IDENTIFYING ETHNIC ENCLAVES USING LINKED EMPLOYER-HOUSEHDOLD DATA

Liliana D. Sousa
Center for Economic Studies, U.S. Census Bureau
April 2, 2012

The research in this paper was undertaken while the author was at Cornell University and the Center for Economic Studies at the U.S. Census Bureau. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau or Cornell University. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation.

Abstract

Using the unique data available through the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program, I identify individuals as part of an ethnic enclave economy based on two dimensions: their neighbors and their coworkers. I create and analyze measurements of immigrant enclaving proclivity based on both residential and employment clustering behavior. These measures of immigrant clustering show that, even among the largest immigrant groups living in five of the biggest immigrant population centers in the U.S., few immigrants are part of ethnic enclaves. I also measure the proportion of immigrant segregation that is attributable to observable demographic characteristics found in public-use data sets. I find that these characteristics combined with geographic area controls explain about half of the variation in residential own-exposure rates and a quarter of workplace own-exposure rates.

Ethnic enclaves as economic shelters first came to the attention of social scientists when Wilson and Portes (1980) wrote about the economic success of the Cuban enclave in Miami. Relying on information on the ethnicity of the employer and workers in the industry/occupation group, they argued that enclave participation paved the way for eventual profitable entrepreneurship within the enclave economy, a potential portal to economic success not found outside of the enclave. Subsequent research in this field has relied primarily on residential clustering, often at the city level, as a proxy for this notion of an ethnic enclave economy (for example, Borjas 2000). Instead of the positive enclave effect found by Wilson and Portes, this line of research often finds that enclave residents have lower wages and lower wage growth than non-enclave residents (for example, Borjas 1995, 2000). Recent research, however, has shown that the magnitude and direction of these enclave effects are sensitive to the data source and methodologies used (for example, Edin, Fredriksson and Aslund 2003; Cutler, Glaeser and Vigdor 2008; Xie and Gough 2011). These papers illustrate that measurement issues arising from data quality have significant effects on the estimation of enclave effects. In this paper, I use uniquely detailed data available through the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program to access both residential and workplace data for a large sample of immigrants. I perform two tasks with these data: 1) I create measurements of the immigrant proclivity to enclave based on residential clustering behavior and employment clustering behavior, as well as the interaction of the two, and 2) I measure the proportion of immigrant segregation in both dimensions that can be explained by observable characteristics found in the standard datasets used in this field of research.

Enclave effects, in essence, are the economic result of social networks defined along cultural and ethnic lines and the spread of information and economic opportunities via these networks. The enclave effect question addressed in this literature can be boiled down to whether these ethnic networks provide economic opportunities or, on the contrary, limit opportunities by limiting the quality or size of the network. However, since collecting data on social networks is prohibitively expensive and intrusive, the scope of such studies is often limited to relatively small networks (for example, the Mexican Migration Project and the Framingham Heart Study). Instead, researchers interested in ethnic network effects must rely on geographic and ethnic identification as proxies.[1] Furthermore, since most public use data sources are limited in geographic detail and sample size, researchers using these data often define enclaves as the total ethnic population in a given city or state. This measure dilutes potential network effects by including individuals who are not, or are only minimally, a part of the ethnic networks. Because of this issue, some recent ethnic networks research has relied on restricted-access data for more detailed geographic identification (for example, Bertrand, Luttmer, and Mullainathan 2000; Edin, Fredriksson and Aslund 2003; Bayer, McMillan and Rueben 2004).

Though using restricted-access residential information better identifies who resides in high co-ethnic areas, it still does not capture the economic connections that are also an integral part of the enclave economy. Using both residential and coworker information from the Longitudinal Employer-Household Dynamics (LEHD) program, I identify individuals as part of an enclave economy based both on their neighbors and on their coworkers. This allows me to distinguish between individuals born in some country *j* living in city *k* who have assimilated (as measured by residence and employment ethnic exposure) versus those from the same country of birth and residing in the same city who are members of an ethnic enclave. Several measures of enclave can be constructed and analyzed using the interaction of the two measures, shedding light on what

[1] Some examples of ethnic identification include country of birth (Cutler, Glaeser, and Vigdor 2008), self-reported

today's immigrant enclaves look like and the significance of the role they play in the lives of contemporary immigrants.

There is a clear tendency for immigrants to cluster in the host country (for example, Bartel 1989; Borjas 2000; Edin, Fredriksson and Aslund 2003). How do these areas of high ethnic clustering emerge? Toussaint-Comeau (2008) provides the following outline of the enclaving process: initial waves of immigrants from a given country settle in a port of entry or an area with some significant immigrant labor demand and, due to mobility costs, many members stay. Due to U.S. immigration policy favoring family reunification, subsequent waves of immigrants will join previous cohorts where they have settled, taking advantage of the familial social networks already available to them in that area. As the number of co-ethnics increases in an area, economies of scale in the production of ethnic goods (such as food, religious institutions and marriage markets) lead to greater availability of ethnic goods, and thus more incentive for co-ethnics to stay near the enclave (Chiswick and Miller 2002). Lazear (1999) shows that ethnic clustering also results in a greater number of potential trade partners for those facing high assimilation costs, such as language acquisition. This clustering leads to more economic opportunities within areas of high co-ethnic density. The resulting ethnic good production and availability of social networks are such that immigrants are willing to pay higher rents to reside in high co-ethnic areas (Gonzalez 1998; Cutler, Glaeser and Vigdor 2008).

Researchers have consistently found that many enclave residents earn less than similar immigrants who are not enclaved, but this effect is mitigated when data allow for a correction of the self-selection into high co-ethnic areas in the empirical mode. Edin, Fredriksson and Aslund (2003), using a pseudo-natural experiment design based on detailed residential data of refugees

---

ethnicity (Borjas 1992), race (Borjas and Bronars, 1989), and language (Bertrand, Luttmer, and Mullainathan 2000).

in Sweden, found that the negative enclave effect on wages found in previous research is actually the result of negative self-selection into enclaves. Similarly, Cutler, Glaeser and Vigdor (2008) find that using restricted-access microdata on place of residence to correct for negative selection into enclaves yields a net positive effect of enclaving, and a negative effect only for groups with very low levels of education.

Other researchers have attempted to study the impact of ethnic employment networks, as opposed to simply residential concentration, on economic assimilation. For example, Xie and Gough (2011) proxy for enclave employment by identifying immigrants who speak a language other than English at work. They find a negative enclave effect on earnings, though they do not address negative self-selection into enclaves. Some research on job networks has been done using restricted-access data on both place of residence and employer information (for example, Bayer, Ross and Topa 2008). Andersson et al. (2010) use the restricted 2000 Decennial Census and the LEHD data to look at the concentration of immigrants in the workplace and residentially. They find that immigrants are more likely to work with other immigrants than with natives, though most immigrants work with some natives. This effect is more pronounced for immigrants with limited English skills. Half of the difference between the probability of a U.S. native working with an immigrant and the probability of an immigrant working with another immigrant is explained by observables, including industry, language skills and residential segregation. Though they find that living in the same neighborhood as other immigrants increases the proportion of coworkers who are immigrants, the magnitude of the estimated effect is not large enough to support the hypothesis that social networks are being used intensively as recruitment networks. They do not extend their analysis to consider the existence of enclave economies or enclave effects.

This paper exploits the rich data on households and employers collected by the LEHD program to more precisely identify and measure ethnic enclaves among contemporary immigrants. By using data on neighbors and on coworkers, this paper is able to capture two important dimensions of ethnic segregation and integration, thus shedding light on just how segregated some large immigrant groups are at the beginning of the 21st century. The remainder of the paper is organized as follows: Section 1 provides a description of the data and empirical strategies used in this paper. Section 2 details the residential and workplace own-ethnic exposure rates and the resulting enclave population estimates. Section 3 provides an analysis of which demographic groups are more likely to be enclaved and what proportion of enclaving behavior is explained by observables. Section 4 concludes with an overview of what is gained with the strategies employed in this study.

Section 1. Data and Empirical Strategy

1.1 Data

I combine detailed microdata from the LEHD program at the U.S. Census Bureau with the confidential version of the 2000 Decennial Census of Population and Housing in order to identify enclave economies using both residential and employment ties between individuals from the same country of birth. The confidential 2000 U.S. Census of Population and Housing is a one-in-six household sample containing detailed residential and demographic data. It provides data on census block of residence, year of immigration, age, gender, educational attainment, English-language skills and other important demographic information for all individuals in sampled households. Using the restricted-access version of this file has three important advantages over using the publicly available version. First, it provides access to census tract and block-level residential data, allowing for a more detailed identification of ethnic residential segregation. Second, its significantly larger sample size allows for the inclusion of smaller country of birth groups. Finally, it is linkable to the employment data used in the LEHD program, largely derived from administrative earnings and employment data. The LEHD microdata files, in turn, provide information on firms, including location and industry, and basic demographic details for all covered employees of these firms, such as place of birth.[2] Combined, these data allow for the construction of residential and coworker ethnic exposure measures, as described below.

This analysis focuses on five of the top destination urban areas in the United States: Chicago, Houston, Los Angeles, Miami and New York City.[3] These five Consolidated Metropolitan Statistical Areas (CMSA's), large urban areas composed of cities and surrounding suburbs connected by extensive commuting patterns, were home to 47% of the immigrant population in the U.S. in 2000. These five urban areas were also selected in order to include different regions of the U.S., an important consideration given the prevalence of regional preferences among different groups of immigrants. Almost 40% of the sample lives in the New York City CMSA, 30% in Los Angeles and 17% in Chicago. Miami and Houston, at 7 and 8% respectively, are relatively small shares of the sampled population.

Information on country of birth, both for the sample and their coworkers, is from the LEHD's Individual Characteristics File (ICF). While most country of birth data are from Social Security Administration records, 4% of records have been imputed by LEHD staff. The imputation model imputed place of birth using the nineteen largest immigrant source countries and ten region of

[2] LEHD microdata are used to create the Quarterly Workforce Indicators and OnTheMap, two U.S. Census Bureau products publically available online. Abowd et al. (2005) provide a detailed description of the LEHD infrastructure files.
[3] In order to protect confidentiality, no results based on any one urban area are reported in this research.

birth groups composed of the remaining source countries.[4] To accommodate this feature of the data, this paper uses the same thirty-four country and region of birth groups in calculating both residential and workplace exposure rates.

The sample used in this paper consists of adults, ages 18 through 70, in sampled households in the 2000 Decennial Census who resided in the five CMSA's listed above. Though the LEHD sample is far larger since it includes almost all workers with state unemployment insurance (UI) records, the universe is limited to those surveyed in the 2000 census because two particularly important variables are only available via the census: English language fluency and years since migration. The labor force sample is a subsample of the above sample limited to individuals who report their primary work as self-employment, or who report that they are otherwise in the labor force *and* match to the LEHD data.[5] Hence, individuals who report being in the labor force in the census but who do not have job records in the LEHD data are excluded from the workplace analysis. Certain jobs, such as federal jobs and self-employment, are not currently included in the LEHD records. Furthermore, an important limitation of using administrative earnings data to study immigrants is that illegal immigrants cannot be identified since they lack valid social security numbers. Since LEHD linkage to the decennial data is partially based on social security numbers, individuals who do not provide valid social security numbers to their employers might also be excluded from the workplace data. However, address and name information are also utilized in the matching algorithm, thus improving the match quality even for illegal immigrants. To the extent that different place of birth groups exhibit varying levels of federal employment or

[4] The U.S. born population is divided by race and ethnicity: white non-Hispanic, black non-Hispanic, Asian non-Hispanic, other non-Hispanic (includes Native American/Pacific Islander groups and those individuals reporting more than one race), and Hispanic.
[5] The sample was matched to LEHD data in the years 1999, 2000 and 2001. When data was not available for 2000, data from one of the other years was used, if available.

