IMMIGRANT COMMUNITIES, HUMAN CAPITAL EXTERNALITIES AND LABOR MARKET OUTCOMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Liliana do Couto Sousa

January 2012

IMMIGRANT COMMUNITIES, HUMAN CAPITAL EXTERNALITIES AND LABOR MARKET OUTCOMES

Liliana do Couto Sousa, Ph.D.

Cornell University 2012

The three essays that encompass this dissertation contribute to our understanding of the economic impact of ethnic communities on immigrants while also addressing issues associated with the identification and measurement of ethnic enclaves. Immigrant enclaves provide access to ethnic goods and trade partners with shared language and culture, potentially resulting in increased job opportunities. However, these same amenities may also decrease incentives to assimilate, or acquire U.S.-specific human capital, and can ultimately keep some immigrants from achieving economic success. The first essay considers whether the human capital of an ethnic community influences the decision to become self-employed, for example by affecting certain costs, such as transaction and information costs, associated with entrepreneurship. I find that immigrants with low levels of human capital are more likely to enter into self-employment if their ethnic communities have higher levels of human capital while immigrants with more human capital, such as those with a college education, enter into self-employment independently of the human capital available in their ethnic communities. These ethnic human capital externalities may play an important role in the economic assimilation of low human capital immigrants by potentially offsetting some of the economic costs associated with low education and limited English skills.

The second and third essays use unique linked employer-household data available through the U.S. Census Bureau's Longitudinal Employer-Household Dynamics program to identify individuals as part of an enclave economy based not only on their neighbors — the strategy employed by the current literature — but also on their coworkers. In the second essay, I create and analyze measurements of immigrant enclaves based on both residential and employment clustering behavior. These measures show that, even among the largest immigrant groups in five of the biggest immigrant population centers in the U.S., few immigrants live or work in neighborhoods and workplaces with high co-ethnic exposure rates.

Though ethnic enclaves can provide economic opportunities for their members by generating or matching individuals to employment opportunities, they may also stifle assimilation and create human capital traps by limiting interactions between enclave members and non-members. In the third essay, I find that higher residential and workplace clustering is consistently correlated with lower earnings. While negative self-selection fully explains the lower earnings attributed to higher co-ethnic exposure for immigrants with a high school education or less, I find evidence of human capital traps for immigrants with more than a high school education who enclave. Their earnings decrease with higher levels of co-ethnic exposure both residentially and in the workplace.

BIOGRAPHICAL SKETCH

Liliana do Couto Sousa graduated from Centereach High School, Centereach, New York in 1999. She received her Bachelor of Arts in economics from Vassar College in 2003. After an academic hiatus to work in public policy research, she completed her Masters of Arts in economics in February 2011 and her Doctorate of Philosophy in economics in January 2012, both from Cornell University.

Dedico este trabalho aos meus pais, Maria de Jesus e José,

não teria conseguido sem vocês.

ACKNOWLEDGMENTS

I would like to sincerely thank my co-chairs: John Abowd, for the years of valuable academic feedback, research opportunities and indispensible advice and mentorship, and Lawrence Kahn, for his consistently insightful and attentive feedback on my research and his thorough draft readings. I am also indebted to Francine Blau for her guidance and thoughtful comments on my research. Many thanks to the Cornell Census research team for the many research talks and animated lunches over the years. Thanks are also due to the fantastic staffs at CISER, both for the reliable and amazing computing resources they provided and their help and friendship during the years, and at the U.S. Census Bureau. Many thanks to the good friends I made in Ithaca and elsewhere for the countless hours of essential non-academic chats and for the almost-as-many hours of academic chats. I especially want to thank Mario Ramírez Basora for his much needed and unflagging support. Finally and most importantly, I wish to express my sincere gratitude and appreciation for my parents, whose sacrifices and determination have given me opportunities that were unimaginable to them, and for my sisters, Milene and Glynis, who have always been my most spirited cheerleaders.

This research was undertaken while I was at Cornell University with the generous support of the Walter Galenson Fellowship in Labor Economics from Cornell University and the Emilie Louise Wells Fellowship from Vassar College. Any opinions and conclusions expressed herein are mine and do not necessarily represent the views of the U.S. Census Bureau or Cornell University. All results utilizing confidential data have been reviewed to ensure that no confidential information is disclosed. Some of 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.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

COMMUNITY DETERMINANTS OF IMMIGRANT SELF-EMPLOYMENT: HUMAN CAPITAL SPILLOVERS AND ETHNIC ENCLAVES

Self-employment plays an important role in the economic assimilation of some groups of immigrants by providing an income stream outside of the formal wage/salary market. This outside option is especially important for immigrants who face barriers to entry in the formal labor market due to foreign education[1] and weak English skills. Though it is an inherently riskier occupational choice defined by less predictability, self-employment results in steeper earnings growth for immigrants relative to wage/salary employment (Lofstrom 2002, 2009). Rates of self-employment, however, vary dramatically between different immigrant groups. Less than 8% of the Mexican-born while nearly a quarter of the Korean-born in the U.S. report being self-employed. Clearly, differences in individual human capital among immigrants from these two source countries provide some explanation for different self-employment rates. But, this paper shows that another factor is the differences in aggregate levels of human capital in immigrant communities. By exploiting the variation in human capital levels between different communities of immigrants from the same country of origin, I find evidence that human capital spillover effects may encourage and facilitate self-employment of community members with low-levels of human capital.

A positive enclave effect on self-employment among Hispanic immigrants has been found in several studies (Borjas 1986; Lofstrom 2002; Toussaint-Comeau 2008). These enclave effects are empirically estimated using the size of the ethnic community in which an immigrant resides.

[1] Friedberg (2000) finds that foreign schooling is valued less by the labor market than similar domestic schooling.

The argument is that the community serves both as a consumer of his goods as well as a source of information and inputs. Borjas (1986) finds that this effect is much stronger for the foreign-born Hispanic population than their U.S.-born counterparts. Similarly, Borjas and Bronars (1989), looking at racial groups rather than immigrant groups, find that the percent of the MSA that is black has a positive effect on black self-employment propensity. On the other hand, Clark and Drinkwater (2002) look at residential concentration of ethnic groups in England and Wales and find that self-employment falls with ethnic concentration. They also find that the educational attainment of a group does not affect self-employment, but does affect other employment outcomes. Yuengert (1995), on the other hand, finds no evidence that self-employment rates are higher in cities with large immigrant populations.

Toussaint-Comeau (2005, 2008) combines the notion of ethnic capital (Borjas 1992) with the neighborhood effects methodology in Bertrand, Luttmer and Mullainathan (2000) to measure the impact of ethnic networks on self-employment. Borjas (1992) argues that the production of human capital can be influenced not just by family human capital, but also by externalities from the human capital of the ethnic group, referred to as ethnic capital. He finds that the average educational level of an individual’s ethnic group in the father's generation affects the individual’s educational attainment. Building on this, Toussaint-Comeau creates an ethnic network measure that combines the size and concentration of the ethnic community in which an immigrant resides with the “entrepreneurial ethnic capital” of the immigrant group. This entrepreneurial ethnic capital value is calculated as the percent of the adult ethnic population that is self-employed in the U.S. Like in Borjas (1992), this measure reflects an ethnic level human capital externality. Members of groups with greater numbers of self-employed have access to more and better self-employment advice and information, thus possibly affecting their occupational choice.

Multiplying this entrepreneurial ethnic capital measure by the availability of contacts, in the spirit of Bertrand, Luttmer and Mullainathan (2000), results in a measure of the availability of entrepreneurial information in one's local ethnic network. The result is a positive effect on self-employment, suggesting that effective ethnic capital transmitted via ethnic networks facilitates self-employment for some groups. She further interacts this ethnic network variable with the individual's education and language skills and finds that immigrants with a high school diploma or lower education benefit from having access to more self-employed co-ethnics while those with higher education do not.

In this paper, I build on Toussaint-Comeau's research by considering how community English skills and educational attainment can impact individual self-employment. While the papers cited above focused on the size of the ethnic community or on the entrepreneurial ethnic capital available via ethnic networks, I consider whether local ethnic capital, measured in English-skills and education, affects members of the community by leading to self-employment possibilities that might not otherwise exist. Specifically, I address two questions: Do immigrants with low English-skills benefit from residing near co-ethnics who speak English? Do immigrants with little formal schooling benefit from access to highly educated co-ethnics?

Human capital spillovers might influence an individual's decision to become self-employed through a number of venues: by lowering transaction costs, by lowering capital costs, by lowering information costs, and by providing better (or worse) job referral networks. Transaction costs incurred by the self-employed include interactions with suppliers, landlords, regulators, customers, and, in larger enterprises, employees. As demonstrated by Lazear (1999), the ability to interact with co-ethnics in these different roles can decrease transaction costs through shared language and culture. Transaction costs are influenced by both the size of the local community (more trade partners implies more possibility for trade) and the quality of the local community (more co-ethnics with business connections can decrease costs, for example). An ethnic community can also serve as a source of informal lending, an especially important consideration for credit-constrained immigrants starting small enterprises.[2] Being able to tap into co-ethnic channels may result in lower borrowing costs, or increased borrowing opportunities, than if one is limited to financial institutions. Co-ethnics with strong labor market attachment can serve as conduits for local market and industry-specific information – better information, in turn, can decrease costs faced by small businesses. On the other hand, better work referral networks can provide better wage opportunities, thus increasing the opportunity costs of becoming self-employed. One last important role that the local ethnic community can play is as a market for ethnic goods. Since co-ethnics have a comparative advantage in providing ethnic goods, the existence of an ethnic market for such goods results in expanded business opportunities.

In order to test these theories, I consider the effect of the community's educational attainment, a measure of human capital and a good proxy for financial capital stocks, and the effect of language skills on the self-employment propensities of immigrants with different levels of schooling and English-skills. Previous research has consistently found that one's English language skills and formal schooling are important in predicting self-employment (Borjas 1986, Borjas and Bronars 1989, Evans and Leighton 1989). I now consider how these two types of human capital at the community-level interact with an individual's own human capital to impact self-employment. I find that immigrants with lower levels of human capital are more sensitive to ethnic spillover than immigrants with higher levels of human capital. I also find that, with the

[2] Bohn and Pearlman (2009) find lower rates of formal banking in areas with higher concentrations of co-ethnics while Bates (1998) documents Chinese and Korean immigrant entrepreneurship's reliance on informal lending and on ethnic credit associations in addition to financial institutions.

exception of college educated immigrants, immigrants are more likely to be self-employed if they reside in communities with higher educational attainment. Similarly, among Spanish-speaking immigrants, individuals opt into self-employment at greater rates if more of their co-ethnics speak English.

Speaking the host country language yields higher returns in the labor market (Chiswick and Miller 1995; Carliner 2000). However, learning a new language can present formidable costs, particularly for individuals who immigrate as adults and for those with little schooling.[3] Similarly, acquiring more education as an adult can also be prohibitively expensive – often requiring at least a partial exit from the labor force in addition to financial expenditures. The human capital spillover effects identified in this paper may present an alternative approach to reaping the rewards of more education and better language skills for immigrants who face high costs of acquiring these skills for themselves.

Theoretical Framework

The decision to become self-employed is a choice between a relatively risky and unpredictable income stream through self-employment and a pre-determined and relatively predictable income stream through wage employment. Building on fundamental models of self-employment (de Wit 1993), I assume a one-period game, where all individuals have preferences that can be represented by a utility function of the form $u_i(y)$, where y is individual income, and u_i is continuous and differentiable such that $u_i'(y) > 0$, and $u_i''(y) < 0$ for all y. Assume that the

[3] Cognitive research has shown that languages are learned more easily by children than by adults (for example, Johnson and Newport 1989). Rosenzweig (1995) finds that an increase in schooling results in an increased ability to absorb new knowledge and learn new skills.

degree of concavity varies by individual, but is symmetrically distributed within each group.[4]

All individuals, having already decided to enter the labor force, can choose between self-employment and wage employment. If the choice is wage employment, each receives wages w_i with certainty.[5] Self-employment income will depend on the investment made by the individual and on exogenous market factors. Prior to choosing between self-employment and employment, each individual will estimate his potential outcome from self-employment by choosing x_i, a vector of the amount of each good or service being provided, so as to maximize expected utility from self-employment. The individual solves the following problem to optimize his self-employment payoff:

$$\max_{x_i} Eu_i[\pi(x_i, \gamma_i, \theta)] = \int u_i[p(\theta)x_i - c(x_i, \gamma_i, \theta)]dF(\theta) \quad \text{(I)}$$

where θ denotes different states of the world, $p(\theta)$ is a random vector of prices for the goods being produced, and γ_i is entrepreneurial capital. Researchers often define this abstract concept of entrepreneurial capital as an individual trait that can lead an individual to be successfully self-employed (de Wit 1993, Clark and Drinkwater 2000, Lazear 2005). Note that costs of inputs also depend on θ. The implied cost function, $c(xi, \gamma i, \theta)$, is assumed to be decreasing in γ_i.

Define S_i as the net utility gain from self-employment for individual i. Suppose x_i^* is the solution to equation (I). An individual will choose to become self-employed if

$$S_i = Eu_i[\pi(x_i^*, \gamma_i, \theta)] - u_i(w_i) > 0 \quad \text{(II)}$$

Now, I extend this basic model to include the question of how social networks can affect the

[4] Note I am not assuming that the mean or variance of risk-aversion is equal between groups.

decision to become self-employed. Suppose individuals are of J types, where $j \in \{1, 2, \ldots J\}$ represents county of birth. Let $k \in \{1, 2, \ldots K\}$ represent the location in the U.S. in which the individual resides. A pair j, k is an ethnic community born in country j and residing in city k, to which I will refer as an enclave or COB-MSA group. An important and reasonable assumption that runs through this research is that immigrants from the same country residing in the same metropolitan area in the U.S. are more likely to have social ties to local co-ethnics than to the rest of the local population.

Equations (I) and (II) imply that there are three ways in which individual i's self-employment likelihood can be increased: 1) higher expected revenue, 2) lower expected costs and 3) lower opportunity cost. Below I show how each of the three can be affected by the COB-MSA group to which the individual belongs.

Higher expected revenue can be achieved through higher prices or higher production. An ethnic enclave can create higher prices by demanding goods that are not supplied outside of the ethnic group or by preferring to do business with co-ethnics (Borjas and Bronars 1989), thus creating a protected market for an ethnic business. To a large extent, the impact of the local community on prices is determined by its demand for goods produced by co-ethnics. This is related to the size of the community and is, in effect, the enclave effects found by Borjas and others, as cited above. However, this demand is also a function of cultural distance (expressed through differences in preferences and tastes) and linguistic isolation between the community and the rest of the local residents. Since I do not measure cultural distance in this research, I focus instead on language barriers. Let Ω_{jk} be the size of the linguistically isolated ethnic community. We can expect the

[5] This is clearly a simplifying assumption. Though uncertainty exists in the labor market, the important detail here is that wage is more easily predictable than returns to self-employment.

following relationship:

$$\frac{\partial p(\theta)x_i}{\partial \Omega_{\mathrm{jk}}} > 0 \qquad \text{(III)}$$

The size of the linguistically isolated ethnic community increases expected revenue by increasing the demand for good x_i. This is due both to a co-ethnic's comparative advantage in producing ethnic goods and to high transaction costs faced by the consumers in a linguistically isolated community. This is only relevant for businesses that choose to cater to co-ethnics.

Another way to affect the likelihood of self-employment is by lowering self-employment costs. Immigrants face higher costs than the U.S.-born when attempting self-employment due to immigrant-specific obstacles such as language and cultural barriers, poor information regarding local regulations or preferences, limited financial knowledge/access, and a limited credit history (Bowles and Colton 2007). Ethnic communities can promote informal business arrangements and lending with relatively low search costs and information costs (Bond and Townsend 1996). Additionally, consider the role of effective ethnic capital in acquiring new information (Borjas 1998; Toussaint-Comeau 2008). Knowing more individuals in your social network with self-employment experience or industry-specific employment experience results in increased access to information about how to run a successful business or industry-specific issues. This access might play an important role in explaining ethnic clustering by industry (for example, Ellis and Wright 1999). Having access to co-ethnics with high levels of human capital implies an increased number of potential trading partners and, thus, lower transaction costs. On the other hand, having access to a low human capital co-ethnic community can lead to access to a low-wage labor pool with low supervisory transaction costs.

Specifically, suppose γ_i captures differences in enclave ethnic capital, s_{jk}, in addition to personal differences in entrepreneurial ability. That is, s_{jk} is an input to the business that decreases production costs. Note that unlike the enclave effects on expected revenue, the enclave effects on expected costs are primarily driven by *quality* of co-ethnics (as measured by human capital and capital stocks) not quantity. We expect the following relationship:

$$\frac{\partial c(x_i, \gamma_i(s_{jk}), \theta)}{\partial s_{jk}} < 0 \qquad \text{(IV)}$$

Finally, the third way an enclave can impact the self-employment decisions of its members is through wages. Forgone wages are the opportunity cost incurred by the self-employed. Evans and Leighton (1989), for example, find that men with poor employment outcomes are more likely to become self-employed since they faced lower opportunity costs in leaving the formal labor market. If an enclave or locality can provide members of a certain group with relatively high wage opportunities, maybe via well-established job referral networks, then we can expect less self-employment in this group.[6] Suppose that s_{jk} is a determinant of the market wage rate for immigrants. We expect the following relationship:

$$\frac{\partial w(s_{jk})}{\partial s_{jk}} > 0 \qquad \text{(V)}$$

This paper focuses on the effects of Ω_{jk}, the size of the linguistically isolated population in an enclave, and s_{jk}, the group-specific supply of community human capital, on S_i. From (IV) and (V) above, we have two opposing effects from an increase in s_{jk}: community human capital decreases production costs but also increases the opportunity cost of self-employment. The relative importance of these two effects is empirically tested below, using educational attainment of the community as a measure of s_{jk}. The relative importance of equations (III), (IV) and (V) is tested using the English-acquisition of the community. The results, as detailed below, show that equation (IV), a decrease in self-employment production costs, dominates the effect of increased opportunity cost of regular employment. Some empirical evidence to support the protected market shown in equation (III) is also found.

The impact of community human capital on an individual will vary by the level of human capital he possesses. For example, consider an immigrant who speaks English and is part of a community with low levels of English skills. He has a comparative advantage in providing goods and services to his linguistically isolated community - both relative to non-English speakers within the ethnic community and to English speakers outside of the community. Additionally, he might have access to cheaper labor, without incurring additional transaction costs, by hiring co-ethnics who do not speak English. However, his ethnic community, being composed of individuals who do not speak English, might also be poorer, resulting in less opportunities for informal lending (thus resulting in higher self-employment costs), less disposable income to spend on new goods and services (resulting in lower demand), and weaker job referral networks (resulting in lower opportunity costs).

Similarly, educational attainment of the community will have different effects on immigrants who are highly educated and those who are not. An individual with higher education might provide better information regarding the local economy or industry to other members of the

[6] Beaman (2007), for example, finds evidence that the social networks of refugees in the U.S. impact the wage draws of their members; communities with longer tenure result in higher wage draws for new members than those with shorter tenure.

ethnic community, resulting in decreased costs of self-employment. Immigrants with little formal education benefit more from having access to highly educated individuals since they might not be able to procure this information otherwise. On the other hand, educated professionals residing in ethnic communities with low educational attainment might be able to profit from the unmet demands for goods and services demanded by their co-ethnics (for example, a lawyer with roots in the ethnic community would have a comparative advantage in providing legal services to co-ethnics).

Empirical Approach

The primary hypothesis of this paper is that the aggregate human capital within an immigrant community can have a direct impact on an individual's propensity to become self-employed – and that this effect depends on the individual's own level of human capital. That is, I am interested in the interaction between individual i from country j living in MSA k and the aggregate levels of human capital, measured as English-acquisition rates and educational attainment, of other individuals born in country j who reside in MSA k.

The terms "enclave" and COB-MSA group are used interchangeably throughout this paper to refer to a community of co-ethnics (as defined by country of birth) living within the same primary metropolitan area in the U.S. Thus, Chinese-born immigrants distributed throughout a suburban MSA are as much part of an "enclave" in this paper as are those who actually live in the ethnic neighborhood of Chinatown.[7] Though they may not live within the same concentrated neighborhood, the underlying assumption is that social ties still connect many immigrant residents in the suburbs or spread throughout non-ethnic neighborhoods of cities. This is

[7] This empirical definition of "enclave" is often found in literature on U.S. immigrants, for example Borjas (1986).

supported by research such as Alba *et al*. (1999), who find that the ability to speak English and years since migration have both become less important in explaining suburbanization patterns of immigrants, showing that suburbanization no longer implies assimilation.

In order to test the predictions discussed above, I use a reduced-form regression, equation (VI), where Z_i is a 0/1 indicator of self-employment and Y_i is the human capital measure being tested, either educational attainment or English language skills. I include the individual's level as Y_i and the aggregate level, measured as a percent within the co-ethnic local community, as Y_{jk}. I also consider how these effects may differ by the individual's own human capital by including an interaction term, $Y_i \times Y_{jk}$. X_i is a vector of observable characteristics that have been shown to be correlated with self-employment: age, age squared, years since migration, years since migration squared, race, Hispanic ethnicity, the presence of a spouse in the household, and American naturalization status.[8] Depending on the regression, either educational attainment or English ability is also included in X_i;

$$Z_i = X_i\beta_1 + \beta_2 C_j + \beta_3 L_k + \beta_4 E_{jk} + \beta_5 E_{jk}^{90} + \beta_6 Y_i + \beta_7 Y_{jk} + \beta_8\left(Y_i \times Y_{jk}\right) + \epsilon_i \qquad \text{(VI)}$$

where the parameters of interest are β_6, $\beta_7$and β_8.

Due to the interaction design of the logit regressions, marginal effects cannot be as easily interpreted as the usual straightforward approaches employed by similar research (Norton, Wange and Ai 2004). Rather than reporting marginal effects, I report the logit coefficients and then present graphed predicted probabilities of self-employment for some of the specifications.

Addressing Self-selection

Individuals make three, non-random choices to select into the universe of interest: whether to immigrate, where to live in the U.S., and whether to become self-employed. In order to control for self-selection and local conditions, four aggregate controls are included in every regression:

1. C_j (the percent of COB group j in the U.S. that is self-employed),
2. L_k (the MSA k self-employment demand index),
3. E_{jk} (the percent of the MSA population born in COB) and
4. E_{jk}^{90} (the share of the 1990 U.S. population from COB who resided in the MSA in 1990).

Country of birth can be endogeneous in the self-employment decision since it is entirely plausible that different rates of individuals with high predisposition for self-employment will emigrate from different source countries due to selection into immigration, source country development and cultural differences. As discussed in Borjas (1987), the population from each country that elects to immigrate to the United States is not randomly selected. Significant variation in skill-distribution among different immigrant groups can result from the income differentials between skill groups within the source and destination countries and the cost of immigration. Additionally, self-employment preferences and entrepreneurial skills might vary based on differences in source country characteristics (Light 1979).

To control for this endogeneity, I include C_j, the average self-employment rate of a COB group in the United States, as a control variable in the regression model. Note that this is not the self-employment rate in the individual's country of birth, but rather among the U.S. population who were born in that country. By using the immigrant-specific rate rather than the country of birth rate, I am implicitly controlling for the country-specific selection mechanisms that created these immigrant populations in the U.S. That is, since immigrants are not drawn randomly from their

[8] Regressions are limited to male immigrants since they have more homogeneous employment patterns across COB

country of birth, I control not for the average of the people who did not emigrate, but rather, the average of those who *did* emigrate. After controlling for this group average, only the individual deviation from the COB mean is left as the unmeasured individual proclivity for self-employment.

The choice of residence within the U.S. is also neither random nor fully explained by observables. Research on enclave effects has long struggled with just how to control for selection into enclaves. One approach, developed by Altonji and Card (1991), uses the co-ethnic concentration in the city from an earlier census as an instrument for movement into this area. This is a good control for city selection because immigrant location choices in the host country are largely determined by the location choices of previous waves of immigrants from the same country of birth (for example, Bartel 1989). Adopting this approach, I include the percent of the country of birth's adult population in the U.S. that was living in the individual's city in 1990, labeled as E_{jk}^{90} above.

To address the potential selection of members of a COB group with high propensity for self-employment into areas with high demand for self-employment, I control for local demand. This can be disaggregated into two different demands: the demand of the ethnic community and the demand in the local market. The demand of the ethnic community, as discussed above, is the result from demand for ethnic goods (in which co-ethnics have a comparative advantage) and consumer preferences to do business with co-ethnics. I control for this demand by using E_{jk}, the concentration of the country of origin group in the MSA.[9]

cells than female immigrants; hence gender is not included as a control.

[9] E_{jk} varies from E_{jk}^{90} not only since one is based on year 2000 data and the other on 1990, but also on how the concentration is measured. Specifically, E_{jk} is measured as a proportion of the total local population while E_{jk}^{90} is a

For non-ethnic demand, I use an MSA self-employment index, L_k. Certain industries, such as manufacturing, require heavy capital investment which means there are high costs to entry. Other industries require relatively little capital investment, making them more attractive to small business owners. In the spirit of Berman, Bound and Griliches (1994), who use a similar index to look at skill distributions within manufacturing, I create an MSA-index of demand for self-employment by multiplying the overall U.S. self-employment rates in each industry by the percent of the local labor force in MSA k employed within each industry. This MSA-level index allows for a comparison of local labor market demand for self-employment, taking the distribution of employment within local industries as exogenous.[10]

Since random selection into self-employment seems particularly implausible, I do not evaluate the relative success of the self-employed in this paper. Such a comparison is subject to bias based on unobserved characteristics, for example, the relationship between the ambiguous notion of entrepreneurial capital and motivation. Additionally, Hamilton (2000) applies Rosen's (1981) super-star theory to self-employment, arguing that samples of self-employed individuals will be made up of a few high-earning long-term entrepreneurial super-stars and many low-earning, failure-prone, new comers to self-employment. This bimodal distribution results from the gradual exit of entrepreneurs who learn, through experience, that they are not super-stars. Though I consider a less biased variable of interest, whether or not the individual reported being self-employed on the census, this is still somewhat affected by the success of self-employment

---

proportion of the ethnic population that lives in MSA k. Thus, while the first measure might be small for a small COB group living in a large city, the second measure might be very large if the majority of that COB group lives in that city.

