Hidden Gems? Differential Hiring and Self-Employment of U.S. Immigrants

Yoonha Kim*

Working Draft: October 2016

(*Click here for the most recent version*)

Abstract

In this paper, I study the effect of linguistic-cultural differences on firms' acquisition of human capital. I show that these differences produce a systematic misallocation of scarce talent: highly educated, foreign-born workers more likely sort out of salaried work, and into self-employment, than otherwise similar U.S.-born individuals. This differential sorting can be theoretically understood as a rational, but flawed, response to the difficulties of credibly signaling capabilities—the cultural distance between the employer and the candidate generates noisy signals, when precise signaling is more critical for applicants to the more demanding jobs. Using the American Community Survey and the Current Population Survey and measuring cultural mismatch with "linguistic distance", I find evidence consistent with theory: not only the highly educated—who apply to more demanding jobs—but also the linguistically distant—who send noisier signals of ability—disproportionately sort into business ownership. I find a mitigating effect for immigrants who have culturally assimilated or who themselves compose a majority. Furthermore, I show that immigrants' lack of English language skills, among other potential drivers, does not, in and of itself, explain the differential sorting. Rather, the pattern appears to reflect inefficient allocation of talent; thus firms might leverage these hidden gems—the untapped talent pool of highly educated immigrants sorting into self-employment—to gain competitive advantage.

* PhD Candidate at the Haas School of Business, University of California at Berkeley; yoonha@berkeley.edu
I am particularly grateful to my PhD advisors, John Morgan, Ross Levine, Noam Yuchtman and David Card for their invaluable support and guidance. I would also like to thank Victor Bennett, Matthew Bidwell, Seth Carnahan, Christian Catalini, Ronnie Chatterji, Bo Cowgill, Ernesto Dal Bó, April Franco, Deepak Hegde, Olenka Kacperczyk, Hyoseok Kang, Do Yoon Kim, Yongwook Paik, Minjung Park, Chris Rider, Yona Rubinstein, Rob Seamans, Jordan Siegel, Steve Tadelis, Romain Wacziarg and Reed Walker, as well as participants at the Consortium on Competitiveness and Cooperation Doctoral Conference, the Wharton People & Organizations Conference and the Institutions and Innovation Conference.

1 Introduction

Mr. Lee[§] had originally planned to become a Certified Public Accountant when he first moved to the United States. After several years, however, he ultimately became disillusioned about making it in corporate America. Instead, he decided to open up a sushi restaurant. Mr. Lee now owns several popular Asian restaurants in downtown Berkeley.

Stories of immigrant entrepreneurs such as Mr. Lee are not uncommon. Asian immigrants often form their own businesses out of necessity, as they struggle to find and maintain jobs. Although Mr. Lee had hoped to become a successful worker in a major accounting firm, events did not unfold as he planned. What can we generalize from Mr. Lee's experience?

Using individual-level data from the American Community Survey (ACS; 2005 – 2012) and the March supplements of the Current Population Survey (CPS; 1994 – 2012), I document that foreign-born workers with higher education have disproportionately greater tendencies to run businesses than otherwise similar U.S.-born individuals. That is, immigrants are more likely to "positively sort" into self-employment with education than their U.S.-born counterparts. This is a remarkable pattern as it suggests that U.S.-born and the foreign born workers do not simply sort differently into business ownership. Rather, their differential sorting systematically increases as a function of educational attainment.

While an extensive literature studies immigrants' propensities to enter into self-employment, existing theories do not account for the pattern of differential sorting based on education. The potential role of ethnic enclaves (Borjas 1986), social networks (Kerr and Mandroff 2015) or taste and norm for self-employment (Slezkine 2006) will not necessarily be stronger for the highly educated. Similarly, racial and ethnic preferences against immigrants (Becker 1957) would not systematically differ by education; or information based theories of perpetuating differences in labor market outcomes, and how that generate differences in human capital investment (Lundberg and Startz 1983, Cornell and Welch 1996), would not explain differences conditional on educational attainment. Hence, the positive sorting pattern suggests that there exist frictions that differentially affect immigrants, beyond the standard forces affecting employment choices in the labor market.

To explain this pattern, I propose a hypothesis of informational asymmetry based on linguistic-cultural differences. In contrast to existing studies that primarily investigate how language skills shape immigrants' labor market outcomes (Ferrer et al. 2006, Peri and Sparber 2009, Lewis 2011, Imai et al. 2014), I demonstrate how language can cause frictions even when an immigrant does not have communication problems in the typical sense. I argue that because language embodies culture and permeates individuals' verbal and non-verbal social interaction, an immigrant may have a harder time effectively conveying her capabilities in her job search. This can especially cause problems for the highly educated; path dependence in hiring practices generate persisting differences in business ownership patterns. I further show that a lack of English language skills, which renders immigrants unable to perform to expectations, does not, in and of itself, explain the differential sorting. Therefore, the pattern appears to reflect frictions in the labor market preventing

§ I am grateful to the anonymous business owner for sharing his experience.

sufficiently capable workers to appropriately match with firms. How can firms successfully recruit from a diverse workforce in the face of such challenges?

To better understand how firms' ability to identify talent shapes workers' employment choices, this paper (1) describes a theoretical framework that illustrates how linguistic-cultural mismatch may shape differential business ownership patterns, (2) tests predictions that emerge from this framework, and (3) evaluates the predictions from possible alternative theoretical explanations. Based on my findings, I discuss broader implications for how firms can develop more effective hiring strategies.

First, I describe a theoretical framework that predicts how linguistic-cultural differences generate positive sorting of immigrants into business ownership. Suppose that in a hiring interview employers receive verbal and nonverbal signals of ability from candidates to form beliefs about their ability to perform the job. Communication is more effective when two people share the same linguistic cultural background; hence, immigrant candidates are more likely to send noisier signals than their U.S.-born counterparts. This implies that, the employer will change her belief less about an immigrant candidate relative to a U.S.-born candidate with the same level of human capital. As a result, the employer may make rational, but flawed, judgements failing to hire immigrant candidates even when they are sufficiently competent. This especially causes problems for the more high-end jobs, in which the employer has to update her belief to a greater degree. Hence, highly educated immigrants—who fail to match with firms—disproportionately form their own businesses.

Among theoretical models based on this economic intuition of cultural mismatch, Morgan and Várdy (2009) show how employment outcomes between two groups may differ solely based on differences in the preciseness of signals of abilities, even when the average beliefs about the abilities are the same. In other words, immigrants can be as equally capable as the U.S.-born candidates, yet, differences in linguistic-cultural backgrounds alone can induce different hiring outcomes. The difference in employment outcomes between the immigrants and U.S.-born candidates grows with employers' uncertainty, which is determined by a) the difficulty of the job and b) the relative noisiness of candidates' signals. These two factors give rise to informational frictions that drive heterogeneous selection into self-employment.

Second, I empirically test predictions of this theoretical framework. In order to do so, I need proxies for the difficulty of the job and the noisiness of signals. To proxy for the difficulty of the job, I assume that highly educated workers have higher propensities to apply to difficult jobs. To measure the noisiness of signals, I exploit the fact that immigrants come from a diverse set of linguistic-cultural backgrounds. Specifically, I apply a country-level measure of "linguistic distance" constructed by Wacziarg and Spolaore (2009). This measure is built on Fearon's (2003) approach of tracing the number of branches that separate two languages in a language tree. Hence, the main explanatory variables of interest are educational attainment and linguistic distance, as well as the interactions between the two sets of variables.

The theoretical prediction that informational friction increases with the difficulty of the job and the noisiness of the signal is supported by my data: the likelihood of immigrant self-employment systematically increases not only with higher education levels but also with greater linguistic distance. I examine linear probabilities of a worker owning a business as opposed to being a salaried worker using the ACS and the CPS. My empirical results suggest that linguistically distant

immigrants are, on average, 23-40% more likely to enter into self-employment than similarly qualified U.S.-born workers and that this effect is larger for the highly educated: with an additional year of education, the likelihood for the linguistically distant to self-employ increases by 3-5%.

I further validate additional predictions of the model in two ways. First, I show that immigrants who have culturally assimilated, and thus would have the same signal precision as their U.S.-born counterparts, would not face this problem. In support of this hypothesis I find a mitigating effect for those who immigrate before the age of 10 or who was exposed to the U.S. education system. Second, I test whether the predictions of the model can be generalized beyond the context of a U.S.-born employer hiring an immigrant worker. Consistent with the framework, I find evidence that immigrants working in industries or residing in areas densely populated by their co-ethnics are less likely to enter into self-employment.

Third, I evaluate the predictions of the framework relative to the predictions from alternative hypotheses. Among other potential drivers, I particularly investigate whether linguistic distance simply capture workers' lack of communication skills. If more difficult jobs were more communication intensive, a higher propensity to sort into self-employment with higher education and linguistic distance may simply reflect sorting based on English proficiency rather than cultural differences.

I address this concern in two ways. First, I build on Autor et al.'s (2003) to decompose occupations by their skill requirements, using the O*Net Skill scores, a normative measure of skills created by the Department of Labor. If communications skills were an important productivity input, then language deficiency would be more likely to damage workers in communication-intensive occupations. If this were true, we should empirically observe a stronger sorting into self-employment for the subset of workers in jobs that require more communication skills. In support of the imprecise signaling hypothesis, I show that the sorting effect for jobs that are less communication intensive is qualitatively similar to that of those that take language ability as an important input. Thus, I reject the hypothesis that linguistic distance simply measures language as a productivity input.

Second, I complement the linguistic distance measure with individuals' self-reported English scores. The border between a lack of communication skills and imprecise signaling owing to cultural differences is often indistinct. However, I show that my results hold on a subsample of immigrants who report to speak English well. The fact that the theoretical predictions hold even when accounting for a more direct measure of English language ability supports the capacity of linguistic distance to measure something other than English proficiency.

By exploiting the variation in job skill requirements and immigrants' language skills, I show that the differential sorting between immigrants and non-immigrants in the labor market does not reflect immigrants' inability to communicate. I argue that linguistic-cultural mismatch importantly accounts for this differential sorting, representing a systematic bias: firms make false negative judgements in evaluating talented immigrants who could perform well. Moreover, the effect of misallocation of talent owing to linguistic-cultural differences is pernicious as immigrants like Mr. Lee self-select into business ownership in anticipation of their outcome; hence, firms can gain competitive advantage by adjusting for such talent misallocation.

Then how should firms effectively screen candidates in the presence of search friction? I argue that in some cases firms should implement hiring practices that apply a more lenient standard to people of different cultural and ethnic backgrounds. This suggestion contrasts that of some prior studies that suggest double blind procedures to resolve social biases (Goldin and Rouse 2000), but resonates with the handicapping principle in the contest literature, in which Ridlon and Shin (2013) suggest that giving a boost to those with a disadvantage yields better outcomes when there is severe heterogeneity. Alternatively, firms may minimize the effect of cultural noise by investing in their HR division to hire people who can decipher immigrant's signal better.

I further estimate the economic gains to society from a firm hiring a highly educated foreigner who would otherwise start her own business. Previous studies have shown that immigrants are prone to become proprietors of businesses such as dry cleaners or motels (Kerr and Mandroff, 2015) and moreover that entrepreneurs with necessity-based motives are less productive than those with positive motives (Fossen and Buttner 2013, Sauermann and Cohen 2010). Estimations assessing the partial equilibrium wage effect of correctly identifying an immigrant worker suggest that society may potentially gain ~$3,000–$6,000 for an average worker and ~$7,000–$15,000 for a highly educated worker, per year. Notably, there is more to gain from alleviating this informational friction problem for the highly skilled.

While policy discussions have focused on seeking talented workers from outside the US, I hope to inform policy makers and firms about the potential gains from searching inside the US. A number of studies have examined the impact of the H-1B program on US innovation (Kerr and Lincoln 2010, Kerr et al. 2015)[1]. While understanding the effect of that particular segment of the immigration labor force is crucial, only 650,000 of the 41 million immigrants are estimated to be H-1B visa holders[2]. I argue that by adjusting their managerial practices, firms can better harness the untapped talent pool of high skilled immigrant workers, who are abundant but hidden within the U.S.

2 Literature Review

This study relates to several strands of the literature. First, this paper contributes to the literature on the economic effects of language (Wacziarg and Spolaore 2009, Bleakley and Chin 2010). In particular, I add to this literature by investigating how imprecise signals in immigrant's job search affect their employment outcomes. More specifically, in terms of language and immigrant employment outcomes, I examine how language causes friction in the discovery of immigrant talent rather than in immigrants' ability to perform a job. Prior work on immigrants' language and occupational choice mainly considers language to be a productivity input (Chiswick and Miller 1995; Davila and Mora 2004). For example, Imai et al. (2014) show an incomplete transfer of foreign skills from the source to host country in jobs that rely more heavily on communication skills, Lewis (2011) shows how language skills drive immigrant and non-immigrant substitutability, and Peri and Sparber (2009) show how immigrants sort into manual tasks, while non-immigrant workers shift to more language-intensive jobs. In contrast to these studies, I exploit variation in job and language characteristics to investigate how immigrants face barriers in the job search due to cultural noise. While English ability in and of itself is an important consideration and explains a

[1] I exclude ~70% of H-1B Visa holders from my analyses as self-employment is not a feasible option for them.
[2] Center for Immigration Studies' estimates for 2009.

large part of the earnings differential (Ferrer et al. 2006), it is important to explore how language affects immigrants' labor market outcomes beyond the traditional channel of their ability to communicate, given that about 70% of immigrants today consider themselves to speak English well. In this evaluation, I argue that the similarity of an immigrant's first language to English matters more than her proficiency.

Second, this paper relates to the literature on manager biases in the hiring process. Autor and Scarborough (2008) examine the impact of a roll out of a hiring technology on hiring and productivity in a national retail firm; Hoffman et al. (2015) examine the introduction of an online job-testing service in low-skill service sector; Giuliano et al. (2009) examine how racial matching affects employee outcomes in large U.S. retail firms; and Oreopoulos (2011) performs an audit study in the spirit of Bertrand and Mullainathan (2004) by measuring call back rates while randomly varying visible signs for immigrants. While these studies conduct well-controlled experiments, the implications have been specific to the test settings of particular types of firms hiring a narrow group of workers. This study, by contrast, tests for generalizable labor market outcomes of imperfect screening, by using data sets representative of the US population and by exploiting different levels of immigrants' linguistic-cultural backgrounds.

Finally, this study speaks to the entrepreneurship literature. As a byproduct of such hiring bias, I uncover an unexplored mechanism that drives highly educated minority workers to open their own businesses. A number of studies have focused on motivations for self-employment (Åstebro, et al 2014) and attributed drivers to non-pecuniary benefits (Hamilton 2000; Hurst and Pugsley 2015), peer effects (Nanda and Sørensen 2010; Kacperczyk 2013) or individual traits (Lazear 2005; Levine and Rubinstein 2016). Furthermore, recent studies have discussed how the choice of entrepreneurship reflects various types of labor market frictions: unobserved ability (Hegde and Tumlinson 2015) or educational mismatch (Stenard and Sauermann 2016), cause imperfect matching between workers and firms (Åstebro et al. 2011). I build on these discussions by investigating how cultural friction, which causes false negative judgements in firms' hiring process, plays a role in sorting highly educated foreign workers into business ownership. While similar types of ethnic group-based biases have been explored in the context of venture capital financing (Hegde and Tumlinson 2014) or R&D alliance formation (Joshi and Lahiri 2014), the role of cultural friction in matching workers to firms in the labor market and its effect in shaping patterns of small business formation, have not been empirically tested.

3 Theoretical Framework

In this section, I describe the underlying economic intuition for how immigrants in the U.S. differentially sort into self-employment in comparison with non-immigrants.

Suppose that during hiring interviews candidates send signals of their ability and employers have to interpret those verbal and nonverbal cues to form beliefs about whether the candidate can effectively perform the job. However, suppose that immigrants send less precise signals than non-immigrants owing to differences in their linguistic-cultural backgrounds: perhaps an immigrant applicant will likely use language differently or adhere to different social norms than a U.S.-born individual. Such linguistic cultural differences make it more difficult for immigrants to accurately signal their true productivity type.

There are different theoretical models building on this intuition, including Lang (1986), Cornell and Welch (1996) and Morgan and Várdy (2009). While Lang (1986) assumes that there is cost of communication among members from different groups, Cornell and Welch (1996) and Morgan and Várdy (2009) are similar in that they both assume that difference in linguistic-cultural backgrounds can be costly because they generate larger noise in productivity signals. However, the two models differ in that Cornell and Welch (1996) presumes that candidates with noisier signals are judged to be worse while Morgan and Várdy (2009) suggest how even when the employer holds the same belief about their ability there may be differential outcomes.

The underlying intuition is as follows. Suppose that the employer screens in an unbiased manner, where she will hire if she believes that the candidate can perform to expectations. Depending on the nature of the job, however, the employer may be more or less selective: when the talent to perform the job is abundant, the employer is less worried about getting the hiring right, while when the talent to perform the job is scarce, the employer becomes more selective as she becomes more concerned about having to incur the cost of firing the candidate. To avoid this cost, the employer has higher demands when screening for more difficult jobs. Typically, these are jobs in which highly educated candidates compete.

In these jobs, although the threshold for the employer's posterior belief is exactly the same for immigrants and non-immigrants, an immigrant needs to send a stronger signal in order to satisfy the same threshold because her signal is noisier. Hence, when imprecise signaling is taken into account, a gap exists in the signal levels needed to induce the required posterior belief between immigrant and non-immigrant candidates. This gap grows with the employer's uncertainty, which is determined by a) the difficulty of the job and b) the relative noisiness of the candidate's signal. Employers are thus more likely to make false negative judgements about highly educated immigrant candidates, who send noisy signals and who apply to difficult jobs.

Suppose that in the case of a failed job search, candidates who failed to match with existing firms enter into self-employment rather than to accept an offer for a salaried job that pays less. Hence, heterogeneous sorting into self-employment arises, where this sorting is linked to the difficulty of the job. Since the more highly educated will apply for the more difficult jobs, this model fits the business formation patterns of immigrants very well:

Prediction 1: Immigrants are more likely to positively sort into self-employment than otherwise similar non-immigrants.

The differential sorting between immigrant and non-immigrant candidates with a given set of abilities will become more pronounced when minorities send noisier signals. Depending on their familiarity with the English language and U.S. culture, the noisiness of immigrants' signals varies. This leads to the following proposition predicting differential sorting across subsets of the immigrant population:

Prediction 2: Immigrants with more noisy signals will have greater tendencies to enter into self-employment.

An employer's belief of whether a candidate can perform to expectation is contingent on the nature of the job, where noisy signals matter more when the employer is hiring for a more difficult job. For these jobs, immigrants need a stronger signal than their U.S.-born counterparts to sufficiently

increase the employer's posterior belief above the hiring threshold. Given that more highly educated individuals compete for more difficult jobs, the theory further predicts differential sorting across subsets of the immigrant population and across education categories:

Prediction 3: Immigrants with both more noisy signals and more education will have greater tendencies to enter into self-employment.

Relative to other models of statistical discrimination, in which differences in population means of the signal generate differences in labor market outcomes, this model depends on differences in the preciseness of signals and hence the variance of the distribution of talent. Thus, immigrants who have completely assimilated culturally—whose signals are just as precise as that of U.S.-born candidates—should not face this problem. This motivates the following prediction:

Prediction 4: Immigrants who immigrate at a young age should not suffer from an imprecise signaling problem and hence should sort into self-employment less than otherwise similar immigrants.

While immigrants are a natural group to associate with noisy signals, the implications of the model can be interpreted more generally as a mismatch in discourse systems that may occur in any dyadic relationship between an interviewer and an interviewee. Thus, in settings where immigrants compose the majority group, and hence where the employer is more likely to be from the same ethnic group, they would not suffer from this informational friction. This leads to the following testable prediction:

Prediction 5: Immigrants are less likely to enter into self-employment when they themselves compose the majority group

While the theoretical framework is specific to an interview setting, the implications of the model are not confined to the hiring process. First, promotion decisions could also be viewed as an organic hiring decision. I argue that a manager-level job requires a different set of skills than an entry-level job; hence, promoting a worker can be viewed as hiring her for a new role. Second, the model also has implications for employee retention. Firing decisions affect minority groups in a similar manner to hiring decisions. Hence, the long-run workforce composition that we observe in the labor market would be a more skewed version of the composition that is initially suggested by the model, which is that minorities are systematically underrepresented in jobs in which talent to perform the job is relatively scarce. While Rivera (2012) shows how cultural matching is an important factor in hiring decisions in elite firms, Petersen and Saporta (2004) note that discrimination is most heightened in the hiring setting. These suggest that understanding cultural frictions in the hiring process would provide important insights about systematic patterns in the labor market.

In the following sections, I describe the data and empirical methodology to test these propositions and rule out potential alternative factors that may be driving the predictions.

4 Data Description

To test the theoretical predictions outlined above, I use two distinct data sets to examine (a) differential selection into self-employment by U.S.-born and foreign born workers, (b) systematic

patterns of selection into self-employment across subgroups of immigrants and (c) potential alternative explanations that may be driving this pattern. I use the ACS for the years 2005 to 2012 along with the March Supplements of the US CPS for the years 1994 to 2012. Both surveys provide baseline characteristics and occupational and productivity information on individuals. While the ACS is used to present the main results, I use the CPS to further check the robustness of the results and to conduct analyses that require metro area-level divisions. The main empirical findings hold across both datasets.

Table 1 provides basic demographics and labor market outcomes for the sample, where Panel A summarizes the ACS data and Panel B summarizes the CPS data. For both surveys, I include male workers aged 18 – 65 years who worked full-time in the entire year for their work year. Calculations for both samples are weighted using the population weights provided by the respective surveys. I identify first-generation immigrants as those who and whose parents were born outside the US for the CPS and those who are indicated as foreign-born for the ACS. The indicator for self-employment versus salaried employment is the main dependent variable of interest and both surveys classify all workers as either salaried or self-employed.

Details on demographics are as follows: Whites are individuals with the race code "White alone" excluding individuals identified as Hispanics. Blacks, Hispanics and Asian are those who answered yes to "Black or African American", "Spanish/Hispanic/Latino origin", and "Asian", respectively. For educational attainment I use actual grade levels or degrees attained as well as years of education. I categorize education into five education categories: below high school, high school dropout, high school degree, some college, and college degree and above. Years of education is imputed based on the actual grade level or degree. In cases where educational attainment spans multiple grades, I take the average year of education.

Three additional observations are worth noting in Table 1. First, the overall propensity to enter into self-employment is not greater for immigrants than for the U.S. born in both samples. However, immigrants are more likely to enter into self-employment after racial categories are taken into account. In other words, whites are more likely to enter into self-employment than non-whites. Second, the difference in years of schooling and the median hourly earnings between self-employed and salaried workers are greater for immigrant workers than native workers. From Panel A (Panel B), a self-employed immigrant has, on average, 0.5 (1) more years of education than an immigrant in salaried work, while a self-employed U.S.-born worker has 0.2 (0.3) more years of education than a salaried worker. This gap is reflected in the median hourly earnings, where from Panel A (Panel B), a median self-employed immigrant earns $0.2 more ($1.1 more) per hour than a salaried immigrant, while a median self-employed native earns $0.1 more ($1.8 less) per hour a salaried worker. The fact that the education gap and earnings gap between the two employment groups is wider for immigrants provides evidence that immigrants are more likely to select into self-employment. Third, while the median earnings of the self-employed are similar to the median earnings of salaried workers for both the U.S. born and immigrants, the mean earnings of the self-employed are higher. This result suggests that the earning distributions of the self-employed have fatter right tails.

To test the specific theoretical predictions, I need proxies for the difficulty of the job and the noisiness of the signal. I proxy for the difficulty of the job with the average education level of workers employed within jobs. Therefore, the higher the worker's educational attainment, the more likely

her job will demand difficult tasks, where the employer believes that the talent to perform the job is scarce. Measuring the noisiness of signals poses a greater challenge. To overcome this challenge, I use a linguistic distance measure developed by Wacziarg and Spolaore (2009) to proxy for the cultural distance (i.e., the difference in discourse systems), between the US and the immigrant's source country. This measure is built on Fearon's (2003) approach of tracing the number of branches that separate two languages in a language tree, which previous studies have used as a summary statistic for intergroup cultural differences (Montalvo and Reynal-Querol 2005). I use a normalized measure between 0 and 1, where all U.S.-born individuals have a linguistic distance of 0 and all immigrants have some positive value of linguistic distance. The linguistic distance between the US and countries such as the UK and Australia is very close, while most Asian countries included will fall on the farthest end. For countries such as Singapore and India, which are de jure English-speaking countries, I assign a value of 0.5. I identify de jure English-speaking countries based on the Central Intelligence Agency's World Factbook. Linking these data to the ACS and CPS data provides the linguistic distance for immigrants from over 150 countries.

My interpretation of language by using linguistic distance is similar to that of Cornell and Welch (1996), where cultural beliefs and shared values are embedded in language, which affects the style of speech even after an immigrant technically acquires English as a communication tool. One concern that arises from using the linguistic distance in this manner is that it confounds immigrants' inability to communicate well with the cultural barrier they face. To address this problem, I corroborate this measure with a self-reported English ability score from the ACS, for which respondents choose among 'very well', 'well', 'not well' and 'not at all'. I use this measure to test whether linguistic distance merely captures immigrants' inability to speak English.

To further address this problem, I run more nuanced tests on subsets of occupation categories that require more or less communication skills. I follow Autor et al. (2003) to characterize jobs by using O*Net Skill scores, a normative measure of the required skill level for each standard occupation created by the Department of Labor. In particular, I use communication skills required for different jobs, which I impute by taking the average scores of reading comprehension, speaking and writing skills required for jobs. Using this measure, I am able to determine whether the occupational sorting occurs only in jobs that have language ability as an important input or whether such sorting also occurs for jobs requiring fewer communication skills.

I also consider institutional factors that shape immigrants' employment choices. In particular, I account for H-1B visa holders, whose career trajectory would likely differ from others because their immigration status ties them to a specific employer. While it is difficult to determine the stock of immigrants under H-1B visas, the annual flow of immigrants with a particular visa status is informed by the U.S. Citizenship and Immigration Services (USCIS). As H-1B visas are allocated disproportionately across countries, industries and occupations, I identify immigrant subgroups that would compose ~70% of the H-1B holders based on USCIS' FY2012 Annual Report to Congress. Specifically, I exclude Indian, Chinese and Canadian immigrants with a college degree working in universities or in computer- or engineering-related occupations from the baseline empirical results.

Finally, given that immigrants are not proportionately distributed across space, I construct two additional measures. First, to determine the different dynamics in ethnic enclaves, I create a proximate indicator for whether an immigrant resides in an enclave. Specifically, for each ethnic group, I rank metro areas by the size of the ethnic group population and identify the metro areas

that are above the 95th percentile and 99th percentile of the distribution. This measure captures slightly over half and one third of the immigrants residing in the US. I assign 1 if an immigrant resides in these metro areas and 0 otherwise. Second, I construct an indicator denoting whether an immigrant composes the majority of their organization. I proxy for an organization by using an occupation category within an industry in a metro area and assess the proportion each origin group represented in each metro area-industry-occupation cluster. If the origin group is the most represented origin group in the cluster, I identify individuals in the origin group to be part of a majority group and denote them to have "majority representation". I assign 1 for "majority representation" if the individual is part of the majority group and 0 otherwise.

5 Empirical Methodology

In this section, I discuss the empirical methodology employed to test the main predictions of the framework. I use linear probability models with an indicator for self-employment as the dependent variable and individual- or origin country-level characteristics as explanatory variables. I use the following specification to test prediction 1, which concerns immigrants' positive selection into self-employment by immigrants:

$$SelfEmp_i = \beta_0 + \beta_1\, EducCategories_i + \beta_2\, Immig_i + \beta_3\, Immig_i \times EducCategories_i + \beta_4\, X_i + \epsilon \quad (1)$$

SelfEmp is an indicator for self-employment that takes a value of 1 for self-employment and 0 otherwise. In all of the regressions using this indicator as the dependent variable, the sample is limited to either salaried or self-employed workers who have worked full-time for the reported year. Hence, the results of the regression can be reflected to indicate the propensity to be a self-employed rather than a salaried employee. *Immig* is an indicator for first-generation immigrant status, that takes a value of 1 if the individual and both of the individual's parents are foreign born and 0 otherwise. *EducCategories* are education groups including less than a high school degree, high school dropout, a high school degree, some college education, a college degree and an advanced degree. In most cases, I group the last two categories as college degree and above. X includes individual-specific controls, such as race categories, years spent in the US and the natural log of the GDP per capita of the origin country. The specification also includes fixed effects for age, year, state, occupation, industry and time spent in the US. For non-immigrants, I assign age for the number of years spent in the US. These fixed effects for year and time spent in the US account for the selection of immigrants from their host countries depending on the year of immigration and the change in immigrant's business ownership rates over time (Borjas, 1987, Clark and Drinkwater 2000, Fairlie and Lofstrom 2013).

In equation (1), a positive value for the combination of coefficients $\beta 2$ and the respective $\beta 3$s for each education category indicates that an immigrant is more likely to self-employ than a U.S.-born individual conditional on the education category, and a positive coefficient of $\beta 3$ indicates that an immigrant is more likely to enter into self-employment for that particular education group. For the purpose of testing prediction 1, concerning positive selection into self-employment, the $\beta 3$ coefficients should significantly increase with higher educational attainment.

Then to test for prediction 2, which suggests that the selection effect should be stronger for the more linguistically distant and the more highly educated, I use the following specification:

$$SelfEmp_i = \beta_0 + \beta_1 LingDist_i + \beta_2 YrsEduc_i + \beta_3 X_i + \epsilon \qquad (2)$$

The main coefficient of interest here is $\beta 1$. The coefficient indicates the additional likelihood that an immigrant who speaks the most distant language will self-employ in comparison with non-immigrants. LingDist is the linguistic distance measure of an immigrant depending on her country of origin. Further, YrsEduc is a continuous variable for education years, and other controls and fixed effects are the same as in the specification in equation (1). I use years of education and education categories interchangeably. In cases such as prediction 1, where the examination focuses on significant differences in the sorting effect with increased educational attainment, I use education categories. In contrast, in cases such as prediction 2, where linguistic distance is the main explanatory variable of interest and education is used as a control, or prediction 3 where the purpose is to assess its interaction with education, I use years of education.

Hence, I run the following specification to test prediction 3:

$$SelfEmp_i = \beta_0 + \beta_1 LingDist_i + \beta_2 YrsEduc_i + \beta_3 Ling.Dist_i \times YrsEduc_i + \beta_4 X_i + \epsilon \qquad (3)$$

Here, I interact linguistic distance with years of education. A positive coefficient for $\beta 3$ would provide evidence for sorting into self-employment by the linguistically distant who are highly educated.

6 Key Empirical Findings

In this section, I test the predictions delineated in section 3 by using the empirical setting and methods discussed in sections 4 and 5.

6.1 Positive selection into self-employment by immigrants

In this section, I test prediction 1, namely, there is positive selection into self-employment by immigrants relative to otherwise similar, non-immigrants. Positive selection into self-employment by immigrants suggests that immigrants are more prone to be business owners with higher education.

Average years of schooling from Table 1 provides suggestive evidence for such selection. The ACS sample suggest that the education gap between self-employed and salaried workers is 0.5 years for immigrants and 0.2 years for the U.S. born. Similarly, the CPS sample suggest that the self-employed have about 1 more year of education than salaried workers among immigrants, while the gap is only 0.3 years among the U.S. born.

Table 2 provides the main evidence for differential selection into self-employment by immigrants depending on education level. Panel A shows the result based on the ACS, while panel B shows result based on the CPS. The increasing coefficients for the interaction terms between immigrants and education level indicate stronger positive selection by immigrants compared with U.S.-born individuals. The coefficients from columns (2) and (5) can be interpreted as follows: immigrants with a high school degree (college degree) are 4.5% (6.3%) more likely to enter into self-employment than immigrants without a high school degree. The coefficients are statistically

significant at the 1% level, and they are statistically significantly increasing with the level of education for immigrants.

Note that the selection effect is mitigated for immigrants with a US education. The similar magnitude of the coefficients for those with some college education in the US and those who completed their degree in the US suggests that the offsetting effect is driven by cultural adjustment rather than credential acquisition, in line with the hypothesis that cultural differences push immigrants to enter into self-employment. This finding contrasts that of Hegde and Tumlinson (2015), who argue that immigrants suffer from sending credible signals of their ability, but resembles that of Ferrer et al. (2006), who argue that immigrants who completed their degrees abroad lack "usable" cognitive skills in the labor market. I argue that it is the imprecise signal owing to cultural differences that affects immigrants' employment outcomes.

My baseline empirical results exclude immigrant subgroups that are most represented among H-1B holders. Columns (1), (2), (4) and (5) report findings excluding this subgroup. For comparison purposes, columns (3) and (6) show the results with H-1B visa holders included. As expected, the differential selection effect between an immigrant with a high school degree and an immigrant with a college degree loses significance, especially in column (6). This result is not surprising, as it is well known that companies such as Google, Amazon, and EBay hire Indian developers under the H-1B visa program. These immigrant workers are specifically tied to the employer who sponsors their H-1B visa.

6.2 Informational frictions and positive selection

Table 3 reports the results testing predictions 2 and 3. Prediction 2 holds that different degrees of informational frictions should account for sorting into self-employment, and prediction 3 argues that such frictions are most acute for the highly educated. Hence, the explanatory variables of interest are linguistic distance and its interaction with years of education. A positive coefficient would suggest that both the linguistically distant and the highly educated linguistically distant are more likely to enter into self-employment. Panel A reports result based on the ACS, while panel B shows result based on the CPS.