Table 1: Distribution of Ethnicity/Place of Birth for the Residential and Workforce Samples

Place of Birth	Proportion of Total Sample	Proportion of Labor Force Sample	Proportion of each POB's population in both
Africa	0.005	0.006	0.819
Caribbean	0.006	0.006	0.784
Central America	0.009	0.009	0.755
Central Asia	0.005	0.004	0.613
Middle East/N. Africa	0.008	0.007	0.694
Oceania	0.001	0.001	0.758
Socialist Europe	0.006	0.006	0.687
South America	0.018	0.018	0.766
Southeast Asia	0.008	0.007	0.661
Western Europe	0.012	0.011	0.711
Asian N.H. U.S.-born	0.009	0.010	0.797
Black N.H. U.S.-born	0.106	0.101	0.728
Hispanic U.S.-born	0.080	0.083	0.797
Other N.H. U.S.-born	0.009	0.010	0.784
White N.H. U.S.-born	0.489	0.506	0.796
Canada	0.004	0.004	0.773
China	0.008	0.007	0.695
Colombia	0.007	0.007	0.763
Cuba	0.016	0.015	0.718
Dominican Rep.	0.012	0.011	0.698
El Salvador	0.010	0.010	0.772
Former U.S.S.R.	0.008	0.008	0.719
Germany	0.005	0.004	0.726
Guatemala	0.005	0.005	0.757
Haiti	0.006	0.006	0.784
India	0.010	0.010	0.785
Iran	0.004	0.003	0.718
Italy	0.006	0.005	0.650
Jamaica	0.009	0.009	0.814
Japan	0.003	0.003	0.691
Mexico	0.057	0.059	0.739
Philippines	0.013	0.014	0.821
Poland	0.006	0.006	0.732
Puerto Rico	0.015	0.012	0.613
South Korea	0.007	0.007	0.718
Taiwan	0.005	0.005	0.728
United Kingdom	0.005	0.005	0.827
Vietnam	0.008	0.007	0.715
Total	30,380,515	23,378,773	0.770

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity.

lack valid SSNs, the data used here might lead to the underestimation of co-ethnic exposure rates in the workplace.

To explore whether this is an issue in these data, the distributions of place of birth for the complete data sample and the labor force sample are shown in Table 1.[6] Approximately 70% of each sample is composed of the U.S. born population, with white non-Hispanics making up about half of each. Of the immigrant groups, the Mexican-born, with over 5% of both samples, is the largest. Every other group accounts for less than 2% of either sample, with most accounting for less than 1% of each sample. The last column in Table 1 reports the workforce sample, composed of the self-employed and those who report being in the labor force and have LEHD earnings, as a percentage of the residential sample. Recall that the residential data include individuals who are out of the labor force, hence one should not expect a 100% match rate between the residential and workforce samples. The U.S.-born groups exhibit a labor force attachment rate of just under 80%, with the exception of the black, non-Hispanic population which has a labor force attachment rate of just under 73%.

Table 1 shows that, as a proportion of the residential sample, groups with higher rates of undocumented migration do not show different shares in employment compared to groups with low rates of illegal migration. Passel (2006) estimates that 80-85% of Mexican immigrants who had been in the U.S. for less than 10 years in 2005 were undocumented. In this sample, which includes large Mexican-immigrant destinations such as Los Angeles and Chicago, Mexican immigrants represent equal shares of the residential and workforce population with a match rate of 73%, well in line with the other immigrant groups. If indeed undocumented immigrants are

[6] The exact sample size is not reported since it has not been released by the U.S. Census Bureau. The total immigrant residential sample is approximately 780,000 while the workforce sample is approximately 550,000.

not being fully captured in these data sources, it is probably the case that they are equally underrepresented in both the LEHD data and in the census data. On the other hand, some of the lowest employment shares belong to groups with low rates of illegal migration: Puerto Ricans, U.S. citizens by birth, have one of the lowest match rates at 63% while Italians, a group composed primarily of earlier immigrant cohorts, has a match rate of 65%.

1.2 Methodology

Demographers and other researchers have developed a variety of indices to measure spatial distributions of different groups and clustering behavior (Massey and Denton 1988; Iceland, Weinberg and Steinmetz 2002). One of the most popular of these measures is the index of dissimilarity. For group j, it is calculated as follows:

$$D_j = \quad {}_{k} \ \frac{N_k \ P_{jk} - P_j}{2NP_j(1 - P_j)}$$

where k denotes a census tract or other measurement area, N is the total population being studied, N_k is the total population in area k, P_j is the proportion of N that belongs to group j, and P_{jk} is the proportion of N_k that belongs to group j. It measures the proportion of group j that would need to change tracts (if k is tract) in order to achieve an even distribution over all tracts. For example, if D_j=0.2 then 20 percent of the population of group j would need to change tracts for it to be evenly distributed over tracts.

Carrington and Troske (1997) note that, along with most indices used to measure segregation, the index of dissimilarity “measure(s) the sample’s distance from evenness rather than randomness.” When applied to large analysis units such as census tracts this may not be a significant problem –

however, it translates into a significant issue when k is measured at smaller units, such as firm-level. To see why, consider a small universe composed of two, equal-sized groups: the blues (B) and the reds (R). Suppose these individuals are randomly allocated into two person firms. The result will be that about 25% of firms employ two B's, 25% employ two R's and the remaining 50% will employ one of each group. However, the resulting index of dissimilarity is 0.5, suggesting that half of the group in question needs to relocate in order for the two groups to be evenly distributed among employers. If we were studying segregation of blues and reds using this result, we would conclude that there is high employment segregation between the two groups, even though it is simply the artifact of random allocation into small firms.

Because of this characteristic and the fact that I am calculating exposure rates for firm-level units, I measure own-exposure as the simple fraction of neighbors or coworkers who are from the same place of birth as the individual. Since half of the R's work with no other R's while the other half work with exactly 100% R's, this simple measure results in an own-exposure rate of 0.5 when applied to the simple workplace problem presented above. This value is the same as the proportion in the overall population, implying that the workplace own-exposure rate is in line with the population proportion. This approach is similar to that used in Bayer, McMillan and Rueben (2004) to calculate the average exposure for racial groups in the San Francisco area by block group. Building on their approach, I calculate the average exposure between each group in both the Census tract of residence and the workplace using the following method:

Let n_{jk} be the number of individuals in ethnic/immigrant group $j \in J$ that live in census tract $k \in K$.[7] Then, $\;_j n_{jk} = N_k$, where N_k is the total population in tract k. Now, from the

[7] J includes the five U.S. born groups and each country or region of origin available in the LEHD data, as described above.

perspective of an individual i who is a member of group j, the proportion of his neighbors that belong to his ethnic group j is:

$$C_k^j = \frac{n_{jk} - 1}{N_k - 1}$$

which I will refer to as his residential co-ethnic exposure rate or residential own-exposure rate. Note that the denominator and numerator always exclude the individual. In other words, the residential co-ethnic exposure rate is the proportion of individual i's census tract that belongs to his group, excluding himself.

The residential exposure rate of an individual i from group j to members of a group different than his, j^*, is similarly calculated as follows:

$$E_k^{j^*} = \frac{n_{j^*k}}{N_k - 1}$$

The average exposure for members of some group j to members of any one group (including their own) in aggregate area *K* is:

$$E_K^{j,j'} = \frac{{}_{k\in K}\; n_{jk} \;\; \mathbf{1}_{j'=j} C_k^j + \mathbf{1}_{j'=j^*} E_k^{j^*}}{N_K^j}$$

Where $N_K^j = \;\; {}_{k\in K} n_{jk}$ is the total number of members of group j living in area *K*. The same methodology is carried out to construct measures of workplace co-ethnic exposure rate. Specifically, by substituting k for $w \in W$ as employer identifiers, the results are workplace exposure rates: $C_w^j, E_w^{j^*}$ and $E_W^{j,j'}$.[8]

Many researchers adjust their segregation or exposure measures to take into account the lower risk of having co-ethnic neighbors for smaller groups. For example, Bertrand, Luttmer and Mullainathan (2000) primarily use the following formula to measure the local network size (which they refer to as contact availability):

$$CA_{jk} = \ln \frac{C_{jk}/A_k}{L_j/T}$$

where the numerator is the proportion of the local population that is part of language group *j* in area *k* and the denominator is the proportion of the overall U.S. population that belongs to group *j*. Their measurement approach corrects for the overall size of a given language group, a characteristic that the own-exposure rates described above lacks. However, consider an extreme hypothetical case in which a very small ethnic group had all 100 of its members living in the same census tract. Though this group might not add up to a sizeable proportion of the tract, they are completely concentrated in a small geographic area - a quality captured by Bertrand, Luttmer and Mullainathan's measure. However, though geographically concentrated, this residential group is still only 100 possible co-ethnic contacts, regardless of the size of the co-ethnic population *not* in the tract. Measures adjusting for total population size capture a dimension of the geographical distribution of an ethnic population that, though important for many questions, does not necessarily inform the question exposure to residential or workplace ethnic networks. I do, however, adjust for census tract or firm size by using own-exposure rates based on proportions rather than raw counts. This controls for systematic tract size differences between

[8] Residential exposure rates are calculated based on the decennial census sample that is 16 years of age and older.

urban and suburban areas.[9]

An important caveat in interpreting these own-exposure rates is that, since these measures are based solely on place of birth rather than ethnic identity, they are sensitive to when these immigrant groups arrived in the United States. Consider a hypothetical census tract composed entirely of Italian immigrants in 1960. Assume that families do not leave the tract and new families do not enter. Even in this extreme hypothetical, the own-exposure rate drops as the U.S.-born children of these Italian immigrants reach the age of 16 and are counted towards the own-exposure rate. If these U.S.-born Italian-Americans continue to draw their social networks primarily from Italian immigrants and their descendants, this result is an underestimation of the actual role of ethnic networks in this population.[10] This same process might also explain why Mexican immigrants, though by far the largest immigrant group in the U.S. and known for large communities in Los Angeles and Chicago, do not have higher rates of co-ethnic residential exposure. It is possible that many second-generation and later Mexican-Americans continue to be part of the enclave, but the own-exposure measures exclude them based on their birthplace.

Section 2. Residential and Workplace Own-exposure Rates

2.1 Residential Ethnic Exposure Rates

The average residential co-ethnic exposure rate, $E_K^{j,j}$, reported in Table 2 shows that, even among the largest immigrant groups in five of the biggest immigrant destinations in the U.S., most

---

Workplace exposure rates are calculated based on coworkers between the ages 16 and 70.

[9] Census tracts are designed to contain between 1,500 and 8,000 people.

[10] This issue may be attenuated by using the decennial's ethnicity variable but, unfortunately, there is no equivalent variable in the LEHD files.

Table 2: Residential Own-Exposure Rates and Estimated Population Proportions Residing in Enclaves

Country of Birth	Residential own-exposure rates				Estimated proportion of POB group living in each type of tract	
	Average	Ratio to group size	90th percentile	Standard deviation	majority co-ethnic	25% or more co-ethnic
Africa	0.0251	5.020	0.0637	0.1076	0.00%	0.20%
Caribbean	0.0741	12.350	0.2074	0.2286	0.00%	3.90%
Central America	0.0568	6.311	0.1429	0.2292	0.00%	5.60%
Central Asia	0.0303	6.060	0.0826	0.1171	0.00%	0.20%
Middle East/N. Africa	0.0243	3.038	0.0596	0.0804	0.00%	0.00%
Oceania	0.0035	3.500	0.0095	0.0159	0.00%	0.00%
Socialist Europe	0.0294	4.900	0.0742	0.1013	0.00%	0.00%
South America	0.0713	3.961	0.1621	0.2339	0.00%	4.60%
Southeast Asia	0.0343	4.288	0.0834	0.125	0.00%	0.90%
Western Europe	0.0397	3.308	0.08	0.1994	0.40%	2.30%
Asian N.H. U.S.-born	0.0384	4.267	0.0848	0.154	.	.
Black N.H. U.S.-born	0.4698	4.432	0.9553	0.9394	.	.
Hispanic U.S.-born	0.1634	2.043	0.3247	0.3277	.	.
Other N.H. U.S.-born	0.0163	1.811	0.0321	0.0501	.	.
White N.H. U.S.-born	0.686	1.403	0.8968	0.5538	.	.
Canada	0.0088	2.200	0.019	0.0235	0.00%	0.00%
China	0.1105	13.813	0.3227	0.4141	5.30%	12.30%
Colombia	0.0411	5.871	0.1016	0.1213	0.00%	0.00%
Cuba	0.367	22.938	0.6929	0.7516	41.80%	60.90%
Dominican Rep.	0.1911	15.925	0.5153	0.5338	10.10%	32.90%
El Salvador	0.0727	7.270	0.1855	0.206	0.00%	3.80%
Germany	0.0077	1.540	0.0156	0.0186	0.00%	0.00%
Guatemala	0.0348	6.960	0.0852	0.1196	0.00%	0.40%
Haiti	0.1266	21.100	0.3189	0.3413	0.00%	16.60%
India	0.0527	5.270	0.1328	0.1932	0.00%	1.80%
Iran	0.0677	16.925	0.2057	0.2291	0.00%	4.60%
Italy	0.0318	5.300	0.083	0.0941	0.00%	0.00%
Jamaica	0.1061	11.789	0.2469	0.2759	0.00%	9.80%
Japan	0.016	5.333	0.0421	0.0787	0.00%	0.00%
Mexico	0.248	4.351	0.4963	0.4682	9.50%	45.50%
Philippines	0.0582	4.477	0.1479	0.2093	0.00%	3.60%
Poland	0.0928	15.467	0.2584	0.3134	0.90%	10.70%
Puerto Rico	0.0927	6.180	0.2308	0.2528	0.00%	6.80%
South Korea	0.0652	9.314	0.1824	0.245	0.00%	6.40%
Taiwan	0.0448	8.960	0.1195	0.1614	0.00%	1.10%
United Kingdom	0.0079	1.580	0.0175	0.0195	0.00%	0.00%
Former U.S.S.R.	0.1325	16.563	0.3639	0.4294	2.80%	21.20%
Vietnam	0.1218	15.225	0.3417	0.3882	2.60%	19.40%
Overall immigrant					4.70%	16.70%

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity.

immigrant groups do not live in neighborhoods of high co-ethnic exposure rates, though these rates are far higher than would be expected under random sorting.[11] For example, the average co-ethnic residential exposure rate for immigrants from India and the Philippines is about 0.05. That is, the average immigrant from India or the Philippines who resides in these five urban areas lives in a neighborhood where the co-ethnic population is only 5% of the total. The Cuban-born population, on the other hand, has an average own-exposure rate of 0.37, making it by far the most enclaved immigrant group in this sample. Recall from Table 1 that Cuban immigrants make up less than 2% of the sample, indicating that, in order to achieve such a high average own-exposure rate, they need to be concentrated in very few census tracts.

In fact, the next column, the ratio to group size, reports the ratio of the average own-exposure rate to the proportion of the total population belonging to that group. For example, the residential own-exposure rate for immigrants from sub-Saharan Africa ("Africa" in the table) is 5 times the proportion of sub-Saharan African immigrants in the sampled population. Even after adjusting for the population proportion, Cuban immigrants remain the most enclaved group, though they are closely trailed by immigrants from Haiti. Both of these groups exhibit residential own-exposure rates that are over 20 times larger than their overall population ratio. Note that the average residential own-exposure rate for Mexican immigrants, second only to Cuban immigrants, is only 4 times larger than their overall population ratio. That is of similar magnitude to the non-Hispanic U.S.-born Asian-American and African-American populations.

Immigrants born in Russia, Haiti, China, Vietnam and the Dominican Republic also exhibit relatively high average own-exposure rates, but still far lower than the Cuban-born or the U.S.-

[11] Note that all values in the column labeled "Average" are statistically different from 0 and from the proportion of the total population that belongs to the corresponding place of birth group.

born. On average, these groups live in neighborhoods where only about 10-18% of the population is from the same country of birth. Though this is a larger share than would be expected if individuals sorted randomly into neighborhoods (as easily verified in the second column), it is not what comes to mind when ethnic enclaves are discussed. At the 90th percentile, reported in the third column of Table 2, there is evidence of enclaving in other groups. Dominican immigrants at the 90th percentile, for example, live in neighborhoods where a majority of the adult population was born in the Dominican Republic. Immigrants from Vietnam, China, the former U.S.S.R., and Haiti stand out as well with rates in the 0.3 – 0.4 range.

In the enclave effects literature, there is no empirical definition of an enclave. The last two columns in Table 2 offer two possibilities: a census tract is an enclave of group j if 1) a majority of the tract population belongs to the foreign-born group in question, or 2) a quarter of the tract population is from that group. Under definition 1, only 5% of the immigrant population is considered enclaved - including 42% of the Cuban-born, 10% of the Dominican-born, 9% of the Mexican-born and 5% of the Chinese-born. Definition 2 results in almost 17% of immigrants living in immigrant enclaves. Under this definition, 61% of Cubans, 45% of Mexicans, 33% of Dominicans, 21% of those born in the former U.S.S.R. and 17% of Haitians live in enclaves.

Table 3 takes a different approach to the question of enclaving and exposure to different ethnic groups. The first column shows the name of the group to which each country of origin group has the highest average exposure while the second column reports this average rate. Only two immigrant groups, Cubans and Dominicans, have higher average exposure to their own group than to any other group, including any of the U.S.-born groups. Most other immigrant groups reside in tracts where the single largest adult group is white, non-Hispanic and U.S. born.

Table 3: Ethnic Group to Which POB has Highest Average Residential Exposure Rate, Over All CMSAs

Country of Birth	Overall maximum exposure group	Average exposure to maximum group	Maximum exposure, immigrant group	Average exposure to max immigrant group
Africa	White N.H. U.S.-born	0.3146	Mexico	0.0321
Caribbean	Black N.H. U.S.-born	0.2773	Caribbean	0.0741
Central America	White N.H. U.S.-born	0.2100	Cuba	0.1362
Central Asia	White N.H. U.S.-born	0.4268	South America	0.0353
MidEast/N Africa	White N.H. U.S.-born	0.5148	Mexico	0.0311
Oceania	White N.H. U.S.-born	0.5223	Mexico	0.0623
Socialist Europe	White N.H. U.S.-born	0.5203	Mexico	0.0311
South America	White N.H. U.S.-born	0.3328	South America	0.0713
Southeast Asia	White N.H. U.S.-born	0.4029	Mexico	0.0620
Western Europe	White N.H. U.S.-born	0.5687	Western Europe	0.0397
Asian N.H. U.S.-born	White N.H. U.S.-born	0.4759	Mexico	0.0464
Black N.H. U.S.-born	Black N.H. U.S.-born	0.4698	Mexico	0.0423
Hispanic U.S.-born	White N.H. U.S.-born	0.3556	Mexico	0.1247
Other N.H. U.S.-born	White N.H. U.S.-born	0.4910	Mexico	0.0632
White N.H. U.S.-born	White N.H. U.S.-born	0.6860	Mexico	0.0266
Canada	White N.H. U.S.-born	0.6145	Mexico	0.0302
China	White N.H. U.S.-born	0.3494	China	0.1105
Colombia	White N.H. U.S.-born	0.3522	Cuba	0.0850
Cuba	Cuba	0.3670	Cuba	0.3670
Dominican Rep.	Dominican Rep.	0.1911	Dominican Rep.	0.1911
El Salvador	White N.H. U.S.-born	0.2190	Mexico	0.1572
Germany	White N.H. U.S.-born	0.6367	Mexico	0.0277
Guatemala	White N.H. U.S.-born	0.2397	Mexico	0.1525
Haiti	Black N.H. U.S.-born	0.2558	Haiti	0.1266
India	White N.H. U.S.-born	0.5193	India	0.0527
Iran	White N.H. U.S.-born	0.5243	Iran	0.0677
Italy	White N.H. U.S.-born	0.6367	Italy	0.0318
Jamaica	Black N.H. U.S.-born	0.2932	Jamaica	0.1061
Japan	White N.H. U.S.-born	0.5158	Mexico	0.0353
Mexico	White N.H. U.S.-born	0.2644	Mexico	0.2480
Philippines	White N.H. U.S.-born	0.4097	Mexico	0.0749
Poland	White N.H. U.S.-born	0.5531	Poland	0.0928
Puerto Rico	White N.H. U.S.-born	0.2692	Puerto Rico	0.0927
South Korea	White N.H. U.S.-born	0.4333	South Korea	0.0652
Taiwan	White N.H. U.S.-born	0.4532	China	0.0537
United Kingdom	White N.H. U.S.-born	0.6055	Mexico	0.0236
Former U.S.S.R.	White N.H. U.S.-born	0.4440	Former U.S.S.R.	0.1325
Vietnam	White N.H. U.S.-born	0.3360	Vietnam	0.1218

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity.

Interestingly, Jamaican, Haitian and Caribbean immigrants,[12] all predominantly black, live in tracts where the largest group is non-Hispanic black Americans, reflecting the important role that race plays in residential choices. European immigrant groups (except for those born in the former U.S.S.R), as well as immigrants from Canada, the Middle East/North Africa, Oceania, India and Japan live in census tracts where more than 50% of the adult population is white, American-born non-Hispanics.