[10] Due to the tendency of different immigrant groups to cluster in particular industries, one might be concerned that the high concentration of an immigrant group in a specific industry might impact the relative size of the labor force in that industry. Indeed, some of the largest COB-MSA cell groups, such as the Mexican-born in El Paso and the Cuban-born in Miami, represent over 25% of their MSA populations. However, the 90th percentile COB-MSA cell

since longer spells are more likely to fall within the period of time being sampled than shorter spells, all else equal.

Specification Testing

To address the endogeneity of entrepreneurial ability and/or preferences, previous researchers have typically included country or region of birth dichotomous variables as controls (Borjas 1986, Lofstrom 2002, Toussaint-Comeau 2008). Borjas (1986) looks at different racial/ethnic groups of immigrants while Lofstrom (2002) collapses country of origin groups into regional groups, arguing that they are relatively homogenous. However, as Toussaint-Comeau (2008) and Table 1.4 below show, there is significant variation in self-employment rates by COB group within aggregated immigrant/ethnic groups such as "Asians." Toussaint-Comeau (2008), adopting a similar approach to Bertrand, Luttmer and Mullainathan (2000), addresses this by using a linear probability model to predict self-employment, thus allowing for the inclusion of a large set of COB dummy variables without sacrificing the validity of the error estimates.

Though including a large array of dichotomous variables for each COB and MSA is a good way to control COB and MSA unmeasured effects, it quickly consumes degrees of freedom, resulting in unreliable test statistics.[11] This is particularly problematic in nonlinear regressions, such as the logistic model employed below. Furthermore, the coefficients on the COB and MSA variables are too numerous to be meaningfully informative. Instead of using this approach, I opted for two continuous variables: the percentage of the COB population that is self-employed (C_j) and the MSA self-employment index (L_k). The validity of this alternative specification, relative to the

represents only 3.46% of the MSA population. Thus, for the vast majority of communities, this index will not suffer from COB endogeneity.

[11] It also introduces computational error from machine approximations of 0, a pertinent concern given the large sample sizes used.

inclusion of COB and MSA dichotomous variables, is explored in detail in Appendix A. These tests show that using the continuous variables results in slightly smaller effects for the education regressions; if anything, my approach underestimates the ethnic spillover effects of education. For immigrants from countries where neither Spanish nor English is spoken, the inclusion of the two vectors of dichotomous variables produces slightly larger coefficients on the impact of enclave English-skill on self-employment of both groups who speak English. For immigrants from Spanish-speaking countries, the vectors of controls decrease the coefficient on English-language enclave effect by about one-third, though the interacted effects (i.e., the difference between the effect for non-English speakers and the effects of the other two groups) remain the same. Overall, however, the continuous variables do a good job of controlling for the heterogeneity addressed in other research projects with the inclusion of COB and MSA dummy variables.

Data

This paper uses data from the 2000 U.S. Census 5% Public Use Microdata Sample. The sample of interest is restricted to foreign-born men between the ages of 25 and 65 who immigrated as adults, are in the labor force and have not been in school for at least 2 months as of April 2000. The sample is limited to those who immigrated as adults so as to minimize sample composition effects due to 1) selection into immigration, since children typically do not make this decision for themselves, and 2) differences in U.S.-specific capital accumulated by the two groups. This also simplifies the interpretation for years since migration and education (which will primarily be completed in the country of origin). Some additional sample restrictions were made limiting individuals to those who reside in a PMSA/MSA with a significant co-ethnic sampled population in the 1990 and 2000 U.S. Censuses. Only immigrants belonging to a COB-MSA group with more than 50 sampled adult men were included since the empirical specification relies heavily on variables measured at the COB-MSA level. This resulted in dropping about 20% of the sample. Appendix B shows that these immigrants look different from those who live in MSAs with larger co-ethnic samples. These restrictions limit the sample to almost 233,000 men, representing 5.1 million immigrant men. Nearly 12% of these 5.1 million men are self-employed. Table 1.1 shows that, as expected, these men are highly clustered in traditional immigrant cities: half of this sample resides in only seven PMSAs.

Table 1.1: Top 20 Primary Metropolitan Statistical Areas, by Size of the Sampled Population

Primary Metropolitan Statistical Area	Estimated Population	%	Sample Size	%
Los Angeles-Long Beach, CA PMSA	767,745	15.1	37,638	16.2
New York, NY PMSA	717,073	14.1	29,421	12.6
Chicago, IL PMSA	326,346	6.4	13,110	5.6
Miami, FL PMSA	223,077	4.4	10,365	4.5
Houston, TX PMSA	191,629	3.8	8,067	3.5
Washington, DC-MD-VA-WV PMSA	188,297	3.7	8,836	3.8
Orange County, CA PMSA	172,060	3.4	9,041	3.9
Dallas, TX PMSA	136,098	2.7	5,957	2.6
San Jose, CA PMSA	129,630	2.5	6,220	2.7
Oakland, CA PMSA	119,093	2.3	5,815	2.5
Riverside-San Bernardino, CA PMSA	113,690	2.2	5,186	2.2
San Diego, CA MSA	103,708	2.0	4,994	2.1
San Francisco, CA PMSA	102,773	2.0	4,867	2.1
Boston, MA-NH PMSA	92,491	1.8	4,279	1.8
Atlanta, GA MSA	90,347	1.8	3,710	1.6
Phoenix-Mesa, AZ MSA	83,713	1.6	4,115	1.8
Newark, NJ PMSA	81,614	1.6	3,812	1.6
Nassau-Suffolk, NY PMSA	77,461	1.5	3,759	1.6
Fort Lauderdale, FL PMSA	67,315	1.3	3,069	1.3
Bergen-Passaic, NJ PMSA	67,078	1.3	3,189	1.4
Total Top 20	3,851,238	75.5	175,450	75.3

Source: Author's calculations based on U.S. Census PUMS 5% sample. The universe is limited to male immigrants who report being in the labor force, not in school and between the ages of 25 and 65 who immigrated as adults.

Table 1.2 presents basic demographic information on the sample used in the analysis. The sample represents about 600,000 self-employed immigrant men and 4.5 million who are in the labor force and not self-employed. On average, these individuals are 41 years old, though the average self-employed individual is nearly 4 years older. White non-Hispanic men make up 23% of the self-employed in this sample, though they are only 14% of the sample. Non-Hispanic black and Hispanic immigrants are underrepresented in the self-employed category. All other races account for the remaining quarter of the sample; this group is slightly overrepresented among the self-employed.

In line with previous research (Borjas 1986; Le 1999; Georgarakos and Tatsiramos 2009), nearly three-quarters of self-employed immigrant men have a spouse in the household compared to just over 60 percent of the employed immigrant men. Over thirty percent of the sample is naturalized; self-employed immigrant men are more likely than employed immigrant men to be naturalized. Overall, the average sampled individual has been in the United States for 14 years. Self-employed men have been in the U.S. for slightly longer. Self-employed immigrant men are less likely to have changed residences in the past 5 years. This residential stability might imply closer ties to the community.

Over a quarter of immigrant men in this sample have 8 years or less of schooling. This group is considerably less likely to be self-employed. Men who completed high school are overrepresented among the self-employed. About 10% of the immigrants in this sample speak only English at home. These are primarily immigrants from English-speaking countries. Roughly 60% who reported speaking a language other than English at home spoke English very well or well.[12] The remaining 30% reported speaking English poorly or not at all.

Table 1.2: Demographic Characteristics of Foreign Born Men, in the Labor Force and not in School, Ages 25-65, who Immigrated as Adults, by Self-Employment Status

	Total	Not Self-Employed	Self-Employed
Sample Size	232,988	205,577	27,411
Weighted Total	5,100,024	4,504,342	595,682
Average Age	41.5	41.0	44.7
White (%)	13.9	12.7	22.6
Black (%)	6.2	6.3	5.3
Hispanic (%)	54.5	56.1	42.5
Other Race (%)	25.4	24.8	29.7
Spouse in Household (%)	64.1	62.7	74.7
Average N of Children in HH	1.2	1.2	1.2
Naturalized (%)	34.3	32.8	45.6
Years since migration	13.8	13.4	16.8
Did not move in past 5 years (%)	41.2	39.9	51.4
No High School	27.8	28.7	21.3
Some High School	17.2	17.4	15.6
High School	17.5	17.4	18.6
Some College	14.2	13.8	17.0
College	12.2	11.9	14.5
Advanced Degree	11.1	10.8	13.1
Speaks English at home	10.3	10.1	11.1
Speaks English very well	25.4	25.0	28.0
Speaks English well	26.7	26.1	30.9
Limited English ability	37.7	38.7	30.1
Speaks Spanish at home	52.5	54.0	41.0
Household Income , Average	64,730	63,352	75,155
Median	49,000	49,000	50,000
Personal Income , Average	32,658	31,409	42,101
Median	22,000	22,000	23,200
Average Wage Income	29,974	31,133	21,209
Average Self-Employed Income	2,684	276	20,892

Source: Author's calculations based on U.S. Census PUMS 5% sample. All monetary values reported in 1999 dollars.

[12] Note that the 2000 U.S. Census was provided in 5 languages, besides English. Furthermore, a guide in 49 languages was provided with both the long-form and short-form censuses.

The average self-employed man in this sample reported total earnings of $42,000 in 1999 (from both self-employment income and wages) while the average wage/salary employee reported earnings of $31,400. Immigrant men who reported being self-employed reported over $21,200 in wage/salary earnings, almost the same as their average reported self-employment earnings. Those who did not report being self-employed yet reported some income from self-employment only reported an average of $300 in self-employment earnings. In this paper, self-employment is defined using the self-reported class of worker variable values for self-employed in own incorporated business and self-employed in own not incorporated business. This approach and these results reflect the fact that many self-employed men supplement their self-employment earnings with part-time or seasonal wage employment.

Immigrants are a bimodal group in terms of educational attainment; they are far more likely to have either very low education or very high education when compared to the U.S. born population. Table 1.3 shows the educational distribution of the twenty largest immigrant groups in the 2000 Census and the native born population, clearly illustrating the drastic differences in educational attainment between COB groups in the United States. Nearly half of Mexican immigrants and two out of every five immigrants from El Salvador and Guatemala had 8 years or less of formal schooling. At the other extreme, over 30% of Indian and Taiwanese immigrants had an advanced degree.

In order to identify spillover effects of the local ethnic community independently of COB-specific differences in preferences and skills, the empirical identification strategy relies on the variation of aggregate human capital at the COB-MSA level. The varying levels of self-employment among immigrant communities in different MSAs but from the *same* country of birth can be seen in Table 1.4. To illustrate these differences along the spectrum of self-

Table 1.3: Distribution of Educational Attainment for the U.S.-Born and the 20 Largest Country-of-Origin Groups

Country of Birth	Estimated U.S. Population	Highest Education Achieved (%)					
		Less than 9 years	Some High School	High School	Some College	College	Grad. Degree
United States	157,471,246	3.0	11.6	29.6	32.2	15.7	7.9
Mexico	7,635,686	44.5	24.6	17.3	9.9	2.3	1.4
Philippines	1,170,239	4.9	5.8	15.0	29.9	36.7	7.7
India	910,668	3.5	7.0	8.8	13.0	32.1	35.7
Vietnam	873,266	16.0	19.3	19.2	26.6	14.4	4.4
China	804,648	15.9	11.5	14.9	14.7	18.5	24.5
El Salvador	733,096	38.2	25.7	18.6	13.2	3.0	1.4
Cuba	676,855	14.8	20.8	21.9	23.0	10.5	9.0
Korea	602,408	4.6	7.1	22.3	24.7	28.3	13.0
Canada	591,563	2.8	9.3	18.0	32.2	22.2	15.4
Russia	581,378	4.1	8.0	18.7	23.3	24.0	21.9
Dominican Rep.	577,948	24.5	24.6	21.9	20.0	5.5	3.5
Germany	524,861	2.8	9.3	28.4	30.2	14.8	14.4
Jamaica	470,427	6.1	19.0	27.7	29.7	11.8	5.7
Colombia	433,861	11.1	14.8	26.7	26.5	12.1	8.8
Guatemala	418,047	41.8	21.4	18.2	13.5	3.4	1.7
Haiti	360,647	12.7	23.3	24.7	26.6	8.7	4.1
Poland	348,854	6.1	12.7	30.7	26.8	10.6	13.1
Italy	333,833	23.9	13.6	28.5	16.8	9.0	8.3
England	329,000	0.8	6.4	22.3	33.8	22.0	14.9
Taiwan	300,495	2.3	3.5	11.4	20.8	28.8	33.1

Source: Author's calculations based on U.S. Census PUMS 5% sample. The universe is limited to all individuals in the labor force in 2000, who were between the ages of 18-70.

employment rates, these ten COB groups were selected by choosing the country of origin group with the largest population in the U.S. at differing self-employment levels.[13] They range in overall self-employment rates from 5.29% for Filipino immigrants to nearly 25% for Korean immigrants.

Although there is substantial variation between different COB groups, there is also significant

[13] Countries of origin (one observation per country) were sorted by their overall self-employment rates in the U.S. They were then split into 10 equally sized groups. The country with the largest population in each group is reported in Table 3.

variation *within* COB groups based on MSA of residence. It is exactly this variation within COB groups that this paper exploits to measure human capital spillover effects on self-employment. The average Filipino-born community (unweighted by population) in an MSA has a self-employment rate of 10.19%. This varies from a low of less than 1% in one community to a high of 64.7% in another. Average MSA level self-employment rates for Taiwanese, Italian and Korean immigrants are roughly 25%. Even these high self-employment groups have communities with self-employment rates below 4%. The three Latin American COB groups included in Table 1.4 show the lowest maximum level of MSA-level self-employment rates, though still showing significant variation between the minimum and maximum percent self-employed.

Table 1.4: Percent Self-Employed at the COB-MSA Level for Ten of the Largest Country of Origin Groups

Country of Birth	Overall	Minimum	Maximum	Average MSA
Philippines	5.29	0.83	64.71	10.19
Mexico	7.69	0.77	37.85	7.88
El Salvador	9.28	1.34	40.84	10.70
Guatemala	9.81	0.47	47.76	14.25
India	10.93	2.24	59.15	16.90
Vietnam	11.33	0.88	74.42	15.26
Canada	13.65	3.37	55.26	14.73
Taiwan	15.27	2.04	87.37	24.55
Italy	18.02	3.44	76.09	24.35
Korea	24.61	2.94	76.00	26.20

Source: Author's calculations based on U.S. Census PUMS 5% sample. The universe is limited to all individuals in the labor force in 2000, who were between the ages of 18-70. Overall reports the overall percent of the COB population that is self-employed. Minimum, Maximum and Average MSA report COB-MSA cell values.

To get a better idea of the immigrant communities being analyzed and the variables used in the regressions, Table 1.5 displays information on the COB-MSA group whose members represented

Table 1.5: Human Capital Measures for Five Representative COB-MSA Groups

Percentile	10	25	50	75	90
Country of Birth	Mexico	Mexico	Yugoslavia	France	Colombia
MSA of residence	Orlando, FL MSA	Phoenix-Mesa, AZ MSA	Phoenix-Mesa, AZ MSA	New York, NY PMSA	Fort Lauderdale, FL PMSA
COB Population	7,635,686	7,635,686	197,632	115,824	433,861
COB % Self-employed	7.69	7.69	9.45	13.80	12.24
MSA Self-employment Index	9.94	10.10	10.10	10.42	10.60
COB-MSA characteristics:					
Population	16,220	226,450	5,155	12,060	27,364
Self-employed (%)	4.92	6.63	9.16	11.97	16.60
COB share in MSA (1990)	0.08	1.61	0.76	11.18	3.50
% of MSA born in COB	1.48	11.38	0.26	0.19	2.62
Schooling (%)					
Less than 9 years	43.55	42.22	13.23	1.70	6.00
Some High School	24.35	27.82	13.71	5.14	12.31
High School	17.12	17.33	36.61	12.08	27.38
Some College	10.22	9.24	27.51	16.67	28.91
College	2.76	2.14	5.51	22.65	16.14
Advanced Degree	2.00	1.24	3.43	41.76	9.26
English Skills (%)					
Only English	6.13	5.74	3.45	17.27	2.80
Strong English	38.85	37.35	62.37	81.06	66.82
Limited English	55.01	56.90	34.18	1.67	30.39

Source: Author's calculations based on U.S. Census PUMS 5% sample. These five MSA-COB groups were selected based on their percentiles in the distribution of self-employment rates. Specifically, all individuals were sorted based on the self-employment rate of their COB-MSA cell. The COB-MSA cells at the 10th, 25th, 50th, 75th and 90th percentile were selected.

the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of self-employment.[14] The first two communities are Mexican-born: those residing in Orlando and those in the Phoenix/Mesa area. The first four variables reported are not community specific. Instead, they report COB and MSA values; about 7.7% of the Mexican-born in the U.S. are self-employed while distributions of local industries imply a higher expected demand for self-employment in Phoenix than in Orlando. The COB-MSA variables depict the differences between these two Mexican-born communities. There are only 16,000 Mexican-born adults in Orlando while there are over 220,000 in the Phoenix area. In 1990, 1.6% of all Mexican-born adults residing in the U.S. lived in Phoenix but less than 0.1% lived in Orlando. Mexican immigrants in Phoenix are also less likely to have at least a high school diploma or to speak English well. Mexican immigrants are slightly more likely to be self-employed in Phoenix (6.6%) than in Orlando (4.9%).

Results

This section reports the results of estimating equation VI. [15] All reported regressions include a constant set of individual and community level controls, as described above. Most of these coefficients remain fairly constant as the specifications change to include different human and community capital measures. In line with previous research, age increases the likelihood of self-employment, though this effect decreases with age. White non-Hispanic immigrants are more likely to be self-employed than all other racial/ethnic groups. Like age, years since migration (YSM) increases the propensity for self-employment, though this effect decreases with time spent in the country, becoming negative after about 28 years of residing in the U.S., depending on the specification. This indicates an initial acclimation period, perhaps in order to accumulate country-specific capital, prior to starting one's own business. Immigrant men with a spouse in the household are more likely to be self-employed. Being naturalized was not statistically significant in any of the regressions. I also included the average years since migration in the COB-MSA cell

[14] The data were sorted by self-employment rate of the COB-MSA group and the individual at each of the percentiles of interest was selected. Data on his COB-MSA are reported in Table 5.
[15] Complete regression results available from author upon request. All regressions in this paper were based on weighted data, and clustered errors at the COB-MSA level.

in order to control for the endogeneity that might arise from the impact of years since migration on language/educational acquisition and self-employment at the community level, but this coefficient was not significant.

Results: English Ability of the Community and Self-Employment

To test the effect of English skills of a community on its members' propensity to become self-employed, I estimated the impact of the percent of the adult COB-MSA population[16] who reported either strong English skills, limited English skills, or who spoke only English at home on an individual's propensity to self-employ.[17] Table 1.6 reports these results. Furthermore, the sample for this set of regressions is limited to men who emigrated from non-English speaking countries.[18] I consider three English ability levels for the individual: limited or no English (the omitted group), strong English skills but speak a different language at home, and those who speak only English at home. The last group represents linguistically assimilated individuals who, I expect, encounter lower transaction costs outside of their co-ethnic community.

I begin by examining the impact of the percent of the community that speaks English well or very well, but still speaks a different language at home. These are the community members who are best able to serve as conduits for information between the enclave, including those with limited English skills, and their English-speaking neighbors. Specification (I) shows that the simple proportion of the community who speak English well or very well does not have a statistically significant impact upon an individual's propensity to become self-employed.

---

[16] These were calculated using the language skills of all adults in the COB-MSA though the regressions are run only on a male subsample.
[17] See appendix C for a distribution of these three values at the COB-MSA level.
[18] English speaking COB is empirically defined as a COB with English as the official language and with over 50% of all adult immigrants in the 2000 Census speaking only English at home, as in Bleakley and Chin (2004) and Blau, Kahn and Papps (2010).

Table 1.6: Testing COB-MSA English-skills Effects: Logit Regression Results

	Type of English-skill COB-MSA Measure					
	% Speak English Well		% Limited English		% Speak Only English at home	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Speaks English	0.107***	0.618***	0.107***	-0.276***	0.111***	0.144**
	(0.027)	(0.099)	(0.027)	(0.074)	(0.027)	(0.060)
Speaks Only English at home	0.153***	0.609***	0.150***	-0.180*	0.151***	0.146*
	(0.043)	(0.147)	(0.043)	(0.097)	(0.042)	(0.082)
English-skill COB-MSA Measure	0.001	0.009***	-0.001	-0.009***	0.002	0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.010)
Interaction: Speaks English		-0.010***		0.009***		-0.005
		(0.002)		(0.002)		(0.009)
Interaction: Speaks only English at home		-0.009***		0.007***		-0.001
		(0.003)		(0.002)		(0.010)
COB % Self-Employed	0.096***	0.095***	0.097***	0.095***	0.096***	0.096***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
MSA Self-Employment Index	0.116***	0.122***	0.116***	0.121***	0.114***	0.114***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
%COB 1990 population in MSA	0.005***	0.004***	0.005***	0.004***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
%MSA from COB	1.024**	1.098**	1.035**	1.098**	0.995**	0.988**
	(0.487)	(0.486)	(0.487)	(0.484)	(0.478)	(0.474)
Observations	218,885	218,885	218,885	218,885	218,885	218,885
Pseudo R-squared	0.069	0.070	0.069	0.070	0.069	0.069

Source: Author's calculations based on US Census PUMS 5% sample. The data universe is limited to men in the labor force, between ages 25 and 65, who emigrated as adults from a non-English speaking country. All regressions controlled for age, age-squared, ethnicity, race, five education groups, years since migration, years since migration squared, spouse in household, naturalized, and the median years since migration in MSA-COB.

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

Clustered robust standard errors are reported in parentheses.

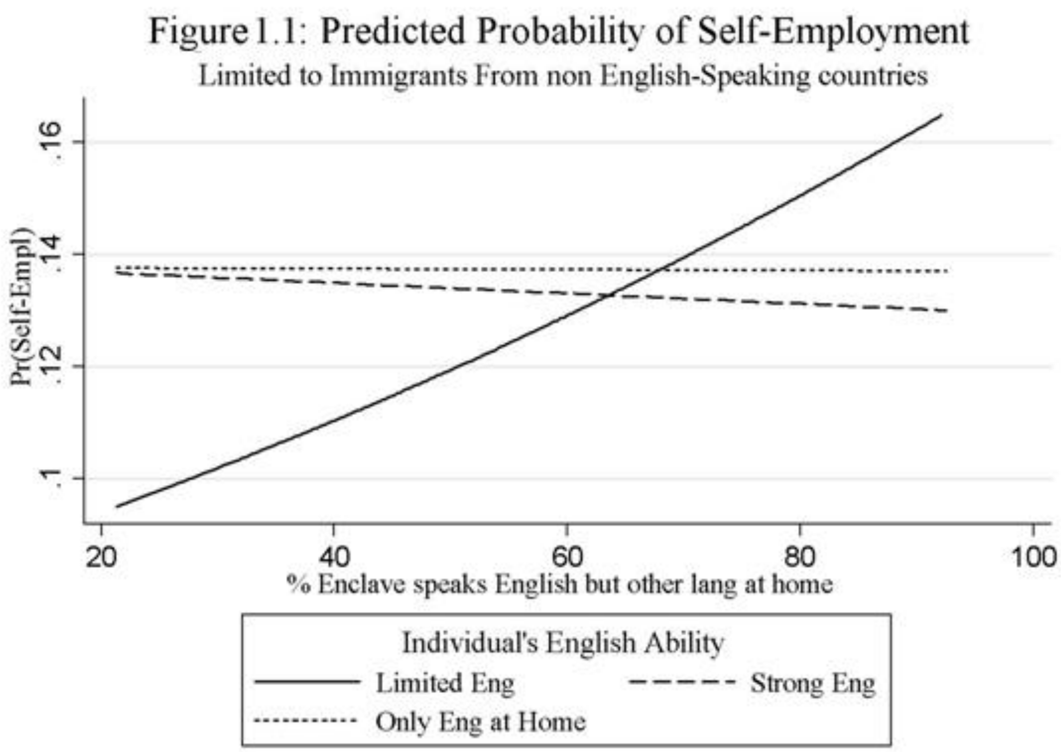
Specification (II), in which the proportion of the community that speaks English is interacted with the individual's ability to speak English, shows that the proportion of co-ethnics who speak English in a community has differing effects on individuals based on whether or not they themselves speak English. The coefficient on the un-interacted enclave fluency level measure indicates that immigrants with limited English skills are more likely to become self-employed as their COB-MSA group’s English-speaking rate increases. The interacted terms of the enclave fluency measure indicate that in communities with few English speakers, immigrants who do not speak English are less likely to become self-employed than those who speak English. But, if these same immigrants reside in a community with a high level of English fluency, they become as likely to be self-employed as immigrants with strong English skills. Interestingly, immigrants with more English skills show almost no sensitivity to their co-ethnics' English-skills when deciding whether or not to become self-employed.

Table 1.6 also reports the results for similar regressions using the percent of the COB-MSA with limited English skills and the percent of the COB-MSA that reports speaking only English at home. As expected from the previous results, I find that immigrants with weak English ability (the omitted group) are more likely to be self-employed when the percent of immigrants with weak English skills is low. I find no significant effect for immigrants who speak English well, but use a different language at home. The estimated net effect for immigrants who speak only English at home, however, decreases from 0 in specification (II) to -0.002 in specification (IV). That is, though the proportion of co-ethnics who spoke English well had no significant impact on immigrants who spoke only English at home, a decrease in the proportion of co-ethnics with limited English skills results in an *increase* in their likelihood of self-employment. For example, residing in a community in which 80% of co-ethnics do not speak English results in a human capital spillover marginal effect of -0.016; the marginal effect is only -0.005 in a community where 25% do not speak English.[19]

Since I have already excluded those from countries where English is the official or primary language, immigrants who speak only English at home represent the most assimilated immigrants in this sample. This is the group that is most likely to belong to social groups outside

[19] Marginal effects are calculated using the mean self-employment rate for immigrants who report speaking only English at home (14.5%) and the net effect of the enclave and interaction effects at the two levels of community limited English rates. Since this logit model uses interaction terms, marginal effects cannot be easily calculated (Norton, Wange and Ai 2004); hence the effects reported here are illustrative approximations.

of their ethnic groups. Specifications (V) and (VI) show that the proportion of the enclave that only speaks English at home does not have a statistically significant impact on the likelihood of self-employment for any of the three language groups. This supports the hypothesis that the effect of the proportion of the enclave that speaks English well or very well is due to human capital spillover effects based on local ethnic interactions, and not due to some other unmeasured human capital effects that are not being captured at the COB-MSA level.