The increasing sorting effect with linguistic distance, which supports prediction 2, holds even when I control only for race and educational attainment and include age fixed effects as shown in Columns (1) and (6). The coefficient decreases when I include fixed effects for industry, occupation, state, and year, as shown in columns (2) and (7). The effect is stronger, however, in columns (3) and (8) when I include time spent in the US and the GDP per capita of immigrants' source country. These results suggest that the noise effect is stronger when I account for immigrants' backgrounds, such as their experience in the U.S. and the characteristics of their host country. The interpretation of the coefficient from the ACS (CPS) results is as follows: the most linguistically distant immigrants are 3.0% (5.1%) more likely to sort into self-employment than comparable U.S.-born individuals. This is a meaningful increase, as it translates into a 23% (40%) increase over the base self-employment rate of 13% (13%).

Furthermore, in support of prediction 3, which suggests that there is a stronger sorting effect for the linguistically distant and the more educated, the coefficient for the interaction term between linguistic distance and education is positive. Columns (4) and (9) show a positive coefficient for the

added interaction term between linguistic distance and years of education in the specification in column (3). While the main coefficient for linguistic distance becomes negative, this is recouped with 8 years of education. I show that the average increase in the likelihood to enter into self-employment by linguistically distant immigrants includes heterogeneous effects across educational attainment. With an additional year of education, the likelihood for the linguistically distant to enter into self-employment increases by 3-5%.

Linguistic distance is a good predictor for the propensity to enter into self-employment—not just between immigrants and non-immigrants but within immigrant subgroups. Columns (5) and (10) report the results for the selection effect within immigrant groups. As the results show, the most linguistically distant immigrants are 5-6% more likely to enter into self-employment than the most linguistically close immigrants. This finding is meaningful as linguistic distance may help explain how self-employment rates systematically differ across ethnic groups in the US and may furthermore serve as a coarse, but simple, summary statistic of the degree of business ownership patterns across ethnic groups.

Existing empirical studies on statistical discrimination, such as Altonji and Pierret (2001) among others, examine how an individual's true ability is revealed over time. By including time spent in the US in my specification—which would correlate with worker experience—I control for such statistical discrimination arising from the mean. The fact that there are abilities that remain uncertain to the employer even after I control for these experiences suggests that statistical discrimination on the variance also plays an important role in the labor market.

To examine whether the results that I find are threshold effects or other nonlinear effects of linguistic distance, I evaluate the differential selection by linguistic distance categories. One way of conducting such an examination is to estimate a degree of differential sorting for each immigrant worker in order to assess the likelihood of self-employment if the immigrant had not faced informational frictions. I examine the degree of differential sorting across both linguistic distance categories and education categories by using the following method:

To estimate the degree of differential sorting for each immigrant, I run the following specification:

$$Degree\,of\,Differential\,Sorting\,into\,SelfEmp_i = SelfEmp_i - \widehat{SelfEmp_i} \qquad (4)$$

where $\widehat{SelfEmp}$ is the fitted value for immigrants using coefficients from running the following regression only for U.S.-born workers:

$$SelfEmp_i = \beta_0 + \beta_1\, EducCategories_i + \beta_2\, X_i + \epsilon \qquad (5)$$

Here, X includes the standard controls including four race categories, time spent in the U.S. and fixed effects for age, industry, occupation, state and year. The underlying idea is that by assigning U.S.-born workers' coefficients to immigrants, I can estimate the likelihood that an immigrant would have entered into self-employment, were it not for her linguistic distance. A positive differential sorting measure would indicate stronger selection into self-employment in comparison with a comparable U.S.-born worker.

I first examine this measure along education to confirm the findings regarding positive selection. Figure 1.1 shows the results based on the ACS. Consistent with Table 2, the coefficients indicating differential sorting into self-employment are significantly greater with increasing education.

I then examine this measure along two dimensions: linguistic distance and education categories. This results are shown in Figure 1.2. Consistent with the theoretical predictions, as well as empirical findings, the differential sorting measure is higher for the more educated and more linguistically distant and lower for the less educated and more linguistically distant. Hence, this visual inspection of the differential sorting measure supports the theoretical predictions. The results based on the CPS, shown in Appendix Figure 1, provide similar results. The same test in regression form is reported in Appendix Table 1. This specification replaces immigrant status in equation (1) with linguistic distance. The coefficients for the interaction terms between linguistic distance and education categories are similar.

Theory predicts that the differential sorting measure would be the lowest for the category in which the talent to perform the job is abundant and the highest for jobs in which the talent to perform the job is scarce. While the overall direction of the measure fits the framework well, for some linguistic distance categories the reported differential sorting measure for those in the 'college and above' category is lower than those with 'some college' or 'a high school degree'. This result is especially driven by those with an advanced degree, beyond college. The reason for this discrepancy between theory and empirics may be that while the theoretical predictions are based on the employer's belief about talent scarcity among the applicants, education attainment may be an imperfect proxy for talent scarcity. For example, the talent to become a surgeon is scarce, but given that the applicant pool for being a surgeon is already a select group of people, screening may not be so demanding if it is conditional on having a medical degree. Conversely, talent for being an effective mid-level manager may be abundant, but if the position does not require postsecondary education, then the applicant pool may be larger, and the employer's belief about the scarcity of talent among the candidate pool may actually be lower than that of an employer hiring a surgeon. In other words, the correlation between education categories and talent scarcity may be loose especially for high-end jobs.

6.3 Self-employment and age of immigration

The model suggests that immigrants who have completely assimilated culturally should not face this problem. I exploit the fact that cultural assimilation naturally interacts with immigration age and that immigrants who immigrate at a young age do not suffer from the noisy signaling problem to test prediction 4: immigrants who come at a young age should self-employ less than otherwise similar immigrants. In Table 4, I examine whether the effects of language discrepancies are different for immigrants who come at a young age by using the following specification for immigrants:

$$SelfEmp_i = \beta_0 + \beta_1 LingDist_i + \beta_2 Immigration\,Age_i + \beta_3 Ling.Dist_i \times Immigration\,Age_i + \beta_4 X_i + \epsilon \quad (6)$$

X includes the standard controls including four race categories and fixed effects for age, industry, occupation, state and year. While Bleakley and Chin (2010) compared social outcomes for immigrants depending on their age of immigration, I study the relation between age of immigration and employment outcomes.

Again, I report the results by using both the ACS and the CPS. I add a variable that indicates whether an immigrant came to the US before the age of 10 in columns (1) and (5) of Table 4 to the standard specification used in columns (2) and (6) of Table 3. The negative coefficients for this variable indicate that coming to the US at a younger age as an immigrant has a mitigating effect on the propensity to sort into self-employment.

I show that the overall mitigating effect among immigrants who came before the age of 10 is mostly driven by the linguistically distant immigrants and that the linguistically distant who come at a young age no longer face the imprecise signaling problem. In columns (2) and (6) of Table 4, I add an interaction term between linguistic distance and age of arrival to the specification. The strong negative coefficient of 3.8% (4.8%) for linguistically distant immigrants who came before the age of 10 mitigates a large part of the selection effect of 4.6% (6.8%). These findings suggest that the findings from Table 3 that immigrants are 3.0% (5.1%) more likely to enter into self-employment masks the heterogeneity across immigrant groups depending on the time of their arrival to the US, particularly from the linguistically distant.

I further argue that immigrants who have been culturally assimilated from a young age may develop a very nuanced but significant skills, which are not reflected in language proficiency. Columns (3), (4), (7) and (8) advance the above analysis by identifying linguistically distant immigrants who came between 10 and 15 years of age and between 15 and 20 years of age. However, in columns (4) and (8), it is unclear whether the decrease is monotonic as one would expect. One observation is that the coefficient difference between immigrants who came before 10 and those who came between 10 and 15 is quite large and significant. While immigrants who came between 10 and 15 years of age are likely to carry somewhat of an accent, their English ability should not be so different from those that came before 10. These results suggest that the skills that immigrants who came before 10 develop, but not those that immigrants who came between 10 and 15 develop, play a significant role in the labor market matching process. This finding supports the main assertion of the theory model that there is statistical discrimination arising from the variance, rather than the quality, of candidates' signal.

6.4 Sorting when immigrants compose the majority group

While I apply theory to the setting of immigrant workers in the U.S., the model can be interpreted more generally as a mismatch between two individuals with a different cultural background. Thus, if an immigrant candidate were interviewing with an immigrant employer of the same ethnic background, she would not face this information problem. Hence, I test prediction 5: immigrants are less likely to enter into self-employment when they themselves compose the majority of their group.

While an ideal data set would identify the ethnicity of both the applicant and the recruiter, this specific information is not available. Instead, I employ two methods to identify immigrants as part of the majority group and use information about the metro area in which individuals live from the CPS.

First, I identify the metro area-industry-occupation cluster of workers and identify immigrants who are the most represented origin group in their cluster. With this information, I am able to categorize whether an immigrant is a majority or a minority within his "firm". Immigrants who are part of the

most represented origin group are assigned a value of 1 for the majority representation variable and 0 otherwise. Along with standard controls, I include fixed effects for time spent in the US, major industry categories, state and year. I do not include fixed effects for occupation categories, as the clusters are defined at the occupation level. Instead, I include a job complexity measure to account for the nature of the occupations in which immigrants self-employ. This measure uses the complex problem-solving skills measure from the O*Net Skill scores.

As shown in Table 6.1, immigrants who are part of the majority group in their workplace are 2.4% less likely to enter into self-employment than those who are not part of the majority group. This result is in line with the theoretical prediction that when immigrants compose a certain critical mass in their organization, they will face less informational friction. This result suggests that there can be path dependence in hiring practices owing to cultural mismatch and indicates how a diverse workforce can beget a diverse workforce.

Second, I examine employment outcomes of immigrants who live in ethnic enclaves. One defining characteristic of immigrants is that they are disproportionately distributed across space, in densely populated ethnic enclaves. If search friction were the only force driving immigrants to enter into self-employment, immigrants living near enclaves would face a lower language barrier, as co-ethnics come from the same discourse system. Hence, the framework would predict that immigrants residing near enclaves would be less likely to enter into self-employment than those living outside enclaves.

I test this reasoning in Table 6.2. The variable enclaves is an indicator for whether an immigrant lives in an enclave or not. For each ethnic group, I rank metro areas by the size of the ethnic group population and identify the metro areas above the 95th percentile of the distribution. I test the sensitivity of this definition by also defining enclave-residing immigrants as those who are above the 99th percentile in columns (3) and (4). The empirical results are mixed: the negative coefficients for the enclave term in column (1) suggest that immigrants near enclaves are less likely to enter into self-employment; however, this result does not hold in larger enclaves, as shown in column (3). Furthermore, the empirical results do not seem to align with the theoretical prediction when the interaction term between enclaves and linguistic distance is included, as shown in columns (2) and (4). While theory predicts that the mitigating effect should be larger for the more linguistically distant, the positive coefficients of 3.2% and 4.6% suggest that there are stronger pull factors into self-employment for more linguistically distant immigrants living near enclaves. This result contrasts findings of Battisti et al. (2016), who show that among immigrants in Germany, those who live in larger ethnic enclaves are more likely to be employed initially.

These results suggest that other dynamics related to proximity to ethnic enclaves are present. It seems likely that the linguistically distant immigrants who face signaling problems may self-employ near ethnic enclaves where they have more ethnic resources to set up a business and where they can cater to people of their own ethnic background. In this way, linguistic distance seems to bind immigrants' options to enter into self-employment outside ethnic enclaves, as it determines which customers they serve better.

7 Potential Alternative Explanations

The above results suggest that highly educated immigrants who face an imprecise signaling problem choose to enter into self-employment as they face a discount from being a salaried employee. How much of this can be explained by language proficiency or other factors? In this section, I compare the predictions of the framework with the predictions of alternative hypotheses. In particular, I investigate whether linguistic distance measures a) a lack of communication skills essential for productivity, b) distaste for unfamiliarity, or c) ethnic group-specific factors. Such factors would confound the main hypothesis that information imprecision arising from differences in discourse systems accounts for immigrants' self-employment decisions. I address these empirical challenges.

7.1 Linguistic distance as a measure for a lack of communication skills

A natural alternative interpretation to the imprecise signaling hypothesis is that immigrants' signals are as precise as non-immigrants but that linguistic distance actually measures lower productivity. If the more educated are more likely to apply to jobs that require more communication skills, immigrants may sort into self-employment with more education and greater linguistic distance because they lack the communication skills to perform the job rather than because they have an imprecise signal.

Throughout my study, I treat imprecise signals and language proficiency as if they were easily separable. In reality, it is impossible to disentangle the level of language proficiency from the noise effect arising from cultural dissimilarities: any miscommunication owing to noise will also affect others' evaluation of the immigrant's communication ability. In this section, I address this challenge in three ways.

First, I build on Autor et al,'s (2003) pioneering work to decompose occupations by their skill requirements, particularly communication intensity, to test whether the selection effect differs across jobs that require different levels of communication intensity.

Suppose that a firm's production is determined by the communication between the employer and the candidate as well as some other tangible and intangible assets, characterized as follows:

$$Y = L \cdot K = C\left(P_m, \lambda P_a\right) \cdot K$$

where Y is the output, K is the other assets of the firm, and Pm and Pa denote the language ability of the manager and the agent, respectively. C(Pm, λPa) characterizes how the complementarity between the manager and the agent is, and λ, which ranges between 0 and 1, denotes the importance of the agent's communication skills for their complementarity. If an agent does not speak English well and if communication skills are important for a job, this complementary term will be low. The first-order condition with respect to the agent's language ability is thus as follows:

$$\frac{\delta Y}{\delta P_a} = \lambda \frac{\delta C}{\delta P_a} \cdot K > 0$$

This expression suggests that the more important communication skills are for a job, the larger the agent's marginal product. Hence, if communication skills are an important productivity input,

employers have much to lose from hiring someone that does not speak English well. Hence, language deficiency would damage workers in communication-intensive occupations to a greater extent. In such a case, we should empirically observe stronger sorting into self-employment for the subset of workers in jobs that require more communication skills.

To test this hypothesis, I build on Autor et al,'s (2003) to decompose occupations by their skill requirements. Specifically, I use the O*Net Skill measure to characterize occupations by their degree of communication intensity. I take the average scores of reading comprehension, speaking and writing skills required for the job, and divide salaried occupations into jobs that require above and below median language skills in order to compare their effects regarding sorting into self-employment. Table 5 reports the results using both the ACS in panel A and the CPS in panel B.

My results show that the effects are qualitatively similar among workers across jobs requiring different degrees of communication intensity. This result is more consistent with the imprecise signaling hypothesis than with the hypothesis regarding language as an important input to productivity. Columns (1) and (3) show the results of regressions run on a subsample of salaried jobs that require low levels of language skills and all self-employment, while columns (2) and (4) compare salaried jobs that require high levels of language skills with self-employment. The coefficients of 2.5% and 5.1% in columns (1) and (3), respectively, indicate that there is a strong sorting effect—12% and 25% when the base rate of self-employment is considered even for jobs that require fewer communication skills. Moreover, when the coefficients in columns (2) and (4) are compared, the sorting effects from jobs that require higher levels of communication skills—12% and 28%—are qualitatively similar. In other words, the sorting effect does not increase as a function of the communication intensity of the job. Thus, I reject the hypothesis that linguistic distance merely measures language as a productivity input.

Second, I complement the linguistic distance measure with individuals' self-reported English scores. If linguistic distance measures technical language skills rather than cultural distance, the sorting effect should hold when self-reported English scores are used instead of the linguistic distance measure. In other words, people who are linguistically deficient should also engage in more self-employment. To test this hypothesis, I use the measure for differential sorting into self-employment by using self-reported English scores and replicate the results of Figure 1 in Figure 2.

The results using self-reported English scores show the opposite results of those using linguistic distance: more linguistically deficient individuals are less likely to enter into self-employment. This result is consistent with findings from previous studies (Portes and Zhou 1996, Fairlie and Meyer 1996). The fact that a more direct measure of English ability does not fit the main theoretical predictions supports the fact that linguistic distance indeed measures something other than English proficiency.

My use of linguistic distance bridges the incongruence between theoretical and empirical discussions on how language proficiency affects immigrants' propensity to enter into self-employment. While the disadvantage theory in the sociology literature (Light 1972, 1979) suggests that a lack of language fluency restrict immigrants' participation in salaried employment, empirical studies find a puzzling result, where an opposite effect is obtained: those who are more proficient in English are more likely to enter into self-employment in the US (Fairlie and Meyer 1996, Portes and Zhou 1996).

I show that the measure of similarity between languages, instead of immigrants' level of proficiency, correctly predicts that those who are more familiar with English are more likely to secure paid employment. In other words, the similarity of an immigrant's first language to English matters more for immigrants' job search than their proficiency in English itself and is thus better suited to assess who gets pushed into self-employment.

I also test the results of Table 2 and Table 5 on a subset of immigrants who report speaking English well in order to ensure that the main results are not driven by people who lack language proficiency. Appendix Table 2 presents results replicating Table 2. The main results hold even when the subgroup of immigrants who speak English well is used. In fact, the positive selection effect is more distinct, confirming that immigrants who do not speak English very well tend to select into self-employment less often. Appendix Table 3 reports results replicating Table 5. Columns (1) and (2) compare immigrants who report speaking English well and very well with non-immigrants, while columns (3) and (4) include only immigrants who report speaking English very well. The empirical findings for both groups confirm that the sorting effect is qualitatively similar across occupation categories requiring different levels of communication skills.

Finally, I assess the effect of cultural distance while controlling for linguistic distance. For this purpose, I employ the cultural distance measure from Wacziarg & Spolaore (2009), which is based on how similarly people from different countries have answered the questionnaires in the World Value Survey. While I could use cultural distance as the main explanatory variable throughout the study, it is less attractive as it includes only 80 countries.

I exploit variance within countries with individuals that speak the same language. For example, while countries such as Argentina or Mexico may have similar linguistic distance with respect to the US, as Spanish is the dominant language for both countries, their cultural distance from the US differs. If the linguistic distance measure serves as a proxy for cultural distance rather than the mere communication barriers that immigrants face, the effect of linguistic distance should be subdued by the inclusion of cultural distance.

Hence, I control for both cultural and linguistic distance in one regression and replicate the results in Table 3 in Table 7 based on the CPS data. As shown in column (2) of Table 7, when I include all the standard controls and fixed effects, the linguistic distance measure becomes nonsignificant while cultural distance explains the selection effect. Column (3) further shows that the positive selection effect also holds with cultural distance. Hence, these results confirm that the linguistic distance measure proxies for the noise effect owing to differences in the discourse system. Column (4) presents the results for the specification for only immigrants. Among immigrants, it seems that linguistic distance is a stronger predictor for self-employment than cultural distance.

A framework that uses communication skills as an important productivity input would not be able to explain (a) the constant selection effect across jobs that require different levels of language ability; (b) the lack of consistency of the selection effect when self-reported English scores are used and the consistency of the selection effect when the analysis is conditioned on immigrants who speak English well; and (c) the effect of cultural distance over linguistic distance. These results support the fact that immigrants' inability to speak the language does not drive the selection effects of the linguistic distance measure. Hence, linguistic distance does not merely proxy for the inability to perform jobs that take language as an important input to production.

7.2 Linguistic distance as a measure for distaste for unfamiliarity

An obvious competing hypothesis for a statistical discrimination model is taste-based discrimination (Becker 1957). Hence, in this section, I argue that linguistic distance does not merely capture distaste for differences. Taste-based discrimination, on its own, would not explain the differential sorting across education levels, as there is no reason to believe that the highly educated are systematically disliked more than the less educated.

However, it is possible that taste-based discrimination can generate positive sorting with the help of additional assumptions. Suppose that immigrants face a discount in their wage when they are salaried employees, while they can earn their ability minus some fixed cost to start a business in self-employment. In this case, there are higher returns to entering into self-employment than seeking salaried work with education. Accordingly, more highly educated immigrants would tend to enter into self-employment more often.

One way to tackle this question is to, again, exploit how language proficiency naturally interacts with acquisition age, as shown in Table 4. Those who immigrate at a young age share the same observable characteristics as those who immigrate at a later age, except they do not suffer from linguistic-cultural barriers. If it were taste-based discrimination, we should see the same effect for this subgroup of immigrants. My results showing that coming before age 10 mitigates selection into self-employment for the linguistically distant is in line with the imprecise signaling hypothesis. This result demonstrates that linguistic distance does not simply measure distaste for immigrant group-specific attributes.

While immigrants who immigrate between 10 and 15 years of age are likely to carry somewhat of an accent, their English ability should not be so different from those that immigrate before 10. One possibility is that there may be biases in the labor market arising from differences in accent. Thus, although linguistic distance does not capture racism per se, it may capture xenophobia toward those with an accent or those who are not entirely Americanized.

While my results may not entirely rule out taste-based discrimination, as taste-based biases may arise from factors other than appearances, at the very least, my results suggest that distaste for observable differences cannot entirely explain the differences in the selection effect.

7.3 Linguistic distance as a proxy for ethnic group specific factors

The last set of alternative explanations relates to ethnic group factors. A large number of studies have examined how ethnic pull factors, including enclave effects (Borjas 1986) and ethnic networks (Kerr and Mandroff 2015), drive immigrant self-employment. However, these factors alone fall short in explaining why the sorting effects vary across education-immigrant subgroups, as ethnic group-specific factors do not necessarily have a stronger effect for the more highly educated. Hence, in general, network effects or ethnic group-specific path dependencies are not a major concern as long as they do not unevenly affect immigrants across education levels. To the extent that ethnic group effects correlate with years of education, however, linguistic distance may potentially mask ethnic group effects, as the measure is defined at the country level. In this section, I show that ethnic

group factors that affect employment choices do not fully explain self-employment decisions. I address this concern in two ways.

First, I exploit within-country variations by analyzing whether the language spoken at home also predicts the employment choices of immigrants from multilingual countries such as Belgium or Switzerland. English belongs to the Indo-European language tree where its specific branches are Indo-European, Germanic, West Germanic, Anglo-Frisian, and Anglic. I exploit the fact that English shares two more branches with German or Dutch than with French. Hence, I test the hypothesis that French-speaking Swiss or Belgian individuals are more likely to sort into self-employment than the German-speaking individuals.

The results reported in Table 8 show weak support for this hypothesis. The sample includes immigrants who were born in either Belgium or Switzerland and those who speak Dutch, German or French at home. In column (1), I find strong support for the hypothesis with a coefficient of 17.2%, when I include standard controls but do not include any fixed effects. Once I include fixed effects for 22 major occupation categories, however, the result loses significance, as shown in column (2). I conjecture that the test may lose variation since immigrants from Belgium or Switzerland may be concentrated in particular occupation categories. Hence, instead, I create 6 categories of occupations constructed based on the complex problem solving skill measure from the O*Net Skill scores, to include them as fixed effects. As shown in column (3), the results regain significance; however, the test loses significance again when I add fixed effects for major industry categories, as shown in column (4).

Overall, I find weak support for the hypothesis that immigrants who speak French, which shares one less branch with English than do German and Dutch, are more likely to enter into self-employment. This result suggests that heterogeneous selection effects may exist even within ethnic groups.

Second, I test whether the positive selection effect holds for a subset of immigrant groups that do not reside in an ethnic enclave. The results for this test are presented in Appendix Table 4. While the positive selection directionally holds, as the interaction terms between the immigrant indicator and education group are significantly positive, it seems that the differential selection occurs only between those with and those without a high school degree but not between those with and those without a college degree. This result again seems to suggest that the highly educated who are pushed out of the salaried workforce self-employ near ethnic enclaves.

The results assessing the employment choices of immigrants from multilingual countries and testing whether positive sorting into self-employment still holds for a subset of immigrants residing in non-enclaves suggest that the heterogeneous selection persists even when ethnic group-specific factors are taken into account.

In this section, I rule out potential alternative explanations, including language as an input to production, taste-based discrimination and ethnic group-specific factors, that may explain why linguistic distance predicts immigrants sorting into business ownership. The series of empirical results suggest that there is systematic bias in the context of hiring immigrant workers that is not fully explained by conventional factors noted in previous work. I attribute such bias to systematic bias arising from cultural mismatch in the context of hiring immigrant workers.

8 Managerial Implications

In this section, I discuss how firms can use the findings of the paper and benefit by adjusting their managerial practices. First, I assess whether firms outperform their peers by correctly identifying the immigrant workers. Second, I discuss how and which firms may be able to benefit by investing in their HR practices. Third, I estimate the productivity gain for society at a partial equilibrium in which immigrant workers who are misallocated as self-employed are employed in firms. Finally, I discuss the limitations of the paper.

8.1 Do firms with higher immigrant representation outperform peers?

I evaluate whether firms that do not appear to have solved the misallocation problem outperform their peers, all else equal. Similar to how Siegel et al. (2014) show how multinational firms can gain competitive advantages from hiring the excluded group to positions of managerial authority, I examine whether firms conducting business domestically can gain competitive advantages by overcoming barriers to attracting immigrant workers.

I empirically test this possibility by considering metro area-industry-occupation clusters as "firms", as in section 6.4. For each cluster, I calculate the weighted average linguistic distance of the employees as well as the average performance of workers in that cluster relative to their peers in the same industry-occupation categories. The underlying idea is that if workers are assumed to earn their marginal product, the average relative performance of workers in each cluster can be interpreted as the relative productivity of a firm. To calculate the relative performance of workers, I use the residual from the following wage regression:

$$(Log)\, Hourly\, Wages_i = \beta_0 + \beta_1\, Race_i + \beta_2\, Educ_i + \beta_3\, Age_i + \beta_4\, Industry_i + \beta_4\, Occupation_i + \epsilon \quad (7)$$

Figure 3 graphs average residual wages of workers by the weighted linguistic distance of metroarea-industry-occupation clusters. I include only clusters in the 31 largest metropolitan areas in the sample, which have a population greater than 6,000 in my sample and include only occupations with O*Net Skill's complex problem-solving skill scores over 50, to account for jobs that are more complex than average. I obtain the results by using log hourly wage as well as hourly wage, where both produce similar results showing that workers in firms with average linguistic distance between 10% and 20% are the most productive. This result is in proportion to the immigrants' population in the working age group—18%, or 13% when it is weighted by linguistic distance. Remarkably, workers in firms that employ foreign workers in proportion to their population in the workforce earn about $9 per hour more than their peers. Although this result cannot be understood to reflect a causal relationship, as there may be other institutional factors that affect the matching of immigrant workers to firms as well as channels of reverse causality in which outperforming firms attract immigrant workers, it provides suggestive evidence that firms that hire immigrants perform fairly well.

8.2 How should firms effectively screen?

Then how should firms improve their hiring strategies to attract the most productive workers? First, I demonstrate that some firms can maximize efficiency not by blindfolding the HR manager or

randomizing the hiring process but rather by implementing a hiring practice that applies a more lenient standard for people of different cultural and ethnic backgrounds. The analyses of this paper suggest that even if employers are unbiased, immigrants face frictions in the labor market owing to their imprecise signals and that they suffer from misallocation, causing them to sort into self-employment. A common managerial practice is to randomly assign candidates. Studies have found that such practices have benefits. For instance, Goldin and Rouse (2000) show how adopting a blind procedure for orchestra auditions serves as a solution to sex-based hiring. My suggestions contrast this common belief; however, they resonate with the handicapping principle in the contest literature: Ridlon and Shin (2013) indicate that giving a boost to those with a disadvantage yields better outcomes in competitions when there is severe heterogeneity.

Second, alternatively, firms may minimize the effect of cultural noise by investing in their HR division to hire people who can better decipher immigrants' signal. Kulchina (2016) shows how foreign entrepreneurs excel by hiring a larger number of foreign workers, which suggests that matching firms' HR representative pool to the candidate pool's cultural mix as closely as possible would alleviate the misallocation problem. I illustrate the tension firms may face between the severity and extensiveness of the misallocation problem.

One factor to consider is that for most firms that hire highly educated workers, employers make very specific searches by conducting campus recruiting at top tier schools rather than searching the local labor market. Hence, I conduct my analysis based on the field of degree. I use the ACS to collapse over 150 fields of degree into 20 major categories, as listed in Appendix Table 5.

I conduct a cost-benefit analysis to identify when it is worthwhile for firms to make the investment to hire an HR representative who speaks the candidate's language. For this purpose, I identify fields from which society may substantially gain from having misallocated workers in salaried work, and I then suggest how costly it would be for firms if a more diverse set of ethnic groups were to pursue their particular fields. The results are visually summarized in Figure 4.

The misallocation problem is more severe if firms are more dependent on fields in which the difference in workers' productivity between salaried employment and self-employment is large. I assume a perfectly competitive labor market where workers are paid the value of their marginal product of labor. I impute the potential productivity gain that firms may face by assessing the additional wage that an immigrant worker makes by being an employee at a firm relative to owning a business. In my setting, productivity differences are driven by the type of employment—salaried work or self-employment—and years of education. In other words, the size of the productivity loss is determined by the sum of the $\beta 2$ and $\beta 3$ coefficients in the following specification:

$$Wages_i = \beta_0 + \beta_1 \, EducCategories_i + \beta_2 \, Salaried_i + \beta_3 \, Salaried \times EducCategories_i + \beta_4 \, X_i + \epsilon \quad (8)$$

Given this difference along with the number of immigrants who major in the different fields, I am able to rank order fields by the acuteness of the misallocation problem. Based on my sample, Engineering and Business majors presents the largest social benefits, while Psychology, Biology and Health Services majors provide the lowest benefits. The solid line demarcates the point where the social gain for having a worker in salaried work becomes positive and hence where immigrants with a Psychology or a Biology major are likely to earn more through self-employment. This rank ordering is plotted along the X-axis of Figure 4.

Furthermore, the problem is more widespread when a more diverse set of ethnic groups pursue those particular fields, as shown by the Y-axis of Figure 4. In this figure, the Y-axis measures the rank ordering of the diversity of the misallocated ethnic pool depending on their field. Specifically, I count the number of misallocated immigrants by 12 country-of-origin categories and then count the number of "peaks" in the distribution. Peaks are defined to have more than one misallocated immigrant in my sample in an origin category. The misallocated immigrants among engineering majors are heavily focused in a few ethnic categories, primarily in the Middle East or Latin America, while misallocated immigrants among business majors occur for a diverse set of ethnic groups. The misallocated measure is a metric for differential sorting into self-employment using equations (4) and (5).

Together, the above analyses offer a cost-benefit analysis where a firm's investment in its HR department will be more worthwhile when social gain is larger and the problem is easier to fix. The X-axis determines the potential social gain, and the Y-axis determines how difficult the problem is to fix.

The framework summarized in Figure 4 suggests that if a firm is in search of a worker in the first quadrant, with an engineering or a computer science degree, it should conduct a targeted search, as the ethnic category span of misallocated workers for those majors is quite narrow, while the productivity gains from having those workers in a firm can be large. Conversely, if a firm is in search of someone in the third quadrant, with a liberal arts or a psychology major, it may be quite difficult to recruit them, as the ethnic category span is too wide to begin with, while it is also difficult to justify the benefits, as those workers are likely to be more productive owning their own business. The implications are more case dependent for majors in the fourth quadrant, such as social science and business, which present a large opportunity for both productivity gains and misallocation over a broad span of ethnic categories. The same can be stated for majors in the second quadrant, such as philosophy or public policy, where misallocation occurs for a targeted ethnic group, but the benefits from hiring are small.

8.3 What is the potential productivity gain from hiring misallocated workers?

I estimate the potential economic gain from hiring a highly skilled foreigner who otherwise would have sorted into their outside option of business ownership. Inefficient sorting of talented immigrant workers may be detrimental for economic growth as immigrants are prone to become proprietors of less competitive businesses, such as dry cleaners or motels (Kerr and Mandroff 2015). In other words, once immigrants are pushed out of the salaried workforce, ethnic factors pull them to own businesses that tend to require less complex problem-solving skills than those owned by similarly qualified U.S.-born business owners. Consistent with this, Fossen and Buttner (2013) have shown how returns to education is 3 percentage points lower for entrepreneurs with necessity-based motives than those with opportunity-based motives and more broadly, Sauermann and Cohen (2010) have found that workers with necessity-based motives tend to be less innovative than those with more positive motives. These suggest that there may be social losses associated with how immigrants sort in the labor market and that society can better leverage their skills. I conduct a productivity analysis to evaluate the potential social gains from correctly identifying self-employed immigrant workers.

The productivity analysis is based on the specification in equation (8), where worker wage is determined by the type of employment—salaried work or self-employment—and their human capital. I assess the difference between salaried employment and self-employment as well as the difference between salaried employment and unincorporated self-employment. According to Levine and Rubinstein (2016), the unincorporated self-employed represent a non-entrepreneurial type of self-employed individuals; hence, I use this group as a lower bound for the productivity analysis.

Estimations based on the ACS for the period from 2005 to 2012 by education group are shown in Figure 5. In summary, the estimates suggest that the potential gains from hiring each talented immigrant who is misallocated in the market may be ~$3,000-$6,000 for an average worker and ~$7,000-$15,000 for a highly educated worker, annually. These results suggest that it could be quite costly to have highly educated immigrants who are linguistically distant sort into self-employment. As shown in the bar charts in Figure 5, the potential social gain is disproportionately large for firms that require more high skilled labor. While the purpose of this analysis is to inform the degree of magnitude of the misallocation problem, this estimates a partial equilibrium effect. The interpretation of the general equilibrium effects may differ and is further discussed in the following section.

8.4 Limitations

In this section I discuss several limitations. First, the estimation of gains from alleviating the misallocation problem in the labor market is imperfect as it does not incorporate a general equilibrium effect. On the one hand, having more immigrant workers in the labor market may lower worker wages, changing immigrants' incentives and productivity gains for working in firms. On the other hand, immigrants may increase wages, as they may have positive spillover effects, as shown in previous research regarding the innovation benefits arising from hiring immigrant workers (Kerr and Lincoln 2010, Kerr et al. 2015) and when workers have more positive motives (Sauermann and Cohen 2010). Given these other forces, there are limitations to assessing the general equilibrium effects of having the self-employed immigrant in salaried work.

