The Impact of Having a Job at Migration on Settlement Decisions: Ethnic Enclaves as Job Search Networks*

Lee Tucker

Boston University†

This version: October 15, 2014

Abstract

Observational evidence has shown that immigrants who choose to locate in enclaves have worse post-migration outcomes than those who locate elsewhere. This suggests that immigrants may be negatively selected into enclaves based on their ability to assimilate or to avoid discrimination. I hypothesize, however, that new immigrants choose to locate in enclaves based on their ability to benefit from the ethnic networks that these enclaves provide. In particular, new immigrants may select themselves into enclaves positively based on their ability to benefit from pre-existing labor market networks. Using data from the New Immigrant Survey, this paper shows that immigrants who arrive without job offers are significantly more likely to locate in enclaves, even after accounting for a wide range of pre-migration and time-invariant characteristics. However, I do not find significant evidence of heterogeneous effects of a job offer. This suggests that highly-skilled individuals and low-skilled individuals may benefit similarly from the job search benefits of locating in an enclave.

*I would like to thank my advisors Daniele Paserman and Sam Bazzi for their feedback and support. Thank you to the BU Institute for Economic Development for funding this project. Additionally, I thank Delia Furtado, Dilip Mookherjee, Nicholas Tilipman, Wesley Yin, and participants of the BU Development Reading Group and Labor Reading Group for their comments.

†Deparment of Economics, 270 Bay State Rd., Boston, MA 02215. Email: tuckerl@bu.edu

1 Introduction

Ethnic enclaves are a persistent phenomenon in America and around the world. They persist over long periods of time, in spite of freedom of movement. They also persist in spite of observational evidence that residents of ethnic enclaves have worse labor market outcomes than non-residents. Thus, enclaves are properly seen as a sorting phenomenon: low-skilled new immigrants cluster in enclaves, while highly-skilled new immigrants more frequently choose to locate in areas where their ethnic group does not constitute a large proportion of the population.

Because enclaves are persistent, new immigrants must be responding to socioeconomic incentives when they choose where to locate. Yet, it remains difficult to separate among the plausible causes of enclave formation. Theoretical justifications for enclaves include discrimination, human capital externalities, and private benefits from avoiding assimilation [Edin et al., 2003, Lazear, 1999, Cutler et al., 2008]. However, one of the most compelling hypotheses is that immigrants sort themselves to take advantage of network effects in the labor market. According to this line of reasoning, immigrants locate in enclaves so that they can use local networks to learn about and apply for jobs. However, relatively little empirical research exists that can directly identify network effects separately from other causes of sorting.

Moreover, none of the existing research looks specifically at network effects in ethnic enclaves for highly-skilled immigrants. I hypothesize that highly-skilled new immigrants benefit less from job search networks than lower-skilled immigrants, and therefore that highly-skilled immigrants are less likely to sort themselves into enclaves based on the benefits of these job search networks. To test this hypothesis, I examine whether immigrants, and highly-skilled immigrants in particular, who have a job offer at the time of migration make different locational decisions than immigrants who do not already have a job offer. If new immigrants who arrive without a job offer are taking advantage of ethnic enclave networks for job search, then they will be more likely to locate in areas where the density of other individuals in the network is high. Meanwhile, immigrants who arrive with job offers may benefit less from these networks, and so these immigrants will be less likely to locate in enclaves.

For my analysis, I use geographic data from the New Immigrant Survey [Jasso et al., 2006], matched to aggregate data from the U.S. Census at the ZIP code level. The New Immigrant Survey (NIS) provides a cross-sectional survey of approximately 8,500 new legal immigrants to the United States, and it is designed to be a representative sample of all new legal immigrants. In addition to asking about employment at the time of migration, it also provides a detailed retrospective survey of pre-migration characteristics that allow me to control for selection on a broad range of observable attributes. Importantly, the NIS includes data on pre-migration salary, hours, occupation and industry that may control for selection on these attributes without the reverse causality issues that would be present if post-migration controls were used in their place.

My empirical analysis shows that immigrants who migrate with a job offer are significantly less likely to locate in neighborhoods with a high proportion of individuals in their language group. Even within a given metropolitan area, my analysis suggests that immigrants who migrate with a job offer locate in less ethnically-dense neighborhoods than those who migrate without a job offer. This result is robust to the inclusion of a variety of covariates, and also to several different methods of defining ethnic groups based on languages spoken at home. In keeping with the predictions of my theoretical model, this suggests that immigrants who arrive without job offers make a trade off between the job-search benefits of enclave membership and commuting costs.

While there do appear to be labor market benefits of locating in an ethnic enclave area, the benefits of this decision need not be homogeneous. Indeed, my model allows for rich selection behavior in locational decisions. Based on plausible assumptions about the nature of commuting costs and the job-matching function, I show that the job search benefits to higher-skilled immigrants of enclave residence may be lower than the job search benefits to lower-skilled immigrants. This would imply that upon receiving a job offer, high-skilled immigrants' locational decisions are impacted less than low-skilled immigrants' locational decisions. To test this hypothesis, I examine whether the impact of a pre-migration job offer is heterogeneous across the distribution of observable skills. However, my analysis fails to find significant evidence of heterogeneous effects, either on education or on other observable characteristics.

Also in keeping with the broader literature on selection in migration, my model suggests that immigrants who arrive with a job offer may be observably different from immigrants who do not arrive with a job offer. I find strong empirical evidence of this assertion. Immigrants who arrive with a job offer are more highly educated than immigrants who do not arrive with one, and they are considerably more highly paid both before and after migration. These differences suggest that selection on unobservables may be a concern. Any unobservable characteristics that are be correlated with the decision to locate in an ethnic enclave may bias regression results if they are also correlated with the probability of receiving a pre-migration job offer. As a robustness check, I perform an analysis using a kernel-based propensity score matching method. This analysis may be more robust to any potential impact of unobservables, as long as the correlation between unobservable characteristics and the probability of receiving a job offer is locally close to 0. The propensity score analysis is suggestive of a treatment effect; the magnitude of the estimated average treatment effect is similar to the results from OLS and probit regression, though the reduced power of this method means that these results are only marginally significant.

This paper proceeds in sections as follows. Section 2 provides an overview of literature related to migration, and in particular to the role of ethnic enclaves. Next, Section 3 introduces a simple theoretical model that supports my hypotheses. This model illustrates that there is an inherent trade-off for new, unemployed immigrants in choosing where to locate that does not exist for immigrants who arrive with a job offer. It also provides a framework for incorporating ethnic enclaves into the broader literature on the selection of immigrants. Then, section 4 provides an empirical framework that allows me to test a key prediction of the theoretical model, that new immigrants who have jobs at the time of migration will be less likely to locate in ethnic enclaves that immigrants who do not have jobs at the time of migration. Section 5 describes the New Immigrant Survey and Census data that I use in this version of the paper, as well as the additional information that I intend to use once I have been granted access to the Restricted Use version of the Survey. Section 6 provides and discusses my main results, and section 7 provides a range of robustness checks, include a propensity score analysis. Finally, Section 8 concludes the paper.

2 Overview of the Literature

Migration is an issue of considerable policy interest, and so there exists a large migration literature. Most of the literature focuses broadly on the issue of selection into migration, or on the impact of migration on native workers throughout the skill distribution [Borjas, 1987, Card, 2001, Borjas and Friedberg, 2009, Peri and Sparber, 2009]. Many of these studies attempt to instrument for endogenous labor market sorting, typically under the assumption that migration patterns are persistent over time [Card, 2001]. However, for this very reason, most of this literature can say little about the nature of this endogenous sorting, or about what mechanisms might cause migrants to sort in a persistent fashion. Apart from broad policies that encourage or discourage migration, there exist many related policy questions that may benefit from an understanding of these mechanisms. For example, to fully understand the impacts of U.S. policies that encourage family reunification, urban planning policies, or welfare programs, it is important to understand the nature of migrant social networks, particularly as they relate to ethnic enclaves.

There is evidence that network effects do matter for job search among the general population. The best such evidence comes from Bayer et al. [2008]. These authors found that an urban resident is more likely to work with a resident of his/her own city block than on neighboring blocks. Since travelling a single city block to a job presumably imposes only very small commuting costs, this is supposed to identify potential network effects. Not only does the study find evidence of such a pattern, but it finds that this effect is stronger when residents have similar socio-economic characteristics. This supports the notion that ethnic enclaves form to take advantage of network effects, and it additionally suggests that enclaves may matter at a very localized level.

Among the immigrant population, Goel and Lang [2012] has proposed a theoretical model in which immigrants take advantage of their ethnic networks to reduce job search costs. Looking at a survey of new immigrants to Canada, they test the predictions of this model empirically by looking at whether new immigrants claim to have a friend or relative in the country at the time of immigration. They also use geographic measures of immigrant density to define ethnic enclaves. This model and results suggest that residents of ethnic enclaves may benefit from the more rapid arrival of job offers, even when the distribution of potential job offers to be received does not change. These predictions are consistent with my hypothesis. However, this theoretical model is based on a population of homogeneous immigrants, and so it is unable to make predictions about effects across different levels of worker skill.

There is also evidence that network effects matter in other decisions that migrants make. For example, Bertrand et al. [2000] uses Census data to look at whether network effects matter in the application for welfare benefits. As in this paper, they group migrants by common language, and they exploit variation in the density of migrants to conclude that network effects cause members of language groups with high existing welfare use to be more likely to apply for welfare benefits themselves. Furtado and Theodoropoulos [2012] uses a similar method to look at the take-up of disability programs. Similarly, research by Patel and Vella [2012] shows that the occupational choices of immigrant groups tend to be persistent over time. The authors show not only that new immigrants tend to choose the same occupations as previous migrants, but also that they earn a wage premium by doing so. This suggests that networks may be used as a way to receive better job offers than would otherwise be available.

While studies such those of Bertrand et al. and Patel and Vella are able to employ careful instrumental variables strategies to achieve clean identification of network effects, their methods are not well-suited to considering initial locational decisions. Their methods can identify network effects only from within-metropolitan area variation in welfare use. In contrast, new immigrants choose where to locate not only within metropolitan areas, but across them as well, and there may be selection in these decisions. It is for this reason that I have adopted a selection on observables strategy, with careful control for pre-migration characteristics.

Another literature related to ethnic enclaves looks not at the causes of their formation, but at the outcomes of their residents. These studies frequently use the initial placement of refugee immigrants, as assigned by government policy, to address the issue of immigrant sorting. The work of Edin et al. [2003] used this method to look at the impact of living in an ethnic enclave on refugee immigrant earnings in Sweden. They found not only that immigrants benefit from initial placement in an ethnic enclave, but they found strong evidence of immigrant sorting after initial placement, particularly among low-skilled refugees. Work by Beaman [2011] on refugees assigned around the U.S. provides additional insight by suggesting that network effects may be heterogeneous. Refugees who were assigned to locations with a large social network of very recent immigrants showed lower incomes, while refugees who were assigned to locations with many longstanding network members appear to benefit from their assignment. These heterogeneous treatment effects support the labor market networks hypothesis, and they also suggest that the persistence of ethnic enclaves benefits new immigrants. However, these studies cannot show whether network effects are also important for non-refugees, or when locations are endogenously chosen.

Finally, looking solely at low-skilled agricultural migrants from Mexico, research by Munshi [2003] finds that migrants with larger networks are more likely to take higher-paying non-agricultural jobs. Because Munshi looks at repeat migrants, he is able to include individual fixed effects, which control effectively for selection on unobservables to cleanly identify network effects. However, it is unclear from this analysis whether such effects might be similar for more highly-skilled migrants, who typically come from countries other than Mexico, and who may be searching for very different types of jobs. It is also unclear whether effects may be similar in largely metropolitan ethnic enclaves.

