Hidden Gems? How Cultural Barriers Lead to Excessive Self-Employment of High Skilled U.S. Immigrants*

Yoonha Kim§
Georgetown University
McDonough School of Business

John Morgan†
University of California, Berkeley
Haas School of Business

Working Draft: July 2018

Abstract

We study how linguistic-cultural barriers produce over-selection into self-employment by highly educated immigrants. A model where talent is more easily discerned in U.S.-born than in immigrants, owing to differences in precise signalling capabilities, can rationalize such over-selection. Using U.S, population survey data, combined with measures of linguistic-cultural differences, we consistently find that the more highly educated and linguistically distant is an immigrant, the more likely is self-employment. Alternative explanations, such as language deficiencies or ethnic factors, cannot, in and of themselves, readily explain the differential sorting we observe. We argue that these employment patterns reflect an inefficient talent allocation; firms, in principle, can better harness these hidden gems—the untapped talent pool of highly educated immigrants sorting into self-employment. We further show that returns to ethnic diversity is context dependent—when the job is difficult, workers in ethnically diverse groups are more productive; when the job is easy, workers in homogeneous groups are more productive.

Keywords: Immigrant, Self-employment, Language, Culture, Human capital, Hiring

---

* We would like to thank David Card, Ernesto Dal Bó, Ross Levine, Yona Rubinstein, Joachim Voth, Romain Wacziarg and Noam Yuchtman for their very helpful guidance and suggestions as well as seminar participants at UC Berkeley, University of Nevada-Reno, Peking University, Georgetown University, UC Davis, University of Maryland and conference participants at UPenn-Wharton, HBS and NYU-Stern.
§ Visiting Assistant Professor, Strategy, Economics and Public Policy Group of McDonough School of Business, Georgetown University; Please direct questions to yoonha.kim@georgetown.edu
† Oliver E. and Dolores W. Williamson Chair of the Economics of Organizations, Haas School of Business and Department of Economics, University of California at Berkeley; morgan@berkeley.edu

1 Introduction

With foreign-born workers composing over 15% of the U.S. workforce, assimilation of immigrants is a global issue that is ever growing in importance. One way to gauge immigrants' assimilation is through their rate of self-employment—sorting into self-employment is one channel which immigrants cope with the disadvantages they face in joining mainstream economic markets (Light 1979). This is one of many factors that give rise to ethnic enterprises; previous studies, including recent work by Fairlie and Lofstrom (2014), have established that immigrants have higher propensities to self-employ. Estimation based on the American Community Survey (2005-2012) confirms this fact—foreign-born workers are on average ~18% more likely to self-employ than U.S.-born workers with same observable traits. However, interestingly, this selection is differentially stronger among the highly educated: the probability that a foreign-born worker with a college education selects into self-employment is ~26% higher. In other words, immigrants exhibit stronger positive sorting into self-employment with educational attainment than their U.S.-born counterparts. This finding is counterintuitive given the plethora of research and policy discussions on the contributions of highly educated immigrants to U.S. productivity—most recently by Hanson & Slaughter (2016); one would expect that high skilled immigrants enter into self-employment less as they better integrate in the workforce.

While an extensive literature studies immigrants' propensities to run businesses, existing theories do not account for the differential sorting patterns 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 2004) will not necessarily be stronger for the highly educated; similarly, racial and ethnic preferences against immigrants (Becker 1957) do not systematically differ by education levels. Furthermore, information based theories on discrimination, originating from Phelps (1972) and Arrow (1973), attributing perpetuating differences in labor market outcomes to unobservable characteristics, or theories of how such beliefs affect endogenous choices in human capital investment (Lundberg and Startz 1983), do not explain differences conditional on educational attainment. Hence, this prompts the question how immigrant status and educational attainment interact to generate frictions beyond the standard channels that affect employment choices in the labor market, to disproportionately sort highly educated immigrants into self-employment.

In this paper, we argue that linguistic-cultural differences or what one might term a "lost in translation effect"—employer misperceptions of a job applicant's aptitude owing to differences in language, communication or culture (Morgan and Várdy 2009)— account for the systematic sorting pattern involving highly educated immigrants. Unlike classical theories of discrimination, based on tastes (Becker 1957) or mean differences in unobserved productivity (Cornell and Welch 1996, Phelps 1972, Arrow 1973, Lundberg and Startz 1983 among others), lost in translation effects operate through differences in signal quality. The main idea is that, in two populations with equal (expected) talent, the population that can communicate individual talent less precisely will suffer more talent misallocation, i.e. talented individuals from that population are more likely to be rejected compared to talented members of the other population. The likelihood of failing to identify talented individuals from the disadvantaged population increases with the degree of the noisiness of the signal as well as with the scarcity of the requisite level of talent in the population. This hypothesis forms the heart of our study.

We test lost in translation effects by comparing the relative propensity to enter self-employment, one indicator of unsuccessful employment, between U.S.-born and immigrant workers. We exploit the diversity of immigrants' linguistic-cultural backgrounds together with differences in educational attainment, which proxies for the scarcity of talent. Measuring linguistic and cultural differences is notoriously subjective and difficult, yet widely conceded to play an important role in economic, and especially hiring, outcomes (Rivera 2012). We quantify these differences by proxying for the effectiveness of communication using Wacziarg and Spolaore's (2009) "linguistic distance" measure and cross-check our findings by adding their "cultural distance" measure as well. The linguistic distance measure is based on Fearon's (2003) approach of tracing the number of branches that separate two languages in a language tree whereas the cultural distance measure is based on a completely different method using the World Values Survey.

The theoretical prediction that informational friction increases with the difficulty of the job and the noisiness of the signal is supported by our data: the likelihood of immigrant self-employment systematically increases not only with higher education levels but also with greater linguistic distance. Our 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 to self-employ increases by 3-5%. Moreover, our results qualitatively hold when we use the cultural distance measure in lieu of linguistic distance, suggesting that linguistic distance contains the degree of cultural acquaintance.

We isolate the impact of linguistic-cultural differences from spurious correlations through other ethnic or linguistic channels that explain immigrants' propensities to run businesses. Critically, we identify a strong interaction between immigrant status and educational attainment that disproportionately sorts highly educated immigrants into self-employment.

We validate two additional patterns consistent with lost in translation effects. First, we 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 we find a mitigating effect for those who were exposed to the U.S. education system or who immigrate before the age of 10. Second, we 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, we 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.

We further evaluate the predictions of the lost in translation framework relative to the predictions from alternative hypotheses. Among other potential drivers, we particularly investigate whether linguistic distance merely captures 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.

We address this concern in two ways. First, we build on Autor et al. (2003) to decompose occupations by their skill requirements by 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, we 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, we reject the hypothesis that linguistic distance simply measures language as a productivity input.

Second, we 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, we show that our results hold on a subsample of immigrants who report speaking English well. The fact that the theoretical predictions hold even when we account 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, we show that the differential sorting between immigrants and non-immigrants in the labor market does not reflect immigrants' inability to communicate. Rather, we argue that linguistic-cultural mismatch importantly accounts for this differential sorting, representing a systematic bias, by providing a set of empirical evidence that is consistent with the "lost in translation effect". Alternative studies that highlight immigrants' self-employment tendencies such as the role of ethnic enclaves or social network effects, do not explain the systematic variation in self-employment by human capital in and of themselves.

This study relates to several strands of the literature. First, this paper studies the effect of language and culture on labor market outcomes in the context of immigrants in the U.S. While prior works on immigrants' language and occupational choice mainly consider how immigrants' lack of communication skills affect their labor market outcomes, we investigate how immigrants face cultural barriers in their job search[1]. We take the epidemiological approach, per Fernández (2011), by studying how variation in job and language characteristics explain differences in rates of entrepreneurship entry between immigrants and the U.S.-born. Previous studies on this approach investigate female labor force participation and fertility (Fernández 2007, Fernández and Fogli 2006, and Fernández and Fogli 2009), political participation (Alesina and Giuliano 2009) or preferences for redistribution (Luttmer and Singhal, 2010), among others. We add to this literature by showing a set of evidence on how linguistic-cultural background shape overrepresentation of high-skilled immigrants in self-employment. While English ability, in and of itself, is an important consideration explaining a large part of the immigrant-native earnings differential (Ferrer et al. 2006), the majority of immigrants today consider themselves to be well versed in English[2], and moreover, cultural differences persist even when an immigrant is proficient in English. Therefore, it is important to explore beyond the traditional channel of immigrants' ability to communicate. In this evaluation, we argue that immigrants' linguistic distance--the familiarity of their first language to English--matters more than their proficiency.

Second, this paper relates to the literature on manager biases in the hiring process and, more broadly, the allocation of human resources. Petersen and Saporta (2004) note that

[1] Among studies on language and immigrants' labor market outcomes, Dávila and Mora (2004) study how English-language fluency affects the earnings of self-employed immigrant workers overtime, 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 substitutability of immigrants, and Peri and Sparber (2009) show how immigrants sort into manual tasks, while non-immigrant workers shift to more language-intensive jobs.

[2] Based on the American Community Survey 2005 - 2012, 70% of immigrants consider themselves to speak English 'well' or 'very well'.

discrimination is most heightened in the hiring setting, and there have been studies on the effect of manager biases on hiring and productivity outcomes: 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) study the introduction of an online job-testing service in low-skill service sector; Giuliano et al. (2009) investigate 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. While studies such as Rivera (2012) show how cultural matching is an important factor in hiring decisions, there have been empirical limitations to studying the consequences of cultural differences. We exploit immigrants' linguistic-cultural backgrounds as well as their educational attainment to study how cultural frictions explain generalizable patterns across the U.S. population. Similar types of ethnic group-based biases in screening have been explored in the context of venture capital financing (Hegde and Tumlinson 2014) and R&D alliance formation (Joshi and Lahiri 2014); we show how such biases affect matching of workers to firms in the labor market. Based on my findings, we further argue how cultural differences can lead to misallocation of highly educated immigrant workers and thus add to the discussion of the efficient allocation of human talent (Bell et al. 2016, Hsieh et al. 2016).

This paper also speaks to the entrepreneurship literature. As a byproduct of cultural friction, we uncover an unexplored mechanism that drives highly educated immigrant 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) or individual traits (Lazear 2005; Levine and Rubinstein 2017). A branch of this literature has discussed how the choice of entrepreneurship reflects various types of labor market frictions: in particular, how unobserved ability (Hegde and Tumlinson

2015) or educational mismatch (Stenard and Sauermann 2016) cause imperfect matching between workers and firms (Åstebro et al. 2011). We build on these discussions by investigating how cultural friction plays a role in sorting highly educated foreign workers into business ownership. Similar types of ethnic group-based biases in screening have been explored in the context of venture capital financing (Hegde and Tumlinson 2014). We show how such biases also affect matching of workers to firms in the labor market.

2 Conceptual Framework

In this section, we explain economic forces that produce differential sorting by immigrants into self-employment.

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 that build 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) agree 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 that there may be differential outcomes. In other words, the framework is meant to illustrate how there may unintendedly biased employment outcomes arising from differences in linguistic cultural backgrounds.

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 worried less 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 apply to difficult jobs and send noisy signals.

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 noisier 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 quality of signals. 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 send precise signals should sort into self-employment less than otherwise similar immigrants.

We proxy for signal precision in two ways: (1) whether an immigrant is exposed to the U.S. education and (2) whether an immigrant came at a young age.

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 they would not suffer from this informational friction as much as the employer is more likely to be from the same ethnic group. 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. We 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.

Furthermore, how immigrants sort in the labor market should not be confined to the theoretical framework. While it is unlikely that the systematic sorting pattern that we observe in the labor market is solely a result of a particular cultural bias in the hiring process, it is quite likely that those biases can have lasting effects on forward-looking immigrants. Immigrants, who anticipate their likely outcome, may shy away from interviews and may more broadly stop making attempts to culturally assimilate. Previous studies have shown how cultural matching is an important factor in hiring decisions in elite firms (Rivera 2012) and how discrimination is most heightened in the hiring setting (Petersen and Saporta 2004). While we should not limit our examination for how workers sort in the labor market to hiring settings, understanding cultural frictions in the hiring process would provide important insights about systematic patterns in the labor market.

The theoretical framework is distinct from theoretical and empirical studies of immigrant selection originating from Borjas (1987). While Borjas (1987)'s selection model concerns the decision to immigrate or not, we focus on the selection into self-employment conditional on being in the U.S. Theoretically, the underlying forces that generate the positive selection pattern may seem similar in that variance parameter drive the different selection patterns[3], our theory relies on the noisiness of the signal conditional on the ability of individuals. Implication of the Borjas (1987)'s selection model in the context of the lost-in-translation model would be that, knowing immigrants are positively selected, employers would adjust their hiring threshold downward for immigrants—this would make it less likely for us to detect the differential sorting pattern.

In the following sections, we describe the data and empirical methodology to test these propositions and rule out potential alternative factors that may be driving the predictions. We empirically operationalize this noisy signalling measure using the linguistic distance measure.

3 Data Description

To test the theoretical predictions outlined above, we use two distinct data sets to examine (a) differential selection into self-employment by U.S.-born and immigrant workers, (b) systematic patterns of selection into self-employment across subgroups of immigrants and (c) potential alternative explanations that may be driving this pattern. We use the American Community Survey ("ACS") for the years 2005 to 2012 along with the March Supplements of the US Current Population Survey ("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, we 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.

[3] Difference in variance in Borjas 1987 arise from the difference in income inequality between the host and the origin country

[Insert Table 1 Here]

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, we 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. We 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, we use actual grade levels or degrees attained as well as years of education. We categorize education into three education categories: below high school degree, high school degree, and college and above. The rationale for this categorization is based on Arcidiacono et al (2010)'s study showing how where one went to college plays a direct role in revealing one's ability in the labor market, while a high school degree only gradually reveals such an ability. Years of education is imputed based on the actual grade level or degree. In cases where educational attainment spans multiple grades, we 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, we need proxies for the difficulty of the job and the noisiness of the signal. We 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, we run our results using two different measures of noisy signaling: (1) an indicator for immigrant status; and (2) a continuous measure of linguistic distance. We further check our results using continuous measure of cultural distance.

The linguistic distance measure is an off the shelf measure developed by Wacziarg and Spolaore (2009) to proxy for the cultural distance 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. For example, English is defined by several branches in a language tree, Indo European – Germanic – West Germanic – Anglo Frisian – English, and the distance of another language can be based on the number of separating branches. Previous studies, including Montalvo and Reynal-Querol (2005), have used this measure as a summary statistic for intergroup cultural differences.

[Insert Figure 1 Here]

Figure 1.1 exhibits a bubble chart that shows the relationship between self-employment rate and linguistic distance, where the size of the bubbles represents the size of the ethnic group. We use a standardized 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 U.S. and countries such as the UK and Australia is closer, while most Asian countries will fall on the farthest end. One thing to note is that English speaking countries such as Australia, United Kingdom or Canada do not have linguistic distance 0. This is because the linguistic distance is created at the population level and aggregated at the country level using the composition of the population. Another thing to note is that there are many countries grouped under linguistic distance 1. This is a feature of the measure as any language not part of the Indo-European language tree will have the furthest linguistic distance from English. For de jure English-speaking countries such as Singapore and India, we assign a mid-value. We 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.

The cultural distance measure, also from Wacziarg & Spolaore (2009), is based on how similarly people from different countries have answered the questionnaires in the World Value Survey. The measure is based on 98 questions asked on opinion polls under the following themes: perception of life, work, family, politics and society, religion and morale and national identity. Figure 1.2 exhibits a bubble chart that shows the relationship between the self-employment rate and cultural distance, where the size of the bubbles again represents the size of the ethnic group. The change in ordinary distance between U.S. and China and U.S. and Korea is interesting, for example. We use linguistic distance as the main explanatory variable throughout the study, however, as the linguistic distance measure directly connects to communication disadvantage in the conceptual framework.

Our 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 linguistic distance in this manner is that it confounds immigrants' inability to communicate well with the cultural barrier they face. To address this problem, we 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'. We use this measure to test whether linguistic distance merely captures immigrants' inability to speak English.

To further address this problem, we run more nuanced tests on subsets of occupation categories that require more or less communication skills. We 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, we use communication skills required for different jobs, which we impute by taking the average scores of reading comprehension, speaking and writing skills required for jobs. Using this measure, we are 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.

We also consider institutional factors that shape immigrants' employment choices. In particular, we account for H-1B visa holders, whose career trajectory would likely differ from others because their immigration status ties them to a specific employer, and exclude them from our baseline empirical results. 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, we identify immigrant subgroups that would compose ~70% of the H-1B holders based on USCIS' FY2012 Annual Report to Congress. Specifically, we exclude Indian, Chinese and Canadian immigrants with a college degree working in universities or in computer- or engineering-related occupations. The inclusion of these immigrants group do not meaningfully change the empirical results, however.

Finally, given that immigrants are not proportionately distributed across space, we construct two additional measures. First, to determine the different dynamics in ethnic enclaves, we create a proximate indicator for whether an immigrant resides in an enclave. Specifically, for each ethnic group, we 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. We assign 1 if an immigrant resides in these metro areas and 0 otherwise. Second, we construct a measure for what proportion of their organizations immigrant composes. We proxy for an organization by using an occupation category within an industry in a metro area and assess the proportion of each ethnic group in each metro area-industry-occupation cluster. Analyses using both of these measures are conducted using the CPS as metroarea information is not available for the ACS.

4 Empirical Methodology

In this section, we discuss the empirical methodology employed to test the main predictions of the framework. We use linear probability models with an indicator for self-employment as the dependent variable and individual- or origin country-level characteristics as explanatory variables. We use the following specification to test the main predictions, which concern stronger positive selection into self-employment by immigrants with higher education and noisy signals:

$$\begin{aligned}\mathbb{1}(SelfEmp)_{i,c,FE} = \beta_0 + \beta_1 Signal\,Noise_c + \beta_2 Education_i + \beta_3 Signal\,Noise_c \times Education_i \\ + \beta_4 X_i + \beta_5 X_c + \lambda_{FE} + \epsilon_{i,c,FE}\end{aligned} \tag{1}$$

For individual i from country c, *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 indicate the propensity to be a self-employed rather than to be a salaried employee. *Signal Noise* is the measure for noisy signals where we use two different measures: (1) an indicator for first-generation immigrant status; and (2) a continuous measure of linguistic distance. *Education* is either years of education or education categories, including less than a high school degree, high school degree, some college and above. X_i includes individual-specific controls, such as race categories and years spent in the US and X_c is the natural log of the GDP per capita of the origin country. The specification also includes fixed effects for age, year, state, industry, and occupation. For U.S.-born individuals, we assign age for the number of years spent in the US. Time spent in the US together with year and age fixed effects 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 2014). Standard errors are clustered at the origin country level.

In equation (1), the β_1 coefficient indicates the additional likelihood that an immigrant who sends the noisiest signal will self-employ in comparison with their U.S.-born counterpart. A positive value for the combination of coefficients β_2 and β_3 indicates that an immigrant is more likely to self-employ than a U.S.-born individual with more education.

To assure that the increase is statistically significant with higher education, we examine significant differences in the sorting effect with higher education by categorizing β_2 and β_3 into three education categories and conducting t-test between coefficients. We further categorize β_1 and β_3 into four different levels of noisy signals for linguistic distance and cultural distance. This test aims to assure that the selection effect is not driven by a particular ethnic group or subset of immigrants.

While signal noise is a country level measure that may be endogeneous to other unobservable factors, we conduct additional empirical test to provide evidence of the existence of the "lost in translation effect". To test for subsequent hypotheses concerning whether the selection effect is mitigated for immigrants who send more precise signals we use the following specification:

$$\mathbb{1}(SelfEmp)_{i,c,FE} = \beta_0 + \beta_1 Signal\,Noise_c + \beta_2 \mathbb{1}(SignalPrecision)_i + \beta_3 Signal\,Noise_c \times \mathbb{1}(SignalPrecision)_i + \beta_4 X_i + \beta_5 X_c + \lambda_{FE} + \epsilon_{i,c,FE} \quad (2)$$

This specification is used to test subsequent hypotheses within immigrants, where we add an indicator for signal precision. There are two tests regarding signal precision. The first relates to cultural assimilation, which is 1 if an immigrant is culturally assimilated and 0 otherwise. We proxy for cultural assimilation by identifying immigrants who (1) are exposed to the U.S. education and (2) have immigrated at a young age. The second relates to how represented immigrants are in their occupation. We use a continuous measure of occupation representation as described in section 3.

The main coefficients of interest here are β_2 and β_3. The β_2 coefficient indicates the effect of a more precise signal on the selection into self-employment. Hence, a negative coefficient indicates that there is a mitigating effect for immigrants who send a more precise signal. β_3 will help further us to assess whether there is heterogeneity across immigrant groups with different degrees of noisy signals.

5 Key Empirical Findings

In this section, we test the predictions delineated in section 2 by using the empirical setting and methods discussed in sections 3 and 4. First, we test the predicted differential sorting pattern across educational attainment and noisy signal. Second, we examine sorting patterns of immigrants who have culturally assimilated. Third, we investigate immigrants' propensities to self-employ when they are surrounded by their co-ethnics.

5.1 Selection into self-employment with noisier signals and more education

In this section, we test the main predictions of the framework, predictions 1 through 3: the noisier an individual's signal, the stronger the selection into self-employment; the noisier the signal and the higher the education, the stronger the selection. Prediction 1 concerns the stronger positive selection into self-employment by immigrants with higher education; 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.

Average years of schooling from Table 1 provides suggestive evidence for selection in prediction 1. 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.

[Insert Table 2.1 Here]

In Table 2.1, we test for broad monotonicity for differential selection into self-employment by individuals depending on the noisiness of the signal and education levels. Panel A shows the results based on the ACS, while panel B shows the results using the CPS; we show results using two measures of noisiness of signals: immigrant status and linguistic distance. While immigrant status is an indicator variable, the linguistic distance measure is a continuous one that allows us to distinguish between immigrants who share some branches of the same language family tree with English from those who do not. While there are different ways to measure similarity between an immigrant origin country and the U.S., we use linguistic distance because the theoretical framework is based on noisy signals created owing to linguistic differences and the linguistic distance measure best maps the theoretical to the empirical setting. This measure is imperfect, however, and hence we further test our results using cultural distance in Appendix Table 4.

The explanatory variables of interests are measures of noisy signals and its interaction with years of education: Prediction 2 would suggest that the coefficient for the noisy signals is positive and predictions 1 and 3 predict that this coefficient is especially large for the highly educated. Other controls hold fixed other observable traits, including race, years of education, time spent in the US, log GDP per capita of origin country and fixed effects for age, year, state, industry and occupation categories. Hence, the interpretation is the likelihood to self-employ for a given occupation in a given industry.

The results in Table 2.1 provide evidence that monotonicity directionally holds. In line with prediction 2, the coefficients for the noisy signal measures in columns (1) and (3) are both significantly positive. The interpretations of coefficients for Panel A using ACS are as follows: the self-employment rate of individuals with the noisiest signal—whether that is being foreign-born or having a linguistic distance of 1—is 2.4% and 3.0% higher than their U.S.-born counterparts respectively. Considering the base rate of self-employment of 13%, this translates into a selection effect where being an immigrant makes one 18% more likely to self-employ, and an immigrant from the most linguistic distant country is 23% more likely to self-employ.

Furthermore, in support of predictions 1 and 3, which suggest that there is a stronger sorting effect for the linguistically distant and the more educated, the interaction terms between the noisy signal measure and years of education added in columns (2) and (4) are also significantly positive. This result indicates that the average increase in the likelihood to enter into self-employment by individuals with noisy signals masks heterogeneous effects across different levels of educational attainment. With an additional year of education, individuals with a noisy signal have a 0.4% - 0.5% higher rate of self-employment. Considering the base rate of self-employment of 13%, this translates into a selection effect of 3 – 4%.

Notably, the predictions hold using the Current Population Survey as well as shown in Panel B. [For the remainder of the paper we conduct our analysis based on the ACS and show results using the CPS in Appendix.] Table 2. While the P-value of the t-test testing the difference between B and C does not hold as significantly as in ACS, results using the CPS is consistent with that using the ACS not only in terms of the direction of the coefficient, but also in terms of their magnitudes.]

[Insert Table 2.2 Here]

In Table 2.2 we include indicators for three different education categories—below high school, high school and college education—rather than years of education in the specification and the interaction between the noisy measure and each education category, respectively labelled as A, B and C. The framework predicts that the selection effect would differ between the high and low educated immigrants, and hence it is crucial to test whether the increase holds across different levels of educational attainment: specifically, coefficients B and C are predicted to be statistically significantly positive; moreover, the difference between coefficients for B and C would also need to be statistically significantly increasing. