Centre for Research and Analysis of Migration

CReAM

Discussion Paper Series

Living and Working in Ethnic Enclaves: Language Proficiency of Immigrants in U.S. Metropolitan Areas

Julia Beckhusen, Raymond J.G.M. Florax, Thomas de Graaff, Jacques Poot and Brigitte Waldorf

Living and Working in Ethnic Enclaves:
Language Proficiency of Immigrants
in U.S. Metropolitan Areas

Julia Beckhusen*, Raymond J.G.M. Florax**, Thomas de Graaff†,

Jacques Poot†† and Brigitte Waldorf‡

* U.S. Census Bureau
** Department of Agricultural Economics, Purdue University and Department of Spatial Economics, VU University and Tinbergen Institute
† Department of Spatial Economics, VU University Amsterdam
†† National Institute of Demographic and Economic Analysis, University of Waikato
‡ Department of Agricultural Economics, Purdue University

Non-Technical Abstract

Learning English is a potentially profitable investment for immigrants in the U.S.: while there are initial costs, the subsequent benefits include the ability to communicate with the majority of the population, potentially leading to better paying jobs and economic success in the new country. These payoffs are lessened if immigrants choose to live and work in ethnic enclaves where the necessity to communicate in English is weak. Ethnic enclaves are widespread and persistent in the U.S. This study uses data from the 2010 American Community Survey to examine the impact of residential and occupational segregation on immigrants' ability to speak English. We allow for heterogeneity in the relationship between segregation and English language proficiency across ethnic groups and focus specifically on Mexican and Chinese immigrants. Our results show that immigrants in the U.S. who live and work among high concentrations of their countrymen are less likely to be proficient in English than those who are less residentially and occupationally segregated. The magnitude of the effect of segregation on language proficiency varies across immigrants' birthplaces and other salient characteristics defining the immigration context.

Living and Working in Ethnic Enclaves:
Language Proficiency of Immigrants in U.S. Metropolitan Areas

Julia Beckhusen, U.S. Census Bureau, Suitland, MD, USA.
E-mail: julia.b.beckhusen@census.gov

Raymond J.G.M. Florax, Department of Agricultural Economics, Purdue University, West Lafayette, IN 47907, USA; Department of Spatial Economics, VU University and Tinbergen Institute, Amsterdam, The Netherlands. E-mail: rflorax@purdue.edu

Thomas de Graaff, Department of Spatial Economics, VU University Amsterdam, Amsterdam, The Netherlands. E-mail: t.de.graaff@vu.nl

Jacques Poot, National Institute of Demographic and Economic Analysis, University of Waikato, New Zealand. E-mail: jpoot@waikato.ac.nz

Brigitte Waldorf, Department of Agricultural Economics, Purdue University, West Lafayette, IN 47907, USA. E-mail: bwaldorf@purdue.edu

Abstract: Learning English is a potentially profitable investment for immigrants in the U.S.: while there are initial costs, the subsequent benefits include the ability to communicate with the majority of the population, potentially leading to better paying jobs and economic success in the new country. These payoffs are lessened if immigrants choose to live and work in ethnic enclaves where the necessity to communicate in English is weak. Ethnic enclaves are widespread and persistent in the U.S. This study uses data from the 2010 American Community Survey to examine the impact of residential and occupational segregation on immigrants' ability to speak English. We allow for heterogeneity in the relationship between segregation and English language proficiency across ethnic groups and focus specifically on Mexican and Chinese immigrants. Our results show that immigrants in the U.S. who live and work among high concentrations of their countrymen are less likely to be proficient in English than those who are less residentially and occupationally segregated. The magnitude of the effect of segregation on language proficiency varies across immigrants' birthplaces and other salient characteristics defining the immigration context.

Introduction

The United States government has a laissez-faire attitude toward assimilation and language acquisition. Although 50 percent of the U.S. population growth between 2000 and 2010 came from immigration[1] and the proportion of immigrants who speak English has declined from 85 percent in 1900 (Lazear, 1999) to 64 percent in 2010,[2] the U.S. government does little to provide immigrants with assistance in learning the new culture or to speak and write in English. The absence of assimilation policies is unfortunate since for immigrants, learning the host language fosters economic advancement (Berman et al., 2003; Chiswick, 1998; Chiswick and Miller, 1999, 2002). Moreover, increasing the upward mobility of new immigrants in turn benefits society as a whole (Borjas, 1994).

In a path breaking study, Lazear (1999) framed language acquisition as an investment facilitating trade between cultures. Any type of trade is based on communication, which can only occur when two individuals share a common language. As a result, revenues from trade increase with the probability of meeting others who speak the same language. Lazear juxtaposed the expected revenue gains from learning another language against the costs of language acquisition and showed analytically that the net benefits decrease when the proportion of own-language speakers increases. As a result, the probability of language acquisition in integrated communities will exceed that in ethnic enclaves. In the empirical validation of this theory, he uses 1900 and 1990 U.S. county data for Chinese, German, Spanish and Swedish immigrants. Florax et al. (2005) applied the theory to empirically explain variations in Dutch language acquisition among immigrants in the Netherlands, using a smaller spatial scale but allowing revenues that arise from trade interactions beyond the own community. Waldorf et al. (2010) analyzed U.S. data to test Lazear's theory at the larger spatial scale of labor markets proxied by metropolitan areas, and highlighted differences by immigrants' educational attainment levels.

Neither of these studies separates residence-based versus workplace-based interactions. In fact, they implicitly assume that the probability of meeting individuals of the host culture is the same regardless whether the place of work or the place of residence is considered. This may be a reasonable assumption when the interaction area is large, e.g., a county or a metropolitan area, but less so at smaller spatial scales. The distinction between residence-based and workplace-based interactions is important since immigrant groups are often disproportionally employed in

---

[1] Source: 2000 and 2010 U.S. Census Bureau American Community Survey
[2] Source: 2010 U.S. Census Bureau American Community Survey

certain occupations (Waldinger 2000). When separating the language acquisition impacts of residential and occupational segregation it is crucial to address the inherent endogeneity of language acquisition and location choice. Typically, the literature assumes that immigrants first choose where to reside and subsequently choose whether to learn the host language. However, it could be the case that immigrants decide to locate in a particular area because of their individual language ability.