The fitted probability of self-employment as a function of the percent of the enclave that speaks a different language at home but reports speaking English well or very well is graphed in Figure 1.1.[20] According to Figure 1.1, in high fluency communities, people with limited English skills

[20] This and the other figures showing the fitted probability of self-employment are calculated for white, naturalized, college educated immigrants who reside with a spouse. All other controls in the regression are set to the sample

are the most likely to be self-employed. On the other hand, in low fluency communities, immigrants with limited English ability are far less likely to be self-employed than similar immigrants who speak English. Since Figure 1.1 displays fitted self-employment probabilities, it is reasonable to ask whether the results are relevant or out of sample. Appendix C contains human capital distributions to go along with each of the figures presented. Table 1.A5 shows that 4.7% of the sample not from Spanish or English-speaking countries falls into the group of immigrants who report limited English skills but reside in communities of over 70% English speakers. This cell is particularly sparse for immigrants from Spanish-speaking countries; only 500 individuals (less than 0.5%) fall into this group. However, almost 10% of individuals from Spanish-speaking countries who report having limited English-skills reside in communities where between 50 and 70% of the adult population speak English well and use a different language at home. This is the region of Figures 1.1 and 1.2 where immigrants with the three different English-skills measure converge in roughly the same propensity for self-employment.

Overall, the story that emerges from the regressions in Table 1.6 and Figure 1.1 is one in which an individual who speaks a different language at home but has strong English skills is not affected by his co-ethnics' fluency rates. An immigrant who speaks English is more likely than someone with limited English to start a business if both reside in communities where under half report speaking English well or very well but use a different language at home. However, as the proportion of the community that speaks English increases past 50%, individuals with limited English ability experience a steep increase in the likelihood of becoming self-employed, showing the same propensity for self-employment as their fluent co-ethnics. Additionally, the proportion

---

averages. Note that these probabilities include own-language and own-education effects, thus enabling direct comparisons between different groups.

of the community that speaks only English at home does not impact self-employment, indicating that the language results stem from network effects based on social interactions within the co-ethnic community.

Table 1.7: Testing COB-MSA English-skills Effects for Immigrants from Spanish-speaking Countries: Logit Regression Results

	Type of English-skill Enclave Measure			
	% Speak English Well		% Limited English	
	Spanish-speaking	Other Non-English	Spanish-speaking	Other Non-English
Speaks English	0.632***	0.406	-0.409***	-0.062
	(0.147)	(0.276)	(0.150)	(0.126)
Speaks Only English at home	0.956***	-0.016	-0.852***	0.129
	(0.222)	(0.369)	(0.204)	(0.147)
English-skill Enclave Measure	0.022***	-0.001	-0.022***	0.000
	(0.005)	(0.005)	(0.005)	(0.004)
Interaction: Speaks English	-0.010***	-0.005	-0.010***	0.009***
	(0.003)	(0.004)	(0.002)	(0.002)
Interaction: Speaks only English at home	-0.017***	0.001	-0.009***	0.007***
	(0.005)	(0.005)	(0.003)	(0.002)
COB % Self-Employed	0.067***	0.087***	0.076***	0.003**
	(0.017)	(0.004)	(0.015)	(0.001)
MSA Self-Employment Index	0.080*	0.161***	0.070	1.195**
	(0.046)	(0.032)	(0.045)	(0.480)
%COB 1990 population in MSA	0.004**	0.002	0.004**	0.002
	(0.002)	(0.003)	(0.002)	(0.003)
%MSA from COB	1.609***	-7.556***	1.559***	-7.589***
	(0.506)	(1.941)	(0.521)	(2.011)
Observations	131,711	88,783	131,711	88,783
Pseudo R-squared	0.041	0.085	0.041	0.085

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to men in the labor force, between the ages of 25 and 65, who emigrated as adults from a non-English speaking country. All regressions controlled for age, age-squared, ethnicity, race, five education groups, years since migration, years since migration squared, spouse in household, naturalized, and median years since migration in MSA-COB.

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

Clustered robust standard errors are reported in parentheses.

Since Spanish is widely spoken in the U.S., I also look separately at the impact of English skills on Spanish-speaking immigrant communities.[21] If social interactions are dictated by language rather than country of origin, a Spanish-speaking immigrant will be less reliant on his or her own COB-MSA group. Speaking Spanish would, for example, increase the number of potential trade partners in the area to include many individuals who are not from the same country. In fact, I find the opposite – Table 1.7 shows that Spanish-speaking immigrants, which make up 60% of the immigrant sample, drive the sensitivity to the enclave's language skills from the previous results. This indicates the importance of the COB-MSA social networks rather than a network based on a common language.

Table 1.7 details how these two groups of immigrants who are not from an English-speaking country differ in terms of enclave effects. The first important difference is that immigrants from Spanish-speaking countries who do not speak English are far less likely to be self-employed than their co-ethnics who speak English. On the other hand, neither the individual's English skills nor those of his enclave have a statistically significant impact on immigrants who are not from Spanish-speaking countries. For immigrants from Spanish-speaking countries, the proportion of the COB-MSA that speaks English increases the self-employment likelihood for all three English-skills groups; the proportion with limited English skills decreases the self-employment likelihood only for immigrants who speak English but use a different language at home and for those with limited English skills. Further tests separate Mexican immigrants, the majority of the Spanish-speaking sample, from all other Spanish-speaking groups, revealing that these results hold for both groups. Another set of tests showed that controlling for the overall percent of the MSA population that spoke Spanish at home only slightly weakened the impact of COB-MSA

[21] The following are included in the group "Spanish-speaking countries": Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama,

English skills, but did not significantly change the results.

Figure 1.2: Predicted Probability of Self-Employment
By English Skills of Enclave

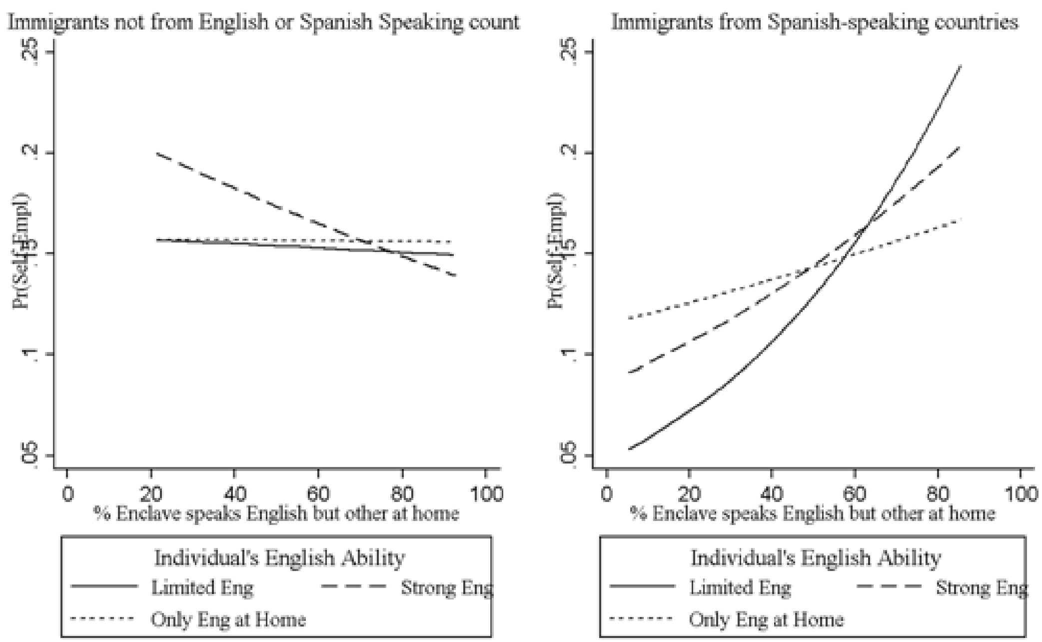


Figure 1.2 shows the predicted self-employment probabilities as a function of the COB-MSA's English skills separately for immigrants from Spanish-speaking countries and those from countries where neither Spanish nor English are the dominant languages. Recall from the regression results that the coefficients for immigrants from non-Spanish speaking countries are not statistically different from zero. However, for the immigrants from Spanish-speaking countries, the propensity for self-employment of immigrants with limited English skills increases dramatically as the proportion of the enclave who reports having strong English skills increases. Interestingly, immigrants who did not have limited English skills also show a sizeable increase in self-employment propensity as the enclave's language skills increased. This is true even for those

Paraguay, Peru, Spain, Uruguay, and Venezuela. Mexican immigrants make up about two-thirds of all sampled immigrants from Spanish-speaking countries.

who only speak English at home, indicating the presence of a protected market in these communities, not only human capital externalities.

Results: Enclave effects by Educational Attainment

As in the English skills analysis above, the individual's education level was interacted with the enclave's education levels, measured as the percent of the COB-MSA adult population with less than a high school degree for regressions (I) and (II) and the percent of the COB-MSA adult population with at least some post-secondary schooling for regression (III).[22] The regressions consider the impact of the enclave's education on five educational groups: those with eight or fewer years of schooling, those with some high school education but no degree, those with a high school degree, those with some post-secondary schooling, and those with a college degree or higher (the omitted group). Note that this set of regressions includes a control for the proportion that speaks English in the COB-MSA cell since English ability and education are highly correlated. Table 1.8 reports the logit coefficients from these regressions, again reporting the effect of each enclave-level education measure through separate regressions.

The first two rows look at the impact of the proportion of the enclave that has not earned a high school degree. Regression (I) reports that individuals in enclaves with a greater share of immigrants who did not complete high school are less likely to be self-employed. Immigrants with fewer than eight years of schooling and those with a college degree or higher were the least likely to be self-employed. Column (II) disaggregates this enclave effect to consider the impact on each educational group separately. For immigrants without a college degree, an increase in

[22] Note that the percent of the COB-MSA cell that has exactly a high school degree is excluded from both aggregate measures of education, thus they are not just inverse images of each other. See appendix C for a distribution of the enclave-level educational attainment variables.

Table 1.8: Testing Enclave Schooling Effects: Logit Regression Results for Five Educational Groups

	Educational Attainment of Enclave Measure		
	% with less than HS diploma		% with more than HS diploma
	(I)	(II)	(III)
8 years or less	0.019	0.407***	-0.699***
	(0.037)	(0.084)	(0.080)
Some HS	0.139***	0.472***	-0.587***
	(0.041)	(0.064)	(0.076)
HS diploma	0.141***	0.379***	-0.531***
	(0.035)	(0.055)	(0.073)
Some college	0.141***	0.278***	-0.358***
	(0.033)	(0.052)	(0.077)
Enclave measure	-0.009	0.000	0.000
	(0.002)	(0.002)	(0.002)
Interaction: 8 years or less		-0.013***	0.015***
		(0.002)	(0.002)
Interaction: some HS		-0.012***	0.014***
		(0.002)	(0.001)
Interaction: HS diploma		-0.011***	0.012***
		(0.001)	(0.001)
Interaction: Some college		-0.008***	0.008***
		(0.001)	(0.001)
COB % Self-Employed	0.087***	0.087***	0.088***
	(0.005)	(0.005)	(0.004)
MSA Self-Employment Index	0.122***	0.126***	0.113***
	(0.033)	(0.031)	(0.031)
%COB 1990 population in MSA	0.003**	0.002	0.002*
	(0.001)	(0.002)	(0.001)
%MSA from COB	1.195**	1.369***	1.166**
	(0.480)	(0.489)	(0.475)
Observations	232,952	232,952	232,952
Pseudo R-squared	0.067	0.068	0.069

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to men in the labor force, between 25 and 65, who immigrated as adults. All regressions controlled for age, age-squared, ethnicity, race, English ability, years since migration, years since migration squared, spouse in household, naturalized, median years since migration in MSA-COB, percent of COB-MSA who speak English fluently or only English at home, percent self-employed in COB, MSA Self-employment index, percent of MSA who was born in COB, and percent of COB 1990 population in the U.S. who was residing in MSA.

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

Clustered robust standard errors are reported in parentheses.

the proportion of the enclave without a high school diploma results in a decrease in self-employment. This negative effect decreases as the individual's educational attainment increases. For those with a college degree, the educational attainment of the enclave does not impact self-employment.

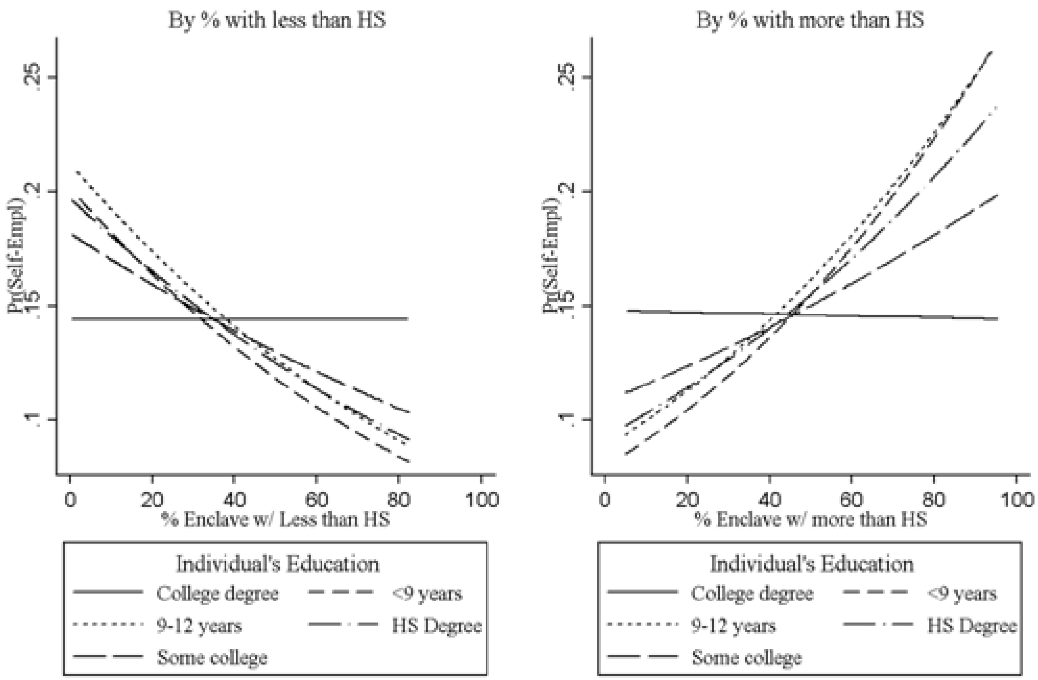
These results are supported by the impact of the proportion of the enclave with some post-secondary education (III). Consistent with the hypothesis, the lower an individual's educational attainment is, the more the enclave's ethnic capital affects his propensity to become self-employed. Comparing the coefficients from column (III) with those from column (II) shows that the propensity to become self-employed for immigrants with less than a high school degree is more sensitive to the proportion of their enclave with some post-secondary education than it is to the proportion of their enclave without a high school degree. Specifically, low education immigrants benefit more from residing among college educated co-ethnics (in terms of self-employment opportunities) than they suffer from residing among other low education co-ethnics.[23]

The fitted probabilities of self-employment are graphed for each educational group on Figure 1.3 by both the proportion of the COB-MSA group that had less than a high school diploma and the proportion with more than a high school diploma. As the regression results showed, in high education COB-MSA communities, immigrants with low education are more likely to become self-employed than immigrants with a college education or more.[24] As the proportion of the COB-MSA with more education falls, so does the likelihood of immigrants without a college

[23] Additionally, this regression was run for immigrants from Spanish-speaking countries separately from all other immigrants. The results, though slightly weaker for the Spanish-speaking COBs, were still statistically significant and in the same direction. Results available from author upon request.

degree of becoming self-employed. When about 40% of the enclave has less than a high school education, then immigrants of all educational attainments have about the same self-employment propensity. As the proportion of immigrants without high school diplomas continues to increase, the probability that an immigrant with less than a college degree becomes self-employed

Figure 1.3: Predicted Probability of Self-Employment



continues to decrease, falling below the probability of self-employment for immigrants with a college degree. Additionally, the flat probability of self-employment for college educated immigrants supports the hypothesis that the changes in probability of self-employment are driven by access to information and capital brought about by human capital externalities, not by catering to ethnic demand. Immigrants with a college degree should be, and empirically are shown to be, making the decision to become self-employed based on their own human capital

[24] As is shown in Appendix C, less than 1,000 immigrants with less than a high school degree reside in an enclave where over 80% of the community has more than a high school diploma. This extreme part of the results is, in essence, out of sample.

and abilities and not on the human capital of their communities.

Conclusion

This paper extends the research already done on ethnic capital and neighborhood effects by considering the impact of human capital externalities, measured as community-level English skills and formal schooling, on the likelihood of self-employment for different groups of immigrants. Both of the community human capital measures tested above support the hypothesis that immigrants with low levels of human capital benefit from the human capital externalities of their co-ethnics. Furthermore, they show greater reliance on their co-ethnic communities than immigrants with either a college education or those who speak English. The empirical results support the existence of protected markets among Spanish-speaking immigrants, as shown in the increased propensity for self-employment even among those who only speak English at home. The language results also indicate the presence of human capital externalities among immigrants from Latin America. I also find that the educational attainment of a community favors self-employment by reducing self-employment costs; this effect is far stronger than the potential increase in opportunity costs of self-employment resulting from the educational attainment of the community.

I did not find significant evidence of language-skills externalities outside of Spanish-speaking country of origin groups. But, among immigrants from Spanish-speaking countries, residing in a COB-MSA group with more English-speakers results in a significantly higher probability of self-employment for immigrants with limited English. This positive effect, though weaker, is also found for immigrants who speak English. This can be interpreted as evidence of human capital externalities playing a large role for those with limited English skills, and the presence of protected markets for ethnic goods and services for those with strong English-skills. When considering the role of education of the enclave and self-employment, I find that college educated immigrants seek self-employment independently of what their enclaves look like. On the other hand, those with less than a college degree show a higher probability of self-employment as the overall human capital of their community increases (measured as the percent of the COB-MSA that has higher education). This effect is stronger for immigrants from non-Spanish speaking countries, though still significant for those from these countries.

Both enclave tests, the English skills and educational attainment of a COB-MSA group, indicate the presence of strong human capital externalities at play within ethnic communities in the United States. These externalities play an important role in the economic assimilation of low human capital immigrants by potentially offsetting some of the economic costs associated with low education and limited English skills. Since acquiring these skills might be prohibitively expensive for some groups, primarily immigrants with the lowest levels of education, having access to a co-ethnic community with higher human capital might serve as an affordable alternative. To the extent that self-employment can serve as a vehicle for economic assimilation for immigrants in the U.S., human capital externalities from co-ethnics can serve as a social tool for economic assimilation as well.

APPENDIX

Appendix A: Specification Testing

In order to test the validity of the specifications containing only continuous variables for $C_j$and L_k, I ran the regressions presented in this paper with four different specifications: (I) using only the continuous controls, (II) using only the two sets of dummy variable controls, (III) using continuous MSA control with the COB dummy variables, and (IV) using the continuous COB control with the MSA dummy variables. Note that the standard errors in specifications (II) through (IV) are not reliable due to insufficient degrees of freedom.

The test shows that, if anything, the specification used in this paper underestimates the enclave effects. Table 1.A1 shows the results of this test for the impact of a community's educational attainment on self-employment. Table 1.A2 shows the results for the English-skill enclave test for immigrants who are not from Spanish or English-speaking countries. The results show that the magnitudes of the coefficients do not change much except that "speaks only English at home" becomes more negative.

Table 1.A3 shows the results for the same set of tests performed as in Table 1.A2 for immigrants from Spanish-speaking countries. Again, the results presented above are robust to different specifications. The only coefficient that changes magnitude by a significant amount is the coefficient on the un-interacted language enclave effect. Using COB and MSA dichotomous controls result in a smaller language enclave effect for immigrants from Spanish-speaking countries. This implies, however, that the overall self-employment rates of the COB groups and the industrial distribution between MSA's is less informative for Latin American/Spanish immigrants than it is for other immigrants from non-English speaking countries. Since the results

Table 1.A1: Specification Testing of Impact of Educational Attainment of Community, as Measured by the Percent of the Local Co-ethnic Community Without a High School Diploma, on Individual's Propensity for Self-Employment

Specification	I	II	III	IV
8 years or less	0.407***	0.422***	0.495***	0.372***
	(0.084)	(0.083)	(0.086)	(0.081)
Some HS	0.472***	0.526***	0.559***	0.474***
	(0.064)	(0.062)	(0.062)	(0.061)
HS diploma	0.379***	0.417***	0.449***	0.364***
	(0.055)	(0.052)	(0.053)	(0.052)
Some college	0.278***	0.328***	0.359***	0.257***
	(0.052)	(0.049)	(0.049)	(0.051)
Enclave measure	0.000	0.008**	-0.002	0.001
	(0.002)	(0.003)	(0.004)	(0.002)
Interaction: 8 years or less	-0.013***	-0.013***	-0.015***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)
Interaction: some HS	-0.012***	-0.013***	-0.014***	-0.012***
	(0.002)	(0.001)	(0.002)	(0.001)
Interaction: HS diploma	-0.011***	-0.011***	-0.012***	-0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Interaction: Some college	-0.008	-0.009***	-0.010***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.001)
Constant	-6.491***	-4.910***	-5.564***	-4.976***
	(0.399)	(0.593)	(0.402)	(0.480)
Observations	232,952	232,794	232,952	232,794
Pseudo R-squared	0.068	0.077	0.072	0.074

Source: Author's calculations based on U.S. Census PUMS 5% sample. Specification (I) controls for COB and MSA effects by using two continuous measures: the percent self-employed in COB and MSA Self-employment index. Specification (II) replaces these two continuous variables with dichotomous MSA and COB identifiers. Specification (III) uses the vector of COB dichotomous identifiers with the MSA Self-employment index. Specification (IV) uses the percent self-employed in COB combined with the MSA dichotomous variables. All regressions controlled for age, age-squared, ethnicity, race, English ability, years since migration, years since migration squared, spouse in household, naturalized, median years since migration in MSA-COB, percent of COB-MSA who speak English fluently or only English at home, percent of MSA who was born in COB, and percent of COB 1990 population in the U.S. who was residing in MSA.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Clustered robust standard errors are reported in parentheses.

in specification (IV) differ most dramatically from those in specification (I), Spanish-speaking immigrants seem to enter self-employment based more on MSA than on the industrial-distribution of the MSA.

Table 1.A2: Specification Testing of Impact of English Language Skill of Community, as Measured by the Percent of the Local Co-ethnic Adult Community that Speaks English Well or Very Well, on Individual's Propensity for Self-Employment, for Immigrants from Non-Spanish and Non-English Speaking Countries

Specification	I	II	III	IV
Speaks English	0.406	0.401*	0.449*	0.405*
	(0.276)	(0.206)	(0.237)	(0.220)
Speaks Only English at home	-0.016	-0.036	0.033	-0.072
	(0.369)	(0.304)	(0.315)	(0.306)
English-skill Enclave Measure	-0.001	-0.001	-0.001	-0.001
	(0.005)	(0.004)	(0.005)	(0.004)
Interaction: Speaks English	-0.005	-0.005	-0.006	-0.005
	(0.004)	(0.003)	(0.004)	(0.003)
Interaction: Only English at home	0.001	0.001	-0.001	0.002
	(0.006)	(0.004)	(0.005)	(0.005)
Observations	88,783	88,783	88,783	88,783
Pseudo R-squared	0.085	0.094	0.090	0.090

Source: Author's calculations based on U.S. Census PUMS 5% sample. Specification (I) controls for COB and MSA effects by using two continuous measures: the percent self-employed in COB and MSA Self-employment index. Specification (II) replaces these two continuous variables with dichotomous MSA and COB identifiers. Specification (III) uses the vector of COB dichotomous identifiers with the MSA Self-employment index. Specification (IV) uses the percent self-employed in COB combined with the MSA dichotomous variables. All regressions controlled for age, age-squared, ethnicity, race, five education groups, years since migration, years since migration squared, spouse in household, naturalized, median years since migration in MSA-COB, percent of MSA who was born in COB, and percent of COB 1990 population in the U.S. who was residing in MSA.

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

Clustered robust standard errors are reported in parentheses.

Table 1.A3: Specification Testing of Impact of English Language Skill of Community, as Measured by the Percent of the Local Co-ethnic Adult Community that Speaks English Well or Very Well, on Individual's Propensity for Self-Employment, for Immigrants from Spanish Speaking Countries

Specification	I	II	III	IV
Speaks English	0.632***	0.670***	0.660***	0.673***
	(0.147)	(0.135)	(0.143)	(0.137)
Speaks Only English at home	0.956***	0.972***	0.988***	0.816***
	(0.222)	(0.217)	(0.221)	(0.211)
English-skill Enclave Measure	0.022***	0.014***	0.019***	0.015***
	(0.005)	(0.005)	(0.006)	(0.004)
Interaction: Speaks English	-0.010***	-0.011***	-0.011***	-0.011***
	(0.003)	(0.003)	(0.003)	(0.003)
Interaction: Only English at home	-0.017***	-0.017***	-0.017***	-0.014***
	(0.005)	(0.005)	(0.005)	(0.005)
Observations	131,711	131,553	131,711	131,553
Pseudo R-squared	0.041	0.051	0.042	0.050

Source: Author's calculations based on U.S. Census PUMS 5% sample. Specification (I) controls for COB and MSA effects by using two continuous measures: the percent self-employed in COB and MSA Self-employment index. Specification (II) replaces these two continuous variables with dichotomous MSA and COB identifiers. Specification (III) uses the vector of COB dichotomous identifiers with the MSA Self-employment index. Specification (IV) uses the percent self-employed in COB combined with the MSA dichotomous variables. All regressions controlled for age, age-squared, ethnicity, race, five education groups, years since migration, years since migration squared, spouse in household, naturalized, median years since migration in MSA-COB, percent of MSA who was born in COB, and percent of COB 1990 population in the U.S. who was residing in MSA.