Second, while I posit that cultural mismatch can cause friction in matching workers to firms, I do not examine how cultural fit may shape organizations' productivity. Prior studies have discussed how cultural fit may facilitate coordination (Van den Steen 2005) and how ethnic ties help generate business leads and meet financing needs (Nanda and Khanna 2010). Conversely, studies have also suggested how firms may benefit from diverse teams, as such teams are more likely to make decisions more carefully and become more open to new ideas (Phillips et al. 2009). While the assessment of cultural fit and its implications for immigrants' labor market assimilation are important considerations, it is outside the scope of this study.

Third, while the margin of adjustments that I consider is between an employment choice between salaried work and self-employment, depending on the employer's attitude towards risk and the nature of the job, employers may cope with market imperfections arising from cultural frictions through wage contracts. While this is an important consideration, I only focus on one particular margin of adjustment—choice of employment—under the assumption that workers are more likely to run businesses when they fail to find the most appropriate match.

9 Conclusion

In this study, I examine how informational frictions in labor markets differentially shape the sorting of workers into either self-employment or salaried work. To conduct this examination, I study the experiences of immigrants, who likely face especially large labor market frictions owing to linguistic cultural barriers. I apply a theoretical framework that presumes that mismatch in the linguistic cultural backgrounds between an interviewer and a candidate in a hiring setting cause minority candidates to be less effective in conveying their ability. The framework predicts that immigrants are less likely to find a good match with existing firms since they send imprecise signals of ability and that highly educated immigrants especially suffer as employers demand more assurance for more difficult jobs. I empirically test these predictions by investigating whether there exist differential patterns of sorting out of salaried work and into self-employment between immigrants and non-immigrants across subsets of education categories.

Consistent with the theoretical framework, I show that immigrants are more likely to sort into self-employment, particularly when they have noisier signals and higher education. I proxy for the degree of imprecise signaling with "linguistic distance," a measure based on how many branches separate two languages in a language tree, and I show that linguistically distant immigrants are, on average, 23-40% more likely to enter into self-employment than similarly qualified U.S.-born workers. Furthermore, there is a heterogeneous effect across educational attainment: with an additional year of education, the likelihood for the linguistically distant to enter into self-employment increases by 3-5%. Relative to previous studies investigating either whether immigrants have a higher propensity to enter into self-employment, or whether the highly educated are more likely to enter into self-employment, this study sets forth an informational friction explanation for how immigrant status and educational attainment interact to generate systematic patterns of immigrant self-employment. This study also advances the literature on the motivations for becoming an entrepreneur by showing how labor market mismatch owing to cultural frictions may motivate highly educated minority workers to become business owners.

A series of empirical results validate that the imprecise signaling hypothesis importantly accounts for the sorting pattern. I show that there is a mitigating effect for immigrants who have culturally assimilated or who compose a majority group and I rule out competing hypotheses, including the hypothesis that language skills may be more important for jobs for which the more highly educated compete.

The findings of this study have implications for the efficient allocation of human capital. I assess how linguistic-cultural differences cause informational frictions in the discovery of immigrant talent, rather than act as a barrier that renders immigrants unable to perform to expectations. Hence, the stronger sorting into self-employment by immigrants reflects inefficient allocation of talent, suggesting that firms systematically make false negative judgements.

I argue that firms should view this labor market imperfection as a source of competitive advantage and adjust their hiring practices to correctly identify the hidden talent. In contrast to the common belief that more objective measures in hiring will overcome social biases, I argue that in some cases, it is better to implement hiring practices that apply a more lenient standard to people of different cultural and ethnic backgrounds. I provide suggestive evidence that firms that hire immigrants in proportion to their representation in the overall population outperform their peers. I also suggest

firms may benefit from hiring people who can decipher immigrants' signal better depending on which fields of degree they are targeting.

The immigrant talent pool composes almost 18% of the working age population, and the potential economic gains for society as a whole from correctly identifying immigrants, especially those who are highly educated, are thus meaningful. Indeed, partial equilibrium estimates suggest that the annual potential gains from hiring each talented immigrant who is misallocated in the market is ~ $3,000-$6,000 for an average worker and ~$7,000-$15,000 for a highly educated worker. While policy discussions have focused on finding the most talented immigrant workers from outside the US, this study suggests that firms may benefit from targeting the talented immigrants who are abundant but hidden in the US.

References

Altonji JG, Pierret CR (2001) Employer learning and statistical discrimination. Q. J. Econ. 116(1):313-350.

Arrow, KJ (1973) The theory of discrimination. In: Ashenfelter, O, Rees, A (Eds.), Discrimination in Labor Markets. Princeton Unversity Press, pp. 3-33.

Åstebro T, Chen J, Thompson P (2011) Stars and misfits: Self-employment and labor market frictions. Manag. Sci. 57(11):1999-2017.

Åstebro T, Herz H, Nanda R, Weber RA (2014) Seeking the roots of entrepreneurship: insights from behavioral economics. J. Econ. Perspect. 28(3):49-69.

Autor D, Levy F, Murnane R (2003) The skill content of recent technological change: An empirical exploration. Q. J. Econ. 118(4):1279-1334.

Autor D, Scarborough D (2008) Does job testing harm minority workers? Evidence from retail establishment. Q. J. Econ. 123(1):219-277.

Battisti M, Peri G, Romiti A (2016) Dynamic effects of co-ethnic networks on immigrants' economic success. NBER Working Paper No. 22389, National Bureau of Economic Research, Cambridge.

Becker G (1957) The Economics of Discrimination, 2nd ed. (University of Chicago Press, Chicago).

Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am. Econ. Rev. 94(4):991-1013.

Bleakley H, Chin A (2010) Age at arrival, English proficiency, and social assimilation among US immigrants. Am. Econ. J. Appl. Econ. 2(1):165-192.

Borjas GJ (1986) The self-employment experience of immigrants. J. Hum. Resour. 21(4):485-506.

Borjas G (1987) Self-selection and the earnings of immigrants. Am. Econ. Rev. 77(4):531-553.

Chiswick BR, Miller PW (1995) The endogeneity between language and earnings: International analyses. J. Lab. Econ. 13(2):246-288.

Clark K, Drinkwater S (2000) Pushed out or pulled in? Self-employment among ethnic minorities in England and Wales. Labour Econ. 7(5):603-628.

Cornell B, Welch I (1996) Culture, information, and screening discrimination. J. Polit. Econ. 104(3):542-571.

Davila A, Mora MT (2004) English-language skills and the earnings of self-employed immigrants in the United States: A note. Ind. Relat. 43(2):386-391.

Fairlie RW, Lofstrom M (2013) Immigration and entrepreneurship. Discussion Paper No. 7669, Institute for the Study of Labor (IZA), Bonn.

Fairlie R, Meyer B (1996) Ethnic and racial self-employment differences and possible explanations. J. Hum. Resour. 31(4):757-793.

Fearon JD (2003) Ethnic and cultural diversity by country. J. Econ. Growth 8(2):195-222.

Ferrer A, Green DA, Riddell WC (2006) The effect of literacy on immigrant earnings. J. Hum. Resour. 41(2):380-410.

Fossen F, Buttner T (2013) The returns to education for opportunity entrepreneurs, necessity entrepreneurs, and paid employees. Econ. Ed. Rev. 37:66-84.

Giuliano L, Levine DI, Leonard J (2009) Racial bias in the manager-employee relationship: An analysis of quits, dismissals, and promotions at a large retail firm. IRLE Working Paper No.178-09, Institute for Research on Labor and Employment Working Paper Series, Berkeley.

Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. Am. Econ. Rev. 90(4):715-741.

Hamilton B (2000) Does entrepreneurship pay? An empirical analysis of the returns to self-employment. J. Polit. Econ. 108(3):604-631.

Hegde D, Tumlinson J (2014) Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in U.S. venture capital. Manag. Sci. 60(9):2355-2380.

Hegde D, Tumlinson J (2015) Unobserved ability and entrepreneurship. Working Paper.

Hoffman M, Kahn L, Li D (2015) Discretion in hiring. Working Paper 16-055, Harvard Business School, Boston.
Hurst E, Pugsley BW (2015) Wealth, tastes, and entrepreneurial choice. Measuring Entrepreneurial Businesses: Current Knowledge and Challenges (University of Chicago Press, Chicago).
Imai S, Stacey D, Warman C (2014) From engineer to taxi driver? Language proficiency and the occupational skills of immigrants. Working Paper 1275, Queen's Economics Department.
Joshi AM, Lahiri N (2014) Language friction and partner selection in cross-border R&D Alliance Formation. J. Int. Bus. Stud. 46(2):123-152.
Kacperczyk AJ (2013) Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry. Organ. Sci. 24(3):664-683.
Kerr SP, Kerr WR, Lincoln WF (2015) Skilled immigration and the employment structures of US firms. J. Lab. Econ. 33(S1):S147-S186.
Kerr W, Lincoln W (2010) The supply side of innovation: H-1B Visa reforms and U.S. ethnic invention. J. Lab. Econ. 28(3):473-508.
Kerr W, Mandroff M (2015) Social networks, ethnicity, and entrepreneurship. NBER Working Paper No. 21597, National Bureau of Economic Research, Cambridge.
Kulchina E (2016) A path to value creation for foreign entrepreneurs. Strateg. Manag. J. 37(7):1240-1262.
Lang K (1986) A language theory of discrimination. Q.J. Econ. 101(2):363-382.
Lazear E (2005) Entrepreneurship. J. Lab. Econ. 23(4):649-680.
Levine R, Rubinstein Y (2016) Smart and illicit: who becomes an entrepreneur and do they earn more?. Q. J. Econ. (in press).
Lewis E (2011) Immigrant-native substitutability: The role of language ability. Card D, Raphael S, eds. Immigration, Poverty, and Socioeconomic Inequality (Russell Sage Foundation, New York), 60-97.
Light I (1972) Ethnic Enterprise in America (University of California Press, Berkeley).
Light I (1979) Disadvantaged minorities in self-employment. Int. J. Comp. Sociol. 20(1-2):31-45.
Lundberg SJ, Startz R (1983) Private discrimination and social intervention in competitive labor markets. Am. Econ. Rev. 73: 340-47
Montalvo JG, Reynal-Querol M (2005) Ethnic polarization, potential conflict, and civil wars. Am. Econ. Rev. 95(3):796-816.
Morgan J, Várdy F (2009) Diversity in the workplace. Am. Econ. Rev. 99(1):472-485.
Nanda R, Sørensen JB (2010) Workplace peers and entrepreneurship. Manag. Sci. 56(7):1116-1126.
Nanda R, Khanna T (2010) Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software industry. J. Econ. Manag. Strateg 19(4):991-1012.
Oreopoulos P (2011) Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. Am. Econ. J. Econ. Policy 3(4):148-171.
Oyer P, Schaefer S (2011) Personnel economics: Hiring and incentives. Ashenfelter O, Card D, eds. Handbook of Labor Economics, Vol. 4b (North-Holland, Amsterdam), 1769-1823.
Peri G, Sparber C (2009) Task specialization, immigration, and wages. Am. Econ. J. Appl. Econ. 1(3):135-169.
Petersen T, Saporta I (2004) The opportunity structure for discrimination. Am. J. Sociol. 109(4):852-901.
Phelps, E (1972) The statistical theory of racism and sexism. Am. Econ. Rev. 62:659-661.
Phillips KW, Liljenquist KA, Neale MA (2009) Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers. Personal. Soc. Psychol. Bull. 35(3):336-350.
Portes A, Zhou M (1996). Self-employment and the earnings of immigrants. Am. Soc. Rev. 61(2):219-230.
Ridlon R, Shin J (2013) Favoring the winner or loser in repeated contests. Mark. Sci. 32(5):768-785.
Rivera LA (2012) Hiring as cultural matching: The case of elite professional service firms. Am. Sociol. Rev. 77(6):999-1022.

Sauermann H, Cohen WM (2010) What makes them tick? Employee motives and firm innovation. Manag. Sci. 56(12):2134-2153.
Siegel JI, Pyun L, Cheon BY (2014) Multinational firms, labor market discrimination, and the capture of competitive advantage by exploiting the social divide. Working Paper 11-011, Harvard Business School, Boston.
Stenard BS, Sauermann H (2016) Educational mismatch, work outcomes, and entry into entrepreneurship. Organ. Sci. 27(4):801-824.
Spolaore E, Wacziarg R (2009) The diffusion of development. Q. J. Econ. 124(2):469-529.
Van den Steen E (2005) Organizational beliefs and managerial vision. J. Law Econ. Organ. 21(1):256-283.

Empirical Supplements

Table 1: Demographics and Labor Market Outcomes by nativity and employment type, ACS and CPS

	All	U.S.-born			1st gen. Immigrants		
		All	Salaried	SelfEmp	All	Salaried	SelfEmp
Panel A: American Community Survey (ACS), 2005 - 2012							
Observations	5,360,837	4,551,230	3,954,587	596,643	809,607	703,568	106,039
		85%	87%	13%	15%	87%	13%
Demographics							
Average age	40.6	40.7	40.0	46.5	39.8	39.2	44.1
% White	68%	79%	78%	88%	16%	15%	26%
% Black	10%	10%	11%	5%	8%	8%	6%
% Hispanic	16%	8%	9%	5%	53%	54%	44%
% Asian	7%	3%	3%	2%	24%	23%	24%
Years of Schooling	13.6	13.9	13.9	14.1	12.2	12.1	12.6
% high school degree	28%	29%	29%	28%	23%	23%	24%
% college degree	30%	31%	30%	35%	28%	27%	29%
Labor Market Outcomes							
Annual hours worked	2,039	2,046	2,031	2,158	2,006	1,993	2,100
Mean earnings	$ 54,654	$ 56,472	$ 54,629	$ 70,209	$ 46,424	$ 45,252	$ 54,885
Median earnings	$ 39,763	$ 41,580	$ 41,580	$ 40,408	$ 30,254	$ 30,234	$ 30,306
Mean hourly earnings	$ 26.8	$ 27.6	$ 26.4	$ 36.1	$ 23.6	$ 22.9	$ 28.5
Median hourly earnings	$ 19.1	$ 19.8	$ 19.8	$ 19.9	$ 15.3	$ 15.3	$ 15.5
Panel B: Current Population Survey (CPS), 1994 - 2012							
Observations	639,774	489,278	424,544	64,608	108,424	97,161	11,224
		76%	87%	13%	17%	90%	10%
Demographics							
Average age	40.0	40.5	39.8	44.8	38.6	38.0	43.4
% White	73%	86%	85%	94%	19%	18%	32%
% Black	9%	10%	11%	4%	7%	7%	5%
% Hispanic	13%	3%	4%	2%	52%	54%	35%
% Asian	4%	0%	0%	0%	22%	22%	28%
Years of Schooling	13.6	13.8	13.8	14.1	12.4	12.3	13.3
% high school degree	33%	34%	34%	31%	28%	28%	28%
% college degree	32%	32%	31%	37%	29%	28%	38%
Labor Market Outcomes							
Annual hours worked	2,333	2,348	2,314	2,587	2,267	2,233	2,541
Mean earnings	$ 61,102	$ 62,895	$ 61,185	$ 75,082	$ 50,837	$ 48,700	$ 68,781
Median earnings	$ 46,153	$ 48,289	$ 48,375	$ 47,551	$ 34,630	$ 33,966	$ 41,556
Mean hourly earnings	$ 26.0	$ 26.7	$ 26.2	$ 29.8	$ 22.2	$ 21.5	$ 27.4
Median hourly earnings	$ 20.2	$ 20.9	$ 21.1	$ 19.3	$ 15.6	$ 15.5	$ 16.6

Notes: Sample summary statistics include male workers, between 18 - 65 old in the survey year, who worked full-time for the entire year.
2005 - 2009 ACS 5 -year estimates and 2010- 2012 ACS 3-year estimates are combined for years 2005 - 2012 of the ACS.
March Annual Demographic Survey files of the Census Bureau's CPS is used for years 1994 - 2012.
1st generation immigrants are defined as those who and whose parents were born outside of the US for CPS and those who are categorized as foreign-born for the ACS. Employment types, either salaried or self-employed, is coded based on classification in the survey.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Table 2: Selection into Self-Employment by Education Categories

	Panel A: ACS			*Panel B: CPS*		
	Self-employment (vs Salaried)			**Self-employment (vs Salaried)**		
	Ex-H1Bs		**All**	**Ex-H1Bs**		**All**
	(1)	(2)	(3)	(4)	(5)	(6)
Education (vs below HS)						
HS dropout	-0.013***	-0.013***	-0.013***	-0.005	-0.005	-0.005
	0.001	0.001	0.001	0.004	0.004	0.004
HS degree	-0.015***	-0.015***	-0.015***	0.001	0.001	0.001
	0.001	0.001	0.001	0.004	0.004	0.004
Some college	-0.011***	-0.011***	-0.011***	0.005	0.005	0.005
	0.001	0.001	0.001	0.004	0.004	0.004
College degree	-0.018***	-0.018***	-0.017***	-0.003	-0.003	-0.002
	0.001	0.001	0.001	0.004	0.004	0.004
1st gen Immigrants (1glmm)	-0.028***	-0.022***	-0.026***	0.001	0.007	0.003
	0.002	0.002	0.002	0.005	0.006	0.005
1glmm x HS dropout	0.035***	0.034***	0.035***	0.026***	0.026***	0.026***
	0.002	0.002	0.002	0.006	0.006	0.006
1glmm x HS degree	0.050***	0.049***	0.050***	0.038***	0.038***	0.039***
	0.002	0.002	0.002	0.005	0.005	0.005
1glmm x Some college	0.063***	0.072***	0.063***	0.039***	0.049***	0.040***
	0.002	0.002	0.002	0.005	0.006	0.005
1glmm x College degree	0.063***	0.069***	0.059***	0.051***	0.056***	0.045***
	0.002	0.002	0.002	0.005	0.005	0.005
1glmm x US Some College[1]		-0.018***			-0.019***	
		0.001			0.005	
1glmm x US College degree[1]		-0.015***			-0.016***	
		0.001			0.004	
Control variables[2]	✓	✓	✓	✓	✓	✓
Fixed effects						
Age	✓	✓	✓	✓	✓	✓
Year, State	✓	✓	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓	✓	✓
Constant	0.121***	0.116***	0.110***	-0.014	-0.019	-0.030**
	0.004	0.004	0.004	0.014	0.014	0.014
Number of Observations	5279173	5279173	5308641	583189	583189	585561

