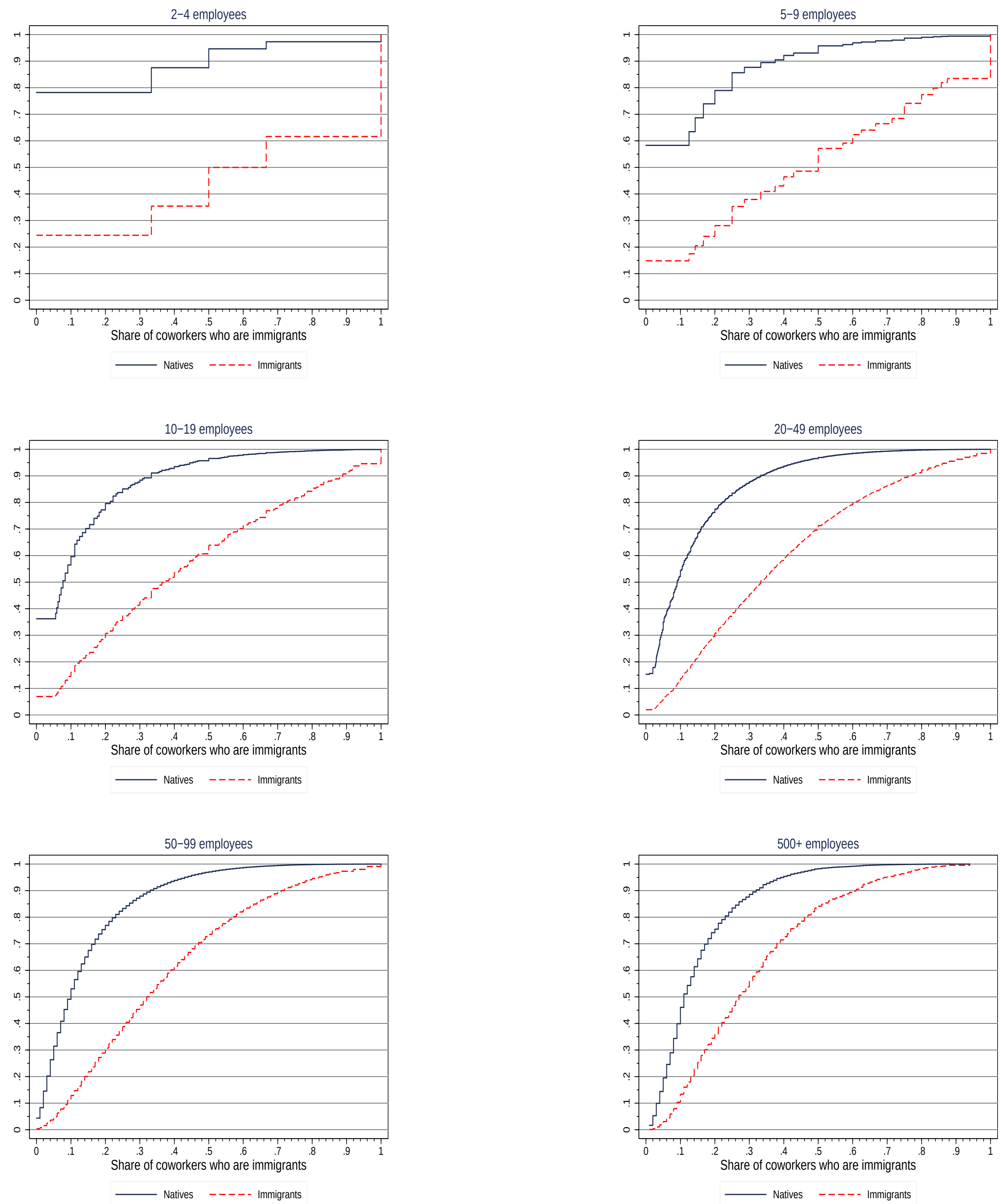
Table A.5: Language cross-effects by country of origin