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

Clustered robust standard errors are reported in parentheses.

Appendix B: Sample selection

Table 1.A4 illustrates the differences between the immigrants who were dropped from the analysis due to COB-MSA cell size. By dropping those who lived in communities of less than 50 sampled individuals, I excluded a disproportionate number of white or Asian immigrants with education exceeding a high school degree. The excluded group was also more likely to be self-employed.

Table 1.A4: Characteristics of Sample, By Sample Size within Enclave, in Percentages

Sample size of Enclave	Overall	Self-Employed	White	Black	Other Race	Hispanic
1000 or more	34.88	10.37	4.75	2.89	15.06	77.29
100-999	35.16	12.54	17.25	8.28	35.22	39.26
50-100	8.06	14.56	35.14	10.06	31.71	23.09
<50	21.90	15.18	42.37	10.03	28.65	18.95
Total	100.00	12.49	19.83	6.93	26.47	46.77
	Less than HS	HS diploma	More than HS	Speaks English	Only English at home	Limited English
1000 or more	59.67	15.97	24.36	44.23	7.00	48.76
100-999	33.35	17.75	48.90	59.34	10.84	29.82
50-100	24.45	18.43	57.12	60.57	18.75	20.69
<50	19.05	17.30	63.65	63.08	21.85	15.07
Total	38.68	17.08	44.24	54.99	12.55	32.46

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to male immigrants in the labor force, not in school and between the ages of 25 and 65 who immigrated as adults.

Appendix C: Distribution of Enclave Human Capital Measures

Tables 1.A5, 1.A6 and 1.A7 illustrate the sampled and estimated number of people who fall into groups of interest based on the fitted probabilities presented in Figures 1.1, 1.2 and 1.3. Particularly thin cells exist for immigrants from Spanish-speaking countries with limited English skills who reside in communities where over 70% of immigrants speak English well and for immigrants with less than a high school diploma who reside in enclaves where over 80% have more than a high school diploma.

Table 1.A5: Percent of Immigrants Residing in MSA-COB Cells With Different Levels of English Ability, by Language Group

% of MSA-COB Cell That Speaks English Well, but Speaks a Different Language at Home	Individual's English Language Skills					
	Non-English and Non-Spanish Speaking COB			Spanish-speaking COB		
	Limited	Strong	Only English	Limited	Strong	Only English
Under 50	3.12	2.16	0.30	43.05	29.51	3.76
50-70	12.17	23.80	1.51	9.68	11.27	0.88
70+	4.71	48.76	3.46	0.41	1.35	0.08

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to male immigrants in the labor force, not in school and between the ages of 25 and 65 who immigrated as adults.

Table 1.A6: Percent of Immigrants Residing in MSA-COB Cells With Different Levels of High School Completion Rates

% of MSA-COB Cell Without a High School Diploma	Individual's Educational Attainment				
	< 9 years	9 - 12 years	High School Diploma	Some College	College Degree
Under 20	0.66	1.45	3.75	5.58	17.00
20 - 40	2.43	3.76	5.21	4.60	4.41
40 - 60	3.22	2.38	2.32	1.61	1.23
60 +	20.17	9.23	5.97	3.42	1.60

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to male immigrants in the labor force, not in school and between the ages of 25 and 65 who immigrated as adults.

Table 1.A7: Percent of Immigrants Residing in MSA-COB Cells With Different Levels of Some College Attendance

% of MSA-COB Cell With Some College Attendance	Individual's Educational Attainment				
	< 9 years	9 - 12 years	High School Diploma	Some College	College Degree
Under 20	19.79	9.05	5.90	3.31	1.53
20 - 40	5.07	4.96	5.81	4.09	3.20
40 - 60	1.48	2.53	4.92	6.26	11.68
60 +	0.14	0.28	0.63	1.55	7.83

Source: Author's calculations based on U.S. Census PUMS 5% sample. The data universe is limited to male immigrants in the labor force, not in school and between the ages of 25 and 65 who immigrated as adults.

REFERENCES

Alba, Richard D., John R. Logan, Brian J. Stults, Gilbert Marzan, and Wenquan Zhang. 1999. Immigrant Groups in the Suburbs: A Reexamination of Suburbanization and Spatial Assimilation. *American Sociological Review* 64:3.

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

Bates, Timothy. 1997. Financing Small Business Creation: The Case of Chinese and Korean Immigrant Entrepreneurs. *Journal of Business Venturing* 12:2.

Beaman, Lori A., 2009. Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S. Revise and Resubmit, *Review of Economic Studies*.

Berman, Eli, John Bound and Zvi Griliches. 1994. Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers. *The Quarterly Journal of Economics* 109:2.

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

Blau, Francine D., Lawrence M. Khan, and Kerry L. Papps. 2010. Gender, Source Country Characteristics and Labor Market Assimilation Among Immigrants: 1980 - 2000. *The Review of Economics and Statistics*, forthcoming.

Bleakley, Hoyt and Aimee Chin. 2004. Language Skills and Earnings: Evidence from Childhood Immigrants. *The Review of Economics and Statistics* 86:2.

Bohn, Sarah and Sarah Pearlman. 2009. Ethnic Concentration and Bank Use in Immigrant Communities. Public Policy Institute of California Working Paper.

Borjas, George J., 1986. The Self-Employment Experience of Immigrants. *The Journal of*

Human Resources 21:4.

--- 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.

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

Bowles, Jonathan and Tara Colton. 2007. *A World of Opportunity*. Center for an Urban Future. New York, NY.

Carliner, Geoffrey. 2000. The Language Ability of U.S. Immigrants: Assimilation and Cohort Effects. *International Migration Review* 34:1.

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

Ellis, Mark and Richard Wright. 1999. The Industrial Division of Labor Among Immigrants and Internal Migrants to the Los Angeles Economy. *International Migration Review* 33:1.

Evans, David S. and Linda S. Leighton. 1989. Some Empirical Aspects of Entrepreneurship. *The American Economic Review* 79:3.

Friedberg, Rachel M. 2000. You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics* 18:2.

Georgarakos, Dimitris and Konstantinos Tatsiramos. 2009. Immigrant Self-Employment: Does Intermarriage Matter? Working Paper 4350, IZA Discussion Paper.

Hamilton, Barton H. 2000. Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment. *The Journal of Political Economy* 108:3.

Johnson, Jacqueline S. and Elissa L. Newport. 1989. Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology* 21.

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

--- 2005. Entrepreneurship. *Journal of Labor Economics* 23:4.

Le, Anh T. 1999. Empirical Studies of Self-Employment. *Journal of Economic Surveys* 13:4.

Lofstrom, Magnus. 2002. Labor Market Assimilation and the Self-Employment Decision of Immigrant Entrepreneurs. *Journal of Population Economics* 15:1.

--- 2009. Low-skilled Immigrant Entrepreneurship. Working Paper 4560, IZA Discussion Paper.

Norton, Edward C., Hua Wange and Chunrong Ai. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal* 4:2.

Rosenzweig, Mark R. 1995. Why Are There Returns to Schooling? *American Economic Review* 85:2.

Toussaint-Comeau, Maude. 2008. Do Ethnic Enclaves and Networks Promote Immigrant Self-Employment? *Economic Perspectives* IV.

de Wit, Gerrit. 1993. Models of Self-Employment in a Competitive Market. *Journal of Economic Surveys* 7:4.

Yuengert, Andrew. 1995. Testing Hypotheses of Immigrant Self-Employment. *The Journal of Human Resources* 30:1.

CHAPTER 2

IDENTIFYING ETHNIC ENCLAVES USING LINKED EMPLOYER-HOUSEHDOLD DATA

Ethnic enclaves first came to the attention of social scientists when Wilson and Portes (1980) wrote about the economic success of the Cuban enclave in Miami as an alternative to the segmented labor market theory. Building on dual labor market theory, they argued that enclave economies mirrored the primary sector by providing immigrants, who would otherwise be in the secondary labor market, with opportunities for promotion and human capital accumulation. This, they argued, paved the way for eventual profitable entrepreneurship within the enclave economy leading to economic success which would not otherwise be found outside of the enclave. Importantly, Wilson and Portes defined the enclave economy based on the employer's ethnicity and ethnic concentration in occupation and industry cells, rather than on residence.

Due to data limitations, subsequent research in this field has relied primarily on residential clustering, typically at the city level, as a proxy for the ethnic enclave economy (for example, Borjas 2000). This empirical identification strategy has not yielded the same rosy picture first reported by Wilson and Portes; instead, 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). However, the magnitude and direction of enclave effects are sensitive to the data and methodologies used. Edin, Fredriksson and Aslund (2003), using a pseudo-natural experiment design based on detailed residential data of refugees in Sweden, found that the negative impact on wages of residing in ethnic enclaves is explained by negative selection into enclaves. Similarly, Cutler, Glaeser and Vigdor (2008) find that using restricted-access micro data 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. These papers illustrate that measurement issues arising from data quality have significant effects on the estimation of enclave effects. In this paper, I combine residential data with workplace data and perform two primary tasks: 1) create measurements of immigrant proclivity to enclave based on both residential and employment clustering behavior, and 2) measure the proportion of immigrant segregation into both dimensions that can be explained by observable characteristics found in typical data sets.

Enclave effects, in essence, are the 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 can be boiled down to whether these ethnic networks provide economic opportunities or, on the contrary, limit the network members to fewer or less successful economic opportunities. However, since collecting data on social networks is prohibitively expensive and intrusive, the scope of such studies is often limited to a relatively small network (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[25] as proxies. 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, 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 who their neighbors are and with whom they tend to work. 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) versus those from the same country of birth and residing in the same city who are members of an ethnic enclave. Using the interaction of the two measures, several measures of enclave can be constructed and analyzed, shedding light on what today's immigrant enclaves look like and the significance of the role they play in the lives of contemporary immigrants.

Researchers have documented 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

[25] Ethnic identification includes country of birth (Cutler, Glaeser, and Vigdor 2008), self-reported ethnicity (Borjas 1992), race (Borjas and Bronars, 1989), and language (Bertrand, Luttmer, and Mullainathan 2000).

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).

Some research on job networks has been done using both place of residence and employer information (for example, Bayer, Ross and Topa 2008). Andersson et al. (2010) use the restricted 2000 Census and Unemployment Insurance data to look at 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 theory that social networks are being used intensively as recruitment networks.

Data

Detailed micro data from the Longitudinal Employer-Household Dynamics (LEHD) program at the U.S. Census Bureau combined with the 2000 Decennial Census allows for the identification of enclave economies using both residential and employment ties. The confidential 2000 U.S. Census of Population and Housing is a one-in-six household sample containing detailed residential and demographic data. This dataset provides data on 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: 1) access to block-level residential data, 2) larger sample size than the public-use 5% sample version of the long form 2000 census, an especially important consideration when studying immigrants by country of birth, and 3) the ability to link to the state Unemployment Insurance data used in the LEHD program. By linking the confidential decennial census to the LEHD files, I am able to identify the employers of all UI-covered jobs for the one-in-six sample. LEHD files also provide basic demographic details for all covered employees in the firm, including place of birth.[26] These data allow for the construction of coworker exposure measures, as described below.

The analysis is limited to five of the top immigrant urban areas in the U.S.: Los Angeles, New York, Chicago, Houston and Miami. These five Consolidated Metropolitan Statistical Areas (CMSA) were home to 47% of the immigrant population in the U.S. in 2000. A CMSA is a large urban area composed of Primary Metropolitan Statistical Areas (PMSA), cities and surrounding suburbs, connected by extensive commuting patterns. For example, the Chicago CMSA is composed of four separate PMSA's: Chicago, IL, Gary, IN, Kankakee, IL, and Kenosha, WI. 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.

[26] Abowd et al. (2005) provide a detailed description of the LEHD infrastructure files.

Data on coworker country of birth is primarily derived from Social Security Administration records, but has been imputed for about 4% of coworkers. These imputes are limited to the 23 largest country of birth groups plus 10 aggregated country groups.[27] To accommodate this feature of the data, both residential and workplace exposure rates will be calculated for this group of countries of birth and regions of birth.

Residential concentrations are calculated based on the population that is 16 years of age and older. Workplace concentration is further limited to those who are also 70 years old or younger and who report being in the labor force at the time of the census (including the unemployed and those working as unpaid family labor). The distributions of countries of birth and CMSA for the residential concentration measures and the workplace concentration measures are included in Table 2.1.[28] The majority of both samples are made up of the U.S. born population: white, non-Hispanics make up over 50% while another 10% of the total sample is black, non-Hispanics. The Hispanic U.S. born population is 7.9% of the sample population, but about 8.5% of the workforce population. Of the immigrant groups, the Mexican-born is the largest with over 5% of the total residential and workforce samples. Every other group makes up less than 2% of either sample, with the majority of these representing less than 1% of the total sample.

The last column in Table 2.1 reports the workforce sample, composed of the self-employed and the employed with LEHD earnings, as a percentage of the residential sample. Recall that the residential data includes all individuals ages 16 and up, including full time students and individuals who are not in the labor force, hence one should not expect a 100% match rate

[27] 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.

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

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

Table 2.1: Distribution of Ethnicity/Place of Birth for the Residential and Workforce Samples

Place of Birth	Proportion of Total Sample	Proportion of Total Work 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.

which has a labor force attachment rate of just under 73%. An important limitation of using administrative earnings data to study immigrants is that undocumented workers cannot be identified since they lack valid social security numbers. Since the linking between the decennial data and the LEHD data is based on social security numbers, this could lead to an underestimation of the co-ethnic exposure rates. Table 2.1 shows that, as a proportion of the residential sample, groups with higher rates of undocumented migration do not show different shares in employment attachment 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. It is probably the case that, if undocumented immigrants are a significant part of the populations of the urban areas chosen, they are equally underrepresented in both the LEHD data and in the residential data. On the other hand, some of the lowest employment shares belong to groups with low rates of illegal migration: Puerto Ricans (who, as U.S. citizens, have no illegal migration) 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%.

Methodology

Demographers and other researchers have developed a variety of indices to measure spatial distributions of different groups and clustering behavior (for an overview, see Iceland, Weinberg and Steinmetz 2002). Many of these measures consider the overall size of the population in determining the local relative size. 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):

$$CA_{jk} = \ln\left(\frac{C_{jk}/A_k}{L_j/T}\right)$$

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*. They chose not to use the simple proportion of the local population (the numerator) as the variable of interest because it underweighs small groups. Their measurement approach corrects for the overall size of a given language group by considering the size of that group relative to the U.S. population. Consider an extreme hypothetical case in which a very small group had all 100 of its members living in the same city tract. Though they still might not add up to a sizeable proportion of the tract, they are completely concentrated in a small geographic area, a trait captured by this measure. However, a group of 100 is still only 100 possible co-ethnic contacts, regardless of the size of the co-ethnic population *not* in the tract. Measures adjusting for small overall population size capture a dimension of the geographical distribution of an ethnic population that does not necessarily inform the question of local residential networks. Instead of adjusting for overall ethnic population size, most of my results are based on a simple proportion of the ethnic population in each Census tract. I do, however, adjust for tract size by using proportion of the tract population rather than just the number of co-ethnics in the tract to control for systematic tract size differences between urban and suburban areas.[29]

Bayer, McMillan and Rueben (2004) calculate a measure of average exposure for racial groups in the San Francisco area by looking at the distribution of the race of the household head by block group. This measure quantifies the expected contact between people belonging to different racial/ethnic groups based on the proportion of their neighbors (as measured within Census tract) that belong to other groups. 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 $k \in K$.[30] Then, $\sum_j n_{jk} = N_k$, where N_k is the total population in census tract k. Now, from the 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 or 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:

[29] Census tracts are designed to contain between 1,500 and 8,000 people.
[30] J includes the five U.S. born groups and each country or region of origin available in the LEHD data, as described above.

$$\bar{E}_K^{j,j'} = \frac{\sum_{k\in K}\left[n_{jk}\left(\mathbf{1}_{j'=j}C_k^j + \mathbf{1}_{j'=j^*}E_k^{j^*}\right)\right]}{N_K^j}$$

Where $N_K^j = \sum_{k\in K} n_{jk}$ is the total number of members of group j living in area K.

The same methodology is carried out to construct measures of co-ethnic coworker concentration. Specifically, by substituting k for $w \in W$ as employer identifiers, the results are coworker co-ethnic exposure rates: $C_w^j, E_w^{j^*}$ and $\bar{E}_W^{j,j'}$.

An important caveat in interpreting these exposure rates is that, since these measures are based solely on location of birth rather than ethnic identity, these measures are sensitive to when these immigrant groups arrived in the U.S. Consider a hypothetical case: a census tract composed entirely of Italian immigrants in 1960. Assume the families do not leave the tract and new ethnic groups 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 since they are counted as U.S. born, not Italian. If these U.S.-born Italian-Americans continue to draw their social networks primarily from Italian immigrants and their descendants, this results in an underestimation of the enclaved population.[31] 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. In order to exhibit high co-ethnic residential rates, it is necessary to be a recently arrived immigrant group whose members cluster in relatively few census tracts.

[31] This issue might be attenuated by using the decennial's ethnicity variable but, since there is no equivalent data in the LEHD files, it cannot be used in calculating both residential and workplace exposure rates.

Residential Ethnic Exposure Rates

The average residential co-ethnic exposure rate, $\bar{E}_K^{j,j}$, reported in Table 2.2, shows that, even among the largest immigrant groups in five of the biggest immigrant population centers in the U.S., most immigrant groups do not live in neighborhoods of high co-ethnic exposure rates. 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 neighborhoods where the co-ethnic adult population is only 5% of the adult 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 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 must be concentrated in relatively few census tracts. Immigrants born in Mexico, 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.-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, it is not what comes to mind when ethnic enclaves are discussed. At the 90th percentile, reported in the second column of Table 2.2, 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.

Table 2.2: Residential Own-Exposure Rates and Estimated Population Proportions Residing in Enclaves

	Residential own-exposure rates			Estimated proportion of POB population living in each type of tract		
Country of Birth	Average over all 5 CMSAs	90th percentile over all 5 CMSAs	Standard deviation over all 5 CMSAs	% of POB living in tracts predominantly co-ethnic	% of POB living in tracts with 25% or more co-ethnic	% of POB living in tracts where own POB group is largest immigrant group
Africa	0.0251	0.0637	0.1076	0	0.0015	0.0763
Caribbean	0.0741	0.2074	0.2286	0	0.0394	0.2269
Central America	0.0568	0.1429	0.2292	0	0.0556	0.0299
Central Asia	0.0303	0.0826	0.1171	0	0.0018	0.0645
Middle East/N. Africa	0.0243	0.0596	0.0804	0	0	0.0807
Oceania	0.0035	0.0095	0.0159	0	0	0.0007
Socialist Europe	0.0294	0.0742	0.1013	0.0001	0.0001	0.1097
South America	0.0713	0.1621	0.2339	0.0001	0.0455	0.2110
Southeast Asia	0.0343	0.0834	0.1250	0	0.0091	0.0325
Western Europe	0.0397	0.0800	0.1994	0.0043	0.0229	0.2551
Asian N.H. U.S.-born	0.0384	0.0848	0.1540	.	.	.
Black N.H. U.S.-born	0.4698	0.9553	0.9394	.	.	.
Hispanic U.S.-born	0.1634	0.3247	0.3277	.	.	.
Other N.H. U.S.-born	0.0163	0.0321	0.0501	.	.	.
White N.H. U.S.-born	0.6860	0.8968	0.5538	.	.	.
Canada	0.0088	0.0190	0.0235	0	0	0.0433
China	0.1105	0.3227	0.4141	0.0532	0.1225	0.2982
Colombia	0.0411	0.1016	0.1213	0	0.0002	0.0622
Cuba	0.3670	0.6929	0.7516	0.4183	0.6085	0.7598
Dominican Rep.	0.1911	0.5153	0.5338	0.1013	0.3293	0.4808
El Salvador	0.0727	0.1855	0.2060	0	0.0380	0.1349
Germany	0.0077	0.0156	0.0186	0	0	0.0351
Guatemala	0.0348	0.0852	0.1196	0	0.0042	0.0149
Haiti	0.1266	0.3189	0.3413	0	0.1662	0.4369
India	0.0527	0.1328	0.1932	0	0.0183	0.3261
Iran	0.0677	0.2057	0.2291	0	0.0455	0.4316
Italy	0.0318	0.0830	0.0941	0	0	0.2398
Jamaica	0.1061	0.2469	0.2759	0	0.0978	0.4247
Japan	0.0160	0.0421	0.0787	0	0	0.0627
Mexico	0.2480	0.4963	0.4682	0.0951	0.4549	0.8435
Philippines	0.0582	0.1479	0.2093	0.0001	0.0356	0.2013
Poland	0.0928	0.2584	0.3134	0.0091	0.1069	0.4172
Puerto Rico	0.0927	0.2308	0.2528	0.0001	0.0676	0.3352
South Korea	0.0652	0.1824	0.2450	0	0.0640	0.2684
Taiwan	0.0448	0.1195	0.1614	0	0.0110	0.1751
United Kingdom	0.0079	0.0175	0.0195	0	0	0.0307
Former U.S.S.R.	0.1325	0.3639	0.4294	0.0277	0.2118	0.4683
Vietnam	0.1218	0.3417	0.3882	0.0258	0.1935	0.3613
Overall immigrant				0.0467	0.1667	0.3719

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.

In the enclave effects literature, there is no empirical definition of an enclave. The last three columns in Table 2.2 offer three possibilities: a tract is an enclave of country of birth group j if 1) a majority of the tract population belongs to the foreign-born group in question, 2) a quarter of the tract population is from that group, or 3) that group is the largest immigrant group in the tract. 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. Interestingly, though 5% of the Chinese-born live in census tracts with over 50% of adults also born in China, only 12% live in tracts with over 25% of adults born in China. This indicates that, though some Chinese immigrants live in very high co-ethnic areas, the vast majority do not. The final enclave definition estimates the immigrant population who were residing in tracts where their immigrant group was the largest foreign-born group. Since the vast majority of the sample is U.S.-born, it is possible that a significant proportion of the U.S.-born population in ethnic neighborhoods is first or second-generation members of the same ethnic group. Under this definition, 37% of immigrants live in enclaves, including 84% of the Mexican-born population. A different issue arises with this definition: consider the hypothetical case in which there are only two immigrants living in census track k and they both happen to be from country j. Under this definition, these immigrants would be classified as enclaved even though they reside in a tract made up almost entirely of U.S.-born Americans.

Table 2.3 takes a different approach to the question of enclaving and exposure to different ethnic groups. The first column shows to which group each country of origin has the highest average

Table 2.3: 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
South East 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.

exposure (again averaged over all five CMSAs) while the second column reports this average exposure 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. Interestingly, Jamaican, Haitian and Caribbean immigrants,[32] all predominantly black, live in tracts where the largest group is non-Hispanic black Americans suggesting that race plays a part in residential choice for predominantly black groups. 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 2nd and 3rd generation Mexican-Americans living in the same neighborhoods as those who were born in Mexico. Guatemalan and Salvadorian immigrants, with residential exposure rates of about 0.15, also show high average exposure rates to Mexican immigrants. The other Hispanic groups, however, do not. Both immigrants from other Central American countries and Colombian immigrants have high rates of exposure to

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

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 tendency, including Vietnamese, Chinese, Korean, Russian, Polish, Haitian and Iranian immigrants.

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 2.4 shows the extent to which Hispanics (U.S. born and foreign born) 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 is the average own-exposure for each group as reported in Table 2.2. By reading across each row, it is easy to compare each group's own-exposure rate to its exposure rate of other 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 is probably the result of recent waves of Hispanic immigrants choosing to settle where previous waves had already settled and adult children remaining in the neighborhoods in which they grew up.

In short, this table illustrates that there is no "Hispanic" enclave though there is extensive regional clustering between some Hispanic groups. Dominican and Puerto Rican immigrants have higher exposure rates to each other than to any other foreign-born Hispanic group while

Table 2.4: 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 on the first column to the "column" country of birth.

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 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.

Workplace ethnic exposure rates

We now move to the second dimension of enclaving: workforce co-ethnic exposure rates. These were constructed using analogous estimation methods at the firm level rather than the census tract level. However, due to small cell size and data limitations, job network ties were measured in three different ways depending on the size of the employer's firm and on whether the individual reported being self-employed on the decennial census:

- In large firms, workplace co-ethnic exposure rate is measured as the proportion of an individual's coworkers in the year of analysis who are co-ethnics.
- In small firms (less than 6 employees), workplace co-ethnic exposure rate is calculated using the ethnicity of workers employed in firms in the same industry located in the same census block group. 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

who work for the same employer.[33] This measure is needed in order to address some of the measurement problems inherent in looking at coworkers in small firms. By construction, in small firms with no other employees a workplace exposure rate cannot be calculated. Furthermore, comparing workplace own-exposure for workers with only 5 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, co-ethnic density of the self-employed by industry and workplace census block group is used to calculate the ethnic composition of their coworkers.

Table 2.5: Distribution of Employer Type and Average Workplace Own-Exposure by Employer Type

	Percent	Average own-exposure
Large firm	84.98	0.3935
Self-employed	10.55	0.4169
Small firm	4.47	0.3676

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.