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants. Panel A uses the ACS, Panel B uses the CPS.
Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Columns (1), (2), (4) and (5) exclude immigrant groups that compose about 80% of H-1B workers, based on USCIS' FY2012 Annual Report.
1. Imputed based on immigration age (less than 22 and 23 for Some College and College degree respectively);
Assign US Some college and US College degree 0 for all natives;
Immigrants who are assigned 1 for US Some College and US College degree are also assigned 1 for some college and college degree.
All other dummy variables defined exclusively.
2. Include four race categories (White, Black, Hispanic, and Asian), time spent in US (for which non-immigrants are assigned their age), and log GDP per capita of source country.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Table 3: Linguistic distance and selection into self-employment

	Panel A: ACS					*Panel B: CPS*				
	Self-employment (vs Salaried)					**Self-employment (vs Salaried)**				
	All sample				**1glmm**	**All sample**				**1glmm**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Linguistic distance (LingD)	0.015***	0.011***	0.030***	-0.030***	0.045***	0.042***	0.024***	0.051***	-0.029***	0.063***
	0	0	0.001	0.002	0.014	0.002	0.002	0.003	0.006	0.005
Years of education	0.005***	0.001***	0.001***	-0.001***	0.002*	0.002***	0.001***	0.001***	-0.001***	0.002***
	0	0	0	0	0.001	0	0	0	0	0
LingD x Yrs of education				0.004***					0.006***	
				0					0	
Time spent in US			0.001***	0.001***	0.001***			0.001***	0.001***	0.001***
			0	0	0			0	0	0
ln GDP per capita of origin			0	0	0.001**			0.004***	0.003***	0.004***
			0	0	0			0.001	0.001	0.001
Race (vs White)										
Black	-0.042***	-0.025***	-0.025***	-0.026***	-0.035***	-0.076***	-0.026***	-0.025***	-0.026***	-0.018***
	0	0	0	0	0.002	0.001	0.001	0.001	0.001	0.005
Hispanic	-0.042***	-0.041***	-0.040***	-0.037***	-0.054***	-0.066***	-0.052***	-0.050***	-0.046***	-0.056***
	0	0	0	0	0.001	0.002	0.002	0.002	0.002	0.003
Asian	-0.009***	-0.004***	-0.002***	-0.005***	-0.014***	-0.030***	0.001	0.009***	0.002	-0.005
	0.001	0.001	0.001	0.001	0.001	0.003	0.002	0.003	0.003	0.003
Fixed effects										
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year, State		✓	✓	✓	✓		✓	✓	✓	✓
Industry, Occupation		✓	✓	✓	✓		✓	✓	✓	✓
Constant	0.206***	0.095***	0.075***	0.106***	-0.013	0.008	0.022***	-0.042***	-0.006	-0.062**
	0.002	0.003	0.004	0.004	0.015	0.006	0.008	0.013	0.013	0.026
Number of observations	5321693	5321693	5279173	5279173	806601	592717	592717	583189	583189	94037
Base rate of self-employment			13%		13%			13%		10%
Selection effect			*23%*		*33%*			*40%*		*61%*

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants. Panel A uses the ACS, Panel B uses the CPS. Columns (5) and (10) based on first generation immigrants only.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively. Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Table 4: Age of immigration and selection into self-employment (within immigrants)

	Panel A: ACS				*Panel B: CPS*			
	Self-employment (vs Salaried)				**Self-employment (vs Salaried)**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic Distance (LingD)	0.039***	0.046***	0.039***	0.045***	0.062***	0.068***	0.062***	0.062***
	0.003	0.003	0.003	0.004	0.005	0.005	0.005	0.006
Age of immigration								
Before 10	-0.010***	0.023***	-0.006***	0.026***	-0.010***	0.027**	-0.004	0.026**
	0.001***	0.001***	0.001***	0.001***	0.003	0.011	0.003	0.011
Between 10 to 15			0.013***	0.020**			0.015***	-0.013
			0.001	0.01			0.003	0.013
Between 15 to 20			0.009***	-0.002			0.009***	-0.002
			0.001	0.008			0.002	0.011
LingD x Age of immigration								
LingD x Before 10		-0.038***		-0.036***		-0.048***		-0.041***
		0.008		0.008		0.014		0.014
LingD x Between 10 to 15				-0.009				0.036**
				0.011				0.016
LingD x Between 15 to 20				0.013				0.014
				0.009				0.014
Other controls	✓	✓	✓	✓	✓	✓	✓	✓
Fixed effects								
Age	✓	✓	✓	✓	✓	✓	✓	✓
Year, State	✓	✓	✓	✓	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓	✓	✓	✓	✓
Constant	-0.016*	-0.022**	-0.021**	-0.026***	-0.067***	-0.069***	-0.074***	-0.074***
	0.009	0.01	0.009	0.01	0.026	0.026	0.026	0.026
Number of observations	806601	806601	806601	806601	94059	94059	94059	94059

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, first generation immigrants, who worked full-time full-year. Panel A uses the ACS, Panel B uses the CPS.
Controls include race (four categories), years of education, and log GDP per capita of source country.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Table 5: Language as input to productivity

	Panel A: ACS		*Panel B: CPS*	
	Self-employment (vs Salaried)		**Self-employment (vs Salaried)**	
Language requirements for salaried jobs:	**Low**	**High**	**Low**	**High**
	(1)	(2)	(3)	(4)
Linguistic Distance (LingD)	0.025***	0.032***	0.051***	0.069***
	0.001	0.001	0.004	0.005
Yrs of education	0.004***	-0.005***	0.008***	-0.007***
	0	0	0	0
Other controls	✓	✓	✓	✓
Fixed effects				
Age	✓	✓	✓	✓
Year, State	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓
Constant	0.090***	0.336***	0.610***	0.089***
	0.005	0.007	0.02	0.024
Number of observations	3323900	2622300	358459	297620
Base rate of self-employment	21%	26%	21%	25%
Selection effect	*12%*	*12%*	*25%*	*28%*

Notes: Table reports linear estimates of the probability of a worker to be self-employed.
Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants.
Panel A uses the ACS, Panel B uses the CPS.
Language requirements for salaried jobs use O*Net Skill, taking the average scores of Reading Comprehension, Speaking and Writing scores. For both ACS and CPS, "Low" compares self-employed with salaried occupations with communication scores less than or equal to 50. "High" compares self-employed with salaried occupations with communication skills over 50.
Controls include race (four categories), years of education, time spent in US, log GDP per capita of source country.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Table 6.1 Selection into self-employment by immigrant's majority representation in the workplace

Dependent variable:	Indicator for Self-employment (vs Salaried)		
Independent variables	(1)	(2)	(3)
Linguistic Distance	0.068***		0.068***
	0.005		0.005
Majority representation		-0.024***	-0.024***
		0.002	0.002
Job complexity	0.003***	0.003***	0.003***
	0	0	0
Control variables	✓	✓	✓
Fixed effects			
Age	✓	✓	✓
Year, State	✓	✓	✓
Major Industry	✓	✓	✓
Constant	-0.161***	-0.070***	-0.152***
	0.025	0.024	0.025
Number of Observations	93976	93976	93976

Souce: Current Population Survey, 1994-2012
Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes immigrants only, who are males between 18 and 65, who worked full-time full-year.
Controls include race (four categories), years of education, time spent in US, and log GDP per capita of source country.
Majority representation indicates immigrants who are the most represented ethnic group in their metroarea-industry-occupation cluster.
Industry and Occupation categories based on the 1950 Census Bureau occupational classification system.
Job complexity uses O*Net Skill scores measuring complex problem solving skills required to perform jobs.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Calcualtions are weighted using the population weights provided by the survey.

Table 6.2:Selection into self-employment in ethnic enclaves

Dependent variable:	Indicator for Self-employment (vs Salaried)			
Enclave definition:	95th percentile		99th percentile	
Independent variables	(1)	(2)	(3)	(4)
Linguistic Distance	0.064***	0.047***	0.063***	0.050***
	0.005	0.007	0.005	0.006
Enclaves	-0.006***	-0.031***	0.010***	-0.026***
	0.002	0.008	0.002	0.008
Enclaves x Ling. distance		0.032***		0.046***
		0.009		0.01
Control variables	✓	✓	✓	✓
Fixed effects				
Age	✓	✓	✓	✓
Year, State	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓
Constant	-0.065**	-0.052**	-0.067***	-0.057**
	0.026	0.026	0.026	0.026
Number of Observations	92532	92532	92532	92532

Souce: Current Population Survey, 1994-2012
Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes immigrants only, who are males between 18 and 65, who worked full-time full-year.
Controls include race (four categories), years of education, time spent in US, dummy variable for whether an immigrant is naturalized, log GDP per capita of source country.
To proxy for ethnic enclaves, I rank metro areas by the size of the ethnic group population and identify the metro areas that are above the 95th percentile and 99th percentile of the distribution. I assign 1 if an immigrant resides in these metro areas, 0 otherwise.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Calcualtions are weighted using the population weights provided by the survey.

Table 7: Cultural distance and selection into self-employment

	Self-employment (vs Salaried)			
	All Sample			Immigrants only
	(1)	(2)	(3)	(4)
Cultural distance (CultD)	-0.032***	0.113***	0.02	0.004
	0.009	0.013	0.012	0.014
Linguistic distance	0.032***	0.002		0.065***
	0.006	0.007		0.009
Years of education	0.009***	0.001***	0	0.001***
	0	0	0	0
CultD x Yrs of education			0.007***	
			0.001	
Control variables		✓	✓	✓
Fixed effects				
Age		✓	✓	✓
Year, State		✓	✓	✓
Industry, Occupation		✓	✓	✓
Constant	-0.109***	-0.182***	-0.149***	-0.065*
	0.006	0.022	0.021	0.034
Number of observations	567637	562874	562874	73722

Source: Current Population Survey, 1994-2012

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants.

Controls include race (four categories), years of education, age, and log GDP per capita of source country.

All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.

Calcualtions weighted using the population weights provided by the survey.

Table 8: Selection into self-employment for immigrants from multilinguial coutries (Belgium & Switzerland)

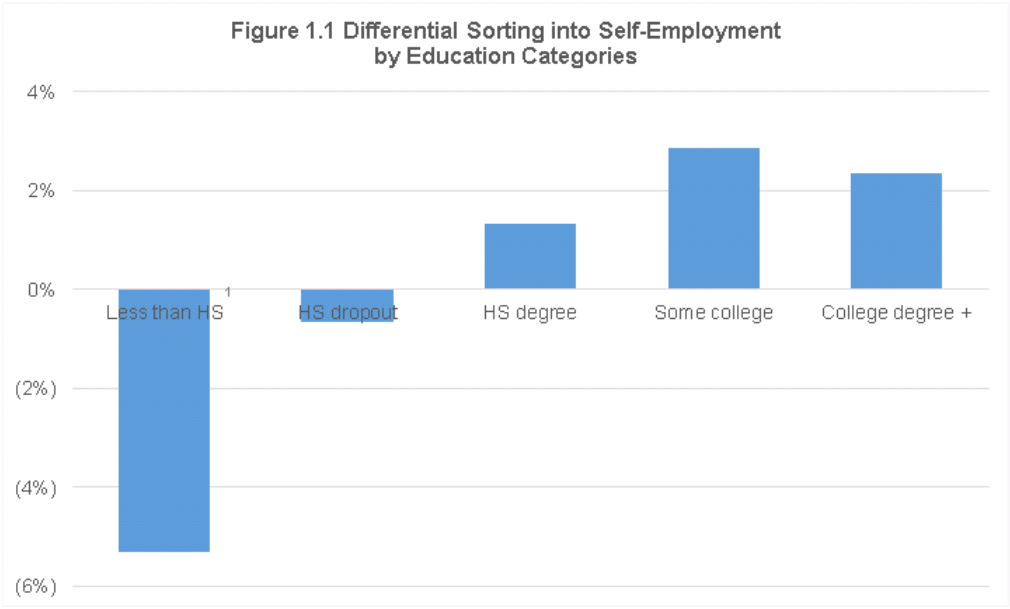
	Self-employment (vs Salaried)			
	(1)	(2)	(3)	(4)
Linguistic distance of language spoken at home	0.172*	0.056	0.174*	0.138
	0.095	0.09	0.096	0.09
Age	0.006***	0.006***	0.007***	0.007***
	0.001	0.001	0.001	0.001
Time spent in US	0	0	-0.001	-0.001
	0.001	0.001	0.001	0.001
Education categories (vs below HS)				
HS dropout	0.481**	0.255	0.557**	0.482**
	0.236	0.222	0.237	0.225
HS degree	0.325*	0.233	0.385**	0.311*
	0.177	0.167	0.177	0.169
Some college	0.273	0.161	0.330*	0.267
	0.175	0.167	0.175	0.169
College degree	0.16	0.095	0.248	0.232
	0.173	0.166	0.174	0.168
Race categories (vs Whites)				
Black	-0.144*	-0.071	-0.171**	-0.089
	0.075	0.074	0.078	0.073
Hispanic	0.146	0.041	0.141	0.143
	0.161	0.155	0.161	0.154
Asian	0.042	0	0.049	-0.014
	0.103	0.097	0.103	0.096
Fixed effects				
Major industry				✓
Major occupation		✓		
Occupation complexity category			✓	✓
Constant	-0.398**	-0.262	-0.499**	-0.087
	0.2	0.193	0.216	0.225
Number of Observations	996	983	983	983

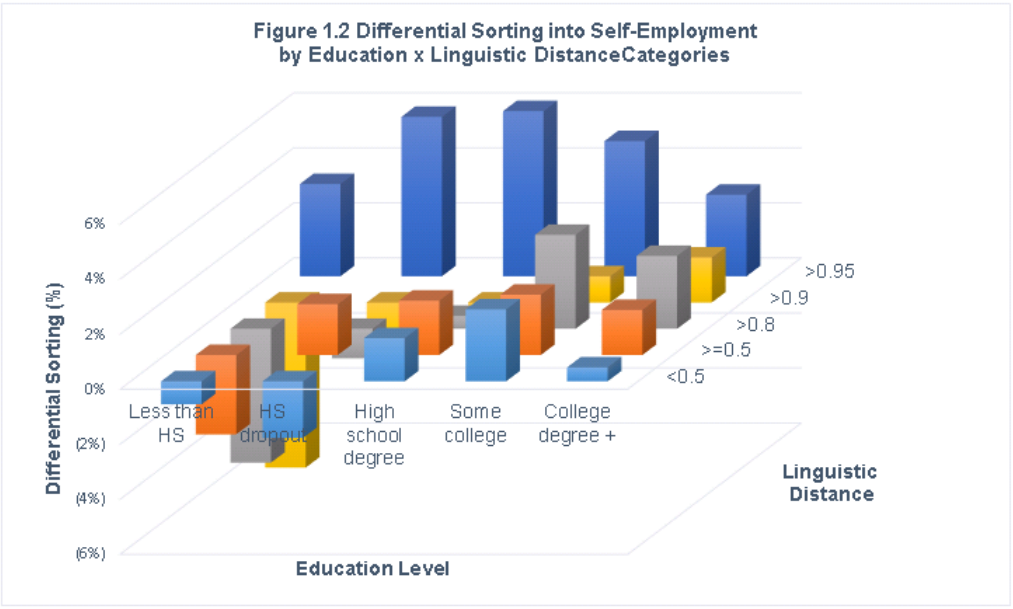
Source: American Community Survey, 2005-2012

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, first generation immigrants who are either born in Belgium or Switzerland, and who speaks Dutch, German or French at home.

All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively. Occupation complexity category sort occupations into 6 categories depending on complex problem solving scores from O*Net Skill. Calcualtions weighted using the population weights provided by the survey

Figure 1: Differential sorting across subgroups of the immigrant population





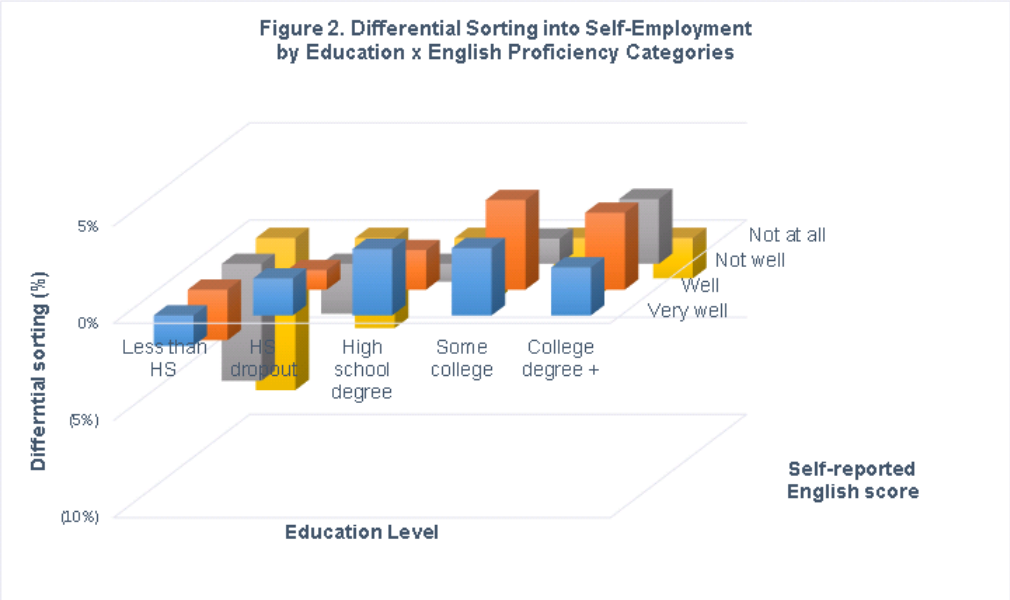
Source: American Community Survey 2005 -2012

Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year
Differntial sorting based on an imputed measure comparing immigrant's likelihood to self-emply to non-natives.
This is calculated as Self-Employed - Estimated Self-Employed, where Estimated Self-Employed is the fitted value for immigrant usingi the coefficients from running the following regression for U.S. born workers:

$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon$$

All education categories and linguistic categories defined exclusively.
1. Immigrants compose 65% of 'Less than HS' category, suggesting that the imputation is driven by a smaller native sample
All education categories and linguistic categories defined exclusively.
Calcualtions weighted using the population weights provided by the survey.

Figure 2: Differential sorting using self-reported English scores



Source: American Community Survey 2005 -2012

Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year

Differntial sorting based on an imputed measure comparing immigrant's likelihood to self-emply to non-natives.

This is calculated as Self-Employed - Estimated Self-Employed, where Estimated Self-Employed is the fitted value for immigrant usingi the coefficients from running the following regression for U.S. born workers:

$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon$$

Self-reported English scores as reported in the American Community Survey.

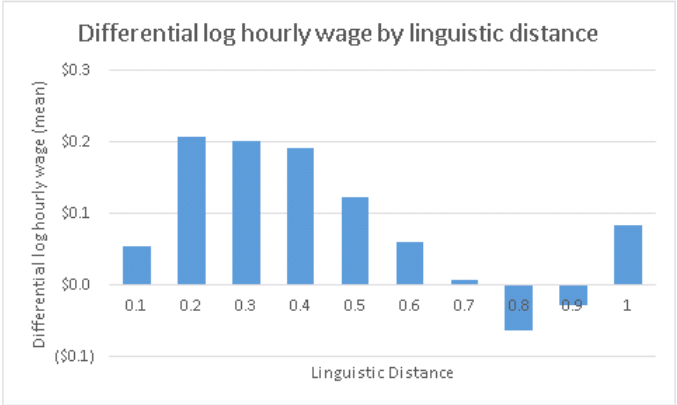
All education categories and self-reported english score categories defined exclusively.

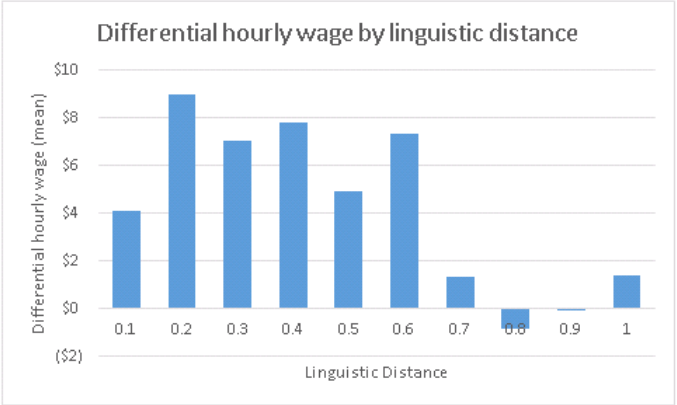
Immigrants compose 65% of 'Less than HS' category, suggesting that the imputation is driven by a smaller native sample

All education categories and linguistic categories defined exclusively.

Calcualtions weighted using the population weights provided by the survey.

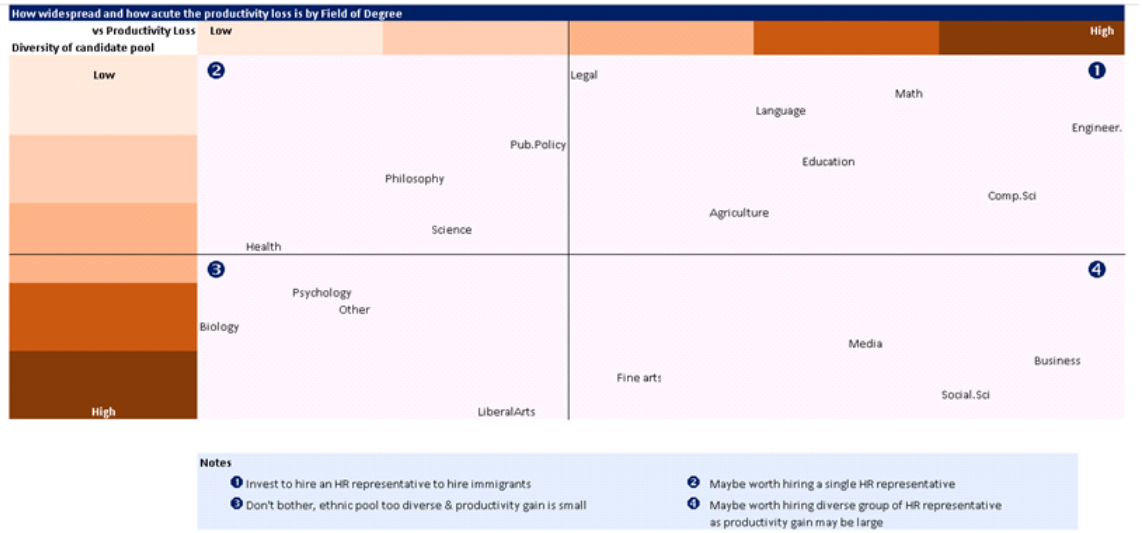
Figure 3: Immigrant representation and firm performance (among difficult jobs)





Source: Current Population Survey, 1994 -2012
Notes: Plots average residuage wage of workers by weighted linguistic distance of Metroarea-Industry-Occupation clusters in 31 largest metropolitan areas in sample, for occupations with complex problem solving skill scores over 50 from O*Net Skill. Industry and Occupation categories based on the 1950 Census Bureau occupational classification system.
Residuage wage of workers calculated as residuals from regressing (log) hourly wage:on four race categories, seven education categories, age as well as industry and occupation fixed effects.
Calcualtions weighted using the population weights provided by the survey.

Figure 4: Diversity vs Importance of potential candidate pool by field of degree



Source: American Community Survey, 2010 -2012

Notes: Over 150 field of degree categories grouped into 20 categories.

X-axis is the rank ordreing of the productivity loss associated with having a worker as self-employed rather tnan salaried work
High indicates high productivity loss, Low indicates low productivity loss
Productivity loss evaluated using the sum of coefficients beta2 and beta 3 from the following equation:

$$Wages_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 Salaried_i + \beta_3 Salaried \times EducCategories_i + \beta_4 X_i + \epsilon$$

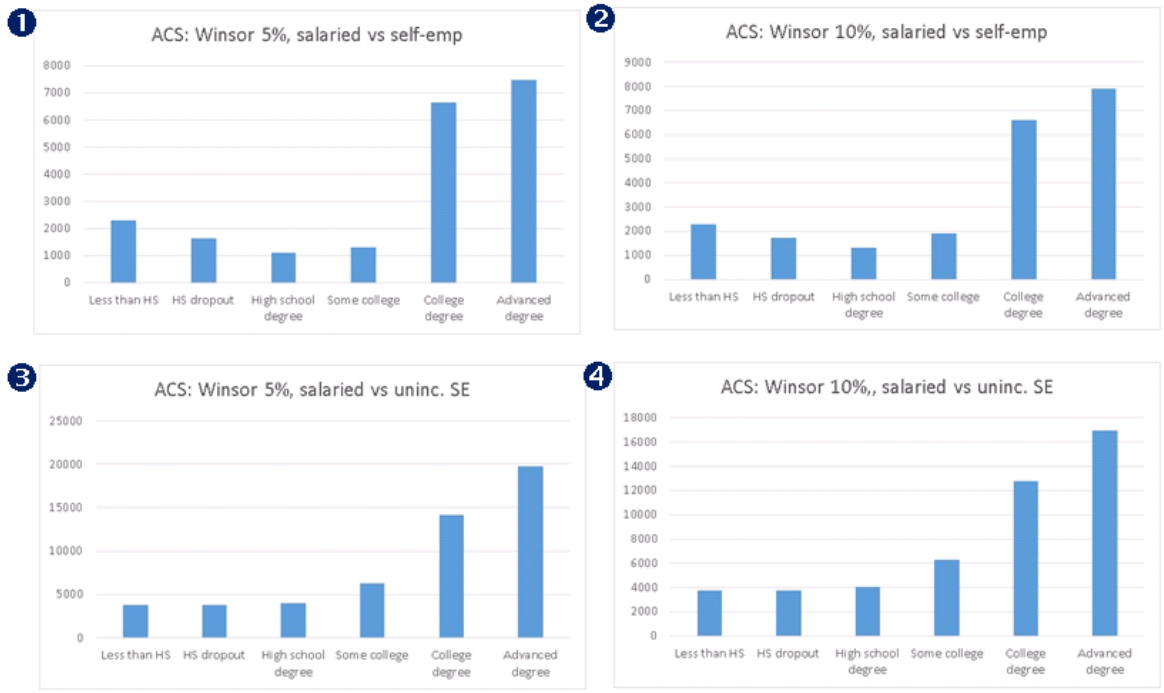
Y-axis measures the rank ordering of how diverse the misallocated ethnic candidate pool is depending on their field of degree. Specifically, I count number of misallocated immigrants by 12 country of origin categories and then count the number of "peaks" in the distribution. Peaks defined as having more than one misallocated immigrant in my sample in an origin category.
The misallocated measure is the metric for differential sorting into self-employment:, SE- estSE, where the estimated self-employed estimates immigrants propensity to self-employ using coefficients from running the following regression only on U.S. born workers:

$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon$$

High indicates that there are many ethnic groups that are misallocated, Low indicates that the misallocation is concentrated in a few ethnic groups.
The solid lines demarcate the point where productivity loss becomes positive for the X-axis, and the midpoint for candidate diversity for the Y-axis.

Notes in blue box summarizes suggestions based on the cost-benefit analysis.

Figure 5: Potential annual social gain per worker by education categories



Source: American Community Survey 2005 -2012

Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year
Productivity loss evaluated the sum of coefficcients beta2 and beta 3 from the following equation:

$$Wages_i = \beta_0 + \beta_1 \, EducCategories_i + \beta_2 \, Salaried_i + \beta_3 \, Salaried \times EducCategories_i + \beta_4 \, X_i + \epsilon$$

Graphs 1 and 3 use winsorized earnings at 5%; Graphs 2 and 4 use winsorized earnings at 10%
Graphs 1 and 2 compare salaried workers against self-employed; Graphs 3 and 4 compare salaried workers against unincorporated self-employed.

Appendix

Appendix Table 1: Selection into Self-Employment by Education Categories (Using linguistic distance)

	Panel A: ACS			*Panel B: CPS*		
	Self-employment (vs Salaried)			**Self-employment (vs Salaried)**		
	Ex-H1Bs		**All**	**Ex-H1Bs**		**All**
	(1)	(2)	(3)	(4)	(5)	(6)
Education (vs below HS)						
HS dropout	-0.012***	-0.012***	-0.012***	-0.002	-0.002	-0.002
	0.001	0.001	0.001	0.004	0.004	0.004
HS degree	-0.013***	-0.013***	-0.013***	0.003	0.003	0.003
	0.001	0.001	0.001	0.004	0.004	0.004
Some college	-0.009***	-0.009***	-0.009***	0.007*	0.007*	0.007*
	0.001	0.001	0.001	0.004	0.004	0.004
College degree	-0.016***	-0.016***	-0.016***	-0.001	-0.001	-0.001
	0.001	0.001	0.001	0.004	0.004	0.004
Linguistic Distance (LingD)	-0.038***	-0.038***	-0.036***	0.005	0.007	0.012*
	0.002	0.002	0.002	0.006	0.006	0.006
LingD x HS dropout	0.039***	0.039***	0.039***	0.029***	0.028***	0.029***
	0.002	0.002	0.002	0.007	0.007	0.007
LingD x HS degree	0.057***	0.057***	0.057***	0.047***	0.047***	0.047***
	0.002	0.002	0.002	0.006	0.006	0.006
LingD x Some college	0.075***	0.072***	0.075***	0.050***	0.061***	0.051***
	0.002	0.002	0.002	0.006	0.007	0.006
LingD x College degree	0.069***	0.067***	0.064***	0.061***	0.066***	0.058***
	0.002	0.002	0.002	0.006	0.006	0.006
LingD x US Some College[1]		0.003**			-0.016***	
		0.002			0.005	
LingD x US College degree[1]		0.002*			-0.010**	
		0.001			0.004	
Control variables[2]	✓	✓	✓	✓	✓	✓
Fixed effects						
Age	✓	✓	✓	✓	✓	✓
Year, State	✓	✓	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓	✓	✓
Constant	0.083***	0.083***	0.072***	-0.02	-0.017	-0.040***
	0.004	0.004	0.004	0.014	0.014	0.013
Number of Observations	5278538	5278538	5307765	583189	583189	585561

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants. Panel A uses the ACS, Panel B uses the CPS.
Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Columns (1), (2), (4) and (5) exclude immigrant groups that compose about 80% of H-1B workers, based on USCIS' FY2012 Annual Report.
1. Assign US Some college and US College degree 0 for all natives;
Immigrants who are assigned 1 for US Some College and US College degree are also assigned 1 for some college and college degree.
All other dummy variables defined exclusively.
2. Include four race categories (White, Black, Hispanic, and Asian), time spent in US (for which non-immigrants are assigned their age), and log GDP per capita of source country.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Appendix Table 2: Selection into self-employment by those who speak English well + very well

	Self-employment (vs Salaried)		
	Ex-H1Bs		All
	(1)	(2)	(3)
Education (vs below HS)			
HS degree	0.003***	0.003***	0.003***
	0.001	0.001	0.001
Some college	0.011***	0.011***	0.011***
	0.001	0.001	0.001
College degree	0.006***	0.006***	0.006***
	0.001	0.001	0.001
1st gen Immigrants (1glmm)	0.005***	0.009***	0.007***
	0.002	0.002	0.002
1glmm x HS degree	0.018***	0.018***	0.018***
	0.002	0.002	0.002
1glmm x Some college	0.032***	0.039***	0.032***
	0.002	0.002	0.002
1glmm x College degree	0.026***	0.031***	0.024***
	0.002	0.002	0.002
1glmm x US Some College[1]		-0.012***	
		0.002	
1glmm x US College degree[1]		-0.009***	
		0.002	
Control variables[2]	✓	✓	✓
Fixed effects			
Time spent in US	✓	✓	✓
Year, State	✓	✓	✓
Industry, Occupation	✓	✓	✓
Constant	-0.070***	-0.070***	-0.068***
	0.009	0.009	0.008
Number of Observations	3136837	3136837	3153057

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants who self reports to speak English well or very well.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Columns (1), (2), (4) and (5) exclude immigrant groups that compose about 80% of H-1B workers, based on USCIS' FY2012 Annual Report.
1. Assign US Some college and US College degree 0 for all natives;
Immigrants who are assigned 1 for US Some College and US College degree are also assigned 1 for some college and college degree.
2. Include race (four categories), age, dummy variable for whether an immigrant is naturalized, log GDP per capita of source country
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Appendix Table 3: Selection into self-employment by those who speak English well + very well

	Self-employment (vs Salaried)			
Incl. immigrants who speak English:	**Well + Very well**		**Very well**	
Language req. for salaried jobs:	**Low**	**High**	**Low**	**High**
	(1)	(2)	(3)	(4)
Linguistic Distance (LingD)	0.074***	0.080***	0.055***	0.066***
	0.002	0.003	0.003	0.004
Yrs of education	0.011***	-0.007***	0.013***	-0.007***
	0	0	0	0
Other controls	✓	✓	✓	✓
Fixed effects				
Time spent in US	✓	✓	✓	✓
Year, State	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓
Constant	-0.114***	0.540***	-0.100***	0.576***
	0.012	0.014	0.016	0.017
Number of observations	1940628	1619786	1851439	1581682
Base rate of self-employment	22%	26%	22%	26%
Selection effect	*34%*	*31%*	*25%*	*26%*

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants who reports to speak English either Very Well or Well. Language requirements for salaried jobs use O*Net Skill measures, taking the average scores of Reading Comprehension, Speaking and Writing scores. For both ACS and CPS, "Low" compares self-employed with salaried occupations with communication scores less than or equal to 50. "High" compares self-employed with salaried occupations with communication skills over 50.
Controls include race (four categories), years of education, age, and log GDP per capita of source country.
All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Calcualtions for both samples are weighted using the population weights provided by the respective surveys.

Appendix Table 4: Testing positive selection for immigrants residing in non-enclaves

Enclave definition:	95th percentile			99th percentile		
	Ex-H1Bs		All	Ex-H1Bs		All
	(1)	(2)	(3)	(4)	(5)	(6)
Education (vs below HS)						
HS degree	0.012***	0.012***	0.013***	0.011***	0.010***	0.011***
	0.002	0.002	0.002	0.002	0.002	0.002
Some college	0.024***	0.023***	0.024***	0.021***	0.020***	0.021***
	0.002	0.002	0.002	0.002	0.002	0.002
College degree	0.027***	0.025***	0.028***	0.022***	0.021***	0.023***
	0.002	0.002	0.002	0.002	0.002	0.002
1st gen Immigrants (1glmm)	-0.021***	-0.013**	-0.020***	-0.017***	-0.011**	-0.016***
	0.006	0.006	0.006	0.005	0.005	0.005
1glmm x HS degree	0.040***	0.039***	0.040***	0.040***	0.039***	0.040***
	0.005	0.005	0.005	0.004	0.004	0.004
1glmm x Some college	0.041***	0.053***	0.041***	0.039***	0.045***	0.039***
	0.005	0.007	0.005	0.004	0.006	0.004
1glmm x College degree	0.032***	0.038***	0.028***	0.041***	0.046***	0.036***
	0.005	0.005	0.005	0.004	0.004	0.004
1glmm x US Some College[1]		-0.025***			-0.013**	
		0.008			0.006	
1glmm x US College degree[1]		-0.019***			-0.017***	
		0.007			0.005	
Control variables	✓	✓	✓	✓	✓	✓
Fixed effects						
Time spent in US	✓	✓	✓	✓	✓	✓
Year, State	✓	✓	✓	✓	✓	✓
Industry, Occupation	✓	✓	✓	✓	✓	✓
Constant	-0.066***	-0.070***	-0.077***	-0.078***	-0.081***	-0.091***
	0.025	0.025	0.025	0.021	0.021	0.021
Number of Observations	529181	529181	530314	551421	551421	553211

Source: Current Population Survey, 1994-2012

Notes: Table reports linear estimates of the probability of a worker to be self-employed. Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants.

To proxy for ethnic enclaves, I rank metro areas by the size of the ethnic group population and identify the metro areas that are above the 95th percentile and 99th percentile of the distribution. I assign 1 if an immigrant resides in these metro areas, 0 otherwise.

All dummy variables defined exclusively. Reported Standard Errors; *, **, and *** indicate significant at 10%, 5% and 1% respectively.

Columns (1), (2), (4) and (5) exclude immigrant groups that compose about 80% of H-1B workers, based on USCIS' FY2012 Annual Report.

1. Assign US Some college and US College degree 0 for all natives;
Immigrants who are assigned 1 for US Some College and US College degree are also assigned 1 for some college and college degree.

2. Include race (four categories), age, and log GDP per capita of source country

Calcualtions weighted using the population weights provided by the survey.

Appendix Table 5 – Field of Degree category, ACS 2010 - 2012

Field of Degree categories from the American Community Survey (2010-2012)

Agriculture
- GENERAL AGRICULTURE
- AGRICULTURE PRODUCTION AND MANAGEMENT
- AGRICULTURAL ECONOMICS
- ANIMAL SCIENCES
- FOOD SCIENCE
- PLANT SCIENCE AND AGRONOMY
- SOIL SCIENCE
- MISCELLANEOUS AGRICULTURE

Architecture
- ARCHITECTURE

Media & Communications
- COMMUNICATIONS
- JOURNALISM
- MASS MEDIA
- ADVERTISING AND PUBLIC RELATIONS

Computer and information systems
- COMMUNICATION TECHNOLOGIES
- COMPUTER AND INFORMATION SYSTEMS
- COMPUTER PROGRAMMING AND DATA PROCESSING
- COMPUTER SCIENCE
- INFORMATION SCIENCES
- COMPUTER ADMINISTRATION MANAGEMENT AND SECURITY
- COMPUTER NETWORKING AND TELECOMMUNICATIONS

Education
- GENERAL EDUCATION
- EDUCATIONAL ADMINISTRATION AND SUPERVISION
- SCHOOL STUDENT COUNSELING
- ELEMENTARY EDUCATION
- MATHEMATICS TEACHER EDUCATION
- PHYSICAL AND HEALTH EDUCATION TEACHING
- EARLY CHILDHOOD EDUCATION
- SCIENCE AND COMPUTER TEACHER EDUCATION
- SECONDARY TEACHER EDUCATION
- SPECIAL NEEDS EDUCATION
- SOCIAL SCIENCE OR HISTORY TEACHER EDUCATION
- TEACHER EDUCATION: MULTIPLE LEVELS
- LANGUAGE AND DRAMA EDUCATION
- ART AND MUSIC EDUCATION
- MISCELLANEOUS EDUCATION

Engineering
- GENERAL ENGINEERING
- AEROSPACE ENGINEERING
- BIOLOGICAL ENGINEERING
- ARCHITECTURAL ENGINEERING
- BIOMEDICAL ENGINEERING
- CHEMICAL ENGINEERING
- CIVIL ENGINEERING
- COMPUTER ENGINEERING
- ELECTRICAL ENGINEERING
- ENGINEERING MECHANICS PHYSICS AND SCIENCE
- ENVIRONMENTAL ENGINEERING
- GEOLOGICAL AND GEOPHYSICAL ENGINEERING
- INDUSTRIAL AND MANUFACTURING ENGINEERING
- MATERIALS ENGINEERING AND MATERIALS SCIENCE
- MECHANICAL ENGINEERING
- METALLURGICAL ENGINEERING
- MINING AND MINERAL ENGINEERING
- NAVAL ARCHITECTURE AND MARINE ENGINEERING
- NUCLEAR ENGINEERING
- PETROLEUM ENGINEERING
- MISCELLANEOUS ENGINEERING
- ENGINEERING TECHNOLOGIES
- ENGINEERING AND INDUSTRIAL MANAGEMENT
- ELECTRICAL ENGINEERING TECHNOLOGY
- INDUSTRIAL PRODUCTION TECHNOLOGIES
- MECHANICAL ENGINEERING RELATED TECHNOLOGIES
- MISCELLANEOUS ENGINEERING TECHNOLOGIES
- MILITARY TECHNOLOGIES

Language
- LINGUISTICS AND COMPARATIVE LANGUAGE AND LITERATURE
- FRENCH GERMAN LATIN AND OTHER COMMON FOREIGN LANGUAGE STUDIES
- OTHER FOREIGN LANGUAGES

Legal
- COURT REPORTING
- PRE-LAW AND LEGAL STUDIES

Liberal arts, humanities
- ENGLISH LANGUAGE AND LITERATURE
- COMPOSITION AND RHETORIC
- LIBERAL ARTS
- HUMANITIES
- LIBRARY SCIENCE
- AREA ETHNIC AND CIVILIZATION STUDIES
- INTERCULTURAL AND INTERNATIONAL STUDIES
- HISTORY
- UNITED STATES HISTORY

Biology
- BIOLOGY
- BIOCHEMICAL SCIENCES
- BOTANY
- MOLECULAR BIOLOGY
- ECOLOGY
- GENETICS
- MICROBIOLOGY
- PHARMACOLOGY
- PHYSIOLOGY
- ZOOLOGY
- NEUROSCIENCE
- MISCELLANEOUS BIOLOGY

Mathematics
- MATHEMATICS
- APPLIED MATHEMATICS
- STATISTICS AND DECISION SCIENCE

Philosophy / Religious study
- PHILOSOPHY AND RELIGIOUS STUDIES
- THEOLOGY AND RELIGIOUS VOCATIONS

Science
- NUTRITION SCIENCES
- MATHEMATICS AND COMPUTER SCIENCE
- COGNITIVE SCIENCE AND BIOPSYCHOLOGY
- PHYSICAL SCIENCES
- ASTRONOMY AND ASTROPHYSICS
- ATMOSPHERIC SCIENCES AND METEOROLOGY
- CHEMISTRY
- GEOLOGY AND EARTH SCIENCE
- GEOSCIENCES
- OCEANOGRAPHY
- PHYSICS
- MATERIALS SCIENCE
- MULTI-DISCIPLINARY OR GENERAL SCIENCE
- NUCLEAR, INDUSTRIAL RADIOLOGY, AND BIOLOGICAL TECHNOLOGIES

Psychology
- PSYCHOLOGY
- EDUCATIONAL PSYCHOLOGY
- CLINICAL PSYCHOLOGY
- COUNSELING PSYCHOLOGY
- INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY
- SOCIAL PSYCHOLOGY
- MISCELLANEOUS PSYCHOLOGY

Public Policy / Administration
- CRIMINAL JUSTICE AND FIRE PROTECTION
- PUBLIC ADMINISTRATION
- PUBLIC POLICY
- HUMAN SERVICES AND COMMUNITY ORGANIZATION
- SOCIAL WORK

Social Science
- FAMILY AND CONSUMER SCIENCES
- GENERAL SOCIAL SCIENCES
- ECONOMICS
- ANTHROPOLOGY AND ARCHEOLOGY
- CRIMINOLOGY
- GEOGRAPHY
- INTERNATIONAL RELATIONS
- POLITICAL SCIENCE AND GOVERNMENT
- SOCIOLOGY
- MISCELLANEOUS SOCIAL SCIENCES
- INTERDISCIPLINARY SOCIAL SCIENCES

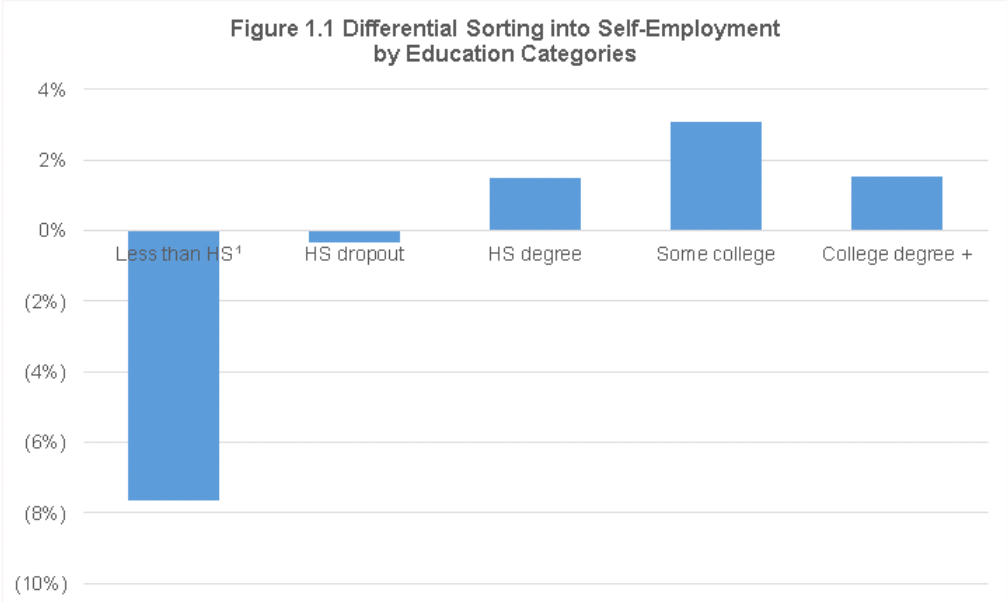
Fine arts
- FINE ARTS
- DRAMA AND THEATER ARTS
- MUSIC
- VISUAL AND PERFORMING ARTS
- COMMERCIAL ART AND GRAPHIC DESIGN
- FILM VIDEO AND PHOTOGRAPHIC ARTS
- ART HISTORY AND CRITICISM
- STUDIO ARTS
- MISCELLANEOUS FINE ARTS

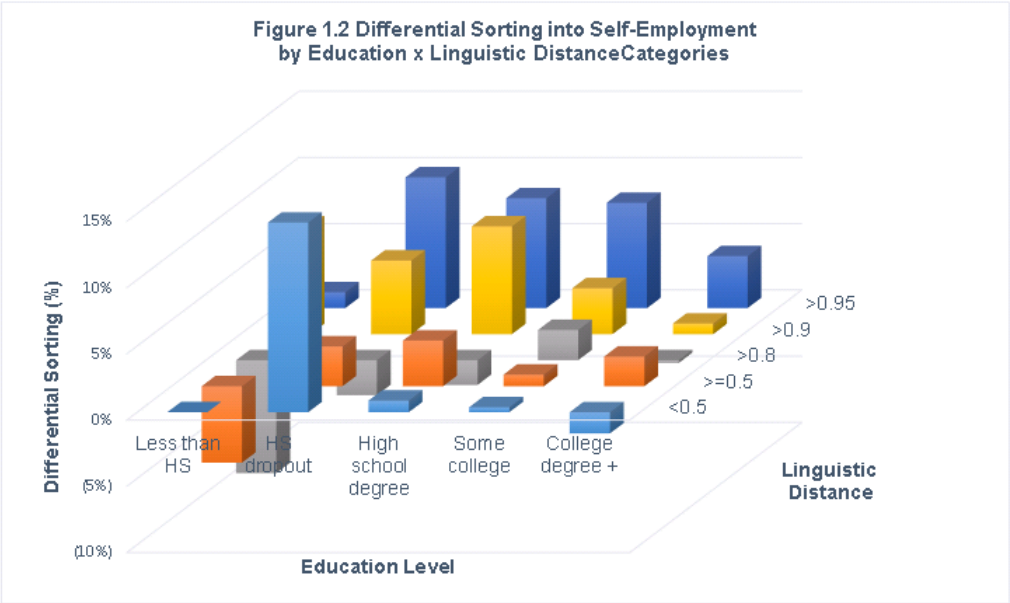
Health services
- GENERAL MEDICAL AND HEALTH SERVICES
- COMMUNICATION DISORDERS SCIENCES AND SERVICES
- HEALTH AND MEDICAL ADMINISTRATIVE SERVICES
- MEDICAL ASSISTING SERVICES
- MEDICAL TECHNOLOGIES TECHNICIANS
- HEALTH AND MEDICAL PREPARATORY PROGRAMS
- NURSING
- PHARMACY PHARMACEUTICAL SCIENCES AND ADMINISTRATION
- TREATMENT THERAPY PROFESSIONS
- COMMUNITY AND PUBLIC HEALTH
- MISCELLANEOUS HEALTH MEDICAL PROFESSIONS

Business
- GENERAL BUSINESS
- ACCOUNTING
- ACTUARIAL SCIENCE
- BUSINESS MANAGEMENT AND ADMINISTRATION
- OPERATIONS LOGISTICS AND E-COMMERCE
- BUSINESS ECONOMICS
- MARKETING AND MARKETING RESEARCH
- FINANCE
- HUMAN RESOURCES AND PERSONNEL MANAGEMENT
- INTERNATIONAL BUSINESS
- HOSPITALITY MANAGEMENT
- MANAGEMENT INFORMATION SYSTEMS AND STATISTICS
- MISCELLANEOUS BUSINESS & MEDICAL ADMINISTRATION

Other
- ENVIRONMENTAL SCIENCE
- FORESTRY
- NATURAL RESOURCES MANAGEMENT
- COSMETOLOGY SERVICES AND CULINARY ARTS
- MULTI/INTERDISCIPLINARY STUDIES
- PHYSICAL FITNESS PARKS RECREATION AND LEISURE
- CONSTRUCTION SERVICES
- ELECTRICAL, MECHANICAL, AND PRECISION TECHNOLOGIES AND PRODUCTION
- TRANSPORTATION SCIENCES AND TECHNOLOGIES

Appendix Figure 1: Differential sorting across subgroups of the immigrant population (Using the CPS)





Source: Current Population Survey 1994 -2012

Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year

Differntial sorting based on an imputed measure comparing immigrant's likelihood to self-emply to non-natives.

This is calculated as Self-Employed - Estimated Self-Employed, where Estimated Self-Employed is the fitted value for immigrant usingi the coefficients from running the following regression for U.S. born workers:

$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon$$

All education categories and linguistic categories defined exclusively.

1. Immigrants compose 64% of 'Less than HS' category, suggesting that the imputation is driven by a smaller native sample

All education categories and linguistic categories defined exclusively.

Calcualtions weighted using the population weights provided by the survey.