Share of coworkers from:	Main effect	Difference in coworker share (relative to share for natives who do not speak English well) for immigrants who do not speak English well and are from:								
		Mexico	**El Salvador**	**Cuba**	**India**	**Philippines**	**China**	**Japan**	**Korea**	**Vietnam**
Mexico	1.9	**1.3**	1.8	-3.0	-1.9	-1.4	-4.4	-2.2	-0.9	-3.0
El Salvador	0.2	0.1	**1.0**	-0.3	-0.2	-0.3	-0.9	-0.8	-0.2	-0.5
Cuba	0.2	-0.2	-0.1	**5.5**	-0.1	-0.2	-0.3	-0.2	-0.3	-0.3
India	0.0	-0.0	-0.0	0.0	**6.5**	0.0	-0.6	-0.4	-0.2	-0.2
Philippines	0.0	-0.3	-0.5	0.3	0.1	**2.6**	-1.0	-0.2	-0.7	-0.6
China	0.0	-0.1	-0.1	-0.1	-0.2	0.1	**6.5**	-0.5	-0.5	1.7
Japan	0.0	-0.0	-0.0	-0.0	-0.1	-0.1	-0.3	**11.4**	-0.1	-0.2
Korea	0.1	0.0	0.1	-0.0	-0.1	-0.0	-0.2	1.1	**8.3**	-0.3
Vietnam	0.1	-0.3	-0.2	-0.2	0.3	0.4	3.0	-0.2	-0.1	**6.7**

Notes: Each of the figures in the table is a coefficient from a regression with the share of coworkers from the country listed in the left-most column as the dependent variable. The main effect column gives the coefficient on a dummy for not speaking English well. (To make the cross-effects manageable, we collapse the English proficiency categories to combine speaking English poorly and not speaking it at all into one category, and combine speaking English well or very well in the omitted category.) The other columns contain the coefficients on the interaction between a dummy variable for the country listed above the column and the dummy for not speaking English well. The coefficients have been rescaled to represent percentages (i.e. they are multiplied by 100). The regression specifications are as in Table 6, except that the language interactions are with our full set of country of origin dummies rather than the own/other grouping used in Table 6.

A.3 Additional analysis of employer size effects

Figure A.1: Cumulative Distribution of Coworker Share by Employer Size



Source=LEHD database. Year 2000 second quarter.

A.3.1 Simulations of employer size effects in a statistical model with segregation

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

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

$$b(i,s) = \begin{pmatrix} i \\ s \end{pmatrix} p_D^i (1-p_D)^{s-i} \tag{A.3.1}$$

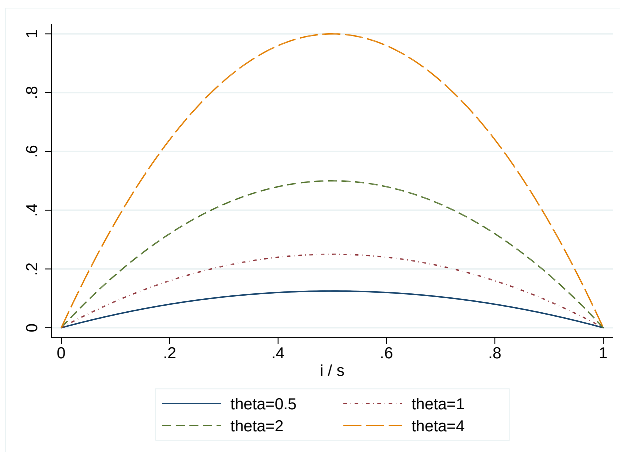
where i represents the number of immigrants in the workforce draw, s represents employer size, and p_D represents the fraction of workers who are immigrants in the group being sampled in draw D. For the initial draw, the parameter p_0 will equal the overall share of immigrants in the workforce.

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

$$d(i; s, \theta) = \frac{i}{s} \left(\frac{s-i}{s} \right) \theta \tag{A.3.2}$$

If an employer draws only immigrants or only natives, then $d = 0$ and the original draw is kept. If there are some of both types of employees, then the workforce is redrawn with probability d. This shifts some of the probability mass from more integrated towards more segregated types of employee mixes. Figure A.2 illustrates the shape of $d()$ for various values of θ.