The third and fourth columns show the largest immigrant group to which each group is exposed and the average exposure rate to that group. Many country of birth groups live in census tracts where Mexican immigrants are the largest immigrant group. The Hispanic U.S.-born population has an exposure rate to Mexican immigrants of 0.1247, meaning that the average U.S.-born Hispanic in this sample lives in a tract where about 12% of the adult population was born in Mexico. This lends support to the argument above that the lack of Mexican-majority neighborhoods might be due to the number of second and later generation Mexican-Americans living in the same neighborhoods as those who were born in Mexico. Guatemalan and Salvadorian immigrants also show high average exposure rates to Mexican immigrants. Other Hispanic groups, however, do not. Both immigrants from other Central American countries and Colombian immigrants have higher rates of exposure to Cuban immigrants; about 0.085 and 0.136 respectively. Along with Dominican, Cuban, and Mexican immigrants, Puerto Ricans live in neighborhoods where they are the largest immigrant group. Several non-Hispanic groups also share this trait, including Vietnamese, Chinese, Korean, Russian, Polish, Haitian, and Iranian immigrants.

[12] The immigrants included in the Caribbean group are predominantly from Barbados, Trinidad and Tobago, and the Bahamas.

Lazear's (1999) model of ethnic segregation relied heavily on barriers to trade imposed by language and cultural differences to explain why immigrant groups cluster in host countries. To consider the impact of common language versus other source country differences on social networks, Table 4 shows the extent to which Spanish-speaking Latin American and U.S. born Hispanics segregate based on country of birth. The exposure rates reported are the average residential exposure rate of the group listed on the left column to the group listed on the top row. For example, the first cell is the average exposure rate of Central American immigrants to white, non-Hispanic U.S. natives. Note that the exposure of group *x* to group *y* is not the same as that of group *y* to group *x* since each group makes up different proportions of each neighborhood. The italicized values are the average own-exposure for each group as reported in Table 2. By reading across each row, it is easy to compare each group's own-exposure rate to its exposure rate of other Latin American/Hispanic groups. One relationship that becomes obvious is that all of the foreign-born Hispanic groups have relatively high exposure rates to the U.S.-born Hispanic population. This probably indicates a tendency of recent waves of Latin American/Hispanic immigrants to settle in areas with pre-existing Hispanic populations or potentially the settlement areas of previous waves of Hispanic immigrants.

Dominican and Puerto Rican immigrants, hailing from neighboring islands, have higher exposure rates to each other than to any other foreign-born Latin American/Hispanic group while Salvadorian and Guatemalan immigrants have higher exposure rates to Mexican immigrants than to other groups. Other Central American immigrants (Panamanians and Hondurans, for example) have relatively high average exposure rates to Mexican immigrants as well but, surprisingly, their exposure rates to Cuban immigrants are double their exposure rates to Mexican immigrants. This is almost certainly a function of some smaller Central American groups having chosen

Table 4: Cross-ethnic Residential Exposure Rates for Latin Immigrants

Place of Birth	White N.H. U.S.-born	Central America	South America	Hispanic U.S.-born	Mexico	Puerto Rico	El Salvador	Cuba	Guatemala	Colombia	Dominican Republic	Western Europe
Central America	0.2100	*0.0568*	0.0358	0.1041	0.0626	0.0285	0.0226	0.1362	0.0115	0.0179	0.0286	0.0084
South America	0.3328	0.0175	*0.0713*	0.0825	0.0247	0.0323	0.0109	0.0444	0.0055	0.0221	0.0361	0.0195
Hispanic U.S.	0.3556	0.0108	0.0179	*0.1634*	0.1247	0.0237	0.0154	0.0190	0.0075	0.0066	0.0182	0.0085
Mexico	0.2644	0.0102	0.0083	0.1984	*0.2480*	0.0084	0.0285	0.0053	0.0139	0.0029	0.0030	0.0048
Puerto Rico	0.2692	0.0165	0.0384	0.1283	0.0304	*0.0927*	0.0080	0.0280	0.0050	0.0123	0.0613	0.0122
El Salvador	0.2190	0.0202	0.0200	0.1365	0.1572	0.0122	*0.0727*	0.0172	0.0284	0.0075	0.0120	0.0079
Cuba	0.1939	0.0651	0.0449	0.0917	0.0161	0.0242	0.0093	*0.3670*	0.0057	0.0335	0.0240	0.0113
Guatemala	0.2397	0.0205	0.0200	0.1315	0.1525	0.0151	0.0569	0.0212	*0.0348*	0.0071	0.0096	0.0082
Colombia	0.3522	0.0228	0.0581	0.0795	0.0229	0.0271	0.0107	0.0850	0.0051	*0.0411*	0.0296	0.0186
Dominican Rep.	0.1508	0.0210	0.0541	0.1269	0.0135	0.0784	0.0098	0.0358	0.0040	0.0168	*0.1911*	0.0107
Western Europe	0.5687	0.0056	0.0274	0.0533	0.0193	0.0143	0.0059	0.0149	0.0030	0.0098	0.0097	*0.0397*

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity. Exposure rates are calculated as the exposure of the "row" country of birth to the "column" country of birth.

Miami as their primary destination. Western European immigrants, of which Spanish immigrants make up a small part, are included as a comparison immigrant group. South American, Puerto Rican, Cuban and Colombian immigrants have the highest exposure rates to Western European immigrants – though these rates are roughly a third to a half of Western European's own-exposure rate. In short, this table illustrates that there is no obvious "Hispanic" enclave though there is extensive regional clustering between some Hispanic groups.

2.2 Workplace ethnic exposure rates

We now move to the second dimension of enclaving: workforce co-ethnic exposure rates. These are constructed using analogous estimation methods as those for residential co-ethnic exposure rates relying on firm level data whenever possible. Workplace exposure rates were measured in three ways, depending on the size of the firm and on whether the individual reported being self-employed on the decennial census:

- In "large" firms (6 or more employees reporting UI earnings during the year), workplace co-ethnic exposure rate is measured as the proportion of an individual's coworkers who are co-ethnics.[13]
- In small firms (less than 6 employees reporting UI earnings), workplace co-ethnic exposure rate is calculated using the ethnicity of workers employed in firms in the same census block.[14] The underlying assumption is that individuals who work in the same geographic area are likely to be part of a labor network in a similar way to individuals

[13] About 38% of in-sample workers who work for large firms work for firms with multiple units. Results are robust to using establishment-level or firm-level own-exposure rate. The results presented here are based on establishment-level data. Firm-level results are available from author upon request.

[14] For the less than 1% of workers for whom firm block location was unavailable or who did not have at least 5 coworkers in the same census block, workplace own-exposure rate was based on workers in the census tract or, if tract unavailable or too small, the "sub-county" (akin to a town).

who work for the same employer.[15] This methodology is necessary to address some of the measurement problems inherent in looking at coworkers in small firms. By construction, exposure rates cannot be calculated for firms with only one employee. Furthermore, comparing workplace own-exposure for workers with only 3 coworkers to those with 50 coworkers would result in skewing the average own-exposure rate measures to the extremes since, with fewer coworkers, workers are more likely to either have 0 or 100% of coworkers be co-ethnics.

- For the self-employed, workplace own-exposure rates are based on the ethnic proportion of the census sample that is self-employed and in the same industry and census tract cell.

Table 5 shows the resulting workplace own-exposure rates by self-employment status and firm size for the employed. The first column shows that 85% of the sampled workforce works for employers with 6 or more employees, while another 11% are self-employed. The remaining 4% work for employers who have less than 6 employees. Using the approach detailed above, the average-own exposure rate is only slightly lower for immigrants in small firms than for those in large firms, indicating that the pseudo-employers created by combining all firms in the census block[16] leads to an acceptable approximation of coworker ties. In line with previous research that has shown significant entrepreneurial ethnic clustering by industry, the self-employed have higher shares of co-ethnics as "coworkers."

[15] This is similar to the assumption used by Bayer, Ross and Topa (2008) who show that individuals who live on the same block are also more likely to work on the same block – thereby indicating the presence of job networks by location of employer.
[16] Smaller industries were collapsed into similar industry groups to address issues arising from too few employers per industry group.

Table 5: Distribution of Employer Type and Average Workplace Own-Exposure by Employer Type

	Percent	Average own-exposure
Large firm	84.81	0.3825
Self-employed	10.55	0.4035
Small firm	4.64	0.3392

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File.

Tables similar to the residential exposure rates tables discussed above have been constructed using workplace exposure rates. Table 6 shows average own-exposure rates in the workplace by country of origin, and the proportion of each sample working with at least 25% and with at least 50% co-ethnics. Mexican, Cuban and Chinese immigrants work in workplaces where a little over 20% of their coworkers are co-ethnics; this is a similar magnitude to that experienced by African-Americans, a much larger group. None of the U.S.-born groups have higher own-exposure rates in their workplaces than in their neighborhoods. Except for Russian and Iranian immigrants, all Asian and European immigrant groups, exhibit the opposite tendency – making up smaller proportions of their neighborhoods than in their workplaces. For the most past, immigrants who are not from Latin America or the Caribbean work in workplaces that are more “ethnic” than their neighborhoods. This is especially pronounced for immigrants from Japan who, on average, live in neighborhoods where only 1.6% of the adults are Japanese-born but work in firms where almost 14% are Japanese-born. Similarly, South Korean and Chinese immigrants have workplace own-exposure rates double that of their residential own-exposure rates. With the exception of Colombian immigrants, Latin American groups have more co-ethnic exposure in their neighborhoods than at their workplaces. This also holds for all non-Latin Caribbean groups, Russians and Iranians.

Table 6: Workplace Own-Exposure Rates and Estimated Population Proportion Working in Enclaved Workplaces