Table 2.5 shows the resulting workplace own-exposure rates by self-employment status and firm size for the employed. The first column of numbers shows that 85% of the 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 as for those in large firms, indicating that the pseudo-employers created by combining all firms in the census

[33] This is supported 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.

block by industry[34] leads to an acceptable approximation of coworker ties. In line with previous research that has shown significant ethnic clustering by industry, the self-employed have higher shares of co-ethnics as coworkers (defined as other self-employed individuals in the same industry and census block)

Tables similar to the residential exposure rates tables have been constructed using workplace exposure rates. Table 2.6 shows the average exposure to co-ethnics in the workplace by country of origin group and the most common co-worker ethnic group for each group. Mexican, Cuban and Chinese immigrants work in workplaces where a little over 20% of their coworkers are co-ethnics, a similar proportion of co-ethnic coworkers as that experienced by African-Americans, a much larger group. All of the U.S.-born groups have higher own-exposure rates in their neighborhoods than at their workplaces. Except for Russian and Iranian immigrants, all Asian and European groups, exhibit the opposite tendency – making up smaller proportions of their neighborhoods than of their workplaces. Though these groups might be too small to compose large proportions of their residential neighborhoods, this is evidence of the existence of job networks leading to ethnic clustering in the workplace. 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 12% 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 groups have more co-ethnic exposure in their neighborhoods than at their workplaces. This also holds for all non-Hispanic Caribbean

[34] Smaller industries were collapsed into similar industry groups to address issues arising from too few employers per industry group.

Table 2.6: Workplace Own-Exposure Rates and Estimated Population Proportions Working in Enclaved Workplaces

	Workplace own-exposure rates			Estimated proportion of POB population working in each type of firm		
Country of Birth	Average over all 5 CMSAs	90th percentile over all 5 CMSAs	Standard deviation over all 5 CMSAs	% of POB working in workplaces predominantly co-ethnic	% of POB working with 25% or more co-ethnics	% of POB in workplaces where co-ethnics are the largest group
Africa	0.0322	0.0744	0.2103	0.0054	0.0193	0.1117
Caribbean	0.0343	0.0805	0.1470	0.0017	0.0094	0.1067
Central America	0.0459	0.1263	0.2380	0.0065	0.0337	0.0651
Central Asia	0.0578	0.1482	0.3891	0.0342	0.072	0.133
Middle East/N. Africa	0.0377	0.0769	0.2640	0.0131	0.0338	0.1032
Oceania	0.0077	0.0090	0.1028	0.0007	0.0049	0.0316
Socialist Europe	0.0445	0.1208	0.2984	0.0162	0.0468	0.121
South America	0.0612	0.1473	0.2491	0.0081	0.0409	0.2250
Southeast Asia	0.0522	0.1250	0.3203	0.0207	0.0515	0.0864
Western Europe	0.0508	0.1314	0.2958	0.0169	0.0522	0.2364
Asian N.H. U.S.-born	0.0336	0.0652	0.1594			
Black N.H. U.S.-born	0.2394	0.4839	0.5358			
Hispanic U.S.-born	0.1504	0.2778	0.2956			
Other N.H. U.S.-born	0.0155	0.0275	0.0864			
White N.H. U.S.-born	0.6228	0.8750	0.5456			
Canada	0.0105	0.0180	0.0887	0.001	0.0034	0.0687
China	0.2076	0.7500	0.8232	0.2104	0.2927	0.3656
Colombia	0.0426	0.1000	0.2295	0.0088	0.0262	0.0805
Cuba	0.2281	0.5556	0.6498	0.1363	0.3856	0.6859
Dominican Rep.	0.1443	0.4124	0.5142	0.0625	0.2013	0.4224
El Salvador	0.0666	0.1667	0.2585	0.0078	0.0485	0.1231
Germany	0.0124	0.0160	0.1309	0.0019	0.0074	0.0874
Guatemala	0.0295	0.0735	0.1458	0.0013	0.0113	0.0389
Haiti	0.0911	0.2401	0.3511	0.0185	0.0946	0.3233
India	0.1087	0.3387	0.6110	0.0784	0.1201	0.3528
Iran	0.0403	0.0833	0.2753	0.0141	0.039	0.096
Italy	0.0340	0.0846	0.2092	0.0054	0.0259	0.1534
Jamaica	0.0587	0.1318	0.2649	0.009	0.0368	0.2809
Japan	0.1365	0.5455	0.6892	0.1153	0.2111	0.3083
Mexico	0.2229	0.4868	0.4953	0.0942	0.3685	0.8575
Philippines	0.0806	0.1939	0.3515	0.0233	0.0633	0.3875
Poland	0.1299	0.4286	0.5849	0.0848	0.1883	0.3747
Puerto Rico	0.0546	0.1333	0.2228	0.0047	0.0354	0.2845
South Korea	0.1324	0.5306	0.6872	0.1088	0.1616	0.2567
Taiwan	0.0775	0.2674	0.4208	0.0412	0.1081	0.1405
United Kingdom	0.0131	0.0217	0.1147	0.0022	0.0058	0.098
Former U.S.S.R.	0.0910	0.3103	0.4916	0.0559	0.1173	0.2696
Vietnam	0.1429	0.4769	0.6310	0.0972	0.1846	0.3367
Overall immigrant				0.0506	0.148	0.3652

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.

groups, Russians and Iranians. The most significant differences between residential own-exposure rates and workplace own-exposure rates among groups with higher residential clustering is among non-Hispanic African-Americans: their residential own-exposure rate is double that of their workplace own-exposure rate reflecting the high level of residential racial segregation between African-Americans and other U.S.-born groups. Of the immigrants, Cubans have the biggest difference between their residential own-exposure rate and their workplace own-exposure rates at 0.37 to 0.23 respectively. Though Cuban immigrants exhibit high rates of enclaving in both measures, it is clear that they have higher exposure to non-Cubans at the workplace than in their neighborhoods.

The last three columns in Table 2.6 show what percentage of each group and the overall immigrant population would be labeled as enclaved using the same three definitions as on Table 2 but applied to the workplace: 1) more than half of one's coworkers are co-ethnics, 2) at least 25% of coworkers are co-ethnics, and 3) co-ethnics are the largest immigrant group in the firm. 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, 15% work in firms where co-ethnics are at least 25% of the workforce and 37% work in firms where their immigrant group is the largest. Of course, the groups contributing to workplace enclaving rates differ somewhat from those contributing to residential enclaving.

Table 2.7, the workplace equivalent of Table 2.3, shows that, on average, individuals from all ethnic groups work for employers where the largest racial/ethnic group is white, U.S.-born non-Hispanics. The average work exposure rate to this group varies from a low of about 0.28 for Dominican immigrants to a high of 0.57 for German immigrants. The third and fourth columns

Table 2.7: 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	0.3893	Africa	0.0322
Caribbean	White N.H. U.S.-born	0.3730	Jamaica	0.0394
Central America	White N.H. U.S.-born	0.3232	Cuba	0.0773
Central Asia	White N.H. U.S.-born	0.3909	Central Asia	0.0578
MidEast/N Africa	White N.H. U.S.-born	0.4480	Mexico	0.0441
Oceania	White N.H. U.S.-born	0.4917	Mexico	0.0580
Socialist Europe	White N.H. U.S.-born	0.4473	Socialist Europe	0.0445
South America	White N.H. U.S.-born	0.3734	South America	0.0612
South East Asia	White N.H. U.S.-born	0.3850	Mexico	0.0579
Western Europe	White N.H. U.S.-born	0.4974	Western Europe	0.0508
Asian N.H. U.S.-born	White N.H. U.S.-born	0.4874	Mexico	0.0418
Black N.H. U.S.-born	White N.H. U.S.-born	0.4273	Mexico	0.0303
Hispanic U.S.-born	White N.H. U.S.-born	0.4415	Mexico	0.0720
Other N.H. U.S.-born	White N.H. U.S.-born	0.5175	Mexico	0.0494
White N.H. U.S.-born	White N.H. U.S.-born	0.6228	Mexico	0.0296
Canada	White N.H. U.S.-born	0.5589	Mexico	0.0378
China	White N.H. U.S.-born	0.3066	China	0.2076
Colombia	White N.H. U.S.-born	0.3659	Cuba	0.0596
Cuba	White N.H. U.S.-born	0.3029	Cuba	0.2281
Dominican Rep.	White N.H. U.S.-born	0.2760	Dominican Rep.	0.1443
El Salvador	White N.H. U.S.-born	0.3136	Mexico	0.1427
Germany	White N.H. U.S.-born	0.5702	Mexico	0.0336
Guatemala	White N.H. U.S.-born	0.3204	Mexico	0.1404
Haiti	White N.H. U.S.-born	0.3310	Haiti	0.0911
India	White N.H. U.S.-born	0.4433	India	0.1087
Iran	White N.H. U.S.-born	0.4191	Mexico	0.0689
Italy	White N.H. U.S.-born	0.5248	Italy	0.0340
Jamaica	White N.H. U.S.-born	0.3795	Jamaica	0.0587
Japan	White N.H. U.S.-born	0.4040	Japan	0.1365
Mexico	White N.H. U.S.-born	0.3392	Mexico	0.2229
Philippines	White N.H. U.S.-born	0.3945	Philippines	0.0806
Poland	White N.H. U.S.-born	0.4335	Poland	0.1299
Puerto Rico	White N.H. U.S.-born	0.3906	Puerto Rico	0.0546
South Korea	White N.H. U.S.-born	0.3711	South Korea	0.1324
Taiwan	White N.H. U.S.-born	0.3835	China	0.0830
United Kingdom	White N.H. U.S.-born	0.5561	Mexico	0.0308
Former U.S.S.R.	White N.H. U.S.-born	0.3957	Former U.S.S.R.	0.0910
Vietnam	White N.H. U.S.-born	0.3487	Vietnam	0.1429

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 2.8: Cross-ethnic Workplace 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	0.3230	*0.0460*	0.0350	0.1000	0.0550	0.0240	0.0180	0.0770	0.0090	0.0180	0.0240	0.0100
South America	0.3730	0.0170	*0.0610*	0.0820	0.0270	0.0260	0.0110	0.0320	0.0050	0.0190	0.0340	0.0160
Hispanic U.S.	0.4420	0.0080	0.0140	*0.1500*	0.0720	0.0120	0.0110	0.0110	0.0050	0.0050	0.0080	0.0070
Mexico	0.3390	0.0090	0.0090	0.1500	*0.2230*	0.0070	0.0260	0.0040	0.0130	0.0030	0.0020	0.0070
Puerto Rico	0.3910	0.0140	0.0340	0.0910	0.0200	*0.0550*	0.0070	0.0230	0.0040	0.0130	0.0390	0.0120
El Salvador	0.3140	0.0170	0.0190	0.1300	0.1430	0.0100	*0.0670*	0.0090	0.0200	0.0070	0.0110	0.0090
Cuba	0.3030	0.0450	0.0400	0.0830	0.0150	0.0250	0.0060	*0.2280*	0.0050	0.0280	0.0180	0.0100
Guatemala	0.3200	0.0160	0.0210	0.1260	0.1400	0.0120	0.0410	0.0140	*0.0290*	0.0070	0.0100	0.0090
Colombia	0.3660	0.0230	0.0510	0.0810	0.0240	0.0280	0.0120	0.0600	0.0060	*0.0430*	0.0340	0.0140
Dominican Rep.	0.2760	0.0190	0.0550	0.0820	0.0110	0.0480	0.0100	0.0240	0.0040	0.0200	*0.1440*	0.0130
Western Europe	0.4970	0.0070	0.0260	0.0650	0.0270	0.0150	0.0080	0.0120	0.0040	0.0090	0.0130	*0.0510*

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 on the first column to the "column" country of birth.

report the largest immigrant group in the workplace for each immigrant group. All groups either work in firms where the largest immigrant group is their own or it is Mexican immigrants. The only group that does not follow this rule is Colombian immigrants who work in firms where the largest immigrant group is the Cuban-born.

Table 2.8 shows that, for all but four of the Hispanic groups, the largest Hispanic group of coworkers is made up of U.S.-born Hispanics. The four exceptions are immigrants from Mexico, El Salvador and Guatemala, who on average work with more Mexican-born, and Cuban immigrants who are more likely to work with other Cubans than with any other Hispanic group.

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 the ethnic social network 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.

Table 2.9: Correlation Between Work and Residential Own-exposure Rates

Africa	0.215
Caribbean	0.270
Central America	0.349
Central Asia	0.162
Middle East/N. Africa	0.115
Oceania	0.065
Socialist Europe	0.173
South America	0.225
Southeast Asia	0.138
Western Europe	0.290
Canada	0.100
China	0.345
Colombia	0.193
Cuba	0.424
Dominican Rep.	0.248
El Salvador	0.196
Germany	0.047
Guatemala	0.183
Haiti	0.187
India	0.198
Iran	0.181
Italy	0.134
Jamaica	0.162
Japan	0.172
Mexico	0.243
Philippines	0.138
Poland	0.251
Puerto Rico	0.259
South Korea	0.171
Taiwan	0.194
United Kingdom	0.097
Former U.S.S.R.	0.222
Vietnam	0.236

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.

As discussed above, 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 2.9 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. A positive correlation coefficient indicates that immigrants exhibiting higher values of one of the own-exposure rates are also more likely to exhibit higher values of the other own-exposure rate. A country of birth group with a high correlation coefficient is one in which people who live with more co-ethnics are also more likely to work with more co-ethnics. 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 0.42 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 2.10 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 also have workplace own-exposure of the value 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 this value of residential own-exposure. For example, the upper left hand corners in each of the three sections show the following: 1) almost 28% 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

Table 2.10: 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	27.54	4.31	2.98	1.88	0.69	0.40	0.27	0.35	0.24	38.65
0.025 - 0.05	6.54	2.38	1.98	1.38	0.51	0.30	0.20	0.25	0.14	13.68
0.05 - 0.1	4.82	2.36	2.29	1.87	0.77	0.47	0.29	0.35	0.17	13.40
0.1 - 0.2	3.36	2.11	2.40	2.51	1.24	0.81	0.53	0.60	0.24	13.81
0.2 - 0.3	1.35	0.92	1.20	1.56	0.90	0.66	0.47	0.47	0.12	7.64
0.3 - 0.4	0.60	0.46	0.69	1.08	0.67	0.51	0.39	0.43	0.10	4.92
0.4 - 0.5	0.28	0.25	0.39	0.72	0.52	0.43	0.34	0.36	0.09	3.37
0.5 - 0.75	0.31	0.31	0.44	0.85	0.63	0.62	0.46	0.63	0.18	4.44
0.75 - 1	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	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	71.26	11.14	7.71	4.86	1.77	1.03	0.70	0.91	0.62	38.65
0.025 - 0.05	47.78	17.41	14.50	10.07	3.76	2.18	1.43	1.82	1.05	13.68
0.05 - 0.1	35.99	17.64	17.12	13.96	5.74	3.49	2.17	2.60	1.28	13.40
0.1 - 0.2	24.33	15.26	17.38	18.15	9.01	5.87	3.87	4.38	1.74	13.81
0.2 - 0.3	17.65	12.01	15.66	20.47	11.77	8.59	6.13	6.12	1.60	7.64
0.3 - 0.4	12.14	9.26	13.99	21.92	13.63	10.45	7.88	8.68	2.04	4.92
0.4 - 0.5	8.23	7.31	11.62	21.31	15.45	12.83	10.01	10.68	2.58	3.37
0.5 - 0.75	7.00	7.01	9.97	19.23	14.11	14.00	10.39	14.19	4.10	4.44
0.75 - 1	5.14	7.16	8.27	14.32	12.97	15.90	12.59	17.86	5.79	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	61.48	32.87	24.06	15.85	11.54	9.46	9.11	10.24	18.49	38.65
0.025 - 0.05	14.59	18.19	16.02	11.61	8.66	7.08	6.64	7.19	11.11	13.68
0.05 - 0.1	10.77	18.05	18.53	15.77	12.95	11.10	9.86	10.09	13.27	13.40
0.1 - 0.2	7.50	16.10	19.38	21.14	20.94	19.26	18.08	17.51	18.65	13.81
0.2 - 0.3	3.01	7.01	9.66	13.19	15.14	15.58	15.85	13.55	9.49	7.64
0.3 - 0.4	1.33	3.48	5.56	9.09	11.28	12.21	13.12	12.37	7.79	4.92
0.4 - 0.5	0.62	1.88	3.16	6.05	8.77	10.26	11.41	10.42	6.73	3.37
0.5 - 0.75	0.69	2.37	3.57	7.20	10.54	14.76	15.60	18.24	14.13	4.44
0.75 - 1	0.01	0.04	0.05	0.09	0.17	0.29	0.33	0.40	0.35	0.08
Total	44.8	13.1	12.38	11.86	5.94	4.21	2.96	3.45	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.

coworkers, and 3) among all workers for whom co-ethnics represent less than 2.5% of their coworkers, 61% also live in neighborhoods with less than 2.5% co-ethnics.

Relying only on the overall percentage section of the table, one can easily gauge the size of the enclaved immigrant population 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 almost 10% 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 the population of immigrants in the U.S., especially when one considers that the sample selected for this analysis is composed of cities with the largest immigrant populations. Indeed, over half of all immigrants in this sample neither live nor work with more than 10% co-ethnics.

Predicting Enclaves: Selection based on observables

An important consideration in designing empirical models for research on enclaving that relies on less detailed data is the amount of enclaving that is driven by unobservable characteristics. When not properly addressed, these result in biased estimates of outcomes such as earnings and children's educational attainment due to omitted variable bias. In order to get a sense of how well observables 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.

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. Allowing for slight abuse of notation, let C_h^j designate either residential own-exposure or workplace own-exposure rate at some geographical level. Specifically,

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

where $h \in \{k, k^*, k^{**}, w, w^*, w^{**}\} \in \{k, k_{PMSA}, k_{CMSA}, w, w_{PMSA}, w_{CMSA}\}$. That is, the higher geographical levels than the ones previously used are the CMSA (e.g., New York City) and the PMSA (e.g., Newark, a Primary MSA within the New York City CMSA). Finally, 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 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 $t \neq h$.

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.

The matrix X_i contains widely-available individual-level 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 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 2.11 shows that the average residential own-exposure rate in this sample is 0.1147. When measured at the CMSA level, this measure drops to 0.0335. That is, the average immigrant in this sample lives in a CMSA where 3.35% of the adult population is from her same country of birth. The PMSA measure of own-exposure rate is higher at 0.0412, illustrating that immigrants do not randomly distribute themselves among the CMSA but rather gravitate towards parts of the CMSA where other co-ethnics already reside. The mean immigrant is almost 44 years old and immigrated over 15 years ago. Only 13% of the sample has never been married, with over 70% currently married. As is well documented with immigrants, this group is bimodal in educational attainment, where just over one in five immigrants have 8 years or less of education while one in four have a college degree or higher. Sixty-one percent of the sample is white, 25% Asian and 12% black. Hispanics account for 43% of the sample. Nearly half of the sample is already a U.S. citizen and over half report speaking English well or very well. For the most part, the workplace sample differs little from the residential sample. The exceptions are unsurprising: higher

Table 2.11: Demographic Information of the Residential and Workplace Samples

Variable	Residential Mean	Residential *S.D.*	Workplace Mean	Workplace *S.D.*
Residential own-exposure	0.1147	*0.1585*	0.1140	*0.0002*
Workplace own-exposure	0.1198		0.1193	*0.0003*
Res. Own-exp, CMSA	0.0335	*0.0470*	0.0334	*0.0001*
Res. Own-exp, PMSA	0.0412	*0.0636*	0.0410	*0.0001*
Work. Own-exp, CMSA			0.0061	*0.0000*
Work. Own-exp, PMSA			0.0078	*0.0001*
Age	43.7910	*0.0156*	43.9614	*0.0159*
Years since migration	15.6727	*0.0141*	15.8553	*0.0144*

	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		80.08
Self-employed		13.61
Small firm		6.32

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.

education groups and men are overrepresented in the workplace sample, as are immigrants who report speaking English well or very well.

Predicting residential own-exposure rates

OLS regressions predicting residential own-exposure rates are reported on Table 2.12. 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

Table 2.12: 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	-0.0103
	(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)	(0.0034)
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***	-0.0305***
		(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**	-0.0035*
			(0.0084)	(0.0065)	(0.0021)	(0.0020)	(0.0021)	(0.0018)
English			-0.0320***	-0.0309***	-0.0227***	-0.0199***	-0.0181***	-0.0174***
			(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)	(0.0174)
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	0.468

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

adds $C^j_{k^{**}}$, 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 $C^j_{k^*}$, 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. Though the R^2 increases by almost 0.03, the workplace own-exposure is not as good a predictor of residential own-exposure as the overall proportion of the CMSA or PMSA population that belongs to the country of birth group. That is, the local size of the ethnic population is more important in predicting own-exposure rates than the very individual's observed tendency to work with other co-ethnics.

Predicting workplace own-exposure

Predicting workplace own-exposure turns out to be much more difficult than predicting residential own-exposure, as is shown on Table 2.13. 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. The first batch of demographic variables (age, gender, marital status, race, and ethnicity) explain less than half as much of the variation in workplace own-exposure rate as they explained for residential own-exposure rate. Including education and immigrant-specific demographic variables further increases the R^2 to 0.1360, less than was explained of the residential own-exposure using just the first model. Adding CMSA and place of birth variables nearly doubles the proportion of the variation that is explained by the observables.

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, that is, 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 of the PMSA and CMSA (not included in Table 2.13) workforce aggregate own-exposure measure have less explanatory power than the residential own-exposure measures implying that the size of the immigrant community is more important than the size of the immigrant workforce in determining how ethnically clustered individuals are at work. Model (VI) offers the most explanatory power of the set of models used to predict workplace own-exposure by including the individual's residential own-exposure at the tract level. The gain in the R^2 between model (VI) and (VIII) is minimal, however. Again, 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.

Table 2.13: OLS Regression Results: Explaining Workplace Own-Exposure Rates

Model	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
Hispanic	0.0871***	0.0514	0.0553**	0.0480**	0.0007	0.0035	0.0036	0.0029	0.0005
	(0.0315)	(0.0285)	(0.0277)	(0.0211)	(0.0082)	(0.0083)	(0.0074)	(0.0075)	(0.0081)
Some High School		-0.0406**	-0.0330**	-0.0346***	-0.0227***	-0.0180***	-0.0221***	-0.0228***	-0.0223***
		(0.0168)	(0.0147)	(0.0086)	(0.0042)	(0.0035)	(0.0043)	(0.0041)	(0.0043)
High School Diploma		-0.0704***	-0.0587***	-0.0598***	-0.0372***	-0.0290***	-0.0346***	-0.0350***	-0.0363***
		(0.0199)	(0.0177)	(0.0110)	(0.0055)	(0.0046)	(0.0055)	(0.0054)	(0.0056)
Some College		-0.0989***	-0.0802***	-0.0838***	-0.0551***	-0.0443***	-0.0511***	-0.0518***	-0.0540***
		(0.0230)	(0.0206)	(0.0146)	(0.0076)	(0.0066)	(0.0076)	(0.0075)	(0.0077)
College Degree		-0.1100***	-0.0913***	-0.0937***	-0.0632***	-0.0503***	-0.0591***	-0.0599***	-0.0621***
		(0.0270)	(0.0234)	(0.0179)	(0.0102)	(0.0090)	(0.0102)	(0.0101)	(0.0103)
Graduate/Professional Degree		-0.1240***	-0.1060***	-0.1060***	-0.0830***	-0.0680***	-0.0794***	-0.0796***	-0.0817***
		(0.0282)	(0.0245)	(0.0185)	(0.0143)	(0.0127)	(0.0144)	(0.0140)	(0.0144)
Citizen			-0.0215***	-0.0209***	-0.0180***	-0.0162***	-0.0170***	-0.0172***	-0.0178***
			(0.0066)	(0.0051)	(0.0022)	(0.0019)	(0.0021)	(0.0022)	(0.0022)
English			-0.0344***	-0.0339***	-0.0327***	-0.0265***	-0.0308***	-0.0300***	-0.0326***
			(0.0088)	(0.0081)	(0.0062)	(0.0057)	(0.0063)	(0.0064)	(0.0062)
Co-ethnic Exposure Rate						0.2790***	1.5210***	1.0620***	0.3760***
						(0.0354)	(0.1120)	(0.1100)	(0.1030)
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	0.0740	0.1170	0.1360	0.1460	0.2610	0.2930	0.2850	0.2890	0.2650

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

Relationship between earnings and own-exposure rates

As a first pass at the relationship between enclaving and the economic success of immigrants, Table 2.14 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. In line with previous research, immigrants who reside in neighborhoods with higher concentrations of co-ethnics report lower earnings (Borjas 2000). The coefficient implies that residing in an all co-ethnic neighborhood implies earnings are 29% lower than those one would receive 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 much 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 has a statistically weak, though still significant at the 10% level, relationship with earnings. Models (V) through (VII) show that, in the absence of neighborhood-level and employer-level data, immigrant own-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. Furthermore, the labor force own-exposure measure is not statistically significant in predicting earnings.

Table 2.14. 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***		-0.273***	-0.559***		-0.547***
		(0.0370)		(0.0285)	(0.1560)		(0.1500)
Workplace Co-ethnic Exp. Rate			-0.137**	-0.095*		-0.165	-0.067
			(0.0560)	(0.0549)		(0.1350)	(0.1090)
Some High School	0.091***	0.086***	0.088***	0.084***	0.091***	0.091***	0.091***
	(0.0075)	(0.0074)	(0.0073)	(0.0072)	(0.0075)	(0.0075)	(0.0075)
High School Diploma	0.190***	0.181***	0.184***	0.178***	0.189***	0.190***	0.189***
	(0.0114)	(0.0113)	(0.0104)	(0.0106)	(0.0113)	(0.0115)	(0.0114)
Some College	0.344***	0.333***	0.335***	0.326***	0.343***	0.344***	0.342***
	(0.0175)	(0.0172)	(0.0164)	(0.0165)	(0.0175)	(0.0176)	(0.0176)
College Degree	0.645***	0.631***	0.633***	0.623***	0.643***	0.645***	0.643***
	(0.0246)	(0.0239)	(0.0228)	(0.0227)	(0.0245)	(0.0247)	(0.0246)
Grad/Prof Degree	0.930***	0.914***	0.912***	0.901***	0.928***	0.929***	0.928***
	(0.0373)	(0.0361)	(0.0343)	(0.0339)	(0.0372)	(0.0374)	(0.0373)
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

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 the largest 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 high degree of cross-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 enclave effects. 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 that belongs to a country of birth group serves as a strong predictor of residential own-exposure and, to a lesser degree, workplace own-exposure. Similar to previous research, these regressions also reveal substantial negative selection into high co-ethnic neighborhoods along formal education and limited English skills.