Figure A.2: Shape of function d



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

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

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

$$p_1 = \frac{\sum_{j=1}^{s} b(j, s; p_0) * d(j; s, \theta) * j}{\sum_{j=1}^{s} b(j, s|p_0) * d(j; s, \theta) * s} \tag{A.3.3}$$

and

$$Pr(i; s, p_0, \theta) = b(i, s|p_0) * (1 - d(i; s, \theta)) + b(i, s|p_1) * \left(\sum_{j=0}^{s} b(j, s|p_0) * d(j; s, \theta) \right) \quad \text{(A.3.4)}$$

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

For the simple case $s = 2$ and $\theta = 4$ (so $d = 1$ for the only integrated workforces—those with 1 immigrant, 1 native), $p_1 = .5$, and the probability of observing a workforce with 1 immigrant and 1 native in the final distribution simplifies to $p_0(1 - p_0)$ (half the binomial probability). Figure A.3 illustrates the difference between the distribution of the coworker mean with segregation and without for employers of varying size. It uses parameter values $\theta = 4$ and p_0 = .25. Smaller values of θ would reduce the shift in the distribution, while smaller values of p_0 shift the weight of both distributions to the left.

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

$$E(cw_I|s) = \sum_{i=0}^{s} \left(Pr(i|I_j = 1; s, p_0, \theta) * \frac{i-1}{s-1} \right) = \sum_{i=0}^{s} \left(Pr(i; s, p_0, \theta) * \frac{i}{sp_0} * \frac{i-1}{s-1} \right) \quad \text{(A.3.5)}$$

and for natives,

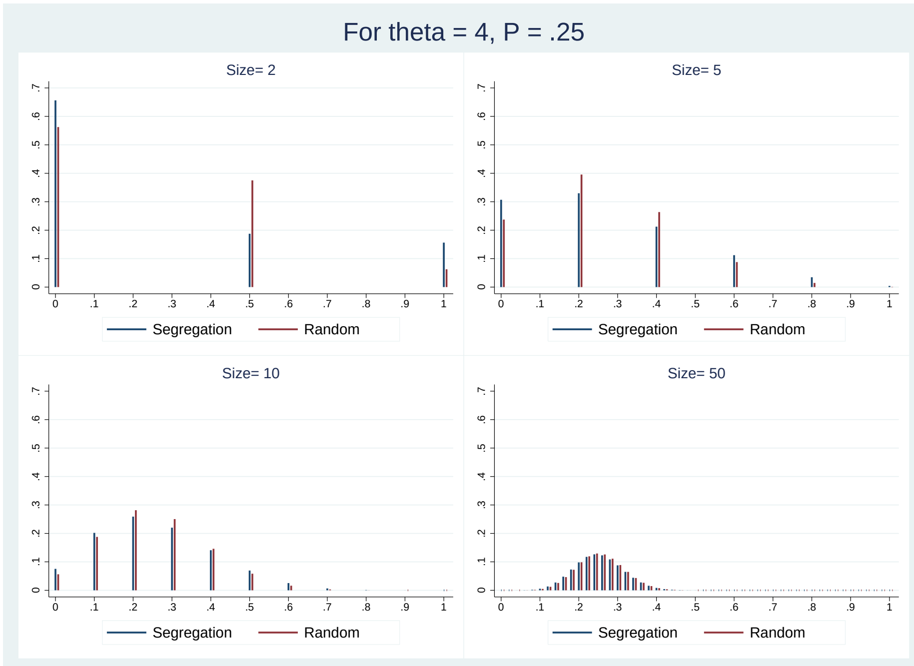
$$E(cw_N|s) = \sum_{i=0}^{s} \left(Pr(i; s, p_0, \theta) * \frac{(s-i)}{s(1-p_0)} * \frac{i}{s-1} \right) \quad \text{(A.3.6)}$$

The difference is then:

$$E(cw_N - cw_I|s) = \sum_{i=0}^{s} \left(Pr(i; s, p_0, \theta) * \frac{i[p_0(s-i) - (i-1)(1-p_0)]}{s(s-1)p_0(1-p_0)} \right) \quad \text{(A.3.7)}$$

Figures A.4 to A.6 plot out the relationship between employer size and coworker means for various values of the immigrant share of the overall workforce p(different colored lines in each

Figure A.3: Immigrant share distribution with and without segregation



graph), using segregation parameter $\theta = 4$. Figure A.4 graph gives the mean by firm size for immigrants, Figure A.5 is for natives, and Figure A.6 gives the difference between them. Figure A.7 repeats Figure A.6, except that it is parameterized to represent a lower level of segregation ($\theta = 1$). Examination of these figures makes a couple of patterns clear: (i) For very small employers (< 10 employees), the model can generate a large difference in coworker means, even with a relatively mild tendency to segregate. (ii) Even for large theta, this model generates essentially no segregation in large firms.