Country of Birth	Workplace own-exposure rates				Estimated proportion of POB group working in each type of workplace	
	Average	Ratio to group size	90th percentile	Standard deviation	predominantly co-ethnic	25% or more co-ethnic
Africa	0.0328	6.5600	0.0769	0.2138	0.57%	2.01%
Caribbean	0.0343	5.7167	0.0780	0.1509	0.15%	1.02%
Central America	0.0483	5.3667	0.1328	0.2481	0.77%	3.55%
Central Asia	0.0590	11.8000	0.1538	0.3953	3.49%	7.47%
Middle East/N. Africa	0.0379	4.7375	0.0833	0.2601	1.31%	3.47%
Oceania	0.0077	7.7000	0.0092	0.0936	-	0.54%
Socialist Europe	0.0448	7.4667	0.1250	0.2972	1.63%	4.84%
South America	0.0621	3.4500	0.1522	0.2519	0.81%	4.19%
Southeast Asia	0.0538	6.7250	0.1326	0.3298	2.07%	5.51%
Western Europe	0.0513	4.2750	0.1356	0.2992	1.75%	5.44%
Asian N.H. U.S.-born	0.0334	3.7111	0.0680	0.1615		
Black N.H. U.S.-born	0.2462	2.3226	0.5058	0.5500		
Hispanic U.S.-born	0.1537	1.9213	0.2967	0.3101		
Other N.H. U.S.-born	0.0151	1.6778	0.0290	0.0860		
White N.H. U.S.-born	0.6172	1.2622	0.8750	0.5552		
Canada	0.0103	2.5750	0.0186	0.0865	0.08%	0.30%
China	0.2151	26.8875	0.7500	0.8303	22.03%	30.60%
Colombia	0.0443	6.3286	0.1041	0.2408	0.98%	2.90%
Cuba	0.2377	14.8563	0.5714	0.6542	14.41%	40.53%
Dominican Rep.	0.1487	12.3917	0.4180	0.5252	6.68%	21.01%
El Salvador	0.0680	6.8000	0.1711	0.2623	0.80%	5.23%
Germany	0.0118	2.3600	0.0165	0.1185	0.15%	0.73%
Guatemala	0.0299	5.9800	0.0769	0.1461	0.11%	1.14%
Haiti	0.0962	16.0333	0.2535	0.3596	1.91%	10.37%
India	0.1076	10.7600	0.3333	0.6120	7.91%	11.85%
Iran	0.0422	10.5500	0.0944	0.2845	1.61%	4.14%
Italy	0.0339	5.6500	0.0864	0.2033	0.43%	2.64%
Jamaica	0.0600	6.6667	0.1366	0.2690	0.97%	3.76%
Japan	0.1384	46.1333	0.5555	0.6945	11.80%	21.43%
Mexico	0.2250	3.9474	0.4878	0.4978	9.42%	37.60%
Philippines	0.0810	6.2308	0.2020	0.3574	2.35%	6.77%
Poland	0.1316	21.9333	0.4390	0.5907	8.73%	19.18%
Puerto Rico	0.0550	3.6667	0.1348	0.2259	0.47%	3.69%
South Korea	0.1331	19.0143	0.5437	0.6839	11.04%	16.26%
Taiwan	0.0750	15.0000	0.2500	0.4047	3.74%	10.27%
United Kingdom	0.0128	2.5600	0.0230	0.1083	0.16%	0.48%
Former U.S.S.R.	0.0954	11.9250	0.3419	0.5093	6.18%	12.75%
Vietnam	0.1382	17.2750	0.4635	0.6168	9.35%	17.88%
Overall immigrant	0.1113		0.2875	0.3963	5.18%	15.24%

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity.

The next column is the ratio of a group's workplace own-exposure rate to the proportion of the labor force represented by that group. A value of 1 signifies that the average workplace has the same proportion of co-ethnics as are present in the general working population, while a value greater than 1 implies higher than expected average own-exposure rates at work. Using this measure, Japanese immigrants again stand out as exceptionally segregated at the workplace: their own-exposure rates are 46 times the rate that would be expected given their relative population size. Chinese and Polish immigrants, with average workplace own-exposure rates of 27 and 22 times their population proportion, also exhibit exceptionally high workplace segregation rates. By comparison, Cuban workplace own-exposure rates are only 15 times the population rate, while, on the opposite end of the spectrum, German, British and Canadian immigrants have own-exposure rates that are less than 3 times their population rates.

The last two columns in Table 6 show what percentage of each group and the overall immigrant population would be labeled as enclaved using the same definitions as on Table 2 but applied to the workplace: 1) more than half of one's coworkers are co-ethnics, and 2) at least 25% of coworkers are co-ethnics. The estimated proportion of all immigrants in enclaved workplaces is almost identical to the proportion found to be enclaved for each definition using the residential-side exposure rates. Overall, only 5% work in firms where co-ethnics are the majority and 15% work in firms where co-ethnics are at least 25% of the workforce. Large proportions of Cuban (41%), Mexican (38%) and Chinese (31%) immigrants work in workplaces where at least a quarter of their coworkers are co-ethnics. Chinese immigrants are particularly likely to work with majority co-ethnics; over a fifth of all Chinese immigrants in the 5 urban areas being studied worked with 50% or more co-ethnics.

Table 7: Ethnic Group to Which POB has Highest Average Work Exposure Rate, Over All CMSAs

Country of Birth	Overall maximum exposure group	Average exposure to maximum group	Maximum exposure, immigrant group	Average exposure to max immigrant group
Africa	White N.H. U.S.-born	0.3833	Africa	0.0328
Caribbean	White N.H. U.S.-born	0.3781	Jamaica	0.0395
Central America	White N.H. U.S.-born	0.4430	Cuba	0.0820
Central Asia	White N.H. U.S.-born	0.3640	Central Asia	0.0590
MidEast/N Africa	White N.H. U.S.-born	0.3105	Mexico	0.0454
Oceania	White N.H. U.S.-born	0.3659	Mexico	0.0592
Socialist Europe	White N.H. U.S.-born	0.3819	Socialist Europe	0.0448
South America	White N.H. U.S.-born	0.4881	South America	0.0621
Southeast Asia	White N.H. U.S.-born	0.6172	Mexico	0.0590
Western Europe	White N.H. U.S.-born	0.4155	Western Europe	0.0513
Asian N.H. U.S.-born	White N.H. U.S.-born	0.4795	Mexico	0.0435
Black N.H. U.S.-born	White N.H. U.S.-born	0.5115	Mexico	0.0308
Hispanic U.S.-born	White N.H. U.S.-born	0.4321	Mexico	0.0739
Other N.H. U.S.-born	White N.H. U.S.-born	0.3354	Mexico	0.0509
White N.H. U.S.-born	White N.H. U.S.-born	0.3879	Mexico	0.0303
Canada	White N.H. U.S.-born	0.3468	Mexico	0.0389
China	White N.H. U.S.-born	0.4378	China	0.2151
Colombia	White N.H. U.S.-born	0.5647	Cuba	0.0617
Cuba	White N.H. U.S.-born	0.3816	Cuba	0.2377
Dominican Rep.	White N.H. U.S.-born	0.3112	DR	0.1487
El Salvador	White N.H. U.S.-born	0.2840	Mexico	0.1441
Germany	White N.H. U.S.-born	0.5507	Mexico	0.0345
Guatemala	White N.H. U.S.-born	0.5540	Mexico	0.1412
Haiti	White N.H. U.S.-born	0.2994	Haiti	0.0962
India	White N.H. U.S.-born	0.3670	India	0.1076
Iran	White N.H. U.S.-born	0.3797	Mexico	0.0713
Italy	White N.H. U.S.-born	0.3186	Italy	0.0339
Jamaica	White N.H. U.S.-born	0.3987	Jamaica	0.0600
Japan	White N.H. U.S.-born	0.3149	Japan	0.1384
Mexico	White N.H. U.S.-born	0.3922	Mexico	0.2250
Philippines	White N.H. U.S.-born	0.3678	Philippines	0.0810
Poland	White N.H. U.S.-born	0.3569	Poland	0.1316
Puerto Rico	White N.H. U.S.-born	0.4317	Puerto Rico	0.0550
South Korea	White N.H. U.S.-born	0.4141	South Korea	0.1331
Taiwan	White N.H. U.S.-born	0.2676	China	0.0838
United Kingdom	White N.H. U.S.-born	0.5250	Mexico	0.0316
Former U.S.S.R.	White N.H. U.S.-born	0.4452	Former U.S.S.R.	0.0954
Vietnam	White N.H. U.S.-born	0.4944	Vietnam	0.1382
Overall Immigrant	White NH US-born	0.3722	Mexico	0.3638

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity.

Table 7, the workplace equivalent of Table 3, shows that, on average, individuals from all ethnic groups work for employers where the largest racial/ethnic group is U.S.-born, Non-Hispanic white. The average work exposure rate to white, non-Hispanic U.S.-born individuals among immigrants is 37, ranging from a high of 62% for immigrants from smaller country of birth groups in Southeast Asia to a low of 27% for Taiwanese immigrants. The third and fourth columns report the largest immigrant group in the workplace for each immigrant group. Most groups either work in firms where the largest immigrant group is their own or it is Mexican immigrants. The only exceptions are Caribbean, Central American, Colombian and Taiwanese immigrants. Colombian and Central American immigrants are more likely to work with Cuban-born coworkers than with co-ethnics while Caribbean immigrants are more likely to work with Jamaican immigrants and Taiwanese immigrants are more likely to work with Chinese immigrants.

Table 8 shows that, for most Hispanic/Latin American groups, the largest Hispanic group of coworkers is made up of U.S.-born Hispanics. The five exceptions are immigrants from Mexico, El Salvador and Guatemala, who on average work with more Mexican-born, and Cuban and Dominican immigrants who are more likely to work with co-ethnics than with any other Spanish-speaking Latin American group.

Table 8: Cross-ethnic Workplace Exposure Rates for Latin Immigrants

Place of Birth	Central America	South America	Hispanic U.S.-born	Mexico	Puerto Rico	El Salvador	Cuba	Guatemala	Colombia	Dominican Republic	Western Europe
Central America	*0.0483*	0.036	0.1033	0.0558	0.0239	0.0182	0.0820	0.0089	0.0188	0.0248	0.0098
South America	0.017	*0.0621*	0.0837	0.0275	0.0264	0.0110	0.0334	0.0052	0.0190	0.0342	0.0158
Hispanic U.S.	0.0081	0.0143	*0.1537*	0.0739	0.0122	0.0115	0.0113	0.0054	0.0050	0.0085	0.0074
Mexico	0.009	0.0094	0.1517	*0.2250*	0.0065	0.0269	0.0035	0.0128	0.0029	0.0022	0.0065
Puerto Rico	0.0139	0.0343	0.0928	0.0205	*0.0549*	0.0067	0.0242	0.0038	0.0132	0.0396	0.0123
El Salvador	0.0167	0.0194	0.1319	0.1441	0.0100	*0.0680*	0.0092	0.0202	0.0075	0.0111	0.0090
Cuba	0.0469	0.0417	0.0852	0.0156	0.025	0.0065	*0.2377*	0.0049	0.0283	0.019	0.0095
Guatemala	0.0161	0.0208	0.1286	0.1412	0.0121	0.0414	0.0143	*0.0299*	0.0075	0.0105	0.0089
Colombia	0.0243	0.0518	0.0827	0.0243	0.0284	0.0123	0.0617	0.0056	*0.0443*	0.0342	0.0141
Dominican Rep.	0.0193	0.0552	0.0834	0.0106	0.0483	0.0103	0.0254	0.0042	0.0206	*0.1487*	0.0132
Western Europe	0.0074	0.0259	0.0664	0.0274	0.0152	0.0077	0.0121	0.0035	0.0086	0.0136	*0.0513*

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. The U.S.-born population is reported by racial/ethnic group where N.H. designates non-Hispanic ethnicity. Exposure rates are calculated as the exposure of the "row" country of birth to the "column" country of birth.

Section 3. The Value of Linked Data in Identifying Enclaves

3.1 Identifying Enclaves

Suppose ethnic social networks are formed via two types of social interactions: residential and workplace proximity. The following matrix captures the possible relationships between two co-ethnic residents of the same CMSA:

	Same Employer	Different Employer
Same Neighborhood	Enclave	Residential Network
Different Neighborhood	Job Network	No Ethnic Network

The traditional notion of an enclave economy is best represented by the top-left cell: co-ethnics live in the same locations and often work for the same firms. The bottom-right cell contains individuals who are not reliant on ethnic social networks for residence or job referrals. Those individuals who live in an ethic neighborhood but work outside of the ethnic labor market and those who live outside of the enclave but work with co-ethnics form two interesting hybrids: one group branching out through the labor market and the other branching out residentially. As discussed above, most empirical research on immigrant enclaves has relied on only residential information – hence, it has been unable to differentiate between those who live and work in an enclave and those who only may only choose to live near co-ethnics.