Each of these existing studies seems to suggest that job search networks may play a significant role in location decisions. However, none of these studies directly addresses the situation of having a job offer at the time of migration. More importantly, none of these studies is able to address some key questions that may be relevant for policymaking. For example: Are the effects of arriving with a job heterogeneous across the skill distribution or across ethnic groups? Furthermore, do migrants with jobs locate in different metropolitan areas altogether, or in different parts of the same metropolitan areas? Could highly localized policies such as urban planning policies that impact the formation of dense enclaves have an impact on immigrant outcomes? For these questions, the detailed retrospective survey data that I use has considerable advantages.

3 Theoretical Model

In this section, I present a model that encapsulates the basic decisions made by potential legal immigrants, and that encompasses both selection in the decision to migrate and selection in the decision to locate in an ethic enclave. This model is a variant on the Roy [1951] model of selection, as applied to migration by Borjas [1987] and others. This model enriches the Borjas model by encompassing two separate problems; the pre-migration decision of the potential migrant, and the decision upon migration of where a migrant should locate. As in the seminal Roy model, the decisions of individuals are a function of individual country-specific shocks, which may be correlated.

Let us first begin with the problem of the potential migrant in his/her native country. In each period, each risk-neutral potential migrant i earns a wage $\mu_0 + \epsilon_{0i}$, where μ_0 is the average wage for a potential migrant in this country, and ϵ_{0i} is a time-invariant, country-specific individual wage shock that is distributed with mean 0 and finite variance σ_0^2.[1] All potential immigrants are assumed to be employed in their native countries. However, upon migration, an immigrant may be either employed or unemployed, and he/she chooses a location in the new country in which to reside. Locations vary by their ethnic density $D \in [0, D_{max}]$, i.e. the fraction of residents in that location that are immigrants. An individual immigrant is said to live in an ethnic enclave if the ethnic density of his/her location is greater than some level $\omega\bar{D}$, where $\bar{D}$ is the average density of that ethnic group in the new country. Once they have settled, immigrants cannot change their location.[2]

An employed immigrant earns a net wage in each period $\mu_1 + \epsilon_{1i} - \gamma(D)$, where μ_1 is the average wage for a migrant, ϵ_{1i} is time-invariant wage shock in the receiving (new) country, also mean-0 with finite variance σ_1^2, and $\gamma(D)$ is a commuting cost that is assumed to be increasing in D, the ethnic density of the location

[1] In the Borjas model, these country-specific shocks are generally interpreted to be the residual wage differential net of the effects of observable characteristics, so that μ_0 is the mean wage for individuals conditional on observable characteristics. To simplify notation and to provide the model with a more intuitive interpretation, I assume that μ_0 is the unconditional mean, and therefore that ϵ_{0i} incorporates the effects of education and other observable characteristics, as well as unobserved heterogeneity in potential migrants. This also implies that the correlation between ϵ_{0i} and ϵ_{1i} is likely to be positive and relatively close to 1. In practice, ϵ_{0i} and ϵ_{1i} may be assumed to take an error components form, and the correlation of residual wage differences net of observables may be positive, negative, or zero.

[2] The basic predictions of this model will hold as long as relocation is costly and/or it is not instantaneous.

in which the immigrant resides.[3] Assume for simplicity and for later use that $\gamma(0) = 0$ and that $\gamma(\cdot)$ is twice continuously differentiable. Earnings of unemployed immigrants in each period are normalized to 0, and they do not depend on D. Immigrants and potential immigrants have complete information in this model, and there is no uncertainty, so each migrant knows the mean wage in each country, as well as his/her individual wage shocks ϵ_{0i} and ϵ_{1i}. Individuals in their native country may choose to migrate to the new country, and upon migration they choose D, the ethnic density of the location in which they choose to live.

In this model, both immigrants and non-immigrants may find receive job offers in the receiving country. However, in each period an immigrant who has already migrated to the receiving country is more likely to find a job than an immigrant who has not yet migrated. In this model, job offers arrive by a Poisson process, where the per-period arrival probability depends on that individual's country-specific shocks and, for unemployed immigrants, the ethnic density of their location. Thus, the probability that a non-immigrant finds a job is $\lambda_0(\epsilon_0, \epsilon_1)$, because it may depend on country-specific shocks in both his/her native country and in the new country. The probability that an unemployed immigrant finds a job is $\lambda_1(\epsilon_1, D)$, because it may depend on his/her country-specific shock in the new country as well as the ethnic density of the location in which the immigrant has chosen to reside. If locating in an ethnic enclave helps a new immigrant to find a job, then $\lambda_1(\cdot)$ will be increasing in D.[4] I assume also that both $\lambda_0(\cdot)$ and $\lambda_1(\cdot)$ are increasing in ϵ_1, and that $\lambda_1(\cdot)$ is twice continuously differentiable.

Migration for immigrants who are not currently employed costs a fraction c of that immigrant's utility in his home county, while migration for immigrants who have a job offer at the time of migration is assumed to be costless.[5] Additionally, I will assume for simplicity that when a potential migrant receives a job offer, that he or she migrates with certainty. Formally, this requires that the supports of ϵ_{0i} and ϵ_{1i} are bounded above and below, respectively, and it requires some restrictions on the relative values of μ_0, μ_1, ϵ_{0i} and ϵ_{1i} which will also depend on the functions $\lambda_0(\cdot)$, $\lambda_1(\cdot)$, and $\gamma(\cdot)$. For this version of this paper, I assume that these restrictions are satisfied.

Define V_{0i} as the value function for a potential migrant i in his/her native country, V_{1Ui} as the value function for an individual i who has migrated but is unemployed, and V_{1Mi} as the value function for an

[3] This assumption simply imposes the requirement that locating in an ethnic enclave is costly for employed individuals. If the jobs performed by employed immigrants are located outside the ethnic enclaves themselves, then this assumption will be met.

[4] My model does not explicitly model either the decisions of firms or the informational content of social network ties, and therefore it does not incorporate the type of asymmetric information problem that would imply that $\frac{\partial \lambda(\cdot)}{\partial D} > 0$. However, this result can be easily obtained from a model with a structure similar to that of Montgomery [1991].

[5] The assumption that migration is costless for individuals with a job at the time of migration simplifies the notation of the problem, but it does not change the intuition of the results. It is also consistent with a world in which employers pay for the migration of the foreign workers they hire.

individual i who has migrated and is employed. Then, the following equations characterize the model:

$$V_{0i} = \mu_0 + \epsilon_{0i} + \lambda_0(\epsilon_{0i}, \epsilon_{1i})\beta V'_{1Mi} + (1 - \lambda_0(\epsilon_{0i}, \epsilon_{1i}))\beta V'_{0i} \tag{1}$$

$$V_{1Ui} = 0 + \lambda_1(\epsilon_{1i}, D)\beta V'_{1Mi} + (1 - \lambda_1(\epsilon_{1i}, D))\beta V'_{1Ui} \tag{2}$$

$$V_{1Mi} = \mu_1 + \epsilon_{1i} - \gamma(D) + \beta V'_{1Mi} \tag{3}$$

Since ϵ_{1i} is time-invariant, and since there is no job separation in this model, the value function for employed immigrants can be rewritten as:

$$V_{1Mi} = \frac{\mu_1 + \epsilon_{1i} - \gamma(D)}{1 - \beta} \tag{4}$$

From (4), it is easy to see that immigrants who arrive with a job offer will seek to minimize their commuting costs $\gamma(D)$, and since $\gamma(\cdot)$ is assumed to be increasing in D, that all immigrants who arrive with job offers will choose $D = 0$. However, immigrants who arrive without job offers will choose D based on both the likelihood that they will find a job, and on the commuting costs that they will incur upon becoming employed.

Substituting the expression in (4) for V_{1Mi} into the equation for unemployed immigrants, we get that:

$$V_{1Ui} = \frac{\frac{\beta}{1-\beta}\lambda_1(\epsilon_{1i}, D)(\mu_1 + \epsilon_{1i} - \gamma(D))}{1 - (1 - \lambda_1(\epsilon_{1i}, D))\beta} \tag{5}$$

Similar substitution into the equation for non-immigrants implies that:

$$V_{0i} = \frac{\mu_0 + \epsilon_{0i}}{1 - (1 - \lambda_0(\epsilon_{0i}, \epsilon_{1i}))\beta} + \frac{\frac{\beta}{1-\beta}\lambda_0(\epsilon_{0i}, \epsilon_{1i})(\mu_1 + \epsilon_{1i} - \gamma(D))}{1 - (1 - \lambda_0(\epsilon_{0i}, \epsilon_{1i}))\beta} \tag{6}$$

A non-immigrant chooses to migrate without a job offer if the value of doing so exceeds the value of remaining in his/her native country, net of his/her migration costs $c \cdot V_{0i}$. Then, from the above, this model implies that migration will occur whenever:

$$\begin{aligned} &\frac{\frac{\beta}{1-\beta}\lambda_1(\epsilon_{1i}, D)(\mu_1 + \epsilon_{1i} - \gamma(D))}{1 - (1 - \lambda_1(\epsilon_{1i}, D))\beta} - \frac{(1+c)(\mu_0 + \epsilon_{0i})}{1 - (1 - \lambda_0(\epsilon_{0i}, \epsilon_{1i}))\beta} \\ &\quad - \frac{(1+c)\frac{\beta}{1-\beta}\lambda_0(\epsilon_{0i}, \epsilon_{1i})(\mu_1 + \epsilon_{1i} - \gamma(D))}{1 - (1 - \lambda_0(\epsilon_{0i}, \epsilon_{1i}))\beta} \geq 0 \end{aligned} \tag{7}$$

Empirically, one cannot observe the decisions of individuals who have not chosen to migrate, because it is not in general possible to observe or identify the value of ϵ_{1i} for anyone who has not migrated.[6] However, we do know that this constraint will be satisfied for any immigrant who is unemployed. Therefore, I will refer to (7) as the "Migration Constraint."

At the time of migration, a new immigrant who has migrated without a job offer chooses where to locate. Unlike a new immigrant who migrates with a job offer, this immigrant is not merely choosing to minimize

[6] Heckman and Honor [1990] demonstrate that under the traditional assumption made in the Roy model that wage shocks are distributed log-normally, and assuming that the correlation of ϵ_0 and ϵ_1 is known, it is possible to identify the distribution of ϵ_0 for non-migrants. However, the log-normal distribution is not bounded above, which is a necessary assumption for this model.

commuting costs. Instead, his/her choice of D reflects an inherent trade-off between the probability of finding employment, and the cost that will be incurred to commute to that employment once it has been found. That is, the new immigrant solves the following problem:

$$\begin{aligned} \max_D \frac{\frac{\beta}{1-\beta}\lambda_1(\epsilon_1, D)(\mu_1 + \epsilon_1 - \gamma(D))}{1-(1-\lambda_1(\epsilon_1, D))\beta}&, \text{ subject to} \\ \text{the Migration Constraint (7)}& \\ 0 \leq D \leq D_{max}& \end{aligned} \tag{8}$$

We can take the first order condition to solve for the optimal choice of D. I will abuse some notation by dropping all subscripts and arguments (ex. $\lambda \equiv \lambda_1(\epsilon_1, D)$). I will also take advantage of the conditions that $\beta \in (0,1)$ and that $\lambda \geq 0$. Then, with some algebra, taking the first order condition yields the following:

$$\frac{\mu + \epsilon - \gamma}{\gamma'} = \frac{\lambda}{\lambda'_D} \cdot \frac{1 - \beta + \beta\lambda}{1 - \beta} \tag{9}$$

From the above equations, the following features of the model are clear:

- *Locational decisions are not impacted* directly *by pre-migration characteristics.* This is true by construction. However, it is important to recognize that in general, since $Corr(\epsilon_0, \epsilon_1) \neq 0$, pre-migration characteristics will be correlated to observable locational outcomes. Additionally, pre-migration characteristics impact the likelihood of receiving a pre-migration job offer. Therefore, any pre-migration characteristics that are correlated to post-migration characteristics may still present endogeneity issues in empirical analysis.

- *All of the possible selection behaviors admissible in the Borjas [1987] model are also admissible in this model.* A key insight of Borjas selection model is that depending on the correlation of country-specific shocks, we may observe that immigrants are positively or negatively selected from the populations of their native countries, or that there may be "refugee" selection of individuals who are at the bottom of the native country's wage distribution but the top of the new country's wage distribution. All of these cases are also admissible in this model.[7]

- *Conditional on migration, the average wage of an immigrant's native country has no impact on his/her locational decision.* This follows naturally from the problem shown above, and from the assumption that ϵ_0 is mean-0. Taken literally, this implies that empirically, any observed correlation between the mean wage of a country and the locational decisions of immigrants who did not arrive with a job offer must result from differences in the distribution of ϵ_1 across countries. However, if the model is relaxed to allow for non-commuting benefits of ethnic density (see below), then it is possible that differences in the extent of non-commuting benefits across ethnic groups or across pre-migration income levels may lead to an observable correlation between average pre-migration wages and average locational decisions.

[7]In particular, the special case in which $\lambda_0(\epsilon_0, \epsilon_1) = 0 \;\forall\; \epsilon_0, \epsilon_1$, $\lambda_1(\epsilon_1, D) = 0 \;\forall\; \epsilon_1, D$, and $\gamma(D) = 0$ collapses neatly into the standard Borjas [1987] model form.

- *Immigrants who arrive with job offers will choose $D = 0$ regardless of their characteristics.* As discussed above, this is a relatively strong assumption of the model made by construction. If the function $\gamma(D)$ is taken more generally to specify "net costs of living in an area of ethnic density," and if the assumption of monotonicity is relaxed, then immigrants who arrive with job offers may choose $D > 0$. This model also does not account for the locational choices of firms, which likely make commuting costs non-monotonic in density below some threshold. However, as long as the non-commuting benefits of ethnic density are not increasing in ϵ_1 quickly relative to the commuting costs of living in an area of high ethnic density, we should observe that immigrants who arrive with job offers are less likely to live in an ethnic enclave than immigrants who do not arrive with job offers.

- *Under some conditions, an increase in ϵ_1 will lead to a decrease in the equilibrium ethnic density chosen for immigrants who migrate without a job offer.*

The predictions of this model suggest a number of testable hypotheses about heterogeneous behavior of immigrants. However, for the purposes of this paper, I will focus my attention on the hypothesis that immigrants who arrive with job offers are less likely to locate in an ethnic enclave than immigrants who do not, as well as the hypothesis that effects may be heterogeneous in observable skill.

4 Empirical Design

The theoretical model in Section 3 demonstrates that the if ethnic enclaves serve as job search networks for new immigrants, then the existence of these networks can be observed by observing the locational decisions of new immigrants. It is a key testable prediction of the model that immigrants who migrate without a job offer are more likely to locate in an ethnic enclave than immigrants who arrive with a job offer. Therefore, I seek to test for the existence of job search networks by testing this prediction.

Consider an individual i of ethnic group k living in geographic area j. Then, I seek to estimate α_1 in a model of the form:

$$E_{ijk} = \Phi\left(\alpha_0 + \alpha_1 Offer_i + \alpha_2 X_i + \delta_k + \eta_j + \nu_i\right) \tag{10}$$

Where X_i are individual characteristics, E_{ijk} is a binary measure of whether migrant i of group k in area j settles in an ethnic enclave, $Offer_i$ is a measure of whether the immigrant had a job offer at the time of his/her migration, δ_k and η_j are group and area fixed effects, and ν_i is a stochastic error term. Ethnic enclaves are defined as areas of substantially greater immigrant density than the average density for that ethnic group.

More formally, define ethnic density as follows:

$$Density_{jk} = \frac{\frac{\text{Number of people from group } k \text{ in area } j}{\text{Total population in area } j}}{\frac{\text{Number of people from group } k}{\text{Total population in country}}} \tag{11}$$

Then, the discrete dependent variable measure of ethnic density is:

$$E_{jk} \quad = \quad \mathbf{1}\left(Density_{jk} \geq \omega \bar{D}_k\right) \tag{12}$$

Each of these variables is constructed using Census geographic data. A more detailed description of the construction of these variables can be found in the Appendix in Table 11 and Table 12.

While a binary definition of ethnic enclaves has some intuitive advantages, there is some evidence to suggest that using a binary dependent variable specification ignores considerable heterogeneity in outcomes. Specifically, there is great heterogeneity in the extent to which different ethnic groups appear to sort themselves into enclaves [Toussaint-Comeau, 2008], which implies that the set of locations that meet a binary threshold E_{jk} based on a fraction of the national average density may be a more or less selected subset of locations for different ethnic groups. To address this concern, I also construct a continuous measure of ethnic density using the inverse hyperbolic sine transformation.[8] Define:

$$D_{jk} \quad = \quad \ln\left(Density_{jk} + \left(Density_{jk}^2 + 1\right)^{1/2}\right) \tag{13}$$

Then, I run OLS regressions, fitting the model:

$$D_{ijk} = \alpha_0 + \alpha_1 Offer_i + \alpha_2 X_i + \delta_k + \eta_j + \nu_i \tag{14}$$

As shown in Section 6, the results of the OLS and probit regressions are broadly similar.

5 Data

5.1 The New Immigrant Survey

The New Immigrant Survey (NIS), administered by by the Princeton University Office of Population Research, is designed to be a nationally-representative multi-cohort longitudinal study of new legal immigrants to the United States. The first cohort of immigrants were sampled using administrative data on new migrants from May to November of 2003, and a baseline survey was conducted between June 2003 and June 2004. In the initial cohort, 12,500 adults were sampled, along with 1,250 children. Of these, there are 8,573 completed interviews, for a response rate of 68.6%. Since the sample is designed to be nationally representative, it includes a sample from each of the top 85 MSAs, each of the top 38 counties, and a random sample of 10 other MSAs and 15 other counties. A follow-up survey of the first wave of immigrants has been conducted and recently released for research use, but data from this follow-up survey is not used in this analysis.[9]

[8] MacKinnon and Magee [1990] show that the inverse hyperbolic sine transformation has similar properties to the logarithmic transformation, except that it is defined at 0. However, my main results are not substantially affected by using the log of ethnic density instead, as Bertrand et al. [2000] and others have done.

[9] See http://nis.princeton.edu/project.html and http://nis.princeton.edu/overview.html for further information.

Each adult immigrant is asked a series of questions about each job that he/she reports having. This includes questions about about occupation, salary, benefits, hours worked, etc. However, most critically for this analysis, I code the variable $Offer_i$ based on whether immigrant i responds yes to the question "Had you been offered this job before coming to the United States to live?" for any of his/her reported jobs. The survey also asks questions such as "Did you get this job with the help of a relative?" and "Do any of your relatives work for this business?" I use these measures as additional covariates in some regression specifications, as they are observable measures of potentially endogenous pre-existing networks. In the Appendix, Table 11 provides some information about the questions that I use to code variables from the NIS.

A key feature of the NIS is that it provides an exceedingly broad retrospective survey. For example, in addition to data on current employment status, the NIS asks for detailed data on such things as pre-migration industries and occupations, pre-migration wages, prior migration history, education, language skills, and social networks. While the Public Use version of the NIS contains limited geographic information, the NIS collected data on immigrants' location at the time of the interview at the ZIP code level, and I match this data to U.S. Census data at the ZIP code level to calculate local ethnic density and identify ethnic enclaves.

The key empirical challenge of attempting to do this type of analysis with a representative sample of the population is that, in general, new immigrants are neither exogenously offered jobs, nor are they exogenously located in ethnic enclaves. Yet, any number of characteristics may be correlated with both of these measures, and these are a potential source of omitted variable bias. For observable characteristics, the detailed retrospective data contained in the NIS provides clear advantages, as it allows me to control for such factors as pre-migration wages. The effects of unobservable characteristics, however, remain of concern. Like other theoretical models of migration behavior, [see, e.g. Borjas, 1987, Borjas and Friedberg, 2009], my theoretical model shows that selection on unobservable characteristics is likely to occur. Yet, the instrumental variables methods in the immigration literature either cannot account for endogenous sorting across metropolitan areas at the time of migration [Bertrand et al., 2000], or their validity is limited to a small subset of the immigrant population [Munshi, 2003].

The selection on observables identification strategy of this paper, then, rests on the breadth of the NIS dataset to account for variables that could otherwise be sources of omitted variable bias. In exchange for this limitation, this paper provides potentially broader external validity and more precise geographic variation than previous research. As shown in Section 6, using NIS covariates I am able to control for pre-existing family networks, occupational/industry sorting, limited measures of cultural attachment, and differences in leisure preferences that yield differences in pre-migration hours or salaries. I have also undetertaken an analysis using the method of propensity score matching, which may be somewhat more robust to selection on unobservables Angrist and Pischke [2008]. Further discussion of the potential benefits of this method, as well as results, follow in Section 7.

5.2 Data on Ethnic Density

Although the NIS sample is designed to be nationally representative, it is relatively small and not suitable for calculating the ethnic density of geographic areas. Therefore, in this paper, I use Census Bureau data from the SF-3 sample of the 2000 Decennial Census to measure the ethnic density of each ZIP code. Specifically, I consider ethnic groups based on language spoken at home, which is available in both the NIS and in Census data. I then define both a continuous measure D_{jk} and a binary measure E_{jk} for each location-language pair based as described in Section 4. Table 12 provides some summary information about the manner in which I define ethnic group membership and density, while Table 13 provides a list of language groups included in the model, along with within-group counts.

6 Baseline Results

6.1 Initial Analysis

I begin my analysis of the new immigrants by looking at overall characteristics of the immigrants in the NIS, and in particular by looking for potential differences in the characteristics of immigrants who arrive with a job offer versus immigrants who do not arrive with such an offer. Table 1 presents summary statistics on a variety of covariates used in my regression specifications, including means and sample standard errors. As the table shows, immigrants who arrive with a job offer do in fact differ on a number of key characteristics. Most noticeably, they are substantially more likely to be highly educated; nearly 80% of immigrants who arrive with a job offer have at least a Bachelors degree, while only about 40% of immigrants who arrive without an offer have that much education.

Immigrants who arrive with jobs are also somewhat less likely to have gotten their current job with the help of a relative. This may suggest that immigrants who do not arrive with a job offer may be more reliant on informal network ties even before migration than immigrants who arrive with an offer, and therefore their decision to locate in an enclave may be the result of existing network ties rather than the potential of an enclave to provide them with new ties. I include this variable, as well as the variable on whether a relative works at the employer, as covariates in my model. The inclusion of these covariates may not make my model robust to the impact of pre-existing network ties connecting non-relatives. However, since the majority of visas granted to new legal immigrants in the United States are awarded based on some sort of family-related sponsorship, and yet my regressions do not find evidence that family relationships impact the locational decision, I argue that the impact of these networks on the likelihood of having a pre-migration offer is likely to be small.

Table 1 also shows summary information on income for the portion of the sample that reports income. All incomes have been converted by the NIS to 2003 U.S. dollars, using PPP where applicable. However, this leads to some individuals reporting very high or very low incomes. Therefore, individuals with incomes

Table 1: Summary Statistics for New Immigrants

		Offered Job Prior To Move			Differences	
	All	Yes	No	Missing	Yes - No	Yes - (No ∪ Miss)
General Characteristics						
Less Than HS	0.281	0.066	0.241	0.353	-0.176	-0.232
	(0.005)	(0.010)	(0.007)	(0.008)	(0.018)	(0.019)
High School	0.227	0.093	0.244	0.232	-0.151	-0.145
	(0.005)	(0.012)	(0.007)	(0.007)	(0.018)	(0.017)
Some College	0.094	0.053	0.109	0.086	-0.056	-0.044
	(0.003)	(0.009)	(0.005)	(0.004)	(0.013)	(0.012)
Bachelors Degree	0.243	0.485	0.234	0.215	0.251	0.260
	(0.005)	(0.020)	(0.007)	(0.006)	(0.019)	(0.018)
Graduate School	0.152	0.302	0.170	0.113	0.133	0.162
	(0.004)	(0.018)	(0.006)	(0.005)	(0.017)	(0.015)
Years of Schooling	12.698	15.997	13.325	11.581	2.672	3.558
	(0.056)	(0.149)	(0.077)	(0.084)	(0.202)	(0.211)
Years of Schooling in U.S.	0.795	0.361	1.157	0.501	-0.797	-0.469
	(0.024)	(0.051)	(0.041)	(0.030)	(0.105)	(0.092)
Female	0.518	0.376	0.405	0.649	-0.029	-0.153
	(0.005)	(0.019)	(0.008)	(0.008)	(0.021)	(0.021)
Year Born	1964	1967	1967	1960	-1	3
	(0.146)	(0.347)	(0.158)	(0.252)	(0.418)	(0.561)
Helped By Relative To Get Job	0.167	0.139	0.174	0.093	-0.035	-0.032
	(0.005)	(0.014)	(0.006)	(0.024)	(0.016)	(0.016)
Relative Works For Company	0.115	0.096	0.114	0.144	-0.018	-0.022
	(0.004)	(0.012)	(0.005)	(0.015)	(0.014)	(0.014)
Offered Job Prior To Move	0.138					
	(0.005)					
N	8,573	625	3,906	4,042		
Salary Information (Where Available)						
First Post-Migration Job	22,879	50,365	19,182	17,230	27,618	31,849
	(674)	(2268)	(707)	(1722)	(2137)	(1571)
	[3,085]	[399]	[2,155]	[531]		
Last Pre-Migration Job	20,418	32,486	18,264	19,485	14,222	13,639
	(831)	(3264)	(1066)	(1297)	(2686)	(2592)
	[3,040]	[350]	[1,404]	[1,286]		
Current	29,147	59,461	24,763	18,498	25,911	35,202
	(756)	(3,894)	(600)	(3,983)	(1,736)	(1,290)
	[3,553]	[470]	[2,966]	[117]		

Notes: Standard errors in parentheses. Counts for salary data in brackets. All salaries are reported in 2003 U.S. dollars; conversions are made in PPP terms where applicable. Salary averages and counts exclude individuals with reported annual incomes less than $100 or greater than $1,000,000.

below \$100 and above \$1,000,000 have been removed from the sample (and they have been removed whenever income is a covariate elsewhere in my analysis as well).[10] Individuals who arrive with a job offer do have incomes that are considerably higher than those of individuals who arrive without a job offer, both pre-migration and post-migration. However, notably, the sample suggests that individuals who arrive with a job offer see, on average, their incomes increase by approximately 55% upon migration. The increase in income for individuals who arrive without a job offer, however, is not statistically significant.

In the Appendix, I provide additional tables showing additional frequency statistics on who receives a pre-migration job offer and who does not. In general, the likelihood of receiving a pre-migration offer is quite heterogeneous across occupations and industries.

6.2 Primary Regression Analysis

Table 2 and Table 3 show my initial regression results using the basic specifications shown in Equations 10 and 14, respectively Table 2 shows the results of probit regression of my binary measure E_{jk} on whether one received a pre-migration offer, as well as a variety of covariates. For reference, the estimated average marginal effect of the indicator $Offer_i$ is also shown at the bottom of the table. Table 3 shows OLS regression results on my continuous measure of enclave density D_{jk}, with similar covariate specifications. For each of these regressions, language group k is assigned using the first non-English language reported to be spoken at home, though the use of the other measures described in Table 12 gives me very similar results. All of these regressions include a full set of language group and geographic area fixed effects, and they are two-way clustered at the MSA by language-group level to be robust to correlated shocks within language groups within an MSA. The estimated effect of having a job offer at the time of migration is consistently negative and significant, implying that immigrants who have received a job offer are in fact less likely to locate in an ethnic enclave. Additionally, the estimated magnitude of this effect varies little when one controls for standard demographic covariates and for measures of family-related job networks. Notably, if one controls for pre-migration salary, hours, occupation and industry as shown in specifications (4) and (5), the magnitude of the effect does not diminish at all. This suggests that the apparent effect of a job offer is unlikely to be driven solely by selection on the skill levels of those that receive a job offer. These regressions suggest that a new immigrant who arrives with a job offer is roughly 8.9% to 10.8% less likely to locate in an ethnic enclave than an immigrant who arrives without a job offer.

These regression results also show that there is a strong negative relationship between one's level of schooling and residence in an enclave, even after controlling for job offers and other covariates. Additionally, individuals who have children appear to be significantly more likely to locate in an enclave.[11] Individuals

[10] All reported incomes have been converted to annual incomes based on reported usual hour/week and usual weeks/year worked. A small number of immigrants report that they are paid daily or per-unit; these individuals have been dropped from the analysis of income.

[11] This binary measure indicates whether the respondent reports that they live with one or more of their own children under the age of 18.

Table 2: Primary regression results: Assignment based on density of first non-English language

	Binary Enclave Indicator (Probit)				
	(1)	(2)	(3)	(4)	(5)
Offer at Migration	-0.268**	-0.256**	-0.254**	-0.336***	-0.349***
	(0.127)	(0.123)	(0.125)	(0.129)	(0.127)
Years School		-0.0177**	-0.0181**	-0.0169	0.00264
		(0.00868)	(0.00808)	(0.0213)	(0.0232)
Age		-0.000722	-0.000809	-0.00462	-0.00703**
		(0.00265)	(0.00256)	(0.00299)	(0.00333)
Female		-0.0525	-0.0537	-0.107	-0.130
		(0.0428)	(0.0428)	(0.107)	(0.109)
Married		-0.0194	-0.0200	-0.0793	-0.0766
		(0.0755)	(0.0717)	(0.149)	(0.162)
Separated		0.0838	0.0831	-0.120	-0.126
		(0.0620)	(0.0603)	(0.320)	(0.342)
Has Children		0.0421	0.0410	0.0340	0.0600
		(0.0478)	(0.0479)	(0.0870)	(0.0861)
Got Job Relative			0.00456	-0.0101	-0.0936
			(0.0913)	(0.0998)	(0.125)
Relative at Emp.			-0.0690	-0.0181	-0.0276
			(0.0578)	(0.137)	(0.128)
Log Pre-Migration Salary				-0.0269	-0.0232
				(0.0266)	(0.0309)
Pre-Migration Hours				0.00156	0.000852
				(0.00194)	(0.00228)
Poor English					0.497***
					(0.0918)
Religious					0.184*
					(0.0970)
Poor Health					-0.101
					(0.437)
Occ. & Industry	No	No	No	Yes	Yes
Observations	3,389	3,372	3,372	1,219	1,219
Avg. Marginal Effect	-0.0889	-0.0847	-0.0839	-0.105	-0.108
	(0.0427)	(0.0413)	(0.0420)	(0.0406)	(0.0393)

Notes: Standard errors in parentheses, clustered at MSA × language-group. All regressions include a full set of MSA and language-group fixed effects. Log salary is measured in U.S. 2003 dollars.
*** p¡0.01, ** p¡0.05, * p¡0.1

Table 3: Primary regression results (OLS Specification): Assignment based on density of first non-English language

	Continuous Enclave Indicator (OLS)				
	(1)	(2)	(3)	(4)	(5)
Offer at Migration	-0.241**	-0.236**	-0.235**	-0.204***	-0.211***
	(0.0995)	(0.0938)	(0.0957)	(0.0622)	(0.0644)
Years School		-0.0168***	-0.0171***	-0.0206*	-0.00851
		(0.00391)	(0.00373)	(0.0124)	(0.0134)
Age		8.66e-05	3.69e-05	0.00219	0.000595
		(0.00174)	(0.00169)	(0.00340)	(0.00288)
Female		-0.00809	-0.00892	-0.0440	-0.0533
				(0.0576)	(0.0567)
Married		0.00645	0.00602	0.0155	0.0196
		(0.0579)	(0.0574)	(0.115)	(0.118)
Separated		0.0980	0.0970	0.142	0.144
		(0.102)	(0.101)	(0.245)	(0.255)
Has Children		0.0784***	0.0776***	0.0615	0.0725
		(0.0243)	(0.0248)	(0.0672)	(0.0664)
Got Job Relative			0.00436	0.0404	-0.0105
			(0.0617)	(0.0747)	(0.0776)
Relative at Emp.			-0.0478	-0.0971	-0.103
			(0.0346)	(0.116)	(0.113)
Log Pre-Migration Salary				-0.0408***	-0.0355**
				(0.0134)	(0.0139)
Pre-Migration Hours				-0.00233*	-0.00282**
				(0.00123)	(0.00124)
Poor English					0.327***
					(0.0865)
Religious					0.0684
					(0.0863)
Poor Health					0.0300
					(0.218)
Occ. & Industry	No	No	No	Yes	Yes
Observations	3,507	3,487	3,487	1,311	1,311
R-squared	0.337	0.343	0.343	0.417	0.427

Notes: Standard errors in parentheses, clustered at MSA × language-group. All regressions include a full set of MSA and language-group fixed effects. Log salary is measured in U.S. 2003 dollars.
*** p¡0.01, ** p¡0.05, * p¡0.1

who report that they speak English or understand English at a poor level are considerably more likely to locate in enclaves. Finally, individuals who report relatively low pre-migration salaries and low pre-migration average hours of work are significantly more likely to locate in an enclave. These coefficients are all consistent with other explanations that have been proposed for why immigrants locate in enclaves. In particular, they suggest that the ability to assimilate, leisure preferences, or other preferences may all play a substantial role in new immigrants' locational decisions. Yet, the significance and robustness of the effect of a job offer suggests that job market network effects are important as well.

Finally, it is noteworthy in my results in Tables 2 and 3 that my measures of family-based networks, whether one got a job from his/her relative and whether one's relative works for the company, do not appear to be significantly related to the decision to locate in an ethnic enclave once other covariates and the reception of a pre-migration job offer have been controlled for. Although I am unable to control for pre-existing non-familial networks as a source of selection on who receives a job offer, this suggests that unobserved networks may be unlikely to a major source of bias in these estimates.

Although the theoretical model shown in Section 3 describes the locational choice using a single ethnic density paramets, in the real world, locational decisions are slightly more complicated. In practice, one can think of an immigrant's decision of where to locate as having two main components: "In which metropolitan area should I locate?" and "Where within this metropolitan area should I locate?" These choices are made simultaneously. However, if individuals choose where to migrate based on labor market networks, then one should be able to empirically identify that individuals with a job offer locate in less ethnically-dense neighborhoods, even within the same metropolitan area. Table 4 repeats the baseline OLS regressions, but instead of including a full set of language group fixed effects plus MSA effects, this analysis includes a fixed effect for each language-group $\times$ MSA combination. Thus, these regressions are identified from only the within-MSA variation in which neighborhoods individuals choose. These results are clustered one-way at the MSA $\times$ language-group level.

As Table 4 shows, even looking within metropolitan areas, individuals who arrived in the U.S. with a job offer locate in less ethnically dense areas. However, the magnitude of these estimates is also smaller than the corresponding baseline estimates in Table 3, because some portion of the effect of a job offer is the result of those who arrive with a job offer choosing to live in different metropolitan areas altogether from those who arrive without an offer. The coefficients on all other covariates are also broadly similar to the baseline regressions. Overall, this provides additional supportive evidence that labor market effects are a significant factor in immigrants' locational decisions.

Next, we will consider the possibility of heterogeneous effects in migration. The theoretical model suggests that individuals who have a high observable type in the receiving country will be less likely to locate in an ethnic enclave because they receive job offers at a relatively higher rate regardless of their location. While the baseline regression results show that more educated individuals are less likely to locate in an enclave, they do not allow for the possibility of heterogeneous effects. Tables 5 and 6 repeat the baseline specification,

Table 4: Primary regression (OLS Specification) Based on Within-MSA Variation Only

	Continuous Language Density (OLS)				
	(1)	(2)	(3)	(4)	(5)
Offer at Migration	-0.110*	-0.113*	-0.111*	-0.101	-0.106
	(0.0627)	(0.0618)	(0.0616)	(0.108)	(0.109)
Years School		-0.0112*	-0.0115*	-0.0179	-0.00542
		(0.00599)	(0.00616)	(0.0109)	(0.0113)
Age		-0.000788	-0.000883	-0.000912	-0.00182
		(0.00196)	(0.00198)	(0.00402)	(0.00411)
Female		-0.0130	-0.0143	-0.0430	-0.0412
		(0.0400)	(0.0401)	(0.0931)	(0.0915)
Married		-0.0141	-0.0145	-0.0432	-0.0467
		(0.0458)	(0.0449)	(0.121)	(0.119)
Separated		0.0386	0.0381	0.0225	0.00515
		(0.0773)	(0.0767)	(0.206)	(0.203)
Has Children		0.0824**	0.0818**	0.104	0.123
		(0.0410)	(0.0410)	(0.0810)	(0.0770)
Got Job Relative			0.0116	0.108	0.0543
			(0.0516)	(0.0999)	(0.0984)
Relative at Emp.			-0.0723	-0.119	-0.114
			(0.0510)	(0.170)	(0.176)
Log Pre-Migration Salary				-0.0371	-0.0355
				(0.0261)	(0.0256)
Pre-Migration Hours				-0.00292	-0.00343
				(0.00266)	(0.00268)
Poor English					0.345***
					(0.113)
Religious					-0.0189
					(0.0718)
Poor Health					-0.171
					(0.255)
Occ. & Industry	No	No	No	Yes	Yes
Observations	3,507	3,487	3,487	1,311	1,311
R-squared	0.557	0.561	0.561	0.629	0.637

Notes: Standard errors in parentheses, clustered at MSA × language-group (one-way clustering). All regressions include a full set of MSA, language-group, and MSA × language-group fixed effects. Log Salary is measured in U.S. 2003 dollars.
*** p¡0.01, ** p¡0.05, * p¡0.1

Table 5: Regression with Heterogeneous Effects Based on Schooling (Probit)

	Binary Enclave Indicator (Probit)				
	(1)	(2)	(3)	(4)	(5)
Offer at Migration	-0.275**	-0.279**	-0.277**	-0.413***	-0.417***
	(0.116)	(0.113)	(0.115)	(0.135)	(0.143)
Offer × School (Demeaned)	0.0125	0.0121	0.0123	0.0386	0.0343
	(0.0197)	(0.0198)	(0.0189)	(0.0338)	(0.0341)
Years School (Lang. Group Demeaned)	-0.0186**	-0.0187**	-0.0191***	-0.0235	-0.00343
	(0.00750)	(0.00778)	(0.00738)	(0.0261)	(0.0281)
Age		-0.000793	-0.000885	-0.00496*	-0.00733**
		(0.00270)	(0.00261)	(0.00285)	(0.00317)
Female		-0.0518	-0.0530	-0.103	-0.126
		(0.0420)	(0.0421)	(0.105)	(0.107)
Married		-0.0194	-0.0198	-0.0816	-0.0790
		(0.0752)	(0.0714)	(0.151)	(0.162)
Separated		0.0843	0.0838	-0.118	-0.123
		(0.0614)	(0.0597)	(0.316)	(0.338)
Has Children		0.0417	0.0406	0.0326	0.0587
		(0.0479)	(0.0480)	(0.0878)	(0.0868)
Got Job Relative			0.00699	-0.00588	-0.0901
			(0.0889)	(0.100)	(0.125)
Relative at Emp.			-0.0703	-0.0143	-0.0235
			(0.0575)	(0.142)	(0.133)
Log Pre-Migration Salary				-0.0282	-0.0242
				(0.0275)	(0.0316)
Pre-Migration Hours				0.00159	0.000887
				(0.00186)	(0.00222)
Poor English					0.492***
					(0.0923)
Religious					0.183*
					(0.0979)
Poor Health					-0.0888
					(0.440)
Occ. & Industry	No	No	No	Yes	Yes
Observations	3,381	3,372	3,372	1,219	1,219
Avg. Marginal Effect	-0.0870	-0.0884	-0.0876	-0.110	-0.112
	(0.0403)	(0.0395)	(0.0401)	(0.0391)	(0.0387)

Notes: Standard errors in parentheses, clustered at MSA × language-group. All regressions include a full set of MSA and language-group fixed effects. Log Salary is measured in U.S. 2003 dollars.Years of Schooling is actual years of schooling less the mean for each individual's language group.
*** p¡0.01, ** p¡0.05, * p¡0.1

Table 6: Regression with Heterogeneous Effects Based on Schooling (OLS)

	Continuous Language Density (OLS)				
	(1)	(2)	(3)	(4)	(5)
Offer at Migration	-0.224**	-0.233**	-0.232**	-0.234**	-0.234**
	(0.105)	(0.102)	(0.103)	(0.0950)	(0.0963)
Offer × School (Demeaned)	-0.000946	-0.00138	-0.00138	0.0146	0.0115
	(0.0189)	(0.0188)	(0.0181)	(0.0261)	(0.0264)
Years School (Lang. Group Demeaned)	-0.0171***	-0.0167***	-0.0170***	-0.0230*	-0.0104
	(0.00363)	(0.00382)	(0.00384)	(0.0124)	(0.0136)
Age		9.58e-05	4.63e-05	0.00208	0.000514
		(0.00177)	(0.00173)	(0.00349)	(0.00299)
Female		-0.00819	-0.00901	-0.0421	-0.0519
				(0.0593)	(0.0581)
Married		0.00643	0.00599	0.0147	0.0190
		(0.0580)	(0.0576)	(0.115)	(0.118)
Separated		0.0979	0.0970	0.141	0.144
		(0.103)	(0.102)	(0.243)	(0.254)
Has Children		0.0784***	0.0777***	0.0608	0.0719
		(0.0244)	(0.0249)	(0.0677)	(0.0670)
Got Job Relative			0.00412	0.0421	-0.00888
			(0.0593)	(0.0739)	(0.0773)
Relative at Emp.			-0.0477	-0.0967	-0.103
			(0.0340)	(0.117)	(0.114)
Log Pre-Migration Salary				-0.0412***	-0.0357***
				(0.0128)	(0.0133)
Pre-Migration Hours				-0.00231*	-0.00280**
				(0.00126)	(0.00129)
Poor English					0.326***
					(0.0886)
Religious					0.0681
					(0.0865)
Poor Health					0.0326
					(0.221)
Occ. & Industry	No	No	No	Yes	Yes
Observations	3,498	3,487	3,487	1,311	1,311
R-squared	0.342	0.343	0.343	0.418	0.427

Notes: Standard errors in parentheses, clustered at MSA × language-group. All regressions include a full set of MSA and language-group fixed effects. Log Salary is measured in U.S. 2003 dollars.Years of Schooling is actual years of schooling less the mean for each individual's language group.
*** p¡0.01, ** p¡0.05, * p¡0.1

but they include an interaction between the reception of a pre-migration job offer and an individual's years of schooling.[12] As discussed in Section 3, under some plausible assumptions, the effect of a job offer on the location decision will be smaller for individuals with more education, because high-skilled immigrants who arrive without a job offer will receive an offer more quickly on average than low-skilled immigrants who arrive without a job offer. However, my empirical results do not appear to support this prediction. The interaction effect of an additional year of schooling is not statistically significant, which suggests that heterogeneity is not present.[13] Yet, when this interaction effect is included, my estimate of the average marginal effect of having an offer on an immigrant with an average level of education is virtually unchanged; an immigrant who receives a job offer before migrating is roughly 8.7% to 11.2% less likely to locate in an ethnic enclave.

7 Robustness Checks

The results shown in Tables 2 and 3 are highly suggestive, but there are a number of potential concerns with the interpretation of these results. For one, immigrants who arrive to the United States legally and who already have a job offer at the time of migration are likely to have arrived through an employer-sponsored visa. The process by which one receives this visa is different than the process for other visa classifications because employers typically have to show that an immigrant provides particular skills that are in demand and not that are otherwise available to the firm. In fact, the majority of immigrants who receive new legal resident status receive it due to family-sponsored preferences or because they are the direct relatives of existing U.S. citizens. [DHS, 2012] To the extent that these employer-sponsored individuals may be different from non-sponsored individuals in unobservable ways (such as unobserved differences in underlying preferences), they may bias our estimates of labor market effects.

Table 7: Frequencies of Offer At Migration vs. Employee Sponsorship in the NIS

	Employee Sponsorship		
Had at Migration	No	Yes	Total
No	2,668	1,238	3,906
Yes	89	536	625
Total	2,757	1,774	4,531

Source: New Immigrant Survey.

One potential solution to this issue would be to run regressions on a sample consisting of only those

[12]Here, each individual's years of schooling is subtracted from the mean for his/her language group. Results are similar whether one uses this measure, a "raw" measure of years of schooling, or a set of discrete indicators for different levels of educational attainment.

[13]In other regressions which are not shown here, I have tested for heterogeneous effects using interactions with English-language ability, as well using interactions with the full set household characteristics included in the baseline regressions. I find no significant evidence of heterogeneous effects along any of these dimensions.

individuals who were not employer-sponsored. Then, the effect of a job offer would reflect the impact of a job offer only for those who are not observably. However, as Table 7 shows, the NIS is too small to have much power in conducting such an analysis. Only 89 individuals in the sample had an offer at migration but were not employer sponsored, and so the baseline results are insignificant when run on this subsample.

Table 8: Regression with Interaction Based on Employer Sponsorship (Probit)

	(1)	(2)	(3)	(4)	(5)
				Binary Enclave Indicator (Probit)	
Employer Sponsored	-0.0819	-0.0722	-0.114*	-0.295***	-0.258**
	(0.0519)	(0.0514)	(0.0683)	(0.109)	(0.123)
Years School		-0.0162**	-0.0176**	-0.0177	0.00111
		(0.00757)	(0.00866)	(0.0217)	(0.0233)
Age		-0.00139	-0.000818	-0.00458*	-0.00673**
		(0.000906)	(0.00267)	(0.00264)	(0.00303)
Female		-0.0599	-0.0612	-0.131	-0.154
		(0.0555)	(0.0439)	(0.114)	(0.117)
Married		-0.0195	-0.00684	-0.0932	-0.0894
		(0.0448)	(0.0705)	(0.145)	(0.158)
Separated		0.0191	0.119	-0.105	-0.106
		(0.0392)	(0.0788)	(0.321)	(0.345)
Has Children		0.0465***	0.0333	0.0312	0.0526
		(0.0164)	(0.0474)	(0.0903)	(0.0907)
Got Job Relative			0.00254	-0.0425	-0.129
			(0.0849)	(0.0968)	(0.123)
Relative at Emp.			-0.0855*	-0.0490	-0.0492
			(0.0505)	(0.135)	(0.125)
Log Pre-Migration Salary				-0.0276	-0.0234
				(0.0250)	(0.0292)
Pre-Migration Hours				0.000606	-0.000100
				(0.00199)	(0.00239)
Occ. & Industry	No	No	No	Yes	Yes
Social Measures	No	No	No	No	Yes
Observations	6,388	6,349	3,413	1,231	1,231
Avg. Marginal Effect	-0.0264	-0.0232	-0.0369	-0.0911	-0.0784
	(0.0168)	(0.0167)	(0.0227)	(0.0334)	(0.0375)

Notes: Standard errors in parentheses, clustered at state/Census division × language-group. All regressions include a full set of state/Census division and language-group fixed effects. Log Salary is measured in U.S. 2003 dollars.
*** p¡0.01, ** p¡0.05, * p¡0.1

An alternative strategy, which I show in Table 8, is to re-run the baseline regressions with an indicator for employer sponsorship instead of an indicator for receiving a job offer. These results are very similar to the baseline regression results. In particular, the effects on other covariates change little relative to the baseline analysis, which we might expect if employer sponsorship is highly correlated with these other covariates. Thus, while it remains possible that employer-sponsored individuals are systematically different from non-sponsored individuals, there is no strong evidence to suggest that these individuals are a source of bias in the baseline estimates.