This paper is organized as follows. The next section discusses the residential and occupational segregation forming the context of language acquisition. This is followed by sections dealing with the discussion of the theoretical and empirical models, the data description and interpretation of estimation results. The last section provides a summary and conclusion of the research.

Context of Language Acquisition

A common language facilitates social interaction or "trade" in Lazear's (1999) terms. For immigrants, the ability to speak the language of their host country enables them to communicate with the majority of the population, thereby lowering information barriers, finding better jobs and, consequently, achieving economic success in the new country. Upon arrival, immigrants' language skills are often limited. Thus, not surprisingly, many immigrants move close to their own countrymen (Bauer et al., 2005; Gross and Schmitt, 2003) in order to take advantage of the informal networks of friends, relatives and acquaintances and be part of the ethnic enclave. The enclaves provide a means for costless communication, labor market opportunities and transportation (Borjas, 1998; Chiswick and Miller, 1994, 1996). Andersson et al. (2009) found that newcomers' probability of finding employment increases if they locate near large numbers of fellow employed expatriates. Finding a job quickly after immigrating reduces the amount of income lost due to the move. Strong social and family network ties in the destination country can also lower many of the monetary and non-monetary costs associated with migration. Residential segregation and the emergence of ethnic enclaves are thus not only the result of individual location decisions, but they also generate externalities. As these enclaves grow, they may become even more attractive to expatriates and hence grow further. Since 1980, immigrant neighborhoods have grown denser and more segregated in the U.S. (Iceland, 2004; Charles, 2003). Logan et al. (2004) found that segregation increased the most for Hispanics with lower incomes relative to whites. In 2000, Hispanic neighborhoods comprised between 16 and 42 percent Hispanics whereas black neighborhoods have between 18 and 59

percent black residents. Asians were the least segregated among all non-white ethnic groups (Charles, 2003).

Within ethnic neighborhoods, immigrants can communicate in their native language and receive revenue from trading without the cost of learning a new language so that the incentive to learn English is much lower in ethnic neighborhoods than in more dispersed settings (Chiswick et al., 2005; Verbon and Meijdam, 2008). One of the most important benefits from locating inside ethnic networks is job market information. Informal networks of friends, relatives and acquaintances are frequently used successfully to find jobs (Ioannides and Loury, 2004). New immigrants obtain job market information on occupations held by those in the neighborhood so that employment is likely found alongside other network members, ultimately contributing to occupational segregation.[3]

Occupational segregation is being perpetuated for two main reasons. First, social networks help channel immigrants into particular jobs. Second, once an ethnic group makes up a good deal of all employees in an occupation, the occupation becomes stereotyped as being typical for that ethnic group. Catanzarite (2000), for example, deemed "brown-collar" jobs as those where the majority of employees are Latino. The ethnic labeling of occupations influences where job seekers look for work as well as employers' hiring preferences. Current employees may even go as far as to prevent non-ethnic group members from being hired (Waldinger, 2000).

Occupational segregation may be exacerbated when workers share a common non-native language. This gives employers an incentive to look for applicants who can speak this language. Hellerstein and Neumark (2002) found that occupational segregation among Mexican immigrants increases when English proficiency levels of workers are low. However, as expected, since jobs often require technical skills that cannot be acquired from one's social network, occupational segregation of ethnic groups is found to be less severe than residential segregation (Ellis et al., 2004).

---

[3] The term "occupational segregation" comes from research on dual labor markets (Anker, 1997). Dual labor market theory poses the existence of two labor market sectors, namely the formal or primary sector comprised of more progressive jobs and the secondary sector of relatively static jobs. The former market typically requires more education or skill and is higher paying. Applicants form job-queues in order of employability where employability is determined not only by education but also by race (Hodge, 1973). In the presence of discrimination, members of the dominant group hold positions at the top of the queue and find jobs in the primary sector leaving minority members to find jobs in the secondary sector. This process contributes to labor markets that are divided along racial lines (Semyonov et al., 2000) and minority members who are more likely to work in the secondary sector (Anker, 1997; Schrover et al., 2007).

Theoretical Background

Lazear (1999) formalized a theory relating immigrants' decisions to acquire the language of the host country and the ethnic composition of one's surroundings. He used a random encounter model in which benefits from trade between two individuals can only be realized if they share a common culture, i.e., if they speak the same language. The derivation begins by assuming a country comprised of two cultures that do not share a common language: a majority and a minority culture. Assume further that p is the proportion of the population speaking the minority language.[4] Hence the probability that a randomly encountered person speaks the minority language is p. Thus, assuming that revenue from trading is normalized to 1, the expected revenue for a monolingual minority member is p, whereas the expected revenue for a bilingual minority member is 1. Becoming bilingual is thus associated with a revenue gain of $1-p$. However, a minority invests in becoming bilingual if and only if the individual j-specific costs of language acquisition, t_j, are smaller than the revenue gain. Suppose that t_j is the outcome of a random variable T with probability density function $g(t)$ and distribution function $G(t)$, then the probability of an individual from the minority becoming bilingual equals:

$$\text{Prob}(T < 1-p) = G(1-p). \tag{1}$$

Note that the distribution function is decreasing in p, and thus the probability of a minority acquiring the majority language is inversely related to the proportion of the population speaking the minority language.[5]

A drawback of Lazear's model is that it does not account for different proportions time being spent in different places, most prominently at home and at work. The probability of encountering a minority language speaker may differ between these two places. Moreover, the time spent in each of these places varies from person to person. Hägerstrand (1970)

---

[4] In most cases, $p < 0.5$ but that is not a necessary condition and will not be the case if many persons belonging to the majority will have chosen to learn the minority language.