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.

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.

CHAPTER 3

HUMAN CAPITAL TRAPS? ENCLAVE EFFECTS USING LINKED EMPLOYER-HOUSEHOLD DATA

An important topic addressed by researchers studying immigrant residential clustering is the impact of ethnic enclaves on the economic assimilation of its members. Does the enclave serve as a “warm embrace” in the American economy or does it, on the other hand, serve to limit immigrant opportunities by reducing incentives and opportunities to assimilate? Immigrants with overly strong reliance on the enclave economy can, in the words of Borjas (2000), become “the victims of a monopsony, a ‘one-company’ town.” Previous studies have yielded mixed results: some have found a negative impact on wage growth (Borjas 2000; Pedace and Rohn 2008), others have found a positive effect (Wilson and Portes 1980; Edin, Fredricksson, Oslund 2003) while others report different effects for high and low education groups (Cutler, Glaeser, Vigdor 2008). These studies report divergent findings primarily due to how each measures “enclave behavior” and how each addresses self-selection into these communities. Though Wilson and Portes (1980) define an enclave using employer ethnicity and ethnic distribution within occupations and industries, most subsequent research has relied on ethnic enclaves defined exclusively using residential information. This study contributes to this line of research by testing enclave effects using a rich linked household employer dataset that allow me to study the effects of both residential and workplace ethnic concentration. Unlike most previous studies in this field, I distinguish between the effects of residential clustering and workplace clustering on the earnings of immigrants, finding evidence that the two types of clustering operate differently on earnings.

Immigrants face extra obstacles when entering the U.S. labor market: the potential devaluation or non-transferability of prior education and work experience are particularly salient concerns for recently arrived immigrants. Because of this, areas with high concentrations of co-ethnics are attractive, particularly as initial location choices upon immigrating. Besides greater availability of ethnic consumer and ethnic goods, these communities also provide access to more trade partners with shared language and culture, two characteristics that substantially decrease transaction costs (Lazear 1999). In fact, Bayer, McMillan, and Rueben (2004) report that language (either speaking a language other than English at home or the ability to speak English) explains almost 40% of Asian segregation and over 30% of Hispanic segregation in the San Francisco Bay Area. For immigrants with limited language skills and limited transferable education, ethnic networks may yield higher initial wages. The problem, however, may be that these same amenities may decrease incentives to assimilate – acquire U.S.-specific human capital – and ultimately drive some immigrants away from assimilation. Furthermore, limiting one's contacts to co-ethnics can also increase the costs of assimilation. Learning English, for example, might be less costly if one must use it daily with coworkers or neighbors whereas an immigrant who lives and works primarily with co-ethnics will have fewer opportunities to practice. A situation such as this may lead to a human capital trap – individuals are able to find work but are not able to accrue the necessary U.S.-specific human capital for their careers to advance.

On the other hand, it may be inefficient for some immigrants to assimilate due to high assimilation costs and low expected returns from the labor market. Costs of assimilation include the costs of learning a new language and of acquiring additional education, training and (re)certification. Learning takes time, implying missed wages, and additional training/schooling can represent a steep financial investment not necessarily feasible for some immigrants. These costs will further vary by an individual's initial human capital stock: individuals with very low levels of education, for example, will require significantly more academic training to earn a high school equivalency than those who already arrive in the U.S. with high school completed in their source country. Additionally, some research has found that more education results in an improved ability to learn new skills (Rosenzweig 1995), suggesting that, for low education immigrants, learning new skills and acquiring U.S.-specific human capital might be more difficult and costly based solely on their lack of initial schooling. For these immigrants, investing into more human capital may not be the optimal choice even in the absence of ethnic communities. For them, the enclave provides important benefits, such as ethnic referral networks and access to ethnic goods, while minimizing everyday transaction costs.

Using linked household employer data, I explore how immigrant clustering in the workplace and in neighborhoods can impact wages and wage growth and how these two different types of clustering behavior can yield different economic results. I document that immigrants who live and work with larger shares of co-ethnics tend to earn less, even after controlling for the individual's human capital and country of origin. Using longitudinal data on earnings, a pattern of consistently lower earnings emerges for immigrants who reside or work with high concentrations of co-ethnics. However, after controlling for residential own-exposure rates, the longitudinal data analysis also indicates some positive returns to working with co-ethnics: immigrants in the 25^{th} through 75^{th} percentile of coworker own-exposure have higher earnings than those with lower concentrations of co-ethnics. Applying instrumental variable analysis to the issue of self-selection on unobservables, I find that sorting on negative unobservable traits may fully explain the lower earnings associated with both higher residential and workplace concentration for immigrants with only a high school education or less. For immigrants with more than a high school education, I find that a third of the decrease in earnings attributed to ethnic residential clustering is explained by sorting. Thus, for immigrants with some college education or higher, living in areas with more co-ethnics depresses earnings. On the other hand, though working with more co-ethnics is associated with lower earnings for immigrants with more than a high school degree even after controlling for selection, I find evidence that negative selection is actually mitigating this earnings penalty. This may indicate that some immigrants who have unobservable qualities that make them relatively unproductive in more integrated workplaces are more productive in workplaces with more co-ethnics.

This study contributes to the literature on enclave effects by considering how ethnic clustering affects highly educated immigrants differently from those without postsecondary education and measuring enclave effects along two dimensions of immigrant clustering: residential and workplace. Both of these dimensions represent important, yet potentially distinct, social networks: one is a source of ethnic goods and social interactions while the other can be a source of economic opportunities. The extent to which these two networks overlap is central to understanding how residential enclaves can lead to economic human capital traps. Immigrants who both work and live with co-ethnics may be too isolated from non co-ethnics and become part of a human capital trap, failing to acquire the necessary country-specific human capital to advance in the labor force. On the other hand, residing in an enclave might be the optimal strategy for some who lack U.S.-specific human capital and for whom the investment into more training is excessively costly. The results below show that both residential and workplace clustering are associated with lower earnings, though the impact on earnings differs between these two types of clustering. However, these negative effects on earnings are partially explained by negative selection into high co-ethnic firms or neighborhoods. Overall, the evidence suggests that co-ethnic clustering has no discernible effect on immigrants with less education but may be leading to human capital traps for immigrants who have more than a high school education.

Literature Review

Research on immigrant settlement patterns and enclaves has consistently documented a significant tendency of immigrants to choose locations within a host country with disproportionately large co-ethnic populations (for example, Bartel 1989 and Borjas 2000). This clustering behavior has led to large immigrant populations in several destination cities throughout the U.S., including the five metropolitan areas in this study. Within these large metropolitan cities, some immigrant communities have evolved into recognizable "ethnic enclaves" – neighborhoods with high concentrations of co-ethnic residents and businesses. Some well know examples include Little Havana in Miami, Chinatowns in Los Angeles and New York, and the Russian-born community in Brighton Beach, Brooklyn. Many other immigrants spill out into the larger metropolitan area and into non-ethnic workplaces in search of better employment opportunities, better schooling and neighborhoods for their children, or more affordable housing. Recent immigration settlement patterns show higher immigrant settlement in the suburbs, even for immigrants with limited language skills who would, traditionally, settle into urban enclaves (Alba et al 1999). The sample of cities chosen for this study purposely includes five of the largest immigrant destination cities and their suburbs. This allows for comparisons between immigrants who choose high co-ethnic neighborhoods or employers within large immigrant populations and those who choose to reside close to co-ethnics but have branched out into more integrated neighborhoods or workplaces.

Both economic and social reasons have been cited for ethnic clustering in host countries.

Economies of scale in the production of ethnic goods, including marriage markets, food, and religious institutions, can lead to the formation of ethnic communities. Also, American immigration policy encourages family migration leading to ethnic residential clustering as individuals choose to settle near their relatives. Residential clustering of co-ethnics can result in an increase in potential trade partners, due to common language and cultural similarities, as discussed by Lazear (1999). Increasing the number of potential trade partners, ceteris paribus, increases economic opportunities for immigrants with limited ability to communicate or trade outside of the ethnic group. This can lead to employment opportunities: McManus (1990), for example, finds a lower earnings penalty associated with not speaking English for workers within enclaves, while Borjas (1986) finds a positive proclivity for self-employment among immigrants who live in cities with more co-ethnics. Ethnic communities may create employment and business opportunities for individuals by generating demand for ethnic labor, products and services.

Unmeasured individual heterogeneity plays a significant role in labor market sorting and worker earnings, even outweighing the effects of unmeasured firm heterogeneity (Abowd, Kramarz and Margolis 1999). The role of these unobserved characteristics is doubly important when we consider the role of social networks in job acquisition. Calvó-Armengol and Jackson (2004) show that the quality of one's social network, measured in terms of labor force attachment, can heavily influence one's own labor market outcomes and can directly affect the growth of inequality between different social groups. They illustrate the existence of positive externalities within referral networks whereby the employment of members of the network leads to higher employment levels throughout the network. Belonging to lower quality networks, thus, limits one's employment prospects. When applied to the context of ethnic enclaves, negative self-selection into ethnic social networks can lead to lower earnings for the members of these networks by limiting the job vacancies available to the network.

Bayer, Ross and Topa (2008) find compelling evidence of referral networks operating between neighbors – specifically, they find that individuals are 33% more likely to work with neighbors who live on the same block as they are to work with neighbors who live in the surrounding blocks. These referral networks result in higher earnings: a one standard deviation increase in potential referrals increases the earnings of men by between 2.0 and 3.7 percentage points. To the extent that residential location informs social networks, negative selection into ethnic neighborhoods can, thus, lead to negative selection into ethnic job referral networks. While acquiring a job via a social network can yield higher than expected earnings, limiting one's social network to immigrants who negatively self-select might result in lower earnings. Andersson et al (2010) look at the proportion of coworkers who are immigrants for both natives and immigrants and find that limited English ability, industry of employment and immigrant composition in the neighborhood account for 40% of total workplace immigrant composition. They find that both residential clustering by country of origin and ethnic clustering in industries contribute heavily to co-ethnic own-exposure in the workplace – though these effects differ substantially between different countries of birth. Though they find evidence of sorting between workplaces by skill (workers with advanced degrees have larger shares of immigrant coworkers), they also document a significant correlation between residential co-ethnic exposure and co-ethnic workplace exposure, indicating the prevalence of neighborhood networks in employment outcomes of immigrants.

Besides the lower transaction costs associated with working with co-ethnics who share a culture and language, another reason for co-ethnic clustering in the workplace might be discrimination in hiring. Using an audit study in Canada, Oreopoulos (2011) finds that individuals with English names were 39% more likely to receive callbacks on their resumes than individuals with foreign-sounding names[35] who also attended college in Canada and had previous work experience in Canada. He finds that work experience outside of Canada substantially lowered call back rates, though employers did not penalize foreign schooling in conjunction with at least 4 years of Canadian work experience. Call back rates for foreign-sounding names with foreign education and foreign work experience (comparable in quality with Canadian counterparts) were 40% those of English names – indicating significant devaluation of education and work experience that occurs in countries deemed to be too different.[36] This labor market discrimination may push immigrants with substantial education and work experience acquired overseas to work with co-ethnics or in ethnic-owned businesses, where their skills might be more appropriately evaluated even if these firms pay lower wages.

Unmeasured individual characteristics lead both to non-random sorting into neighborhoods and non-random sorting into workplaces. Researchers have attempted several approaches to mitigate the effects on the earnings estimates of self-selection into immigrant location choice. One approach has been to look at children or refugees, individuals who typically have their location in the host country chosen for them. Borjas (2002) finds that limiting the analysis to immigrants from source countries with high refugee rates did not significantly impact the effect of ethnic enclaves on immigrant home ownership. Furthermore, Borjas (2000) finds that refugees are even less distributed than other immigrant groups: nearly 60% of the 1980 refugee population in the U.S. was clustered in 5 metropolitan areas compared to 49% of non-refugees. Similarly, using

[35] Greek, Indian, Pakistani and Chinese names were used.
[36] The author notes no statistically significant penalty for employment that occurred in the U.K. compared to Canadian employment.

longitudinal data on detailed location, Edin, Fredriksson and Aslund (2003) find that 46% of refugees in Sweden had left their initial assigned municipality within 8 years and moved to an area with more immigrants.[37] Due to this high internal migration of refugees, it is not clear that limiting the analysis to countries with relatively large refugee populations in the U.S. is a successful tool to address self-selection without having access to data on where they were initially placed in the host country. Similarly, Cutler, Glaeser and Vigdor (2007) limit the scope of their analysis to neighborhood effects on young adults and teenagers, arguing that their location is more likely to be exogenous since it was chosen by their parents. Limiting the analysis to children and young adults severely limits the ability to study earnings and other labor market outcomes. Also, young immigrant adults are a highly self-selected group since college enrollment rates vary substantially between different immigrant groups.

The most prevalent approach used to address selection into areas of high co-ethnic concentration is to employ an instrumental variable analysis (for example, Altonji and Card 1991; Bertrand, Luttmer and Mullainathan 2000; Cutler, Glaeser and Vigdor 2008). While some researchers, such as Cutler, Glaeser and Vigdor (2008), instrument for neighborhood-level segregation indices using a demographic characteristic aggregated to a larger geography, others have used instruments relying solely on geographical aggregation, both contemporaneous and lagged. Altonji and Card (1991) use the proportion of immigrants from the previous decennial living in the city as an instrument for the immigrant population 10 years later in the same city, arguing that immigrant location is significantly influenced by the settlement patterns of previous waves independently of current labor market conditions in the area. Their instrument, thus, captures the migration that occurs into the given metropolitan area based on immigrant migration networks

[37]Along the same lines, earlier research found significant return-migration to Miami by Cuban refugees who had

alone. Bertrand, Luttmer and Mullainathan (2000) use the contemporaneous proportion of co-ethnics[38] at the city-level to instrument for the neighborhood level residential concentration in order to distinguish between sorting and network effects. They argue that, since costs are lower for within-city moves than between-city moves, the effects of sorting will be larger at the neighborhood level than at the city level. Instead, they find evidence of stronger sorting into cities than into neighborhoods.

Data[39]

This study draws its sample of analysis from immigrants residing in five of the largest immigrant destination urban areas in the U.S.:[40] Los Angeles, New York, Chicago, Miami, Houston and their suburbs. The sample is drawn from the confidential 2000 U.S. Census of Population and Housing, a one-in-six household sample containing detailed residential and demographic data, including English language proficiency and census block of residence. Adults, ages 18 – 70, who report being in the labor force are matched to state Unemployment Insurance (UI) data provided through the Longitudinal Employer Household Dynamics (LEHD) program.[41] The LEHD data contain basic demographic characteristics and earnings histories for all employees in UI-covered jobs, as well as basic employer characteristics such as industry and locations. These demographic characteristics in the LEHD data include place of birth and ethnicity, enabling the construction of employer-level ethnic and immigrant composition measures. One limitation of the LEHD data is

---

been placed in other cities in the U.S. (Wilson and Portes 1980).

[38] Bertrand, Luttmer and Mullainathan (2000) look at language groups (for those who speak a language besides English in the home) rather than country of origin groups.

[39] The sample used throughout this paper contains approximately 500,000 observations. Exact sample sizes are not being released for this draft due to confidentiality concerns.

[40] Consolidated Metropolitan Statistical Areas (CMSA) were used to identify these five urban areas. These are large urban areas that include several cities and their suburbs.

[41] For more information about the LEHD infrastructure files, please see Abowd et al (2006). Currently, UI data are available for all states except Massachusetts.

that they do not provide earnings information for federal employees and jobs that are not covered by UI, such as those who are informally employed or paid "off the books" and the self-employed. Note, however, that though the earnings data are from administrative records, matches between the 2000 census and the earnings data are based on name and address matches, not solely on social security number, allowing for a higher coverage of undocumented immigrants than would otherwise be the case.

Table 3.1: All Immigrants with LEHD Earnings Records in 2000 Who Immigrated as Adults and Reside in the Five Metropolitan Areas

	%
Male	54.49
Education	
8 years or less	17.48
Some High School	14.21
High School Diploma	18.64
Some College	19.91
College	17.24
Graduate/Professional Degree	12.51
Speaks English	58.08
English-speaking POB	27.45
Hispanic	42.57
Citizen	47.27
Employer type	
Large firm	86.09
Self-employed	7.12
Small firm	6.79

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 sample of analysis is all adult immigrants in the labor force and residing in the five metropolitan areas listed above who arrived in the U.S. as adults and who had valid UI records in either 1999, 2000 or 2001.[42] Some basic demographic statistics for this sample are reported in Table 3.1. Almost 55% of the sample is male. The sample is evenly distributed between educational groups: half have a high school diploma or less and half have at least some post-secondary schooling. Like other studies have reported, immigrants are more likely than natives to have either very low levels of education or very high levels of education – over 17% have less than 9 years of schooling while over 12% have a professional or graduate degree. Almost 60% of the sample reports speaking English well or very well while over a quarter of the sample is composed of immigrants from countries in which English is an official language. Hispanic immigrants account for 43% of the sample. Nearly half of the sample is composed of naturalized

Table 3.2: Mean Co-ethnic Exposure Rates and Earnings for Immigrants in the Workforce and who Reside in the Five Metropolitan Areas, Full Sample and by Education

	Full Sample		More than High School Diploma		High School Diploma or Less	
	Mean	*S.E.*	Mean	*S.E.*	Mean	*S.E.*
Residential Exposure Rate	0.1178	*0.0002*	0.0752	*0.0003*	0.1602	*0.0004*
Workplace Exposure Rate	0.1270	*0.0003*	0.0868	*0.0004*	0.1669	*0.0004*
Residential Exposure Rate (1990), PMSA-level	0.0351	*0.0001*	0.0199	*0.0001*	0.0503	*0.0002*
Residential Exposure Rate (2000), PMSA-level	0.0421	*0.0001*	0.0279	*0.0001*	0.0562	*0.0002*
Log Earnings (2000)	9.9077	*0.0016*	10.2038	*0.0023*	9.6134	*0.0019*
Age	43.4159	*0.0165*	42.7496	*0.0231*	44.0783	*0.0235*
Years since migration	15.341	*0.0147*	14.5351	*0.021*	16.1422	*0.0205*

Source: With the exception of the residential exposure rate in 1990, all values are the result of the 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. Residential exposure rate is calculated at the Census tract level. Workplace exposure rate is calculated at the state-employer level. Residential exposure rate in 1990 is calculated at the Primary Metropolitan Statistical Area, using the 1990 U.S. Census of Population and Housing, 5% Sample.

[42] Individuals who reported being self-employed but matched to the UI data were kept in the sample, though their self-employment status was controlled for in the regressions. All earnings are adjusted to year 2000 USD.

U.S. citizens. The bottom of Table 3.2 also reports the average log of annual earnings, age and years since migration for the data universe as well as by education group. Immigrants with a secondary education or less earn less and are slightly older and have been in the U.S. for slightly longer than immigrants with more schooling. Table 3.3 reports the distribution of the country of birth of the data sample used in this analysis – showing that Mexican immigrants make up the largest single group in the sample, though they represent a smaller share of the immigrant population in these five urban areas than nationally.

Table 3.3: Country or Region of Birth Distribution for All Immigrants Residing in the Five Metropolitan Areas in 2000 who had LEHD Earnings Records in 2000 and who Immigrated as Adults

	%		%
Canada	1.09	Puerto Rico	2.61
China	3.32	South Korea	2.08
Colombia	2.61	Taiwan	1.96
Cuba	4.59	United Kingdom	1.63
Dominican Republic	3.91	USSR Core	3.14
El Salvador	3.59	Vietnam	2.49
Germany	0.89	Regions of Birth:	
Guatemala	1.86	Africa	2.45
Haiti	2.54	Caribbean	2.27
India	4.63	Central America	3.05
Iran	1.15	Central Asia	1.57
Italy	1.03	Middle East/North Africa	2.21
Jamaica	3.20	Oceania	0.30
Japan	1.13	Socialist Europe	2.12
Mexico	15.93	South America	6.83
Philippines	6.01	South East Asia	2.02
Poland	2.25	Western Europe	3.55

Source: Author's calculations based on 2000 U.S. Census of Population and Housing. Immigrants from smaller country of origin groups are aggregated to region of births group. These region of birth groups exclude the country of birth groups listed above.

Two measures of co-ethnic exposure rates are used to identify ethnic enclaves: residential own-exposure and workplace own-exposure.[43] Individual i's residential own-exposure rate is the proportion of adults in his census tract of residence, k, made up of co-ethnics, i.e. others who were born in the same country of origin, j. C_k^j, the residential own-exposure rate for group j living in census tract k, is calculated as follows

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

where n_{jk} is the total number of adults in k who were born in j, and N_k is the total population in k.[44]

Similarly, C_w^j is the workplace own-exposure rate, calculated as above where w is the individual's workplace. Workplace is defined differently for three groups: 1) for individuals who work in firms with at least six employees, w is the dominant employer in year 2000, 2) for individuals whose dominant employer has less than six employees, w refers to a pseudo-firm made up all employers in the same collapsed industry group and located in the same census block, and 3) for self-employed individuals, the workplace own-exposure is calculated over all other self-employed individuals in the same census block workplace (as reported in the census) and the same collapsed industry. Andersson *et al* (2010) show that the mechanics of calculating coworker shares at the firm-level leads to lower variance in coworker shares for small firms – in order to mitigate this issue, I measure the ethnic composition in pseudo-firms defined by industry and census block. A similar identification strategy, based solely on census block location of

[43] See Sousa (2011a) for a detailed description of how these two measures were calculated and how they compare between different immigrant groups.
[44] This exposure rate was also used by Bayer, McMillan and Rueben (2004) and Andersson *et al* (2010).

workplace, was employed successfully with less detailed data to show the existence of referral networks by Bayer, Ross and Topa (2008). The self-employed are included in this analysis for two reasons: 1) selection into self-employment can vary dramatically by country of birth groups (Sousa 2011b), hence their exclusion from analyses of labor outcome among immigrants can result in serious distortions, and 2) a large fraction of the self-employed report significant proportions of their income earned through employment rather than their own business. Table 3.1 reports the distribution of these three employer types: 86.1% of the sample works for firms with six or more employees, 7.1% are self-employed and, 6.8% of the sample works in firms with 5 or less employees. Table 3.4 reports the average workplace co-ethnic exposure rate by employer type for the data universe as well as for each educational group. Workers in large firms (those with more than five employees) have a higher average level of co-ethnic exposure at the workplace than the other two employer types, probably due to the methodology used to calculate these exposure rates. To control for the different methodologies, control variables for firm size/employer type are included in all regressions.

Table 3.4: Average Workplace Co-ethnic Exposure Rates, by Employer Type, Full Sample and by Education

	Full Sample		More than High School Diploma		High School Diploma or Less	
	Mean	*S.E.*	Mean	*S.E.*	Mean	*S.E.*
Large firm	0.1355	*0.0003*	0.0911	*0.0003*	0.1783	*0.0004*
Self-employed	0.0549	*0.0001*	0.0379	*0.0001*	0.0741	*0.0002*
Small firm	0.0785	*0.0002*	0.0669	*0.0002*	0.0889	*0.0003*
Overall	0.1270	*0.0003*	0.0854	*0.0003*	0.1651	*0.0004*

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. A large firm is defined as an employer with more than 5 employees in 2000. A small firm is defined as an employer with at most 5 employees in 2000. Workplace exposure rate is calculated at the state-employer level for large firms, at the Census block of workplace and industry level for small firms, and at the PMSA-industry cell for the self-employed.

The first two rows of Table 3.2 report the average residential co-ethnic exposure rate, C_k^j, and the average workplace co-ethnic exposure rate, C_w^j, for the full sample and for high and low education groups. Workplace co-ethnic exposure rates are slightly higher than residential co-ethnic exposure rates for each of the three samples. On average, immigrants without post-secondary education both lived and worked with double the proportion of co-ethnics (16.0% and 16.7% respectively) as immigrants with more than a high school diploma (7.5% and 8.7% respectively). Overall, the average immigrant in this sample of five cities with large immigrant populations lived in neighborhoods with about 11.8% co-ethnics and worked in workplaces also with about 12.7% co-ethnics. Additionally, Table 3.2 reports the co-ethnic exposure rates at the PMSA[45] level in the years 1990 and 2000. These are calculated as the proportion of the PMSA in-sample population (including the native-born) that belongs to each individual's ethnic group. As expected, these values are significantly lower than the neighborhood and workplace co-ethnic exposure rates with an average of 3.5% in 1990 and 4.2% in 2000. As above in the neighborhood and workplace, immigrants without post-secondary education have PMSA-level co-ethnic exposure rates about twice the size as immigrants with post-secondary education.