7.1 Propensity Score Analysis

As discussed above, the types of individuals who receive job offers are on average more highly educated and with higher incomes than individuals who do not receive offers. Large differences in average observable covariates may complicate the usual concerns about endogeneity in regression results. For example, when the true effects of observable covariates on an outcome are nonlinear, then the endogeneity bias from omitted variables that are correlated with these covariates will be magnified.

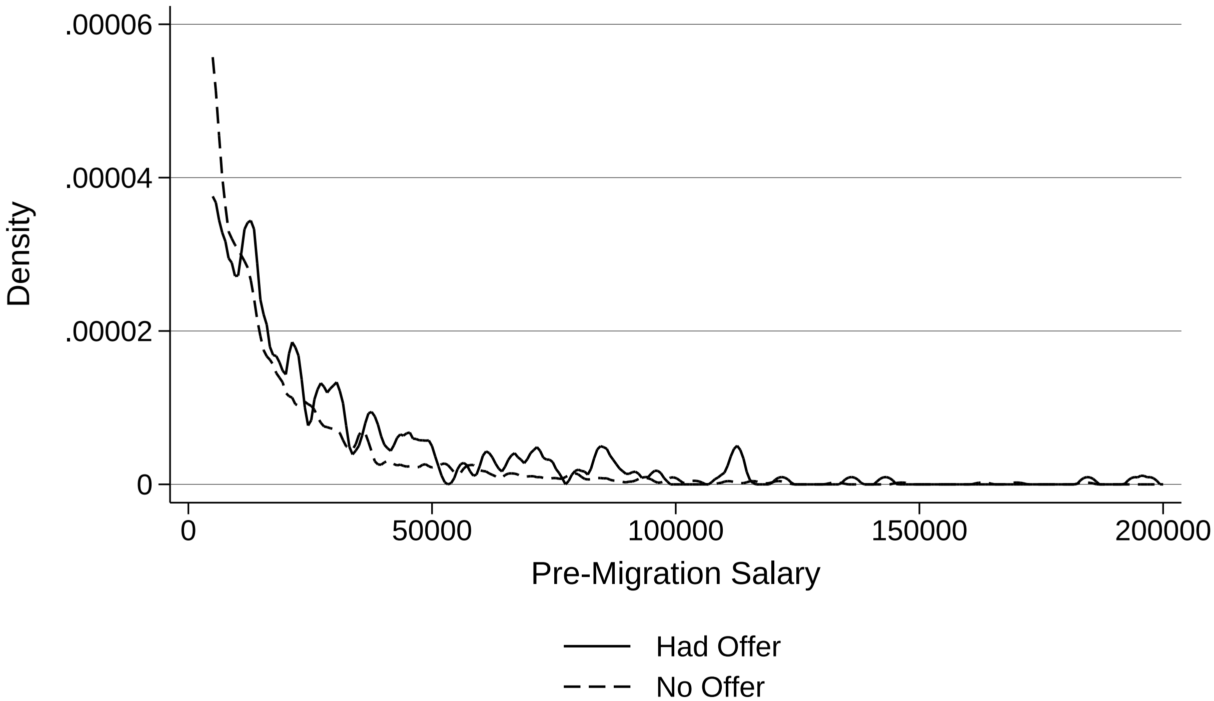
The method of propensity score matching is a close relative of traditional regression analysis that may be beneficial in situations where these concerns exist. More properly, OLS can be understood as a matching estimator with a particular weighting scheme [Angrist and Pischke, 2008]. In essence, propensity score matching creates a unique set of control observations for each treatment observation (treatment being defined in this case as the receipt of a job offer before migration). This puts the highest weight in analysis on individuals with the highest probability of treatment. Ordinary Least Squares, in contrast, puts the highest weight on individuals with the highest variance in regressors.

Some analysis, especially in the literature on job training programs, has suggested that the method of propensity score matching may be beneficial in situations where the broader population differs from the treatment group [Dehejia and Wahba, 2002]. In the job training case, endogeneity bias arises from the fact that individuals seeks job training are disproportionately likely to have received an unobservable negative income shock in the recent past. In my model, the potential for endogeneity arises from the possibility that pre-migration immigrant networks are simply less strong for the types of immigrants who tend to receive pre-migration job offers, or that immigrants who receive pre-migration job offers have systematically different preferences from those who do not. Thus, in this subsection I undertake a preliminary application of propensity score matching estimators to identify the average treatment effect of receiving a job offer prior to migration on the decision to locate in an ethnic enclave.

A key requirement of propensity score matching methods is that there be common support over the distributions of covariates. If the assumption of common support is violated, then the set of suitable control observations for each treatment observation will be small or nonexistent, causing these observations to be dropped from the analysis. Figure 1 and Figure 2 provide kernel density estimates of the pre-migration salaries and total current salaries of individuals who received a pre-migration offer and those who did not. Figure 3 provide a dual bar chart showing the density of the distribution of years of schooling for individuals who received a pre-migration offer and those who did not. As expected from the results in Table 1, the distributions of these variables are not identical, and individuals who received a pre-migration offer have more schooling and higher salaries on average. However, it appears from these graphs that the assumption of common support is generally met.

Figures 4 and 5 show histograms of the propensity scores produced by my initial probit regressions of receiving a pre-migration job offer or having an employer sponsored visa on the covariates reported in Table

Figure 1: Kernel Density Estimation of Pre-Migration Salary

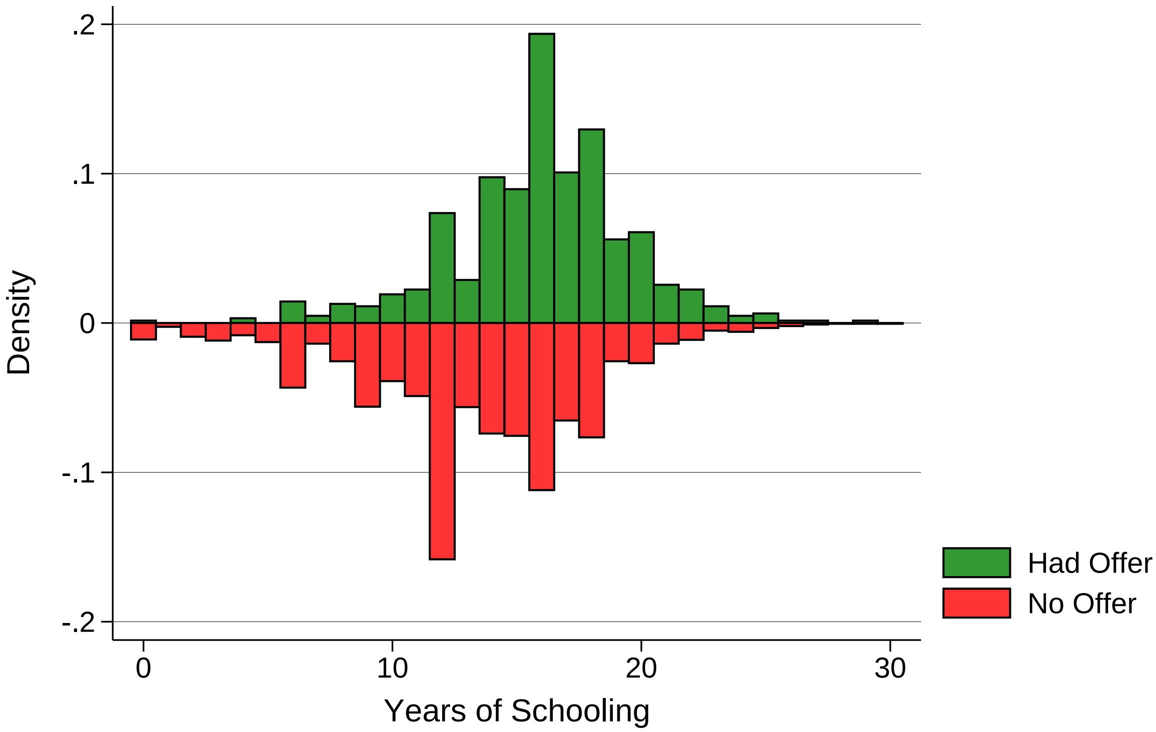


Kernel density plot using an Epanechnikov kernel with bandwidth of $1,000. Excludes individuals earning less than $5,000 or more than $200,000.

Figure 2: Kernel Density Estimation of Current Salary

Kernel density plot using an Epanechnikov kernel with bandwidth of $1,000. Excludes individuals earning less than $5,000 or more than $200,000.

Figure 3: Distribution of Years of Schooling



Excludes individuals reporting more than 30 years of school.

2, as well as the square of the continuous variables years of school, age, and pre-migration income.[14] Again, it does appear that the assumption of common support is met, but the distribution of propensity scores for individuals who did not receive a job offer is comparatively skewed to the less in both cases and is relatively less dense at high propensity scores. While this does suggest that the method of propensity score matching may lead to improved estimates in this case, it also suggests that the power of propensity score estimates may be relatively low.

Table 9 shows the results of propensity score matching methodology, using an Epanechnikov kernel matching method with default bandwidth assumptions. Because the standard errors produced by propensity score estimates make strong assumptions about homoskedasticity, I provide bootstrapped standard errors for these propensity score estimates. In general, my preliminary propensity score estimates are reasonably similar in magnitude to the estimates of average marginal effects in my initial probit and OLS regressions, and in my regressions based with an employer sponsorship interaction. However, the standard errors of these estimates are larger, and so these results are in general no longer statistically significant at standard levels of confidence. Once I have access to more disaggregated geographic data, I expect that further analysis will

[14]The inclusion of these polynomial terms is suggested by Angrist and Pischke [2008] and employed by Dehejia and Wahba [2002].