An ethnic enclave should be thought of as a social network composed of both residential and labor connections. As a first step to identifying ethnic enclaves in this sample, Table 9 lists the Pearson correlation coefficients of residential own-exposure rate to workplace own-exposure rate for each of the immigrant populations identified in these data. Over all immigrants, the

Table 9: Correlation Between Work and Residential Own-exposure Rates

Africa	0.2145
Caribbean	0.2571
Central America	0.3619
Central Asia	0.1673
Middle East/N. Africa	0.1283
Oceania	0.0636
Socialist Europe	0.1863
South America	0.2275
Southeast Asia	0.1448
Western Europe	0.2920
Canada	0.0983
China	0.3594
Colombia	0.1922
Cuba	0.4391
Dominican Rep.	0.2534
El Salvador	0.1978
Germany	0.0553
Guatemala	0.1942
Haiti	0.2021
India	0.2000
Iran	0.1872
Italy	0.1433
Jamaica	0.1652
Japan	0.1790
Mexico	0.2455
Philippines	0.1391
Poland	0.2523
Puerto Rico	0.2557
South Korea	0.1939
Taiwan	0.1761
United Kingdom	0.1006
Former U.S.S.R.	0.2403
Vietnam	0.2493
All Immigrants	0.4307

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File.

correlation coefficient is 0.43 – indicating a high degree of concordance between residential and workplace own-exposure rates. That is, without taking any individual characteristics into account, individuals with low levels of residential own-exposure also exhibit low-levels of workplace own-exposure and vice versa. All listed groups exhibit positive correlation rates, though once again the Cuban-born population shows a unique tendency to enclave. The correlation coefficient for this group is a strong positive value of 0.44 indicating that Cuban immigrants who reside in high co-ethnic density neighborhoods also work with a large share of co-ethnic coworkers. Chinese immigrants and those from Central America also exhibit a high, positive correlation between workplace and residential own-exposure rates.

Table 10 expands this correlation analysis by showing the percentage of immigrants by their values on both dimensions of co-ethnic exposure: residential and workplace. The top section of the table reports the percentage that the combination of residential and workplace own-exposure represents in the total sample. The second section of the table, labeled row percentage, reports what percentage of individuals with residential own-exposure of that value have workplace own-exposure of the values along the top row. The third section is the column percentage, reporting what percentage of the workplace own-exposure group along the top row has the value of residential own-exposure along the column. For example, the upper left hand corners in each of the three sections show the following: 1) 27% of all immigrants have less than 2.5% of their neighbors or coworkers belonging to their country of birth group, 2) for those who live in neighborhoods with less than 2.5% co-ethnic neighbors, 71% also have less than 2.5% co-ethnic coworkers, and 3) among all workers for whom co-ethnics represent less than 2.5% of their coworkers, 62% also live in neighborhoods with less than 2.5% co-ethnics.

Table 10: Distribution of Immigrants, by Residential and Workplace Own-Exposure Rates

%	Workplace own-exposure rate									
Residential own-exposure rate	< 0.025	0.025 - 0.05	0.05 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.75	0.75 - 1	Total
< 0.025	27.46	4.29	3.00	1.92	0.71	0.40	0.28	0.37	0.23	38.65
0.025 - 0.05	6.50	2.34	2.00	1.40	0.54	0.30	0.20	0.26	0.14	13.68
0.05 - 0.1	4.77	2.31	2.31	1.89	0.80	0.48	0.30	0.37	0.17	13.40
0.1 - 0.2	3.34	2.03	2.41	2.50	1.28	0.84	0.55	0.63	0.24	13.81
0.2 - 0.3	1.34	0.88	1.18	1.57	0.92	0.67	0.48	0.48	0.12	7.64
0.3 - 0.4	0.60	0.41	0.68	1.07	0.68	0.53	0.41	0.43	0.10	4.92
0.4 - 0.5	0.28	0.20	0.39	0.72	0.52	0.45	0.35	0.37	0.09	3.37
0.5 - 0.75	0.31	0.22	0.42	0.88	0.63	0.65	0.48	0.65	0.19	4.44
0.75 - 1	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.08
Row %	Workplace own-exposure rate									
Residential own-exposure rate	< 0.025	0.025 - 0.05	0.05 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.75	0.75 - 1	Total
< 0.025	71.05	11.11	7.76	4.96	1.84	1.04	0.72	0.95	0.58	38.65
0.025 - 0.05	47.53	17.09	14.61	10.24	3.96	2.16	1.49	1.88	1.03	13.68
0.05 - 0.1	35.62	17.25	17.22	14.14	5.94	3.55	2.26	2.73	1.29	13.40
0.1 - 0.2	24.16	14.70	17.44	18.13	9.24	6.10	3.97	4.54	1.72	13.81
0.2 - 0.3	17.49	11.54	15.38	20.54	12.06	8.80	6.30	6.28	1.61	7.64
0.3 - 0.4	12.16	8.28	13.87	21.82	13.86	10.87	8.25	8.82	2.08	4.92
0.4 - 0.5	8.36	6.04	11.48	21.43	15.41	13.25	10.45	10.85	2.72	3.37
0.5 - 0.75	7.05	4.97	9.40	19.83	14.24	14.68	10.92	14.69	4.23	4.44
0.75 - 1	4.89	4.21	8.27	15.56	13.06	17.08	13.37	17.67	5.88	0.08
Column %	Workplace own-exposure rate									
Residential own-exposure rate	< 0.025	0.025 - 0.05	0.05 - 0.1	0.1 - 0.2	0.2 - 0.3	0.3 - 0.4	0.4 - 0.5	0.5 - 0.75	0.75 - 1	Total
< 0.025	61.56	33.83	24.22	16.00	11.67	9.25	9.08	10.27	17.57	38.65
0.025 - 0.05	14.58	18.42	16.15	11.70	8.91	6.82	6.66	7.20	10.98	13.68
0.05 - 0.1	10.70	18.22	18.63	15.82	13.07	10.99	9.85	10.29	13.40	13.40
0.1 - 0.2	7.48	16.00	19.45	20.91	20.96	19.43	17.87	17.62	18.48	13.81
0.2 - 0.3	3.00	6.95	9.49	13.11	15.13	15.52	15.70	13.48	9.55	7.64
0.3 - 0.4	1.34	3.21	5.51	8.97	11.19	12.34	13.22	12.18	7.95	4.92
0.4 - 0.5	0.63	1.60	3.12	6.03	8.53	10.31	11.48	10.27	7.12	3.37
0.5 - 0.75	0.70	1.74	3.37	7.35	10.38	15.04	15.80	18.31	14.60	4.44
0.75 - 1	0.01	0.03	0.05	0.10	0.17	0.30	0.34	0.38	0.35	0.08
Total	44.61	12.69	12.38	11.97	6.09	4.33	3.07	3.56	1.29	100

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File.

Relying only on the overall percentage section of the table, one can easily gauge the size of the "enclaved" immigrant population for different empirical definitions of enclave by selecting cut-off values for residential and workplace own-exposure. Let us consider some potential cut-off values for own-exposure rates and the resulting sizes of the enclave population. Selecting only immigrants who have both own-exposure rates of over 0.5 (they live and work with mostly co-ethnics) results in less than 1% of the population being enclaved. Extending the definition of enclaves to individuals who both work and live with 20% or more co-ethnics increases the reach of enclaves to include 9% of all immigrants. Including all individuals who live with 20% or more co-ethnics, regardless of where they work, expands the enclave definition to include just over 20% of all immigrants in these 5 metropolitan areas. On the other hand, including all individuals who work with at least 20% co-ethnics results in about 18% of immigrants being categorized as enclaved. This exercise confirms that enclaving is relatively rare among immigrants in the U.S., particularly when one considers that the sample selected for this analysis is composed of immigrant destination cities, which are the most likely hosts for high levels of neighborhood ethnic enclaves. Indeed, nearly 55% of all immigrants in this sample neither live nor work with more than 10% co-ethnics.

3.2 Predicting Enclaves: Selection based on observables

The tract-level and firm-level residential and workplace own-exposure rates discussed throughout this paper can only be calculated using restricted access linked data such as the LEHD. This next section quantifies how much of the variation in own-exposure rates calculated at such fine levels of detail are captured by research using other, more aggregated data sources. In order to get a sense of how well observable characteristics predict who lives and/or works with co-ethnics, two sets of OLS regressions, each predicting either the value of residential own-exposure or workplace own-exposure, are reported below.[17]

Recall that C_k^j is an individual's tract-level own-exposure rate and C_w^j is workplace own-exposure rate at the firm level. Let C_h^j designate either residential own-exposure or workplace own-exposure rate at some geographical level. The following formulate is used to calculate workplace and residential own-exposure rates at different levels of aggregation:

$$C_h^j = \frac{n_{jh} - 1}{N_h - 1}$$

where $h \in \;\; k, k^*, k^{**}, w, w^*, w^{**} \;\; \in \;\; k, k_{PMSA}, k_{CMSA}, w, w_{PMSA}, w_{CMSA}$. The higher geographical levels used here are the CMSA (e.g., New York City) and the PMSA (e.g., Newark, a Primary MSA within the New York City CMSA). These two additional geographical levels are being included since they can be estimated using public-use data easily. Hence, their inclusion will allow for a measurement of how much variation in neighborhood clustering is being captured with other data sources. Analogous to our previous tract-level notation, n_{jh} is the number of individuals in immigrant/ethnic group j in the geographical area h, and N_h is the total population in geographical area h.

The generic regression model is as follows:

$$C_t^j = X_i\beta_1 + \beta_2 CMSA + \beta_3 POB + \beta_4 C_h^j + e_i$$

where $t \in \;\; k, w \;\;, \;\; h \in \;\; k, k^*, k^{**}, w, w^*, w^{**}$ and $t \neq h$.

[17] Note that causality is not being established in these regressions. An important consideration in designing empirical models for research on enclaving is being able to control for unobservable characteristics. When not properly addressed, these may result in biased estimates of outcomes such as earnings and children's educational attainment due to omitted variable bias. For example, if immigrants who are more likely to choose to live in areas of high co-

The matrix X_i contains widely-available individual explanatory variables including age, gender, race, ethnicity, marital status, years since migration, English language skills, country of birth, self-employment status, and educational attainment which are used to explain each measure of co-ethnic exposure. $CMSA$ and POB are vectors of CMSA and place of birth dichotomous variables to control for CMSA-level and place of birth characteristics, including country-specific differences in selection into migration (Borjas 1987).

The aim of this exercise is not to establish causation, but rather, to identify which variables offer explanatory power for own-exposure rates and to identify how much variation can be explained by the proposed empirical model. The magnitude and significance of the estimated coefficients in the regression indicate which variables lend explanatory power to this model. Furthermore, the coefficient of determination, the R^2, calculated by OLS provides a simple measure of how much variation in residential and workplace clustering is explained by the observables. This implies that the variation *not* explained by the observables is simply $(1 - R^2)$.