[5] As argued in Florax et al. (2005), if G is linearly decreasing in p, then T is uniformly distributed and $g(1-p) = 1$ for all p. A more realistic assumption would be that T is normally distributed, implying that G is nonlinearly decreasing in p and hence $\partial G(1-p)/\partial p < 0$. While not stated explicitly in Lazear's paper, this was the most likely reason why Lazear included both p and p^2 in his operational econometric model. The choice of a second-degree polynomial in p is obviously rather arbitrary. One should also note that its parabolic shape can be at odds with the assumption that $\partial G(1-p)/\partial p < 0$. Florax et al. (2005, p. 1881) recognize this, but point to the unlikeliness of sample observations being in the upward-sloping part of the curve. Hence we have maintained the use of a second-degree polynomial in the empirical analyses in order to facilitate comparison with Lazear's results.

demonstrated the importance of positioning individuals in time as well as space, since a location is only important if an individual has time to spend there. We therefore propose an expansion of the meeting probability p that takes a minority's allocation of time at home and work into account.

Let p^h and p^w be the proportions of minority language speakers near home and at work, respectively. Furthermore, let μ and $1-\mu$ be the proportions of time spent at home and at work, respectively. The composite probability of encountering a minority language speaker is then given by:

$$\eta = \mu p^h + (1-\mu)p^w. \tag{2}$$

The expected revenue from trade for a monolingual minority member is η. Hence the minority member's expected revenue gain from becoming bilingual is $1 - \eta$ and a minority person j decides to become bilingual if and only if $t_j < 1 - \eta$. The proportion of bilingual minority members is therefore $G(1-\eta)$.

Empirical Model

The empirical model focuses on the Mexican and Chinese immigrants in the U.S. and explains the variation in English language proficiency (defined as speaking English well or very well) among these ethnic groups. Mexican and English immigrants were selected because they form an interesting juxtaposition with respect to residential segregation (which is much higher for the Mexicans) and human capital (with the Chinese having on average much higher levels of education). We will allow for full heterogeneity between these groups in the English language proficiency modeling. The key explanatory variable is the encounter probability η, decomposed into its residential and workplace probabilities of workers speaking the minority language only, p^h and p^w respectively. It is hypothesized that the relationship between each of the two components of the overall probability of meeting a speaker of one's own minority language will be inversely related to English language proficiency, potentially at different magnitudes for different immigrant groups. Given the assumption that learning costs, T, are normally distributed, the relationship between the probability of speaking English and the encounter probabilities and is nonlinear and, following Lazear, specified as a second-order polynomial. In addition, the empirical model controls for variations due to immigrant's personal

characteristics, socio-economic status, immigration-related attributes and linguistic family context variables.

Unlike previous studies on language acquisition, we address the inherent endogeneity that comes into play when establishing the connection between an individual's selection of a residence and occupation on the one hand, and the choice to learn English on the other. The pay-offs of language acquisition may be a function of the spatial concentration of fellow immigrants, but simultaneously the location choice behavior of immigrants may be determined by their level of proficiency in the host language.

This endogeneity has not been sufficiently addressed in the past, but the problem becomes apparent when reviewing the literature: one group of studies assumes language ability to be the dependent variable, while concentration is an independent variable (Chiswick and Miller, 1996; Lazear, 1999; Florax et al., 2005), while another literature (Chiswick et al., 2001; Bauer et al., 2005) assumes the opposite. Ignoring the endogeneity fails to determine the direction of causality and leads to the estimator being biased.

Lazear (1999) briefly addresses the dual causality between location choice and learning the language. He argues that the issue is a question of 'when' as opposed to 'why' immigrants learn English. If location choice is exogenous then the effect of immigrants' length of stay abroad on their English speaking ability will decline with increasing immigrant concentration. For 1900, the data suggest that the sojourn effect does indeed decrease with increasing immigrant concentration. For 1990, however, the sojourn effect does not vary by immigrant concentration, suggesting that immigrants sort into counties depending on their English-speaking ability.

In this study, we address the issue via instrumental variables (IV) for the composite encounter probability η. Thus, the empirical models consist of a series of binary probit models and binary IV probit models. In the first stage of the IV regression, special care is devoted to instrumenting all components of the meeting probability η as well as the squared meeting probability, η^2 (with $\frown$ indicating predicted values from the first stage regression):

$$\hat{\eta} = \widehat{\mu p^h} + (1 \widehat{-\mu})p^w \tag{3}$$

and

$$\widehat{\eta^2} = (\widehat{\mu p^h})^2 + 2\widehat{\mu p^h}(1 \widehat{-\mu})p^w + ((1 \widehat{-\mu})p^w)^2. \tag{4}$$

6

Note that in the first stage, instrumentation is needed for time-allocation-weighted proportions rather than for the unweighted proportions. Failing to do so implies estimating a "forbidden regression" where estimates in the first stage are inconsistent and the *t*-statistics from the second stage are invalid, even asymptotically (Wooldridge, 2002).

Data

The empirical results are based on a sample of Mexican and Chinese-born adult immigrants living and working in metropolitan areas of the United States, obtained from the 2010 American Community Survey (ACS).[6] The focus on Mexican immigrants is motivated by their large representation and long history in the U.S. According to the 2010 ACS, the U.S. is home to 11,711,103 persons born in Mexico, making up 3.8 percent of the total population and 29.3 percent of the immigrant population. Traditionally, Mexican immigrants were heavily represented in the states along the U.S.-Mexican border but recently the Mexican population also increased substantially in the areas further away from the border. Although about 60 percent of the Mexican immigrants do not hold a high school diploma, the immigrants are positively selected compared to the population in Mexico (Feliciano, 2005). Chinese immigrants provide an interesting comparison group to Mexican immigrants. They account for 0.7 percent of the US population and 5.4 percent of the immigrant population. The cultural and linguistic dissimilarities with the U.S. are much stronger for Chinese than for Mexican immigrants. However, Chinese immigrants have, on average, higher educational attainment levels with 44 percent having a tertiary education.