Figure 4: Comparison of Propensity Scores for Pre-Migration Job Offer

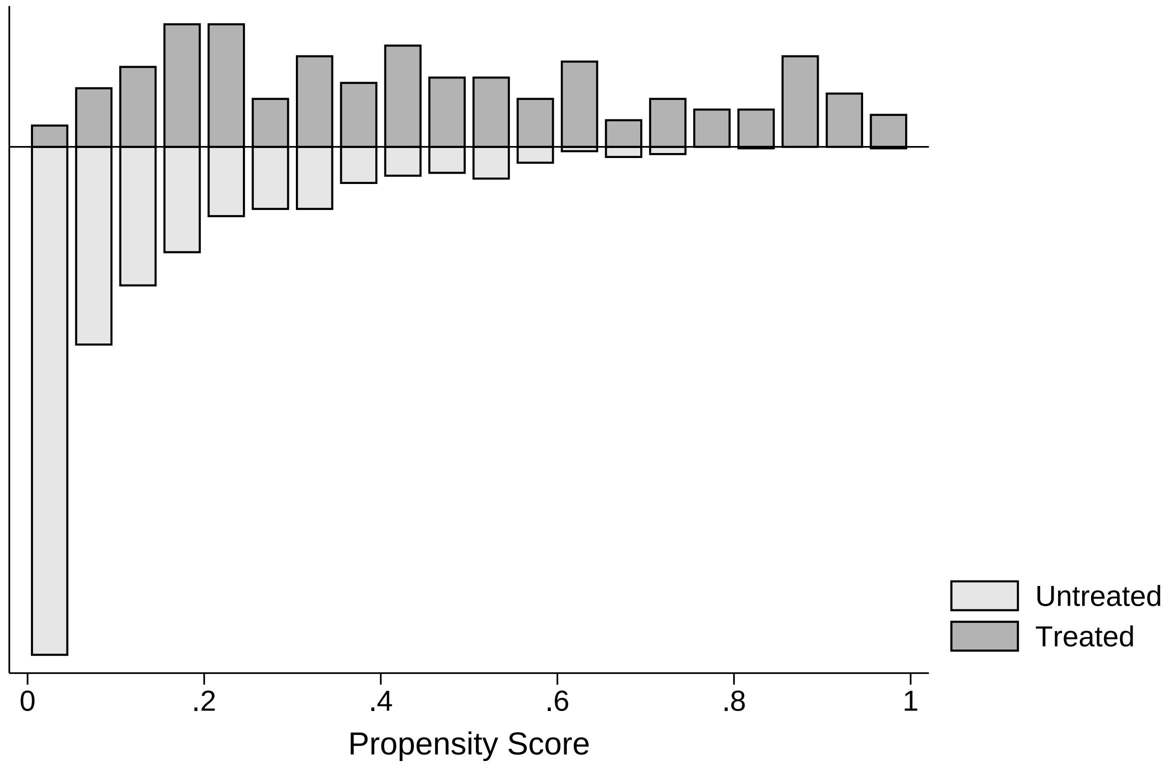


Figure 5: Comparison of Propensity Scores for Employer Sponsorship

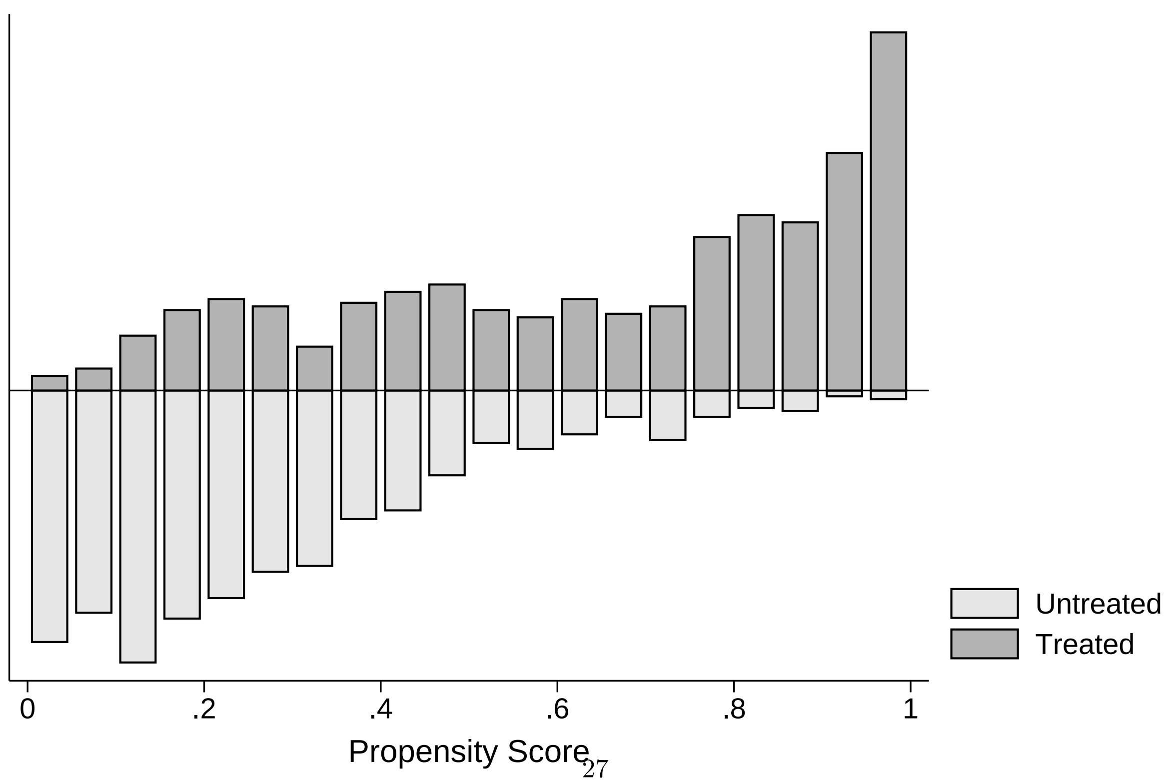


Table 9: Estimated Avg. Treatment Effects Using Propensity Score Analysis

	(1)	(2)	(3)	(4)
	Offer At Migration		Employer Sponsored	
Outcome Measure:	Binary	Continuous	Binary	Continuous
Avg. Treatment Effect	-0.0251	-0.209	-0.112***	-0.202**
	(0.0746)	(0.145)	(0.0363)	(0.0877)
Observations	1,150	1,102	1,343	1,261

Notes: Bootstrapped standard errors in parentheses. Estimates of the average treatment effect of having a job offer at migration or being employer sponsored are computed using the Epanechnikov kernel with default bandwidth. Propensity scores are computed via probit regression including all covariates presented in column (5) of the primary regression specification table, as well as the square of each continuous variable.
*** p<0.01, ** p<0.05, * p<0.1

provide me with additional insights.

Finally, I consider as a robustness check the suggestion of Angrist and Pischke [2008] that the initial probit scores used in propensity score matching may be used to limit the sample considered by more traditional regression methods. They suggest that when selection is a concern, that trimming from the sample individuals with very high or very low propensity scores may improve OLS estimates, and they show this phenomenon using an example from the literature on job training programs. Accordingly, Table 10 replicates my primary regression specifications, but excluding those individuals with a propensity score less than 0.1 or greater than 0.9, similar to what Angrist and Pischke did. The estimates that I produce using this method are generally quite similar to those produced using the baseline standard probit and OLS regressions. Even though the sample size is much smaller in these sample-trimmed regressions, they continue to show that individuals who received a job offer are roughly 10% less likely to locate in an ethnic enclave, and that this result is statistically significant.

8 Conclusion

In order to improve our understanding of ethnic enclaves, we need to understand why they form, and how and why some individuals elect to locate in them upon migration, while others do not. However, isolating the factors that influence the locational decision of new immigrants poses multiple empirical challenges. Living in an ethnic enclave may provide new immigrants not only with improved job search prospects, but also a wide range of social and economic benefits.[15] Furthermore, it is reasonable to expect that many of these benefits may accrue to immigrants regardless of whether they are employed or unemployed, and it is also plausible that some of these benefits may be correlated with the the likelihood of receiving a job offer in the first place. Yet, new immigrants are not randomly assigned to jobs, nor are they typically randomly assigned

[15] See, e.g., Bertrand et al. [2000] and Patel and Vella [2012].

Table 10: Regressions With Trimmed Sample Excluding Individuals with Low and High Propensity Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Binary Enclave Indicator (Probit)					Continuous Language Density (OLS)				
Offer at Migration	-0.309**	-0.317**	-0.317**	-0.373***	-0.382***	-0.236***	-0.235***	-0.233***	-0.154***	-0.159**
	(0.151)	(0.146)	(0.144)	(0.129)	(0.120)	(0.0886)	(0.0804)	(0.0819)	(0.0586)	(0.0629)
Years School		-0.000153	0.000872	0.0258*	0.0572***		-0.0198	-0.0166	-0.00981	0.00655
		(0.0195)	(0.0212)	(0.0146)	(0.0160)		(0.0194)	(0.0194)	(0.0216)	(0.0214)
Age		-0.0147***	-0.0148***	-0.0235***	-0.0254***		-0.00701*	-0.00730*	-0.00734*	-0.00831*
		(0.00430)	(0.00429)	(0.00642)	(0.00693)		(0.00424)	(0.00440)	(0.00423)	(0.00481)
Female		-0.152	-0.151	0.0185	0.0441		-0.0263	-0.0224	-0.00983	0.0116
		(0.103)	(0.103)	(0.213)	(0.243)		(0.0452)	(0.0455)	(0.141)	(0.152)
Married		0.299	0.302	0.613*	0.661*		0.166	0.197	0.202*	0.223*
		(0.277)	(0.307)	(0.367)	(0.361)		(0.116)	(0.133)	(0.120)	(0.116)
Separated		0.0398	0.0256	-0.0206	-0.0291		-0.198	-0.154	-0.178	-0.200
		(0.828)	(0.794)	(0.859)	(0.831)		(0.460)	(0.467)	(0.594)	(0.552)
Has Children		0.0687	0.0721	0.0605	0.117		0.130**	0.126**	0.150***	0.173***
		(0.0738)	(0.0764)	(0.0875)	(0.0938)		(0.0576)	(0.0596)	(0.0179)	(0.0247)
Got Job Relative			0.0234	-0.176	-0.302			0.177	0.0475	-0.0186
			(0.186)	(0.184)	(0.194)			(0.157)	(0.151)	(0.149)
Relative at Emp.			0.0704	0.221	0.224			-0.0621	-0.0429	-0.0403
			(0.152)	(0.142)	(0.158)			(0.172)	(0.117)	(0.108)
Log Pre-Migration Salary				0.0391	0.0709				-0.0440	-0.0284
				(0.0582)	(0.0529)				(0.0350)	(0.0332)
Pre-Migration Hours				0.0108**	0.0102*				-0.00431	-0.00426
				(0.00503)	(0.00553)				(0.00442)	(0.00459)
Poor English					0.737***					0.393**
					(0.167)					(0.160)
Religious					0.0522					0.0621
										(0.0626)
Poor Health					-1.499***					-0.832***
					(0.185)					(0.190)
Observations	589	589	589	574	574	601	601	601	601	601
R-squared						0.391	0.401	0.403	0.460	0.472
Avg. Marginal Effect	-0.0940	-0.0951	-0.0950	-0.100	-0.0999					
	(0.0439)	(0.0414)	(0.0405)	(0.0327)	(0.0292)					

Notes: Robust standard errors in parentheses. All regressions include a full set of state/Census division and language-group fixed effects. Log Salary is measured in U.S. 2003 dollars. Excludes individuals with a propensity score from probit estimation of less than 0.1 or greater than 0.9, using the same propensity scoring method as in Table 9.
*** p¡0.01, ** p¡0.05, * p¡0.1

anything else that may strongly influence the locational decision. As I show in this paper, individuals who arrive in the U.S. with a job offer already in hand have different characteristics from those who arrive without an offer.

By taking advantage of the detailed data and the representative sample of the New Immigrant Survey, this paper shows that immigrants who arrive with a job offer are considerably less likely to locate in an ethnic enclave area, even after carefully controlling for a wide range of observable pre-migration characteristics. These results are also robust to numerous alternative specifications, and they are relatively free of concerns related to reverse causality. In keeping with the theoretical predictions of my model, this provides strong evidence that immigrants do locate in ethnic enclaves for reasons related to job search, and that this affects locational decisions both within and across metropolitan areas.

Perhaps unsurprisingly, regression analysis also suggests that several other factors influence the locational decision of migrants. Individuals who report that they speak English poorly or who have relatively low levels of education are more likely to locate in enclaves. So are individuals who had relatively low salaries or who worked relatively few hours pre-migration, given their occupation and industry. These findings lend evidence to the notions that immigrants sort themselves into enclaves in part based on their personal preferences, their ability levels, and/or their ability to assimilate.

However, there appears to be no significant evidence of heterogeneous effects of a job offer, as might be expected if higher-skilled individuals benefit less from enclave job search networks than lower-skilled individuals. While there is no evidence of positive selection on the ability to benefit from migrant networks, there is also no clearly isolated evidence of negative selection on these benefits. I consider this to be an avenue for continued research.