Because the change in variance with sample size falls off quite quickly as size increases, we think that the statistical effect is unlikely to account for size effects among firms with more than 20 employees. Thus it might be reasonable to think of size effects based on the portion of our sample with at least 20 employees as representing the economic size relationship, while in smaller firms the size effect combines the economic and statistical relationships.

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

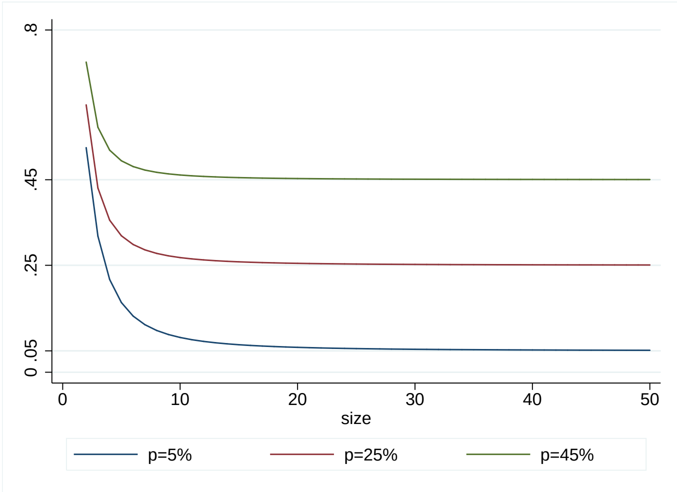


Figure A.5: Native coworker mean and employer size ($\theta = 4$)

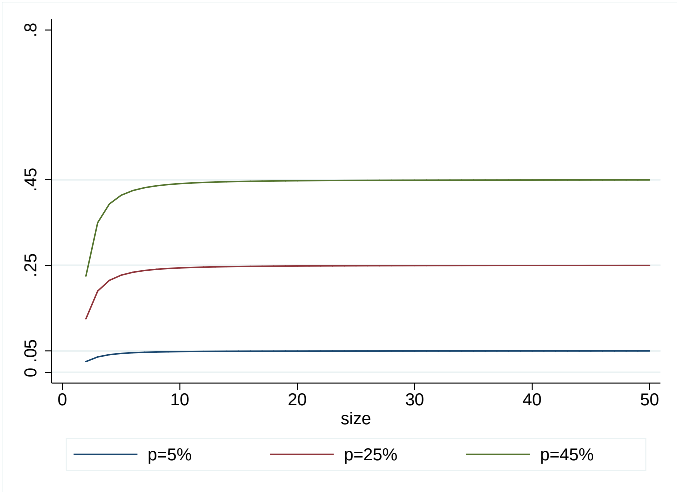


Figure A.6: Immigrant-native difference in coworker mean and employer size ($\theta = 4$)

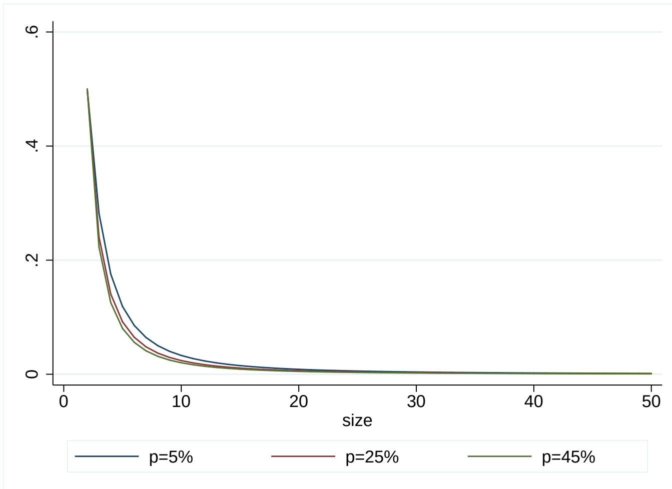


Figure A.7: Immigrant-native difference in coworker mean and employer size ($\theta = 1$)