Table 11 shows that the average residential own-exposure rate in these samples are 0.1180 and 0.1147.[18] When measured at the CMSA level, this measure drops to 0.0343 and 0.0335. That is, the average immigrant in this sample lives in a CMSA where just over 3% of the adult population is from her country of birth. The PMSA measure of own-exposure rate is higher at 0.0425 and 0.0412, illustrating that immigrants do not randomly distribute themselves within the CMSA but rather gravitate towards parts of the CMSA where other co-ethnics already reside. The table also reports averages for the explanatory variables used in the regressions.

ethnic exposure are also less likely to invest in U.S.-specific human capital, some of the negative effects attributed to residential location would, in fact, have been present regardless of where they lived.

[18] The workplace sample is the subset of the residential sample that is in the labor force and reports positive net earnings.

Table 11: Demographic Information of the Residential and Workplace Samples

Variable	Residential Mean	Residential *S.D.*	Workplace Mean	Workplace *S.D.*
Residential own-exposure	0.1180	*0.1581*	0.1147	*0.1554*
Workplace own-exposure			0.1215	*0.1829*
Res. Own-exp, CMSA	0.0343	*0.0470*	0.0335	*0.0464*
Res. Own-exp, PMSA	0.0425	*0.0647*	0.0412	*0.0631*
Work. Own-exp, CMSA			0.3383	*0.4547*
Work. Own-exp, PMSA			0.0413	*0.0595*
Age	45.3633	*11.9462*	43.7910	*10.7545*
Years since migration	16.3587	*10.6745*	15.6727	*9.7717*

	Residential %	Workplace %
Male	48.36	54.86
Married	72.00	72.33
Was married	15.26	14.22
Education		
8 years or less	20.78	17.44
Some high school	15.12	14.25
High school diploma	19.40	18.92
Some college	18.64	19.92
College degree	15.30	17.00
Graduate/Professional degree	10.75	12.48
Race		
White	61.18	59.70
Black	11.63	12.66
Native American	0.58	0.57
Asian	25.16	25.62
Pacific Islander/Hawaiian	0.18	0.17
Other/Multiple Races	1.26	1.28
Hispanic	43.03	42.10
U.S. Citizen	47.59	47.65
Speaks English	53.96	58.00
Employer type		
Large firm		79.99
Self-employed		13.79
Small firm		6.21

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File.

3.3 Predicting residential own-exposure rates

OLS regressions predicting residential own-exposure rates are reported on Table 12. The data universe for the regressions (I) to (VII) on residential own-exposure rates is all adult immigrants over the age of 18, regardless of their labor force participation. Model (I) uses only the individual's demographic characteristics, excluding any immigrant-specific variables, to explain residential own-exposure. The resulting R^2 indicates that 14% of the variation is explained using just these variables, with the bulk of the explanatory power belonging to the Hispanic indicator. Interestingly, neither race nor age affected the propensity of immigrants to reside in high co-ethnic areas. The inclusion of education in model (II) results in a modest increase in the variation that is explained. It also indicates that immigrants without a high school diploma are more likely to live in areas with higher own-exposure rates. Model (III) adds immigrant specific demographic variables on years since migration, citizenship and English ability. Of these, only English ability has a statistically significant coefficient indicating that immigrants who do not speak English live in areas with higher own-exposure rates. At this point, the R^2 is up to 0.18 – more than one-sixth of the variation in residential own-exposure rates is explained by individual-level demographic variables.

The inclusion of CMSA and place of birth variables boosts the R^2 to almost 0.44, with half of the model's explanatory power coming from controlling for place of birth. Including place of birth also decreases the magnitude on the coefficients of all the demographic variables indicating that failing to control for country of origin can lead to serious omitted variable bias. Model (VI) also adds $C_{k^{**}}^{j}$, the residential own-exposure rate at the CMSA level. This additional variable pushes the model's explanatory power over 50% and also decreases the magnitude of the coefficients on the demographic variables. Though this is a powerful addition to the model, replacing it with the more exact $C_{k^*}^{j}$, the residential own-exposure rate at the PMSA level, results in an R^2 of 0.54. Thus, by using only variables available in most publicly available data sets, more than half of the variation in predicting who lives in areas with more co-ethnics can be explained.

Model (VIII) replaces the residential own-exposure rate variables at the larger geographic area with the workplace own-exposure variable. Why might this variable matter? We know from previous research that individuals are more likely to work with their neighbors (Bayer, Ross and Topa 2008; Andersson et al. 2010) even without considering any ethnic connections. Hence, if an individual works with many co-ethnics, it is also likely that some of those co-ethnics also live in his neighborhood. Also, we saw above that there is substantial correlation between the two own-exposure measures. Though the R^2 increases by 0.03 over model (V), the workplace own-exposure is not as good a predictor of residential own-exposure as the aggregate residential own-exposure rates in models (VI) and (VII). That is, the local size of the ethnic population is more informative when predicting own-exposure rates than the individual's observed tendency to work with other co-ethnics.

Table 12: OLS Regression Results: Explaining Residential Own-Exposure Rates

Model	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Hispanic	0.1240***	0.1000**	0.1020**	0.0865***	-0.0105	-0.0060	-0.0068	-0.0102
	(0.0392)	(0.0400)	(0.0401)	(0.0264)	(0.0075)	(0.0072)	(0.0068)	(0.0075)
Some High School		-0.0242	-0.0183	-0.0245**	-0.0165***	-0.0152***	-0.0156***	-0.0134***
		(0.0207)	(0.0200)	(0.0095)	(0.0035)	(0.0036)	(0.0035)	(0.0035)
High School Diploma		-0.0526**	-0.0432**	-0.0497***	-0.0286***	-0.0246***	-0.0243***	-0.0237***
		(0.0220)	(0.0214)	(0.0107)	(0.0045)	(0.0046)	(0.0045)	(0.0043)
Some College		-0.0659***	-0.0513**	-0.0614***	-0.0367***	-0.0307***	-0.0307***	-0.0302***
		(0.0237)	(0.0239)	(0.0133)	(0.0052)	(0.0052)	(0.0051)	(0.0048)
College Degree		-0.0734***	-0.0576**	-0.0663***	-0.0437***	-0.0379***	-0.0377***	-0.0364***
		(0.0263)	(0.0257)	(0.0142)	(0.0055)	(0.0056)	(0.0055)	(0.0050)
Graduate/Professional Degree		-0.0799***	-0.0638**	-0.0703***	-0.0510***	-0.0459***	-0.0449***	-0.0413***
		(0.0270)	(0.0266)	(0.0133)	(0.0059)	(0.0062)	(0.0058)	(0.0052)
Citizen			-0.0039	-0.0037	-0.0059***	-0.0046**	-0.0045**	-0.0034*
			(0.0084)	(0.0065)	(0.0021)	(0.0020)	(0.0021)	(0.0018)
English			-0.0320***	-0.0309***	-0.0227***	-0.0199***	-0.0181***	-0.0171***
			(0.0088)	(0.0071)	(0.0036)	(0.0037)	(0.0036)	(0.0030)
Co-ethnic Exposure Measure						2.1410***	1.6990***	0.1530***
						(0.2110)	(0.1070)	(0.0176)
Years since migration			X	X	X	X	X	X
Years since migration squared			X	X	X	X	X	X
CMSA				X	X	X	X	X
POB					X	X	X	X
R-squared	0.140	0.167	0.181	0.217	0.439	0.508	0.543	0.469

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. All regressions include controls for gender, marital status, race, age, and age-squared.

Model (VI) uses residential co-ethnic exposure measured at the CMSA level for Co-ethnic Exposure Measure. Model (VII) uses residential co-ethnic exposure measured at the PMSA level while model (VIII) uses workplace co-ethnic exposure.

Robust, clustered standard errors in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

3.4 Predicting workplace own-exposure

Predicting workplace own-exposure rates using only publicly available data is significantly more difficult than predicting residential own-exposure, as is shown on Table 13. The first batch of demographic variables (age, gender, marital status, race, and ethnicity) explain half as much of the variation in workplace own-exposure rate as they explained for residential own-exposure rate.[19] Including education and immigrant-specific demographic variables further increases the R^2 to 0.119, less than was explained of the residential own-exposure using just the basic batch of demographic characteristics. Adding CMSA and place of birth variables nearly doubles the proportion of the variation that is explained by the observables, though the explanatory power of this model still pales in comparison to the same model applied to residential own-exposure.

Models (VI) through (IX) explore which aggregate measures are the best predictors of workforce own-exposure. The candidates are residential own-exposure at the tract level, residential own-exposure at the CMSA level, residential own-exposure at the PMSA level, and the workplace own-exposure rate at the PMSA level (the proportion of the labor force in the individual's PMSA who is from his/her country of origin). One might expect that, of the measures utilizing aggregated geographies, ones based on the workforce would serve as superior explanatory variables since they exclude the non-labor force population. However, both the PMSA and CMSA (not included in Table 13) workforce aggregate own-exposure measures have the same explanatory power as their residential counterparts. Model (VI), which includes the individual's residential own-exposure at the tract level, offers the most explanatory power of the set of models used to predict workplace own-exposure. The gain in the R^2 between model (VI) and

[19] Because workplace own-exposure is calculated using different methodologies for each of the three types of employers, the employer type variables are included in each of the models predicting workplace own-exposure.

(VIII) is minimal, however. As with the residential own-exposure rates, the place of birth group as a proportion of the PMSA population proves to be a powerful variable in explaining neighborhood-level and workplace-level own-exposure rates.

3.5 Relationship between earnings and own-exposure rates

As a first pass at the relationship between enclaving and the economic success of immigrants, Table 14 reports the coefficients from regressing the log of self-reported earnings in 1999 on both of the own-exposure rates as well as the exposure rates calculated at the PMSA level. As in the earlier models explored above, these regressions do not establish causality since self-selection has not been addressed. Specifically, it is entirely plausible that unmeasured personal heterogeneity leads immigrants with lower earning potentials to seek out neighborhoods and/or workplaces with higher co-ethnics. In line with previous research, immigrants who reside in neighborhoods with higher concentrations of co-ethnics report lower earnings (Borjas 2000; Xie and Gough 2011). The coefficient indicates that residing in an all co-ethnic neighborhood is associated with earnings that are 29% lower than if living with no co-ethnics. A neighborhood of 10% co-ethnics, thus, implies expected earnings are 2.9% lower than would otherwise be expected. Similarly, immigrants with greater proportions of co-ethnic coworkers also report lower earnings. Working in a firm with 10% co-ethnic coworkers, close to the sample mean, is associated with earning 1.4% less than working with no co-ethnics. Model (IV) shows that some of the wage decrease associated with workplace own-exposure is explained by residential own-exposure. Once the residential enclaving has been taken into account, workplace own-exposure decreases earnings by 1% and is still significant at the 90% confidence level. Models (V) through (VII) show that, in the absence of neighborhood-level and employer-level data, immigrant own-

Table 13: OLS Regression Results: Explaining Workplace Own-Exposure Rates

Model	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
Hispanic	0.0886***	0.0527	0.0566	0.0486**	-0.0001	0.0029	0.0029	0.0022	0.0026
	(0.0320)	(0.0295)	(0.0287)	(0.0217)	(0.0083)	(0.0084)	(0.0075)	(0.0075)	(0.0076)
Some High School		-0.0399**	-0.0321**	-0.0342***	-0.0224***	-0.0174***	-0.0218***	-0.0224***	-0.0223***
		(0.0172)	(0.0151)	(0.0088)	(0.0041)	(0.0033)	(0.0041)	(0.0040)	(0.0040)
High School Diploma		-0.0701***	-0.0582***	-0.0599***	-0.0372***	-0.0286***	-0.0345***	-0.0349***	-0.0347***
		(0.0203)	(0.0181)	(0.0113)	(0.0056)	(0.0046)	(0.0056)	(0.0055)	(0.0055)
Some College		-0.0989***	-0.0799***	-0.0841***	-0.0553***	-0.0440***	-0.0511***	-0.0519***	-0.0515***
		(0.0234)	(0.0211)	(0.0149)	(0.0077)	(0.0066)	(0.0076)	(0.0075)	(0.0075)
College Degree		-0.110***	-0.0915***	-0.0945***	-0.0641***	-0.0507***	-0.0599***	-0.0607***	-0.0604***
		(0.0277)	(0.0241)	(0.0185)	(0.0106)	(0.0093)	(0.0105)	(0.0104)	(0.0104)
Graduate/Professional Degree		-0.124***	-0.106***	-0.107***	-0.0836***	-0.0679***	-0.0799***	-0.0800***	-0.0798***
		(0.0289)	(0.0252)	(0.0192)	(0.0148)	(0.0131)	(0.0149)	(0.0145)	(0.0145)
Citizen			-0.0216***	-0.0211***	-0.0180***	-0.0162***	-0.0170***	-0.0172***	-0.0171***
			(0.0066)	(0.0051)	(0.0021)	(0.0018)	(0.0020)	(0.0021)	(0.0021)
English			-0.0350***	-0.0343***	-0.0334***	-0.0269***	-0.0314***	-0.0306***	-0.0304***
			(0.0090)	(0.0082)	(0.0062)	(0.0057)	(0.0063)	(0.0064)	(0.0064)
Co-ethnic Exposure Rate						0.291***	1.577***	1.102***	1.188***
						(0.0371)	(0.114)	(0.114)	(0.112)
Years since migration			X	X	X	X	X	X	X
Years since migration squared			X	X	X	X	X	X	X
CMSA				X	X	X	X	X	X
POB					X	X	X	X	X
R-squared	0.076	0.119	0.138	0.148	0.266	0.300	0.292	0.296	0.297

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. All regressions include controls for working in a small firm, being self-employed, gender, marital status, race, age, and age-squared. Model (VI) uses residential co-ethnic exposure measured at the tract level for Co-ethnic Exposure Measure. Model (VII) uses residential co-ethnic exposure measured at the CMSA level, model (VIII) uses residential co-ethnic exposure measured at the PMSA level, and model (IX) uses workplace co-ethnic exposure measured at the PMSA level.
Robust, clustered standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

exposure based on the overall proportion of the PMSA population offers approximately the same explanatory power as the measures based on census tract and employer. As expected, the coefficients on these two measures differ significantly – a 10% increase in the co-ethnic share at the tract-level is associated with a 3% decline in earnings while a 10% increase in the PMSA co-ethnic share implies a 5.6% decline in earnings. Model (VII) simply shows that, as one might guess, the residential and workplace own-exposure rates at the PMSA level do not vary sufficiently to use together in a regression model.

Table 14. The Role of Residential and Workplace Own-Exposure Rates in Reported Earnings

Model	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Residential Co-ethnic Exp. Rate		-0.296***		-0.271***	-0.559***		1.436
		(0.0370)		(0.0280)	(0.1560)		(1.715)
Workplace Co-ethnic Exp. Rate			-0.144**	-0.100*		-0.605***	-2.118
			(0.0567)	(0.0554)		(0.162)	(1.848)
Some High School	0.091***	0.086***	0.0877***	0.0841***	0.091***	0.0908***	0.0906***
	(0.0075)	(0.0074)	(0.0073)	(0.0072)	(0.0075)	(0.0075)	(0.0074)
High School Diploma	0.190***	0.181***	0.184***	0.178***	0.189***	0.189***	0.188***
	(0.0114)	(0.0113)	(0.0104)	(0.0106)	(0.0113)	-0.0113	(0.0113)
Some College	0.344***	0.333***	0.335***	0.327***	0.343***	0.343***	0.342***
	(0.0175)	(0.0172)	(0.0164)	(0.0165)	-0.0175	(0.0174)	(0.0174)
College Degree	0.645***	0.631***	0.633***	0.623***	0.643***	0.643***	0.643***
	(0.0246)	(0.0239)	(0.0228)	(0.0227)	(0.0245)	-0.0245	(0.0245)
Grad/Prof Degree	0.930***	0.914***	0.912***	0.901***	0.928***	0.928***	0.928***
	(0.0373)	(0.0361)	(0.0343)	(0.0340)	(0.0372)	-0.0372	(0.0371)
CMSA	X	X	X	X	X	X	X
POB	X	X	X	X	X	X	X
R-squared	0.249	0.25	0.251	0.252	0.249	0.249	0.249

Source: Author's calculations using the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. All regressions include controls for age, age-squared, race, years since migration and its square, citizenship status, English ability, and employer type, Models (II) through (IV) use residential co-ethnic exposure measured at the tract level for Co-ethnic Residential Exposure Measure and workplace co-ethnic exposure measured at the employer while models (V) through (VII) use residential and workplace co-ethnic exposure rates measured at the PMSA level.
Robust, clustered standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Section 4. Conclusion

This paper develops a two-dimensional approach for studying immigrant enclaving behavior by measuring both the residential and workplace concentration of immigrants in five U.S. cities with large immigrant populations. Using linked employer-household data, I am able to estimate the proportion of co-ethnic neighbors and co-ethnic coworkers for immigrants in the labor force. The results show that very few immigrants live and/or work in highly co-ethnic neighborhoods and employers. Most immigrants, in fact, live and work with less than 10% co-ethnics. Though somewhat higher than would be expected under random sorting, this suggests a low degree of co-ethnic exposure even for immigrants living in cities with large co-ethnic populations. Less than 1% of the immigrant population both lives and works with more than 50% co-ethnics. Additionally, analyses conducted on Hispanic immigrants reveal that common language alone is not sufficient for enclaving. Instead, different country of origin groups cluster together with Hispanic groups that are more similar. For example, Mexican, Salvadorian and Guatemalan immigrants are more likely to work and live near each other than to other Hispanic groups.

One of the primary goals of this paper is to explore how well previous research that has relied on larger geographic definitions and did not have access to linked employer-household data was able to measure enclaves. OLS regressions reveal that half of neighborhood-level ethnic clustering can be explained using commonly available demographic information combined with city and place of birth controls. Workplace concentration, however, is more difficult to predict. Only a quarter of the variation is explained by observables and place of birth and CMSA controls. Additionally, the proportion of the population in the PMSA from a given place of birth serves as a strong predictor of residential own-exposure and, to a lesser degree, workplace own-exposure. PMSA-level co-ethnic measures also provide similar explanatory power as firm-level or neighborhood-level exposure rates for reported earnings, suggesting these might be adequate proxies for the more detailed measurements. However, to separate the effects of workplace and residential own-exposure rates, one cannot rely on data at the PMSA-level alone. Using firm-level and neighborhood-level own-exposure rates, an initial, naïve pass at enclave effects that does not address self-selection shows that wages are lower for immigrants with more co-ethnic coworkers, even after controlling for their residential co-ethnic exposure. This provides preliminary evidence of lower earnings for immigrants who are limited to working in high co-ethnic workplaces, independently of whether they reside in a high co-ethnic neighborhood.

REFERENCES

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2005. The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. *Center for Economic Studies Technical Paper*, TP-2006-01.

Altonji, Joseph G. and David Card. 1991. The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives. Immigration, Trade and Labor, ed. John M. Abowd and Richard B. Freedman. Chicago, University of Chicago Press.

Andersson, Fredrik, Monica Garcia-Perez, John Haltiwanger, Kristin McCue and Seth Sanders. 2010. Workplace Concentration of Immigrants. *Center for Economic Studies Working Papers*, CES 10-39.

Bartel, Ann P. 1989. Where do the new U.S. immigrants live? *Journal of Labor Economics* 7:4.

Bayer, Patrick, Robert McMillan and Kim S. Rueben. 2004. What Drives Racial Segregation? New Evidence Using Census Microdata. *Journal of Urban Economics* 56.

Bayer, Patrick, Stephen Ross and Giorgio Topa. 2008. Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy* 116:6.

Bertrand, Marianne, Erzo Luttmer, and Sendhil Mullainathan. 2000. Network Effects and Welfare Cultures. *The Quarterly Journal of Economics* 115:3.

Borjas, George. 1987. Self-Selection and the Earnings of Immigrants. *The American Economic Review* 77:4.

--- 1992. Ethnic Capital and Intergenerational Mobility. *The Quarterly Journal of Economics* 107:1.

--- 1995. Ethnicity, Neighborhoods, and Human-Capital Externalities. *The American Economic Review* 85:3.

--- 2000. Ethnic Enclaves and Assimilation. *Swedish Economic Policy* Review 7.

Borjas, George J., and Stephen G. Bronars. 1989. Consumer Discrimination and Self-Employment. *The Journal of Political Economy* 97:3.

Carrington, William J. and Kenneth R. Troske. 1997. On Measuring Segregation in Samples With Small Units. *Journal of Business & Economic Statistics* 15:4.

Chiswick, Barry R. and Paul W. Miller. 1995. The Endogeneity Between Language and Earnings: International Analyses. *Journal of Labor Economics* 13:2.

Cutler, David, Edward Glaeser and Jacob Vigdor. 2008. When are Ghettos Bad? Lessons from Immigrant Segregation in the United States. *Journal of Urban Economics* 63.

Edin, Per-Anders, Peter Fredriksson, and Olof Aslund. 2003. Ethnic Enclaves and the Economic Success of Immigrants – Evidence From a Natural Experiment. *The Quarterly Journal of Economics* 118:1.

Gonzalez, Arturo. 1998. Mexican Enclaves and the Price of Culture. *Journal of Urban Economics* 43:2.

Iceland, John, Daniel H. Weinberg, and Erika Steinmetz. 2002. Racial and Ethnic Segregation in the United States: 1980-2000. U.S. Census Bureau, Series CENSR-3.

Lazear, Edward. 1999. Culture and Language. *The Journal of Political Economy* 107:6, Part II.

Passel, Jeffrey. 2006. The Size and Characteristics of the Unauthorized Migrant Population in the U.S.: Estimates Based on the 2005 Current Population Survey. Pew Hispanic Center: Washington, DC. http://pewhispanic.org/files/reports/61.pdf.

Toussaint-Comeau, Maude. 2008. Do Ethnic Enclaves and Networks Promote Immigrant Self-

Employment? *Economic Perspectives* IV.

Wilson, Kenneth and Alejandro Portes. 1980. Immigrant Enclaves: An Analysis of the Labor Market Experience of Cubans in Miami. *The American Journal of Sociology* 86:2.

Xie, Yu and Margaret Gough. 2011. Ethnic Enclaves and the Earnings of Immigrants. *Demography* 48.