References

2011 yearbook of immigration statistics. Technical report, Office of Immigration Statistics, U.S. Department of Homeland Security, September 2012. URL http://www.dhs.gov/sites/default/files/publications/immigration-statistics/yearbook/2011/ois_yb_2011.p

Joshua D. Angrist and Jrn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press, 2008.

Patrick Bayer, StephenL. Ross, and Giorgio Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196, December 2008. ISSN 0022-3808, 1537-534X. doi: 10.1086/595975. URL http://www.jstor.org/stable/10.1086/595975.

L.A. Beaman. Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US. *The Review of Economic Studies*, 2011.

Marianne Bertrand, Erzo F. P. Luttmer, and Sendhil Mullainathan. Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055, 2000. ISSN 00335533. URL http://www.jstor.org/stable/2586902. ArticleType: research-article / Full publication date: Aug., 2000 / Copyright 2000 Oxford University Press.

George J. Borjas. Self-selection and the earnings of immigrants. *The American Economic Review*, 77(4):531–553, 1987. ISSN 0002-8282. URL http://www.jstor.org/stable/1814529. ArticleType: research-article / Full publication date: Sep., 1987 / Copyright 1987 American Economic Association.

George J. Borjas and Rachel M. Friedberg. Recent trends in the earnings of new immigrants to the united states. *National Bureau of Economic Research Working Paper Series*, No. 15406, 2009. URL http://www.nber.org/papers/w15406.

David Card. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64, January 2001. ISSN 0734-306X. URL http://www.jstor.org/stable/10.1086/209979. ArticleType: research-article / Full publication date: January 2001 / Copyright 2001 The University of Chicago Press.

D.M. Cutler, E.L. Glaeser, and J.L. Vigdor. Is the melting pot still hot? explaining the resurgence of immigrant segregation. *The Review of Economics and Statistics*, 90(3):478497, 2008.

Rajeev H. Dehejia and Sadek Wahba. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161, February 2002. ISSN 0034-6535. doi: 10.1162/003465302317331982. URL http://dx.doi.org/10.1162/003465302317331982.

Per-Anders Edin, Peter Fredriksson, and Olof slund. Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment. *The Quarterly Journal of Economics*, 118(1):329–357, February 2003.

ISSN 0033-5533. URL http://www.jstor.org/stable/25053906. ArticleType: research-article / Full publication date: Feb., 2003 / Copyright 2003 Oxford University Press.

Delia Furtado and Nikolaos Theodoropoulos. Immigrant networks and the take-up of disability programs: Evidence from US census data. Technical Report 6649, Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, 2012. URL http://www.econstor.eu/handle/10419/62380.

D. Goel and K. Lang. Social ties and the job search of recent immigrants! 2012.

James J. Heckman and Bo E. Honor. The empirical content of the roy model. *Econometrica*, 58(5):1121–1149, 1990. ISSN 0012-9682. doi: 10.2307/2938303. URL http://www.jstor.org/stable/2938303. ArticleType: research-article / Full publication date: Sep., 1990 / Copyright 1990 The Econometric Society.

Guillermina Jasso, Douglas S. Massey, Mark R. Rosenzweig, and James P. Smith. The new immigrant survey 2003 round 1 (NIS-2003-1) public release data. Technical report, March 2006. URL http://nis.princeton.edu.

EdwardP. Lazear. Culture and language. *Journal of Political Economy*, 107(S6):S95–S126, December 1999. ISSN 0022-3808. URL http://www.jstor.org/stable/10.1086/250105. ArticleType: research-article / Full publication date: December 1999 / Copyright 1999 The University of Chicago Press.

James G. MacKinnon and Lonnie Magee. Transforming the dependent variable in regression models. *International Economic Review*, 31(2):315–339, May 1990. ISSN 0020-6598. doi: 10.2307/2526842. URL http://www.jstor.org/stable/2526842.

James D. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review*, 81(5):1408–1418, December 1991. ISSN 0002-8282. doi: 10.2307/2006929. URL http://www.jstor.org/stable/2006929. ArticleType: research-article / Full publication date: Dec., 1991 / Copyright 1991 American Economic Association.

K. Munshi. Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, 118(2):549, 2003.

Krishna Patel and Francis Vella. Immigrant networks and their implications for occupational choice and wages. *Review of Economics and Statistics*, October 2012. ISSN 0034-6535. doi: 10.1162/REST_a_00327. URL http://dx.doi.org/10.1162/REST_a_00327.

Giovanni Peri and Chad Sparber. Task specialization, immigration, and wages. *American Economic Journal. Applied Economics*, 1(3):135–169, July 2009. ISSN 19457782. doi: 10.1257/app.1.3.135.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146, June 1951. ISSN 0030-7653. doi: 10.2307/2662082. URL `http://www.jstor.org/stable/2662082`. ArticleType: research-article / Full publication date: Jun., 1951 / Copyright 1951 Oxford University Press.

Maude Toussaint-Comeau. Do ethnic enclaves and networks promote immigrant self-employment? *Economic Perspectives*, 32(4):30–50, 2008. ISSN 1048115X.

Appendix

Table 11: Description of Key Variables

Variable	NIS Questions	Response Coding
Had At Migration	"Had you been offered this job before coming to the United States to live?"	1 if yes for any current job
Got Job Relative	"Did you get this job with the help of a relative of yours or your husband/wife?"	1 if yes for any current job
Relative At Emp.	"Do any of your, or your husband/wife's relatives work for this business"	1 if yes for any current job

Table 12: Description of Enclave Measures Using U.S. Census and NIS Data

Key Definitions	
List of Languages (NIS)	"What languages do you currently speak at home?"
Binary threshold	$2\times$ National Avg. % Spoken
Continuous measure	$\ln\left(\frac{\text{Local \% Spoken}}{\text{National Avg. \% Spoken}} + \left(\left(\frac{\text{Local \% Spoken}}{\text{National Avg. \% Spoken}}\right)^2 + 1\right)^{1/2}\right)$

Methods of Individual Group Assigment

Measure used	Method of group assignment
Results shown	
First Lang.	The first non-English language reported to be spoken by each immigrant (results reported).
Robustness checks (available on request)	
Lang. Density	The reported language with the highest local percentage density relative to its national average
Abs. Lang. Density	The reported language with the highest local percentage density

Notes: All enclaves are defined at the ZIP code level using data from the 2000 Census, SF-3 Sample, available at http://factfinder2.census.gov/.

Table 13: New Immigrant Survey Language Groups

Group ID	Language	Number of Speakers as First Language
2	Spanish	2,312
3	Portuguese	51
4	Russian	284
5	French	83
6	Arabic	226
7	Chinese	501
8	Tagalog	257
9	Vietnamese	181
10	Korean	126
11	Bengali	57
12	Hindi	107
13	Polish	173
14	Amharic	162
15	Creole	111
16	Other European	410
17	Other Non-European	586
18	Other Spoken in Philippines	75
19	Other Spoken in India	541
20	Other	2

Notes: Counts are as provided by NIS in response to the question "What languages do you currently speak at home?". English is excluded from analysis.

Table 14: Key Variables By Education Group

	Offered Job Prior To Move	Helped By Relative To Get Job	Relative Works For Company
Less Than HS	0.038	0.267	0.186
	(0.006)	(0.014)	(0.012)
	[923]	[940]	[1,060]
High School	0.096	0.272	0.140
	(0.018)	(0.026)	(0.019)
	[281]	[287]	[321]
Some College	0.072	0.157	0.093
	(0.012)	(0.017)	(0.013)
	[447]	[458]	[486]
Bachelors Degree	0.249	0.105	0.077
	(0.012)	(0.009)	(0.007)
	[1,215]	[1,263]	[1,359]
Graduate School	0.222	0.051	0.053
	(0.014)	(0.007)	(0.007)
	[851]	[895]	[944]
Unknown/No Answer	0.048	0.242	0.157
	(0.007)	(0.015)	(0.012)
	[814]	[839]	[903]

Notes: Standard errors in parentheses. Counts in brackets.

Table 15: Frequency of Pre-Migration Offers by Occupational Group

	Received Pre-Migration Offer			
Occupation	Yes	No	Missing	Total
OFFICE AND ADMIN. SUPPORT	4.0%	12.3%	13.5%	12.1%
SALES AND RELATED	7.1%	11.4%	13.1%	11.8%
EXEC., ADMIN. AND MANAGERIAL	11.5%	9.3%	10.8%	10.2%
TEACHERS	4.7%	7.0%	8.9%	7.6%
SETTER, OPERATORS, AND TENDERS	3.3%	7.5%	7.1%	7.0%
MATH. AND COMPUTER SCIENTISTS	16.9%	6.8%	2.0%	5.6%
HEALTH DIAG. AND TREAT. PRACTIT.	18.8%	3.5%	4.7%	5.4%
TRANSP. AND MAT. MOVING	1.3%	4.5%	3.9%	4.0%
MANAGEMENT RELATED	2.9%	3.3%	4.6%	3.8%
CONSTRUCTION TR. AND EXTRACTION	0.9%	4.8%	3.2%	3.7%
ENGINEERS, ARCHITECTS, SURVEYORS	6.4%	3.1%	2.7%	3.2%
FARMING, FISHING, FORESTRY	0.4%	3.6%	3.4%	3.2%
FOOD PREPARATIONS AND SERVING	4.9%	3.3%	2.3%	3.0%
INSTALLATION, MAINTENANCE, AND REPAIR	1.1%	3.5%	2.7%	3.0%
Total (including omitted)	100.0%	100.0%	100.0%	100.0%
Percentage Reporting Pre-Migration Occupation	72.2%	60.5%	55.9%	59.2%

Table 16: Frequency of Pre-Migration Offers by Industry Group

	Received Pre-Migration Offer			
Occupation	Yes	No	Missing	Total
EDUC., HEALTH, SOCIAL SVCS.	25.9%	14.0%	17.9%	16.8%
MANUFACTURING	12.6%	13.2%	13.0%	13.1%
PROFESSIONAL AND RELATED SVCS.	22.4%	11.6%	9.1%	11.4%
RETAIL TRADE	4.9%	10.4%	10.5%	10.0%
ENTERT., ACCOM., AND FOOD SVCS.	6.9%	7.7%	5.8%	6.8%
AGRICULTURE, FORESTRY, FISHERIES	0.7%	6.8%	7.3%	6.5%
OTHER SERVICES	5.1%	5.6%	6.1%	5.8%
PUBLIC ADMINISTRATION	1.6%	4.9%	6.5%	5.3%
CONSTRUCTION	2.0%	6.4%	4.7%	5.3%
FINANCE, INSURANCE, REAL ESTATE	4.7%	4.5%	4.5%	4.5%
TRANSPORT. AND WAREHOUSING	2.0%	4.2%	4.3%	4.1%
WHOLESALE TRADE	3.5%	3.2%	4.0%	3.6%
INFO. AND COMMUNICATION	4.2%	2.7%	2.3%	2.7%
UNCODABLE	2.4%	2.7%	2.6%	2.6%
UTILITIES	0.2%	0.8%	0.5%	0.6%
MINING	0.4%	0.5%	0.5%	0.5%
ARMED FORCES	0.4%	0.5%	0.3%	0.4%
NOT IN LABOR FORCE	0.0%	0.0%	0.1%	0.1%
Total	100.0%	100.0%	100.0%	100.0%
Percentage Reporting Pre-Migration Industry	72.2%	60.5%	55.9%	59.2%

Table 17: Frequency of Pre-Migration Offers by Social Indicators

	Received Pre-Migration Offer			
	Yes	No	Missing	Total
English is Poor	0.198	0.415	0.585	0.480
	(0.016)	(0.008)	(0.008)	(0.005)
Poor Health	0.010	0.025	0.051	0.036
	(0.004)	(0.002)	(0.003)	(0.002)
Regular Religious Attendance	0.598	0.620	0.603	0.611
	(0.020)	(0.008)	(0.008)	(0.005)