We restrict the sample to employed immigrants who were at least 25 years of age when they moved to the U.S. and who were not in school at the time of the survey. These selection criteria help ensure that respondents received the majority, if not all, of their education in their home country. The total number of observations in the sample is 8,223 Mexican and 2,517 Chinese immigrants.

The spatial scale used to assess the impact of residential immigrant concentration on language proficiency is the Public Use Microdata Area (PUMA) designed by the U.S. Census. PUMAs follow boundaries of county groups, single counties, or census-defined "places" with populations of at

---

[6] Extracted from the Minnesota Population Center's IPUMS database (Ruggles et al., 2010).

least 100,000 (but not more than 200,000). Only immigrants living in metropolitan areas with at least 2000 PUMAs are included to ensure that there is sufficient neighborhood variation.

Occupations are classified into 23 groups (see Table 1). Chinese immigrants are relatively dispersed with a somewhat higher representation in Food Preparation and Serving Related occupations (12.5 percent), Computer and Mathematical occupations (9.16 percent), Management occupations (8.82 percent) and Life, Physical, and Social Science occupations (8.53 percent). Mexican immigrants, however, are concentrated in a small number of occupations. Nearly half are employed in only three categories: Building and Grounds Cleaning and Maintenance occupations (18.1 percent), Construction and Extraction trades (14.9 percent) and production occupations (13.9 percent). Another 20 percent are employed in Food Preparation and Serving (13.1 percent) and Transportation and Material Moving occupations (9.27 percent).

To measure the occupational enclave, the smallest spatial units available are census-delineated place-of-work PUMAs (PWPUMAs), which are either identical to PUMAs or a combination of multiple PUMAs. For every Mexican (Chinese) immigrant, j, with occupation k in PWPUMA n, the occupational concentration is measured by expressing the number of Mexican (Chinese) immigrants with occupation k working in n as a proportion of all employees with occupation k working in n.

Table 2 displays the summary statistics of the data. The majority of the Mexican immigrants (over 63 percent) do not speak English well.[7] In contrast, almost 63 percent of Chinese immigrants are fluent in English. This disparity in English proficiency is quite remarkable, particularly since Mexican immigrants have, on average, stayed in the U.S. for almost the same number of years as Chinese immigrants (13.3 versus 14.0 years respectively). The English proficiency disparity is consistent with the differences in naturalization rates. Almost half of the Chinese immigrants have acquired U.S. citizenship, compared to less than 20 percent of the Mexican immigrants.

Chinese and Mexican immigrants also differ with respect to the personal attributes and socio-economic characteristics. Compared to Mexican immigrants, Chinese immigrants are substantially more likely to be married, more likely to be female, and are about two years older on average. More than half of the Chinese immigrants have a bachelors degree. The opposite is

---

[7] The English language proficiency question of the U.S. census questionnaire is: "how well does this person speak English?" Respondents can choose between four possible answers: "very well," "well," "not well," and "not at all." In this study, the first two responses are categorized as being proficient in English.

true for Mexicans, whose majority does not have a high school diploma. Furthermore, and related, the mean individual income for Chinese immigrants is more than twice that of Mexican immigrants.

The next set of variables represents the immigrant's linguistic family context. We define linguistic family context as the composition of the household in terms of household members' ability to speak English. Two potential roles exist for English-speaking family members: they can serve as a "translator" or as a "teacher". The translator decreases the incentive to learn English, whereas family members taking on the role of a teacher lower the costs of investing in language acquisition. Chiswick and Miller (2005) view a family member's proficiency as an externality to other members' proficiency and found evidence of language learning between adults but not between adults and children. This suggests that English-speaking adult family members act as teachers for other adult family members. Chiswick et al. (2005) found that a child's English language fluency has a negative influence on their mother's, but not their father's English language proficiency. Since over 80 percent of immigrant children speak English well, their influence on parents' fluency will not be trivial (Mather, 2009).

We include three indicator variables signifying whether the household includes English-speaking persons, disaggregated into children under 12, teenagers (age 12 to 17) and adults (age 18+). In general, the linguistic family context for Chinese immigrants is heavily influenced by more of 70 percent having another adult in the household who is proficient in English. For Mexicans, this is only true for 60 percent, but they more frequently have a English speaking child or teenager in the household.

The main variables of interest correspond to the meeting probability. Lazear (1999) used the proportion of an immigrant's expatriates in the place of residence which we define as p^h. Mexican immigrants live in neighborhoods composed of, on average in 2010, 13 percent of their expatriates whereas Chinese immigrants reside in much less concentrated areas, on average having around 7 percent Chinese expatriates. With respect to occupational concentration, p^w, the differences are even more pronounced with only 6 percent for Chinese and 29 percent for Mexicans. The ten-year lagged residential and occupational concentration, ℓp^h and ℓp^w, were slightly lower for Mexican immigrants, but substantially lower for Chinese immigrants. Hence residential and occupational segregation are increasing over time, particularly for Chinese immigrants.

Taking into account the amounts of time spent at home and at work, the differential time allocations into account – Chinese immigrants spend on average 58 percent of their time away from work compared to 61 percent for Mexicans – has only a small impact on the difference between average Chinese and Mexican meeting probabilities. [8] The encounter probability at work, $(1-\mu)p^w$, is more than four times bigger for Mexicans than for Chinese, and the encounter probability at home, μp^h is twice as high for Mexicans.

Results

Tables 3 and 4 display the marginal effects for four binary probit models for Chinese and Mexican immigrants, respectively.[9] Model I replicates the specification used by Lazear (1999). The second model extends the first by adding basic personal and immigration-related characteristics. Model III includes the new definition of meeting probability from equation (4). Model IV takes advantage of instrumental variable estimation using the ten-year lagged values of p^h and p^w as instruments.[10]

The effect of the immigrant's demographics, education, income and immigration related traits on language proficiency is consistent in both sign and statistical significance across all six models. For both Mexican and Chinese immigrants, length of stay, citizenship, income and higher education positively impact the probability of speaking English while the absence of a high school diploma has a negative impact. With respect to sojourn it should be noted that the effect for Chinese immigrants is not statistically significant except in the original Lazear specification of the model. Chinese females are more likely to speak English, but Mexican females are less likely to learn the language. Older Chinese immigrants are less likely to be proficient whereas the sign and statistical significance of age varies by model for Mexican immigrants; its magnitude is quite small, and not significantly different from zero. Marriage is

---

[8] The parameter μ is calculated using the usual hours worked per week, among the weeks the person actually worked, during the last 12 months. We subtract that number from 98 (given that there are approximately 14 hours in a day, seven days per week, for a person to allocate between work and other activities besides sleep and self-maintenance) and divide the answer by 98 to get the fraction of available time at home.

[9] We report the marginal effects to make interpretation more straightforward since the coefficients of the original probit only provide information on the sign but not the magnitude of the relationship. The reader should note that marginal effects depend on the observed values for the explanatory variables, because of the nonlinear relationship implied by the probit model. In addition, an infinitely small change in the case of a dummy variable is obviously not appropriate, and the marginal effect is therefore defined as the effect of a change in the explanatory variable from 0 to 1.

[10] We assume that μ is exogenous and instrument μp^h by $\mu \ell p^h$ and $(1-\mu)p^w$ by $(1-\mu)\ell p^w$ where ℓ indicates the ten-year lag of the variable.

not a strong indicator for language acquisition, except among the Chinese, for whom it actually decreases the probability of English fluency. The magnitude of the effect of educational attainment is the same among the immigrant groups of different nativity when it comes to the groups with less than a high school degree, but the effect of having a bachelor's degree is much stronger among Chinese than among Mexicans. The remainder of our discussion focuses exclusively on the salient variables that speak directly to the influence of scale, residential and occupational segregation, and family linguistic context.

Model I estimates the effect of the meeting probability as defined by Lazear (1999). The expected negative relationship between residential concentration and English acquisition holds for both groups, although the marginal effect is larger in magnitude for Chinese immigrants. In addition, the relationship is non-linear and increases at a decreasing rate. When further controls are included in model II, both the sign and magnitude of the coefficients on p^h and $(p^h)^2$ remain very similar. The final variable included in models I and II is the interaction between sojourn and residential concentration. Lazear included this variable to determine the importance of sorting. As explained previously, an insignificant coefficient indicates that residential concentration is indeed endogenous. This is the case among Chinese immigrants but not among Mexican immigrants. Although we correct for endogeneity for both groups, these results indicate that sorting may only be present among Chinese immigrants.

Model III expands the definition of meeting probability to one that includes a spatial component and distinguishes between interactions in the residential neighborhood and at work.[11] Interestingly, neighborhood meetings have a stronger effect than meetings at work for Chinese immigrants, whereas the opposite is the case for Mexican immigrants. Employment in an occupation concentrated with other Mexican immigrants has a stronger impact on English ability than does living in a neighborhood in which Mexicans are concentrated. Overall, the impact of meeting someone of one's own ethnic group is much stronger on the probability of a Chinese immigrant speaking English well than a Mexican immigrant. It can be easily shown with the estimated coefficients that language proficiency is a declining function of meeting probability near home or at work. The results also indicate how important workplace networks are to Mexican immigrants. Because Mexican immigrants have lower education levels, they must rely on their family and friends to find work. However, the higher skilled Chinese immigrants are able to obtain jobs with their personal qualifications instead of networks.

---

[11] Given that the components of the meeting probability are fractions, the marginal effects indicate how the probability of speaking English will approximately change given a 0.01 unit change in the variable, *ceteris paribus*.

The final model employs instrumental variable techniques to correct for the endogeneity of meeting probability. We use a ten-year lagged value of residential and occupational concentration as instruments. Valid instruments have two basic qualities. First, they should be highly correlated with the variable they are instrumenting and second, they should not be correlated with the residuals from the original model, model III in this case.

Table 5 displays these correlations, which are in general satisfactory. In addition, following Stock and Watson (2005), we tested the joint significance of the instruments on each endogenous variable and were able to further validate the strength of the instruments. Finally, the Stata package for IVProbit includes the Wald test of exogeneity,[12] which we included in our tables and the results indicate that endogeneity may not in fact be an issue given the available data for either immigrant group (Wooldridge, 2002).

The results of model IV only solidify those of model III. The role of residential and occupational meeting probabilities depends on the immigrant group in question. For Chinese immigrants, only meeting probabilities within the neighborhood impact English ability whereas occupational meeting probability is the strong factor of English ability among Mexican immigrants.

Models III and IV also include the family linguistic context variables. The results indicate that English speaking adults in the household have a positive influence, that is, the presence of an English-speaking adult in the household fosters immigrants' English language proficiency. English-speaking children and teens do not seem to have an important effect on the language proficiency of adults. The presence of English-speaking adults has a clear "teacher" effect that reduces the cost of language acquisition. Using the results of model IV, adding one fluent adult increases the probability of speaking English by 0.48 for Mexican immigrants and 0.64 for Chinese immigrants. These results are similar to the findings of Chiswick and Miller (2005) discussed above, but more conclusive on the effect of children in the household.

---

[12] The Wald test for exogeneity tests whether the residuals from the first stage are correlated with those from the final model. The correlation of an exogenous model is zero. The test statistic has a χ^2 distribution.

Conclusion

This study expands Lazear's seminal work on language acquisition. Lazear posed that the decision to learn English can be analyzed as an investment decision and the investment will only be made if the benefits of proficiency in the language of the host country exceed the costs of acquiring the language. Benefits from language acquisition come from the ability to communicate with others. Communication leads to trade and increased economic mobility. The degree to which immigrants need to communicate with the host population is related to the language spoken by those around them.

Our extensions focus on three areas, namely the inclusion of occupational segregation in addition to residential segregation, the endogeneity between location choice and language acquisition and an explicit consideration of the household's linguistic context. The first extension leads to the central hypothesis that the immigrant's decision to learn English is influenced by the ethnic segregation of the immigrant's interaction space that is centered about residence and work place as the key anchor points.

Our results confirm that both residential and occupational segregation are inversely related to language acquisition, but that the relationship varies by immigrant group. Three main results can be derived. First, residential as well as occupational segregation negatively affect the English language proficiency of both Chinese and Mexican immigrants. Second, occupational segregation has a stronger impact than residential segregation for Mexican immigrants. Third, for Chinese immigrants, residential segregation plays a stronger role in their English language proficiency than does occupational segregation.

Regarding the third extension, our results suggest that the linguistic context is of pivotal importance for immigrants' English language proficiency. Immigrants living together with English-speaking adults (for example, the case of a migrant marrying a native born person) are much more likely to speak English fluently than those who do not.

While the U.S. has essentially no policy in regards to assimilation, some European countries focus more heavily on the social and economic integration of immigrant minorities. The strongest versions of assimilation policies dictate where immigrants are allowed to locate within a city; this applies, for instance, in the city of Rotterdam in The Netherlands. Immigrants are thus spread out among the natives instead of amassing in an ethnic enclave. These "spreading policies" are an attempt to increase the benefits of assimilation and reduce any

negative impacts from immigration (Verbon and Meijdam, 2008). According to the results of this paper, in order to maximize the effectiveness of such policies, spreading is relevant both with respect to residential location and occupational choice. For some immigrant groups, ethnic concentrations at work play a stronger role in the decision to assimilate than the existence of a residential ethnic enclave.

For future research we would like to point out that the spatial scales used in our analysis are arguably large. A PUMA region is a census-defined place with at least 100,000 persons. In order to measure ethnic concentrations at the neighborhood level, a finer spatial unit should preferably be used. Similarly, more specific knowledge on where individuals work would improve the accuracy of our measure of occupational concentration.

References

Andersson, P. F. D., S. Burgess, and J. Lane (2009). Do as the Neighbors Do: The Impact of Social Networks on Immigrant Employment. SSRN eLibrary.

Anker, R. (1997). Theories of occupational segregation by sex: An overview. *International Labour Review 136* (3), 315–39.

Bauer, T., G. Epstein, and I. Gang (2005). Enclaves, language, and the location choice of migrants. *Journal of Population Economics 18* (4), 649–662.

Berman, E., K. Lang, and E. Siniver (2003). Language-skill complementarity: returns to immigrant language acquisition. *Labour Economics 10* (3), 265–290.

Borjas, G. (1994). The economics of immigration. *Journal of economic literature 32* (4), 1667–1717.

Borjas, G. J. (1998). To ghetto or not to ghetto: Ethnicity and residential segregation. *Journal of Urban Economics 44* (2), 228–253.

Catanzarite, L. (2000). Brown-collar jobs: Occupational segregation and earnings of recent-immigrant latinos. *Sociological Perspectives 43* (1), 45–75.

Charles, C. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology 29* (1), 167–207.

Chiswick, B., Y. Lee, and P. Miller (2001). The determinants of the geographic concentration among immigrants: Application to Australia. *Australasian Journal of Regional Studies 7* (2), 125–150.

Chiswick, B., Y. Lee, and P. Miller (2005). Family matters: The role of the family in immigrants' destination language acquisition. *Journal of Population Economics 18* (4), 631–647.

Chiswick, B. and P. Miller (1994). Language choice among immigrants in a multilingual destination. *Journal of Population Economics 7* (2), 119–131.

Chiswick, B. and P. Miller (1996). Ethnic networks and language proficiency among immigrants. *Journal of Population Economics 9* (1), 19–35.

Chiswick, B. and P. Miller (1999). Language skills and earnings among legalized aliens. *Journal of Population Economics 12* (1), 63–89.

Chiswick, B. and P. Miller (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics 15* (1), 31–57.

Chiswick, B. and P. Miller (2005). Do enclaves matter in immigrant adjustment? *City & Community 4* (1), 5–35.

Chiswick, B. R. (1998). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics 11* (2), 253–271.

Ellis, M., R. Wright, and V. Parks (2004). Work together, live apart? Geographies of racial and ethnic segregation at home and at work. *Annals of the Association of American Geographers 94* (3), 620–637.

Feliciano, C. 2005. Educational selectivity in U.S. immigration: how do immigrants compare to those left behind? *Demography* 42(1): 131-152.

Florax, R., T. de Graaff, and B. Waldorf (2005). A spatial economic perspective on language acquisition: segregation, networking, and assimilation of immigrants. *Environment and Planning A 37* (10), 1877–1897.

Gross, D. and N. Schmitt (2003). The role of cultural clustering in attracting new immigrants. *Journal of Regional Science 43* (2), 295–318.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science 24* (1), 6–21.

Hellerstein, J. K. and D. Neumark (2002). Ethnicity, language, and workplace segregation: Evidence from a new matched employer-employee data set. Working Paper 9037, National Bureau of Economic Research.

Hodge, R. (1973). Toward a theory of racial differences in employment. *Social Forces 52* (1), 16–31.

Iceland, J. (2004). Beyond black and white metropolitan residential segregation in multi-ethnic America. *Social Science Research 33* (2), 248–271.

Ioannides, Y. and L. Loury (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature 42* (4), 1056–1093.

Lazear, E. (1999). Culture and language. *Journal of Political Economy 107* (S6), 95–126.

Logan, J., B. Stults, and R. Farley (2004). Segregation of minorities in the metropolis: two decades of change. *Demography 41* (1), 1–22.

Mather, M. (2009). Children in U.S. immigrant families chart new path. Technical report, Population Reference Bureau.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010). Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Schrover, M., J. van der Leun, and C. Quispel (2007). Niches, labour market segregation, ethnicity and gender. *Journal of Ethnic and Migration Studies 33* (4), 529–540.

Semyonov, M., Y. Haberfeld, Y. Cohen, and N. Lewin-Epstein (2000). Racial composition and occupational segregation and inequality across American cities. *Social Science Research 29* (2), 175–187.

Stock, J.H. and M.W. Watson (2007). *Introduction to Econometrics.* Pearson.

U.S. Census Bureau, Population Division (2007). 2007 Population Estimates. U.S. Census.

Verbon, H. and L. Meijdam (2008). Too many migrants, too few services: a model of decision-making on immigration and integration with cultural distance. *Journal of Population Economics 21* (3), 665–677.

Waldinger, R. (2000). Still the promised city?: African-Americans and new immigrants in postindustrial New York. Harvard University Press.

Waldorf, B., J. Beckhusen, R. Florax, and T. de Graaff (2010). The role of human capital in language acquisition among immigrants in U.S. metropolitan areas. *Regional Science, Policy and Practice 2* (1), 1757–7802.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data. The MIT Press.

Table 1. Distribution of employed Chinese and Mexican immigrants across occupations, in percent.

	Chinese Immigrants	Mexican Immigrants
Food Preparation and Serving Related Occupations	12.50	13.10
Computer and Mathematical Occupations	9.16	0.28
Management Occupations	8.82	3.21
Life, Physical, and Social Science Occupations	8.53	0.19
Sales and Related Occupations	7.78	5.19
Office and Administrative Support Occupations	6.60	4.08
Production Occupations	6.35	13.90
Architecture and Engineering Occupations	5.62	0.43
Education, Training, and Library Occupations	5.29	1.26
Personal Care and Service Occupations	4.91	3.42
Business and financial operations occupations	4.83	0.71
Transportation and Material Moving Occupations	4.43	9.27
Health Diagnosing and Treating Practitioners Occupations	3.76	0.32
Healthcare Support Occupations	3.43	1.20
Building and Grounds Cleaning and Maintenance Occupations	2.47	18.10
Construction and Extraction Occupations	1.58	14.90
Installation, Maintenance, and Repair Occupations	1.07	3.07
Arts, design, entertainment, sports, and media occupations	0.88	0.66
Health Technologists and Technicians	0.86	0.32
Legal Occupations	0.50	0.08
Fire Fighting and Law Enforcement Workers	0.38	0.22
Community and Social Service Occupations	0.23	0.31
Farming, Fishing, and Forestry Occupations	0.00	5.66
Total	100.00	100.00

Notes: The source is the 2010 American Community Survey (ACS). The data includes those who immigrated at age 25+, who are not in school and reside in metropolitan areas. Census based person weights were used to calculate the percentages. The total number of weighted observations is 260,470 Chinese immigrants and 1,109,295 Mexican immigrants.

Table 2. Summary Statistics from the American Community Survey (ACS) for Chinese and Mexican Immigrants.

Variable	Description	Chinese Immigrants		Mexican Immigrants	
		Mean	Std Dev	Mean	Std Dev
english	=1 if speaks English well or very well	0.622	0.485	0.366	0.482
citizen	=1 if citizen	0.485	0.500	0.179	0.383
sojourn	Years in US	14.03	8.626	13.32	8.832
married	=1 if married	0.797	0.403	0.621	0.485
female	=1 if female	0.484	0.500	0.372	0.483
age	Age in years	47.96	9.722	45.22	9.868
inctot	Total personal income in $10,000	5.324	5.424	2.390	2.448
lhs	=1 if less than high school	0.193	0.394	0.583	0.493
bd	=1 if bachelors degree	0.517	0.500	0.084	0.277
echild	=1 if there is an English-speaking child in the household	0.239	0.426	0.382	0.486
eteen	=1 if there is an English-speaking teenager in the household	0.174	0.379	0.310	0.462
eadult	=1 if there is an English-speaking adult in the household	0.706	0.455	0.607	0.488
p^h	Residential concentration	0.067	0.081	0.132	0.099
p^w	Occupational concentration	0.061	0.060	0.287	0.209
μ	Fraction of time spent away from work	0.584	0.119	0.612	0.108
μp^h	Residential meeting probability	0.039	0.049	0.081	0.063
$(1-\mu)p^w$	Occupational meeting probability	0.025	0.026	0.111	0.090
ℓp^h	Residential concentration in 2000	0.050	0.063	0.126	0.110
ℓp^w	Occupational concentration in 2000	0.036	0.043	0.218	0.193

Notes: All variables except residential and occupational concentration in 2000 are from the 2010 ACS while the latter are from the 2000 ACS. The sample includes those who immigrated at age 25 or older, who are not in school and reside and work in metropolitan areas. In order to compute statistics that are representative of the population, Census based person weights are used to calculate means and standard deviations. The total number of weighted observations is 260,470 Chinese immigrants and 1,109,295 Mexican immigrants. Partly due to these large sample sizes, all observed differences in means between Mexican and Chinese immigrants are statistically significantly different from zero.

Table 3. Marginal Effects[a] from Probit (Model I, II, III) and IVProbit (Model IV) Models for Chinese Immigrants.[b]

	Model 1		Model II		Model III		Model IV	
citizen			0.168	***	0.176	***	0.440	***
			(0.033)		(0.036)		(0.088)	
sojourn	0.016	***	0.008		0.005		0.013	
	(0.003)		(0.006)		(0.006)		(0.016)	
sojourn²	0.0001		0.0001		0.0002		0.0005	
	(0.0002)		(0.0001)		(0.0002)		(0.0004)	
age	-0.056	***	-0.041	***	-0.049	***	-0.125	***
	(0.004)		(0.012)		(0.013)		(0.032)	
age²			0.0002	**	0.0003	***	0.0009	***
			(0.0001)		(0.0001)		(0.0003)	
married			-0.019		-0.102	***	-0.255	***
			(0.035)		(0.036)		(0.088)	
female			0.066	**	0.06	**	0.144	**
			(0.027)		(0.028)		(0.073)	
inctot			0.034	***	0.034	***	0.086	***
			(0.006)		(0.006)		(0.012)	
lhs			-0.271	***	-0.268	***	-0.680	***
			(0.038)		(0.038)		(0.093)	
bd			0.485	***	0.504	***	1.278	***
			(0.031)		(0.033)		(0.088)	
echild					0.056	*	0.140	
					(0.033)		(0.09)	
eteen					0.009		0.023	
					(0.035)		(0.092)	
eadult					0.255	***	0.641	***
					(0.031)		(0.079)	
p^h	-3.705	***	-3.734	***				
	(0.492)		(0.702)					
$(p^h)^2$	7.593	***	10.291	***				
	(1.819)		(2.605)					
sojourn × p^h	0.019		0.013					
	(0.019)		(0.018)					
μp^h					-4.014	***	-12.113	***
					(1.078)		(4.219)	
$(1-\mu)p^w$					-2.636	**	-6.95	
					(1.317)		(7.578)	
$(\mu p^h)^2$					12.918	**	43.362	*
					(5.811)		(23.32)	
$((1-\mu)p^w)^2$					13.642	*	26.265	
					(8.036)		(75.094)	
$\mu p^h \times (1-\mu)p^w$					18.181		72.401	
					(14.059)		(53.625)	
Log Like	-1364.8		-890.92		-901.49			
Pseudo-R²	0.16		0.45		0.47			
Wald Test[c]							1.51	

Notes:

[a] Marginal effects are calculated for 45 year-old married males who have lived in the United States for 14 years and reside with at least one English-speaking adult at home.

[b] *, **, and *** indicate significance at α=0.1, 0.05 and 0.01 levels, respectively.

[c] Tests the null hypothesis of exogeneity in the model. In model IV, we do not reject this hypothesis (prob=0.91).

Table 4. Marginal Effects[a] from Probit (Model I, II, III) and IVProbit (Model IV) Models for Mexican Immigrants.[b]

	Model 1		Model II		Model III		Model IV	
citizen			0.28	***	0.267	***	0.675	***
			(0.017)		(0.017)		(0.042)	
sojourn	0.018	***	0.013	***	0.012	***	0.031	***
	(0.002)		(0.003)		(0.003)		(0.007)	
sojourn²	-0.0001		0.0001		-0.0001		-0.0001	
	(0.0001)		(0.0001)		(0.0001)		(0.0002)	
age	-0.008	***	0.002		-0.005		-0.014	
	(0.0008)		(0.006)		(0.006)		(0.016)	
age²			0.0001		0.0001		-0.0001	
			(0.0001)		(0.0001)		(0.0002)	
married			0.017		-0.012		-0.030	
			(0.013)		(0.013)		(0.034)	
female			-0.038	***	-0.075	***	-0.190	***
			(0.013)		(0.014)		(0.034)	
inctot			0.028	***	0.024	***	0.061	***
			(0.005)		(0.005)		(0.008)	
lhs			-0.259	***	-0.234	***	-0.587	***
			(0.013)		(0.013)		(0.034)	
bd			0.231	***	0.210	***	0.531	***
			(0.024)		(0.024)		(0.060)	
echild					0.012		0.031	
					(0.013)		(0.034)	
eteen					-0.019		-0.049	
					(0.014)		(0.035)	
eadult					0.193	***	0.484	***
					(0.013)		(0.034)	
p^h	-1.179	***	-1.054	***				
	(0.191)		(0.211)					
$(p^h)^2$	2.112	***	1.777	***				
	(0.513)		(0.569)					
sojourn × p^h	-0.018	***	-0.011					
	(0.006)		(0.007)					
μp^h					-0.549	*	-1.212	
					(0.324)		(0.889)	
$(1-\mu)p^w$					-1.085	***	-2.446	***
					(0.207)		(0.613)	
$(\mu p^h)^2$					-0.997		-2.982	
					(1.254)		(3.419)	
$((1-\mu)p^w)^2$					1.101	***	1.579	
					(0.429)		(1.314)	
$\mu p^h \times (1-\mu)p^w$					0.207		-0.452	
					(1.206)		(3.465)	
Log Like	-5311.17		-4540.38		-4384.88			
Pseudo-R²	0.03		0.17		0.20			
Wald Test[c]							7.54	

Notes:

[a] Marginal effects are calculated for 45 year-old married males who have lived in the United States for 14 years and reside with at least one English-speaking adult at home.

[b] *, **, and *** indicate significance at α=0.1, 0.05 and 0.01 levels, respectively.

[c] Tests the null hypothesis of exogeneity in the model. In model IV, we do not reject this hypothesis (prob=0.18).

Table 5. Correlations of the instruments with the original variables and residuals of the initial model

	Chinese Immigrants			Mexican Immigrants		
	Residual	p^h	p^w	Residual	p^h	p^w
2000 p^h	-0.059	0.955		-0.021	0.972	
2000 p^w	-0.013		0.610	-0.064		0.915