Earnings Growth Analysis

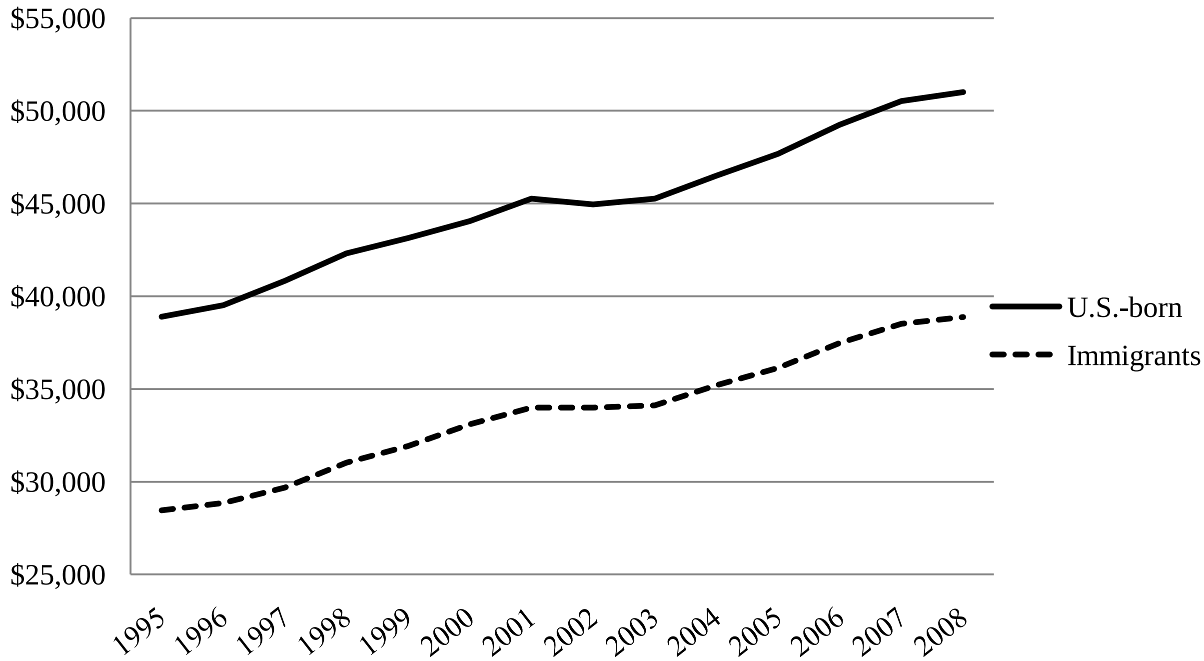
Using the LEHD annual earnings data, this section examines earnings trends for natives and immigrants by own-exposure rates. This longitudinal analysis uses the LEHD and their UI earnings from 1995 through 2008 for the sample described above.[46] Figure 3.1 shows a

[45] A PMSA (Primary Metropolitan Statistical Area) is what is commonly thought of as a city. As described above, the data universe for this study is based on 5 CMSAs. These 5 CMSASs are composed of 26 PMSAs, 15 of which comprise the New York City CMSA.

[46] Earnings are the total earnings reported by employers to state UI programs. To calculate the means in the following figures, annual observations with less than $1,000 or more than $1 million, and those for individuals who were less than 18 years old during the year of reported earnings were eliminated. These restrictions trim annual

consistent earnings gap of about $10,000 between the earnings of the U.S.-born and immigrants for the 13 years plotted. Figure 3.2 shows that this earnings gap is caused primarily by the lack of earnings' growth in immigrant earnings between the ages of 30 and 60.

Figure 3.1. Earnings for U.S. Natives and Immigrants Ages 18 - 70 Residing in Five U.S. Urban Areas in 2000

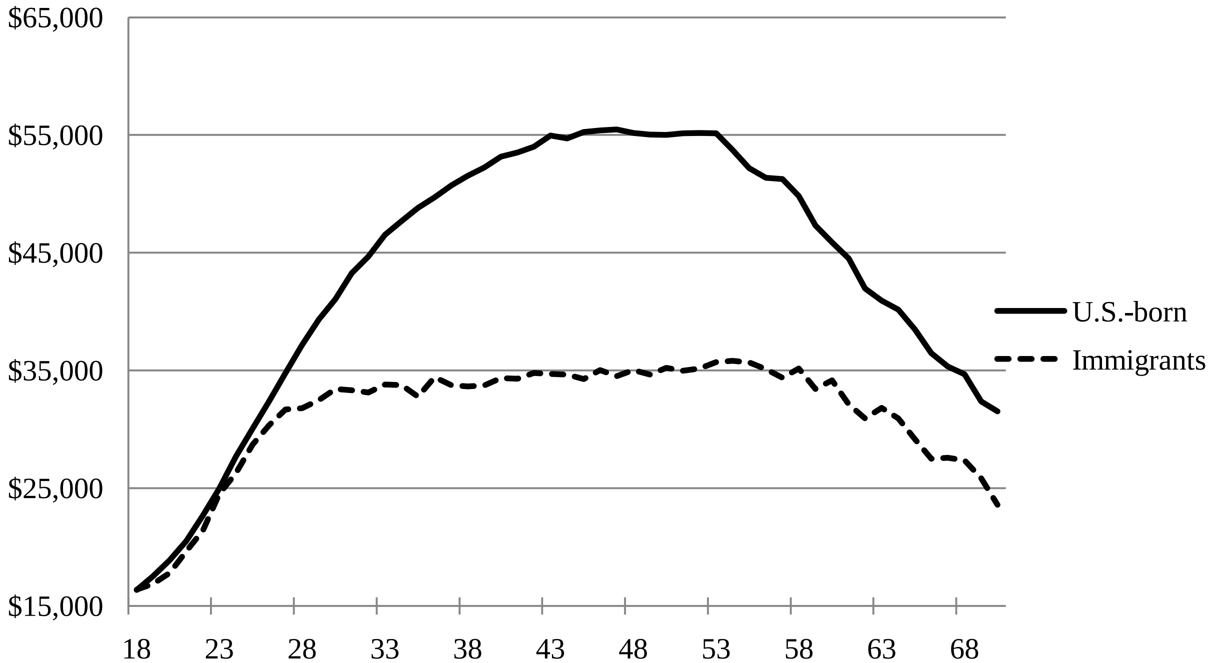


Source: Author's calculations from the LEHD Employment History File. All values have been adjusted for inflation to reflect year 2000 dollars. Only those immigrants who were at least 18 when they first arrived in the

The lack of earnings growth for prime earning years is partially due to compositional factors: as immigrants arrive in the U.S. at different ages and enter the labor market with less U.S.-specific human capital, they bring down the average earnings for immigrants at that age group. Figure 3

observations by less than 8%. Additionally, annual observations based on UI records from more than 5 states or from more than 10 employers were eliminated – together these two data quality restrictions accounted for less than 0.01% of observations.

Figure 3.2: Earnings Between 1995-2008 for U.S. Natives and Immigrants, by Age



Source: Author's calculations from the LEHD Employment History File. All values have been adjusted for inflation to reflect year 2000 dollars. Only those immigrants who were at least 18 when they first arrived in the U.S. are reported.

addresses this issue by plotting earnings by age separately for six different arrival cohorts: 1) 1968 and earlier, 2) between 1969 and 1975, 3) between 1976 and 1982, 4) between 1983 and 1987, 5) between 1987 and 1994, 6) 1995 and later. These cohorts are designed to correspond to two important immigration policy changes in the U.S.: 1) the immigration act of 1965 (which went into effect in 1968) eased restrictions on the legal immigration of non-European immigrants, and 2) the Immigration Reform and Control Act of 1986 (IRCA), which, while granting amnesty for undocumented immigrants who had arrived prior to 1992, also instituted penalties on employers hiring undocumented labor. The country of origin groups in these cohorts vary substantially: the first is made up primarily of Western European immigrants while the most recent cohorts are composed of large majorities from Latin American. Figure 3.3 shows the earnings trajectory by age for each of these cohorts. Even for cohorts that had been in the U.S. for over 20 years as of 1995, earnings are still notably lower than for the native population.

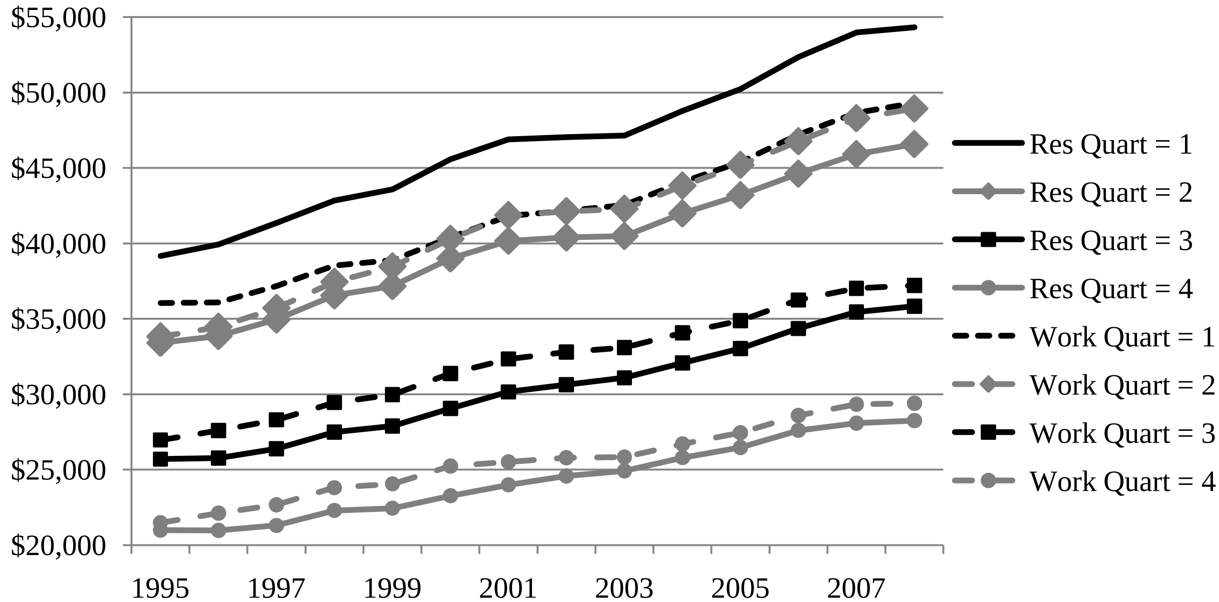
Figure 33: Earnings Between 1995-2008 for U.S. Natives and Immigrants, by Arrival Cohort and by Age

Source: Author's calculations from the LEHD Employment History File. All values have been adjusted for inflation to reflect year 2000 dollars. This figure has been smoothed by using rolling 2 year average income. Only those immigrants who were at least 18 when they first arrived in the U.S. are reported.

The native-immigrant wage gap can be attributed to various factors – but the one that is explored in this study is the role of enclaving, residentially and at the workplace. In order to look at earnings growth by co-ethnic exposure rates, both residential and workplace own-exposures were condensed to quartiles and, for each group, annual earnings were plotted in Figure 3.4.[47] Immigrants with the lowest residential co-ethnic own exposure rates, the lowest quartile of

[47] Quartiles are calculated based on 2000 residence and workplace so each year is composed of the same individuals (allowing for absences from the labor market and excluding any annual observations that occurred before 18 years of age). Standard errors are not included in the figure so as not to clutter it. They range from a maximum of 400 for the first quartile to a minimum of 113 for the fourth quartile.

Figure 3.4: Mean Annual Earnings by Quartile of Residential and Workplace Own-



Source: Author's calculations from the LEHD Employment History File. All values have been adjusted for inflation to reflect year 2000 dollars. Only those immigrants who were at least 18 when they first arrived in the U.S. are reported.

residential own-exposure,[48] report the highest earnings – by 2008 their average earnings were over $54,000, slightly higher than average earnings for the overall U.S.-born population shown in Figure 3.1. On the other hand, the highest quartile of residential own-exposure, those who live in neighborhoods with the largest shares of co-ethnics, had exceptionally low earnings, just barely surpassing $28,000 in 2007. Also notable is the lower earnings growth of immigrants living in high residential own-exposure communities. Whereas earnings grew by 39% over the 13 years of analysis for the three lower quartiles of residential own-exposure, earnings only grew by 34% for the highest quartile. Note that all four quartiles of residential own-exposure report higher earnings growth than the 31% seen among the U.S.-born sample, evidence of gradual economic assimilation.

[48] Quartile cut off values are not reported since they have not yet been reviewed for disclosure avoidance by the U.S. Census Bureau.

This relationship between residential own-exposure quartile and earnings survives the inclusion of demographic factors related to earnings. Table 3.5 shows that even with the inclusion of controls for personal characteristics, place of birth and city of residence, immigrants in the first quartile of residential own-exposure earn significantly more than all other quartiles. Since the dependent variable is log of earnings, the OLS coefficients indicate that immigrants in quartile 4 earn 10.8% less than similar immigrants in the first quartile. Immigrants in quartiles 2 and 3 also earn slightly less than those in quartile 1, with effects on the order of 2.8% and 8.6% respectively.

Table 3.5: The Relationship Between Immigrant Earnings in 2000 and Co-ethnic Exposure Rates in the Neighborhood and in the Workplace

	Residential Quartiles		Workplace Quartiles	
Quartile = 2	-0.0284	***	0.0720	***
	(0.0107)		(0.0148)	
Quartile = 3	-0.0862	***	0.0407	
	(0.0155)		(0.0220)	
Quartile = 4	-0.108	***	-0.0411	
	(0.0202)		(0.0333)	
Constant	7.5540	***	7.5240	***
	(0.1060)		(0.1040)	
R-squared	0.268		0.269	

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 values reported in this table are the OLS coefficients on the residential and workplace quartiles from separate wage regressions on log of wages in 2000. Controls were included for CMSA of residence, place of birth, age, age-squared, gender, Hispanic ethnicity, years since migration and its square, citizenship, employer size and type, English skills, English is an official language in the country of birth, education, estimated minimum education in the U.S., and the proportion of the co-ethnic 1990 population in the U.S. who was residing in the individual's PMSA of residence.
Clustered robust standard errors are reported in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Though the relationship between earnings and quartile of residence is consistently negative –

individuals in higher quartiles of own-exposure have lower earnings – the relationship between earnings and workplace own-exposure is not consistent. The first two quartiles of workplace own-exposure result in overlapping earnings trends. The other two quartiles, however, mirror the high quartiles of residential own-exposure: they exhibit lower earnings than the low quartiles. Three of the workplace quartiles have earnings growth rates of 37-38%; the second quartile, however, shows a higher earnings growth rate of 45% over the 13 years plotted. Table 3.5 shows that, when log of earnings are regressed against workplace own-exposure with the inclusion of human capital controls, immigrants in the second quartile of workplace own-exposure earn 7.2% more than immigrants in the first quartile. Immigrants in the third quartile earn 4% more than those in the first quartile while immigrants in the fourth quartile earn 4% less (though these two coefficients are not statistically significant). As suggested by the earnings trends in Figure 3.4, ethnic segregation operates differently in the neighborhood and in the workplace. While higher concentrations of co-ethnic neighbors implies lower earnings, having some co-ethnic coworkers might result in higher earnings than either working with almost no co-ethnics or with many co-ethnics.

To further explore the role of ethnic own-exposure in the labor market and residential areas, I combine the two own-exposure rates to create a two-dimensional measure of enclave proclivity. Table 3.6 details the distribution of the interaction of the two measures. The two largest groups are the groups at the extremes: immigrants who do not live or work with large proportions of co-ethnics and immigrants who both live and work in high co-ethnic areas.[49] Conversely, the least likely combinations are people who live with very few co-ethnics but work with a large proportion of co-ethnics and the individuals who live in high co-ethnic neighborhoods but work

Table 3.6: Distribution of Quartile of Residential Own-exposure Interacted with Quartile of Workplace Own-exposure

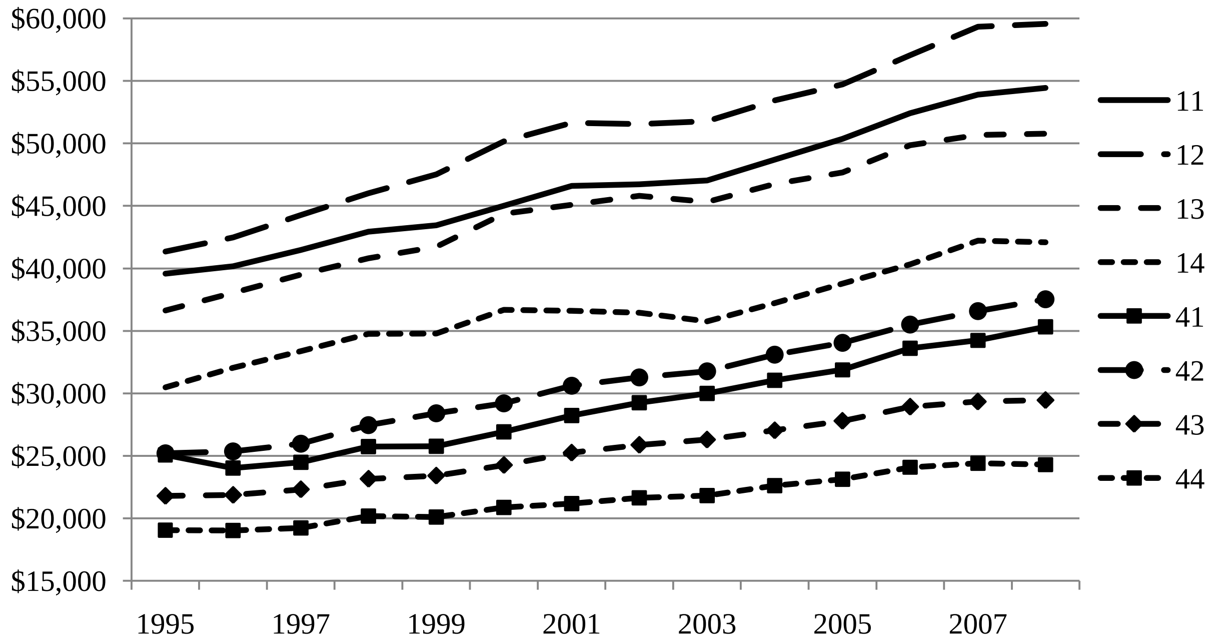
Quartile of Own-exposure		
Residential	Workplace	%
1	1	13.01
1	2	6.73
1	3	3.45
1	4	1.82
2	1	7.08
2	2	8.36
2	3	6.19
2	4	3.38
3	1	3.57
3	2	6.61
3	3	8.38
3	4	6.44
4	1	1.34
4	2	3.31
4	3	6.99
4	4	13.36

Source: Author's calculations based on 2000 U.S. Census of Population and Housing and LEHD Employment History File and Employer Characteristics File. For both the residential and workplace own-exposure rates, each individual is assigned to a quartile where quartile 1 includes the 25% of individuals with the lowest values of co-ethnic exposure rates and each subsequent quartile assigned to individuals with higher values of co-ethnic exposure rates.

with few co-ethnics. Based on the trends reported in Figure 3.4, Figure 3.5 focuses on the first and fourth quartiles of residential own-exposure and dissects each by the quartile of workplace own-exposure. It shows a considerable earnings gap between immigrants who do not live or work with high proportions of co-ethnics and immigrants who live in high co-ethnic areas. Regressing log of earnings on the interacted own-exposure quartiles confirms that immigrants in the first quartile of residential own-exposure who are in either the second or third quartile earn more than similar immigrants who are in the lowest residential and workplace quartiles. Another

[49] For a more detailed analysis of the interaction between residential and workplace own-exposure rates among this sample, see Sousa 2011a.

Figure 3.5: Annual Earnings Between 1995-2008, for Immigrants in High and Low Co-ethnic



Source: Author's calculations from the LEHD Employment History File. All values have been adjusted for inflation to reflect year 2000 dollars. Only those immigrants who were at least 18 when they first arrived in the U.S. are reported. Trend lines show annual earnings for 8 types of immigrants by quartile of residential own-exposure (only the 1st and 4th) and workplace own-exposure quartile. The first number designates the

interesting finding on Table 3.7 is that this pattern holds for all residential own-exposure quartiles: within each quartile, immigrants who were in either the first or last workplace own-concentration quartiles earned less than immigrants in the middle quartiles. Working in workplaces with own-exposure rates between the 25th and 75th percentiles is associated with higher earnings, all things equal, for each quartile of neighborhood co-ethnic exposure rate.

This first pass at the data confirms that earnings are lower among immigrants who live or work in high co-ethnic areas or firms. However, it also indicates that the relationship between workplace co-ethnic exposure and earnings is not monotonic – instead, it appears that working in firms with some co-ethnics may lead to higher earnings than working in firms with exceptionally low levels of co-ethnics or those with exceptionally high levels of co-ethnics.

Table 3.7: The Relationship Between Co-ethnic Exposure Rate Quartiles (in the Neighborhood and the Workplace) and Immigrant Earnings in 2000

Residential Quartile = 1			Residential Quartile = 3		
Work Quartile = 1	*omitted*		Work Quartile = 1	-0.0868	***
				(0.0198)	
Work Quartile = 2	0.0978	***	Work Quartile = 2	-0.0173	
	(0.0185)			(0.0215)	
Work Quartile = 3	0.0668	**	Work Quartile = 3	-0.0331	
	(0.0279)			(0.0256)	
Work Quartile = 4	-0.103	***	Work Quartile = 4	-0.1110	***
	(0.0296)			(0.0407)	
Residential Quartile = 2			Residential Quartile = 4		
Work Quartile = 1	-0.0441	***	Work Quartile = 1	-0.1210	***
	(0.0131)			(0.0268)	
Work Quartile = 2	0.0419		Work Quartile = 2	-0.0382	
	(0.0217)			(0.0305)	
Work Quartile = 3	0.0396		Work Quartile = 3	-0.0671	**
	(0.0269)			(0.0324)	
Work Quartile = 4	-0.0676	*	Work Quartile = 4	-0.106	***
	(0.0370)			(0.0391)	

Source: Author's calculations based on 2000 U.S. Census of Population and Housing and LEHD Employment History File and Employer Characteristics File. The values reported in this table are the OLS coefficients on a categorical variable representing each of the 16 different combinations of residential and workplace quartiles from one regression (with R-squared equal to 0.2700). The dependent variable is log of wages in 2000. Controls were included for MSA of residence, place of birth dummy variable, age, age-squared, gender, Hispanic ethnicity, years since migration and its square, citizenship, employer size and type, English skills, English is an official language in the country of birth, education, estimated minimum education in the U.S., and the proportion of the co-ethnic 1990 population in the U.S. who was residing in the individual's PMSA of residence. Clustered robust standard errors are reported in parentheses.

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

Regression Models and Analysis

The next step in this analysis uses the own-exposure rates detailed above to study how different levels of exposure to co-ethnics affects individuals' earnings. The figures and regressions based on quartile of own-exposure reported above, in addition to the annual earnings figures, show that workplace and residential networks do not have the same relationship with earnings. Instead, some degree of co-ethnic workplace exposure is correlated with higher earnings – implying that residential and workplace ethnic networks are different and operate differently with respect to labor market outcomes. I investigate these mechanisms below by running regressions on earnings using own-exposure rates as explanatory variables.

Estimating Human Capital Accumulation

Economic research on immigrants relies heavily on cohort analysis (Borjas 1985) and/or the inclusion of a measure of years since migration to estimate the effects of country-specific human capital accumulation on earnings. Since data limitations prevent this project from utilizing cohort analysis, I cannot directly estimate rates of human capital accumulation such as education or English skills acquired after immigration. Instead, I control for years since migration and two basic estimates of U.S.-specific human capital accumulation: minimum education completed in the U.S. and whether English is an official language in the country of birth. Ideal data would include time variant measures of education and English-skills to capture human capital accumulation but, unfortunately, data on education and English-skills are limited to one point in time: the 2000 Census. Some identification from these data is still possible: for example, we know who did not learn English and we know who emigrated from a country where English is not spoken but now reports speaking English.

The decision to learn English is motivated by a desire to increase the number of potential trade partners to include those who speak English. Lazear (1999) argues that "those who learn English after coming to the United States perform the same calculation, but do so at a later stage" as those who learned English prior to immigration. Abstracting from any concerns regarding self-selection, this rationale applies even to immigrants for whom English was compulsory in school since, like those who made the choice for themselves, they also learned English so as to be able to access a larger pool of potential trade partners. In order to capture this learning, the analyses below include both the self-reported language skills of the immigrant in the year 2000 and an indicator variable equal to 1 if English is the official language in the country of birth (as determined in Bleakley and Chin 2004). In this manner, I am able to estimate the value of learning English as a second language versus the value of speaking English.

Including immigrants who arrive as children in models of human capital accumulation complicates the interpretation of several important effects, most notably the value of education (since this education is primarily received in the U.S.) and the value of years since migration (since the effect of time in the U.S. may be different during childhood). Some previous research has relied on samples limited to immigrants who arrived after age 25 since, for the most part, individuals have completed their education by this age. However, limiting the sample to those who immigrate after education has been completed can result in biased samples if the process by which individuals select into immigration varies by country and by age group. Indeed, in countries with low educational attainment and relatively low immigration/transportation costs, especially Mexico, Guatemala and El Salvador, individuals who elect to emigrate in early adulthood might be choosing to do so for different reasons and with different expectations than those who choose to immigrate after the age of 25. Table 3.8 shows that 34-38% of immigrants from these three countries immigrated between the ages of 18 and 25, far higher than the average of about 27% for other country of birth groups. If those who immigrate after age 25 differ from those who emigrate earlier in unobservable characteristics, then limiting the sample in this manner will yield biased and unrepresentative results with this bias being more significant for groups with high rates of emigration in early adulthood. Instead, in this paper, I limit the

Table 3.8: Proportion of Immigrants by Age at Arrival, for Different Country and Region of Birth Groups

	Under 18	18 -24	25 and over
Canada	0.4293	0.2024	0.3684
China	0.1465	0.2173	0.6362
Colombia	0.2673	0.2750	0.4577
Cuba	0.3734	0.1472	0.4794
Dominican Republic	0.3256	0.2679	0.4065
El Salvador	0.3409	0.3591	0.3001
Germany	0.5538	0.2129	0.2333
Guatemala	0.2953	0.3805	0.3242
Haiti	0.2549	0.2701	0.4749
India	0.1461	0.3043	0.5496
Iran	0.2918	0.2647	0.4435
Italy	0.5116	0.2321	0.2563
Jamaica	0.3156	0.2157	0.4687
Japan	0.2515	0.2410	0.5075
Mexico	0.4465	0.3408	0.2128
Philippines	0.2202	0.2374	0.5424
Poland	0.2246	0.2318	0.5436
South Korea	0.3799	0.1645	0.4555
Taiwan	0.2699	0.2347	0.4954
United Kingdom	0.2653	0.2238	0.5110
USSR Core	0.1800	0.1558	0.6643
Vietnam	0.3446	0.2564	0.3990
Africa	0.1551	0.3028	0.5421
Caribbean	0.3027	0.2629	0.4344
Central America	0.3369	0.2882	0.3749
Central Asia	0.2718	0.2615	0.4667
Middle East/North Africa	0.2941	0.2815	0.4244
Oceania	0.2794	0.2487	0.4719
Socialist Europe	0.2642	0.2370	0.4988
South America	0.2759	0.2672	0.4568
South East Asia	0.3065	0.2306	0.4630
Western Europe	0.3407	0.2907	0.3685
Total	0.3233	0.2697	0.4070

Source: Author's calculations based on 2000 U.S. Census of Population and Housing.

universe to immigrants who first arrive in the U.S. at the age of 18 or later – in this way, I exclude immigrants who spent their childhoods in the U.S. but allow for immigration by younger immigrants (who immigrate either for work or to attend college).

With the inclusion of younger immigrants, I must also address where education was completed. Since many individuals do not complete their formal education until their mid-20's, I create a new variable measuring estimated education in the U.S. using a similar approach as language above; specifically, given the age at arrival in the U.S. and the total education completed, a measure of maximum source country education can be developed. For example, an individual who emigrates at age 18 but reports having a college education is assumed to have a maximum source country education limited to high school and a U.S. college education. On the other hand, an individual who emigrates at the age of 40 and reports having an 8th grade education has a maximum source country education of 8th grade and no U.S. education. By construction, education levels of high school diploma or less are assumed to have been completed prior to immigration since only those who immigrated at age 18 or later are included in the sample. For immigrants with more than a high school diploma, age at arrival directly determines the value of maximum education completed in the U.S. – this approach fails to identify individuals who continue their education as non-traditional students later in life. Any resulting bias in the estimate of U.S. human capital accumulation will be negative since this measure is purposely conservative in estimating education in the U.S.

OLS Regression Analysis

The log of earnings from employment is a function of standard human capital and demographic characteristics (age, gender, race, ethnicity, marital status, city of residence, and education), plus immigrant-specific traits (English ability, years since migration, and country of birth). A full set of country or region of birth and CMSA of residence indicators are included to address some of the systematic differences between country of birth groups (including the differing selection processes by which immigrants select *into* immigration) and to control for differences in earnings and employment opportunities in the five urban areas included in this study. Additionally, since earnings data are limited to that reported to state UI offices by employers, an indicator for whether the person reported also being self-employed in 2000 (SE_{it}) is included.[50] C_t^j is the share of co-ethnics either in the neighborhood (when $t = k$) or in the workplace (when $t = w$). Hence, the effect of co-ethnic concentration on the log of earnings is β_7.

$$\ln(y_i) = X_i\beta_1 + \beta_2 ysm_i + \beta_3 ed_i + \beta_5 eng_i + \beta_6(eng_i) + \beta_7 C_t^j + POB_j + CMSA_k + e_i$$

$$\text{where } X_i = [age_i, age_i^2, male_i, race_i, Hisp_i, married_i, SE_i]$$

$$t \in \{k, w\}$$

These OLS regressions yield consistently negative and significant coefficients for the residential and workplace own-exposures as reported in Table 3.9. This is the case even after controlling for education and other individual human capital measures and taking certain precautions against selection by including a vector of country of birth dummy variables, another vector of CMSA of residence dummy variables, and the country of birth distribution in 1990, as detailed above. The first set of regression results in Table 3.9 is based on the full sample while the remaining four

[50] The self-employed with UI earnings records in 2000 have not been dropped from the sample of analysis since many small business owners earn large shares of their income from seasonal or yearlong secondary employment in the formal labor market. This income might be especially important for small immigrant businesses whose proprietors may rely on seasonal work for supplemental earnings.

Table 3.9: The Effects of Residential and Workplace Co-ethnic Exposure Rates on Log of Earnings in 2000

	Full Sample		More than High School Diploma		High School Diploma or Less	
Residential Exposure Rate	-0.2701***		-0.4886***		-0.1599***	
	(0.053)		(0.070)		(0.040)	
Workplace Exposure Rate		-0.2933**		-0.2774**		-0.2554**
		(0.096)		(0.096)		(0.095)
Years Since Migration (YSM)	0.0234***	0.0227***	0.0243***	0.0234***	0.0201***	0.0199***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
YSM - squared	-0.0003***	-0.0003***	-0.0002**	-0.0002*	-0.0003***	-0.0003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Speaks English	0.1461***	0.1413***	0.1571***	0.1571***	0.1318***	0.1266***
	(0.015)	(0.014)	(0.023)	(0.023)	(0.009)	(0.010)
English-speaking POB	0.2078***	0.1986***	0.2053***	0.1967***	0.2156***	0.2074***
	(0.019)	(0.017)	(0.017)	(0.016)	(0.030)	(0.028)
Some High School	0.0478***	0.0456***	-	-	0.0632***	0.0604***
	(0.009)	(0.008)			(0.007)	(0.006)
High School Diploma	0.1029***	0.0999***	-	-	0.1315***	0.1274***
	(0.016)	(0.014)			(0.013)	(0.011)
Some College	0.2497***	0.2439***	-	-	-	-
	(0.023)	(0.020)				
College	0.5791***	0.5729***	0.3034***	0.3061***	-	-
	(0.031)	(0.028)	(0.017)	(0.017)		
Graduate Degree	0.8607***	0.8502***	0.5831***	0.5844***	-	-
	(0.047)	(0.042)	(0.026)	(0.027)		
Some College in U.S.	-0.0177	-0.0191	-0.026	-0.0286	-	-
	(0.016)	(0.016)	(0.021)	(0.022)		
College in U.S.	0.0591*	0.0557*	0.0405	0	-	-
	(0.025)	(0.025)	(0.024)	(0.000)		
Graduate School in U.S.	0.0328	0.0319	-0.0034	-0.005	-	-
	(0.026)	(0.026)	(0.023)	(0.000)		
Self-employed	-0.5600***	-0.5793***	-0.6377***	-0.6494***	-0.4650***	-0.4876***
	(0.014)	(0.015)	(0.016)	(0.018)	(0.019)	(0.018)
Small firm	-0.4851***	-0.5013***	-0.5078***	-0.5179***	-0.4523***	-0.4718***
	(0.016)	(0.018)	(0.018)	(0.020)	(0.017)	(0.019)
Observations	~500,000	~500,000	~250,000	~250,000	~250,000	~250,000
R-squared	0.267	0.268	0.222	0.221	0.190	0.192

Source: Author's calculations based on 2000 U.S. Census of Population and Housing and LEHD Employment History File and Employer Characteristics File. All regressions also include age, age-squared, sex, Hispanic ethnicity, U.S. citizenship status, country of birth and MSA of residence identifiers. Data are constructed from the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. Workplace exposure rate is calculated at the state-employer level for large firms, at the Census block of workplace and industry level for small firms, and at the PMSA-industry cell for the self-employed. Note that only wages reported by employer are included for the self-employed: that is, only the wages that were not from self-employment are included in this analysis for those self-employed who were also employed.
Clustered robust standard errors are reported in parentheses.
*** p<0.001, ** p<0.01, * p<0.05

columns report the results for these same specifications limited to the high education groups (more than a high school diploma) and the low education groups (high school diploma or less).

For each specification, the standard human capital results emerge: age[51] and years since migration (YSM) are both concave and speaking English and having more formal schooling are both associated with higher earnings.

Speaking English either well or very well is associated with a 14 to 15% increase in reported earnings for immigrants born in countries where English is not an official language. This premium is slightly higher for immigrants with more than a high school education compared to those with at most a high school diploma (15.7% versus 13%). Those who speak English and are from a country where English is an official language (for example, the U.K., Canada, India, and Jamaica) report earnings that are approximately 35% higher than similar immigrants who do not speak English and are not from a country where English is an official language. The earnings premium associated with emigrating from a country where English is one of the official languages is about 20% - that is, these results imply that, all else equal, an immigrant who does not speak English but is from an English-speaking country (such as a French Canadian) will earn 20% more than an immigrant who also does not speak English but is from a country where English is not an official language (such as France). This implies the presence of other benefits of being born in an English-speaking country in transitioning into the American labor market – perhaps similarity in social or labor force institutions.

The returns to formal schooling are consistent with previous research – the returns increase exponentially as schooling increases. Hence, the increase from an 8th grade education to a 10th grade education implies an increase of only 5% in earnings, whereas an increase from an 8th grade education to a high school diploma is a 10% increase, to a college degree is a 58% increase

[51] Age and its square are not included in this table of results. These coefficients can be obtained from the author upon request.

and to a graduate degree is an 86% increase in earnings. Attending college in the U.S. yields a small earnings premium compared to arriving with a college degree from abroad of about 6%, but no statistically significant earnings premium is found for obtaining a graduate degree in the U.S. as opposed to arriving with one. As discussed above, these variables are inexact and miscode individuals who may have returned to school as adults after immigrating. Because this measure of U.S. education is purposely conservative, it underestimates the impact on earnings of receiving a U.S. college or graduate degree.

The final two coefficients listed in Table 3.9 are control variables for whether an individual reports self-employment as his or her primary employment and whether the individual works in a firm with 5 or fewer coworkers. Reported earnings for the self-employed are 56% lower, overall, then for similar individuals employed in firms with 6 or more employees. This stark earnings differential is a result of the earnings in this analysis being derived solely from employment – explicitly excluding self-employment earnings. Higher educated immigrants who report being self-employed have earnings that are about 64% lower than their non-self-employed counterparts while those with just a high school diploma or less have earnings that are 47-49% lower than similar non-self-employed individuals. This might indicate greater reliance on outside work for self-employed immigrants with low levels of education relative to self-employed immigrants with high levels of education. Immigrants working in small firms, those with 5 or fewer employees, report earnings roughly 50% lower than those who work in large firms. While some of these smaller firms may be underreporting earnings, it is also likely that most of this discrepancy is based on real earnings differences. Either way, the inclusion of the small firm indicator will serve to control for the differences between large employers and small employers.

All six regressions reported in Table 3.9 show that increased rates of own-exposure (both in the neighborhood and at the workplace) have a negative impact on earnings. The coefficients on workplace exposure rate report a nearly identical effect on earnings for both educational attainment groups: moving from a firm with 0% co-ethnics to one with 100% co-ethnics implies new earnings that are between 26 to 29% lower. As Table 3.2 reports, immigrants with more than a high school degree work in firms with lower co-ethnic exposure rates. At the mean workplace co-ethnic exposure rates for each group, an immigrant with some post-secondary education earns about 2% less than he would in a workplace with no co-ethnics while an immigrant with a high school education or less earns 4% less than if he or she worked with no co-ethnics. The effects for residential co-ethnic exposure differ dramatically between educational groups: while the average effect of going from a neighborhood with no co-ethnics to one composed entirely of co-ethnics is a 27% decrease in reported earnings, it is a 49% decrease for those with more than a high school education while only a 16% decrease for those with at most a high school education. Again, using the average residential co-ethnic exposure rates reported in Table 3.2, this implies an average decrease in earnings of almost 4% for high education immigrants and 3% for those with less education. Living in neighborhoods composed primarily of co-ethnics has a large, negative effect on the earnings of immigrants with higher education – implying a far larger opportunity cost of living in an ethnic enclave for highly educated immigrants than for those with less education. This can be attributed to two processes: the impact of limiting social networks to co-ethnics resulting in limited job opportunities and country-specific human capital accumulation *and* the result of self-selection. This is discussed in more detail in the next section.

The regressions using quartiles of own-exposure suggest a nonlinear relationship between own-exposure rates, especially in the workplace, and earnings. Indeed, subsequent regressions[52] using own-exposure and its square reveal that residential own-exposure has a statistically significant convex effect on earnings – its negative effect on earnings gradually weakens until, at very high, out of sample levels of co-ethnic residential exposure rates, it has a positive effect on earnings. That is, the negative earnings effect of living in a neighborhood with more co-ethnics starts off relatively large and gradually becomes smaller as the neighborhood becomes more co-ethnic. On the other hand, specifications using workplace co-ethnic exposure rates and its square resulted in neither coefficient being statistically significant. Additionally, preliminary specifications including both measures of own-exposure and their squares indicate possibly opposing earnings effects, especially for low-education immigrants, warranting further analysis.

Instrumental Variable Analysis

Previous research has struggled with the self-selection problems inherent in looking at residential choice patterns. Immigrants do not sort randomly into ethnic enclaves; rather, observed and unobserved traits influence an individual's residential choice. Certain observable traits are known to lead to higher co-ethnic concentration measures: not speaking English and the individual's country of birth are two of the most important. Problems of selection arise if an unobserved trait, such as ability or proclivity to assimilate, influences both residential choice and earnings outcomes. The issues that arise in studies of ethnic enclave effects, akin to Manski's (2000) "reflection problem," arise from the question of whether the individual outcome is influenced by his social network or, rather, are both the individual and the network being affected by some exogenous trait? To establish enclave effects, exogenous traits that influence

[52] These regression results have not yet undergone disclosure review and so cannot be reported in more detail.

both the individual and the network must be addressed. For starters, area fixed effects are included in all regressions: the city in which you live will affect the labor opportunities to which you are exposed, thus affecting your labor outcome. Additionally, an important approach in immigration research is to control for country of birth since, as discussed in Borjas (1987), the selection into immigration can vary substantially between different countries. Sousa (2011a) shows that half of the individual variation in residential own-exposure and a quarter of workplace own-exposure is explained by observables. Though this addresses country of origin and metropolitan differences, it does not address potential unobserved traits that differ between immigrants from country j who choose to enclave and their co-ethnics who choose not to enclave.

The approach taken in this study combines the strategies employed by Altonji and Card (1991) and Bertrand, Luttmer and Mullainathan (2000), both discussed above, by using the 1990 PMSA-level of residential ethnic concentration as an instrument for census tract level concentrations, both residential and at the workplace, in 2000. The proportion of the PMSA population that is co-ethnic is an important predictor of residential and workplace own-exposure rates since immigrants belonging to a group with more members in the city of residence are at greater risk of having more co-ethnic coworkers or neighbors, even if individuals were randomly sorted into neighborhoods and firms.

By using the lagged value of this variable, I also address issues of simultaneity while incorporating well-established patterns of immigrant settlement. As shown in Blanchard and Katz (1992), local labor markets adjust to labor supply shocks within a decade. Hence, the previous decade's share of co-ethnic labor has, by 2000, already resulted in adjusted earnings or local labor supply/demand changes. Using the 1990 co-ethnic share rather than the 2000 co-ethnic share allows for local labor market adjustments to new labor inflow. By using the PMSA share of co-ethnic population, I am taking advantage of the fact that a significant factor in immigrants' location choice is the location choices of his or her co-ethnics (Bartel 1989). The effect being studied in this paper, however, is the segregation in either neighborhood or workplace *within* five urban areas with high immigrant concentration. Instrumenting at the PMSA level allows for the correction of unobservable traits in the selection into specific neighborhoods and employers while allowing for selection into high co-ethnic metropolitan areas. This approach does not address the selection into destination cities – in fact, the research sample used in this paper purposely limits the scope of analysis to cities of high immigrant concentration. Instead, what is addressed with this instrumental variable approach is the difference between immigrants with social networks limited to co-ethnics (in neighborhood and workplace) and immigrants who have access to co-ethnics but whose social networks are not made up primarily of co-ethnics. In essence, we are not interested in the effect of living or working in New York City on the earnings of a Dominican immigrant, though there are many Dominican immigrants in New York. Instead, we are interested in the effect of living or working in areas of high Dominican concentration on the earnings of Dominican immigrants, allowing for the fact that many Dominican immigrants live in New York.

Instrumental Variable Regression Results

Table 3.10 reports results for six regressions: one for each co-ethnic exposure rate in the neighborhood and at work for the full sample, those with more than a high school education, and

Table 3.10. The Effects of Residential and Workplace Co-ethnic Exposure Rates on Log of Earnings in 2000: Instrumental Variable Analysis Using 1990 PMSA Co-ethnic Exposure Rate

	Full Sample	
Residential Exposure Rate	-0.1739	
	(0.093]	
Workplace Exposure Rate		-0.2984
		(0.162]
R-squared	0.268	0.27
First-stage F-test	5941.5***	2770.17***
First-stage T-test	179.92	91.33
	More than High School Diploma	
Residential Exposure Rate	-0.3474*	
	(0.152]	
Workplace Exposure Rate		-0.5979*
		(0.243]
R-squared	0.222	0.219
First-stage F-test	1818.27***	656.97***
First-stage T-test	114.9	58.22
	High School Diploma or Less	
Residential Exposure Rate	-0.0591	
	(0.098]	
Workplace Exposure Rate		-0.1008
		(0.172]
R-squared	0.191	0.192
First-stage F-test	4285.3***	2157.94***
First-stage T-test	138.96	70.27

Source: Author's calculations based on 2000 U.S. Census of Population and Housing and LEHD Employment History File and Employer Characteristics File. All above coefficients are from two-stage least square regressions where residential co-ethnic exposure rates at the neighborhood level are instrumented using either the 1990 PMSA-level co-ethnic exposure rate. These regressions control for sex, age, age-squared, years since migration and its square, Hispanic ethnicity, citizenship status, educational attainment, estimated U.S. educational attainment, employer type, English-language ability, identifier for English is an official language in country of birth, country of birth and MSA of residence. Data are constructed from the 2000 U.S. Census of Population and Housing 1-in-6 sample and the LEHD Employer Characteristics File and Employment History File. Workplace exposure rate is calculated at the state-employer level for large firms, at the Census block of workplace and industry level for small firms, and at the PMSA-industry cell for the self-employed.

Clustered robust standard errors are reported in parentheses.

*** p<0.001, ** p<0.01, * p<0.05

those with a high school education or less. Using the lagged proportion of the PMSA population that is co-ethnic as an instrument for the workplace or residential own-exposure mitigates the negative effects of clustering for most specifications. This suggests negative selection into high co-ethnic areas or workplaces, given selection into high immigration areas.

Two concerns arise in the use of an instrumental variable approach: instrument relevance and instrument exogeneity. While exogeneity cannot be empirically verified, instrument relevance is easily established by looking at the results from the first stage regression. The high F-statistics associated with each of the first stage regressions reported in Table 3.10 show that, for all specifications, the instruments in these models are highly predictive of the endogenous variable being instrumented (Stock, Wright and Yogo 2002). The regression tables also report the *t*-statistic of the excluded variable, showing that it is a consistently important predictor of the endogenous variable being instrumented.

The results obtained from using the 1990 proportion of the population that is co-ethnic as an instrument for either the residential or the workplace exposure rates support the conclusion that self-selection plays a significant role in explaining the negative impact on earnings that was found using the OLS regressions, especially with regards to immigrants with low educational attainment. The top rows on Table 3.10 show that, for the full sample, controlling for sorting decreases the negative impact of residential clustering by over a third while maintaining the estimated effect of workplace clustering. However, the instrumental variable analysis results in larger standard errors, hence both of these coefficients are only statistically significant at the 90^{th} percent confidence interval. These results imply that negative sorting explains over a third of the negative effect of residential clustering but does not explain the negative returns to working with more co-ethnics.

Once the sample is stratified by educational attainment, however, the results highlight differing roles of sorting for low education and high education immigrants. Instrumenting for co-ethnic exposure rates results in smaller coefficient estimates of the own-exposure effects for immigrants with a high school education or less – the decreases in the coefficients imply that negative self-selection into ethnic neighborhoods and co-ethnic workplaces explains over 60% of the decrease in expected earnings associated with living in ethnic neighborhoods and with working with higher concentrations of co-ethnics. After controlling for sorting on unobservables, the expected loss in earnings of going from a workplace with no co-ethnics to one full of co-ethnics drops from 25% to 10% for immigrants with a high school education or less while the estimated earnings decrease from moving from a neighborhood with no co-ethnics to one fully composed of co-ethnics drops from 16% to 6%. Neither of these coefficients is statistically different from 0, allowing for the possibility that negative self-selection fully explains the negative effects of ethnic clustering for immigrants with less education.

However, for those with more than a high school education, the 1990 instrument results in a negative effect statistically significant at the 95th percentile for both living and working with more co-ethnics. While correcting for sorting mitigates the earnings penalty of residential co-ethnic exposure from 3.7% to 2.6%, it more than doubles the earnings penalty associated with working in firms with higher concentrations of co-ethnic employees to 5.2% from 2.4%.[53] Self-selection in neighborhood choice explains about a third of the earnings penalty found among highly educated immigrants. On the other hand, working with co-ethnics has a significant negative effect on earnings and this effect is only augmented once self-selection is addressed.

[53] The estimated wage penalties reported here are measured at the average residential co-ethnic exposure for immigrants with more than a college diploma, as reported in Table 3.2. Similarly, the average workplace co-ethnic

These results show that, to some extent, negative self-selection is mitigating the earnings penalty associated with working with more co-ethnics. Or, in other words, self-selection is masking larger negative earnings effects of working with more co-ethnics for immigrants with higher levels of education. These results suggest different employment and human capital accumulation mechanisms within the neighborhood and within the workplace. Residing in areas of relatively high co-ethnic exposure may decrease earnings by limiting social interactions with individuals who are not co-ethnics, thus decreasing the accumulation of country-specific human capital. In general, immigrant clustering in the workplace decreases the earnings of immigrants with more than a high school education – however, immigrants with unobservable traits that are less valuable in the general labor market, perhaps an inability to assimilate sufficiently leading to lower productivity levels in more integrated firms, are selecting into firms with higher co-ethnic concentrations where they are more productive.

Conclusion

Immigrants who live and work in high co-ethnic areas and firms earn less. But, would these immigrants earn more if they did not live in high co-ethnic areas? The counterfactual, of course, is not observed. However, the results from the instrumental variable estimation addressing self-selection into high co-ethnic neighborhoods and high co-ethnic employers suggest that the findings of negative enclave effects are partially due to negative selection. I find that negative selection into co-ethnic neighborhoods and workplaces explains a larger portion of the earnings penalties associated with more co-ethnic neighbors or coworkers for immigrants with a high school education or less than for those with higher levels of education. Negative selection

---

exposure rate for immigrants with more than a high school education is used to estimate the wage penalties in the workplace.

accounts for 30% of the earnings penalty associated with higher concentrations of co-ethnic neighbors for immigrants with more than a high school education while it explains 60% of the earnings penalty for immigrants with a high school education or less. The negative effect of residential ethnic clustering on the earnings of low education immigrants that remains after sorting is addressed is not statistically different from 0, implying that negative self-selection may fully explain the lower earnings within residential enclaves. On the other hand, even after addressing sorting on unobservables, an earnings penalty of about 2.6% remains for immigrants with more than a high school education who live in neighborhoods with 7.5% co-ethnics, the average own-exposure rate for this education group. For these immigrants, lower earnings attributed to residential ethnic exposure are only partially explained by self-selection – the remainder may well be due to limited referral networks and human capital traps.

There is no question that negative self-selection is leading to higher co-ethnic concentrations in ethnic neighborhoods and workplaces. Relying on earlier work on social networks, this negative selection can lead to lower earnings and less employment opportunities since the quality of the network will lead to externalities for its members (Calvó-Armengol and Jackson 2004). However, given the limited employment opportunities for immigrants with low educational attainment or limited English skills, this negative self-selection does not seem to yield lower earnings than would otherwise be expected for immigrants with low levels of schooling. For immigrants with some postsecondary education, however, I do find evidence of possible human capital traps. Immigrants with some post-secondary schooling who work with more co-ethnics earn less than they would if they worked in more integrated workplaces. After controlling for sorting, this effect is responsible for an earnings penalty of 5% for the average individual in this data set with more than a high school education. At the same time, I find evidence that negative selection into workplaces is yielding higher earnings than would be the case if these workers were only sorting on observables, resulting in an earnings penalty of only 2% on average. This may indicate that some workers are more productive, or more highly valued, in firms with more co-ethnics perhaps due to lower transaction costs as argued in Lazear (1999). Though negative selection affects both residential and workplace clustering, the impact of ethnic segregation in these two realms operates differently on the earnings of immigrants based on their educational attainment. The evidence suggests that enclaves are not creating a "warm embrace" for immigrants with low levels of education, though they are not necessarily being hurt by ethnic clustering either. On the other hand, immigrants with more than a high school education face earnings penalties due to both types of ethnic clustering, suggesting that ethnic enclaves might be creating human capital traps.

REFERENCES

Abowd, John M., Francis Kramarz and David Margolis. 1999. High Wage Workers and High Wage Firms. *Econometrica* 67:2.

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

Alba, Richard D., John R. Logan, Brian J. Stults, Gilbert Marzan, and Wenquan Zhang. 1999. Immigrant Groups in the Suburbs: A Reexamination of Suburbanization and Spatial Assimilation. *American Sociological Review* 64:3.

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. U.S. Census Bureau, Center for Economic Studies Working Papers CES 10-39.

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.

Blanchard, Olivier Jean and Lawrence F. Katz. Regional Evolutions. *Brookings Papers on Economic Activity* 1992:1.

Bleakley, Hoyt and Aimee Chin. 2004. Language Skills and Earnings: Evidence from Childhood Immigrants. *The Review of Economics and Statistics* 86:2.

Borjas, George J., 1985. Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics* 3:4.

--- 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.

--- 2002. Homeownership in the Immigrant Population. *Journal of Urban Economics* 52:3.

Calvó-Armengol, Antoni and Matthew O. Jackson. 2004. The Effects of Social Networks on Employment and Inequality. *The American Economic Review* 94:3.

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.

Manski, Charles. 2000. Economic Analysis of Social Interactions. *The Journal of Economic Perspectives* 14:3.

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

Oreopoulos, Philip. 2011. Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. *American Economic Journal: Economic Policy 3*.

Pedace, Roberto and Stephanie Rohn. 2008. A Warm Embrace or a Cold Shoulder: Wage and Employment Outcomes in Ethnic Enclaves. U.S. Census Bureau, Center for Economic Studies Working Papers, CES 08-09.

Rosenzweig, Mark. 1995. Why are there Returns to Schooling? *The American Economic Review* 85:2.

Sousa, Liliana. 2011a. Identifying Ethnic Enclaves Using Linked Employer-Household Data. Doctoral Dissertation, Cornell University, Chapter 2.

--- 2011b. Community Determinants of Immigrant Self-Employment: Human Capital Spillovers and Ethnic Enclaves. Doctoral Dissertation, Cornell University, Chapter 1.

Stock, James H., Jonathan H. Wright and Motohiro Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 20:4.

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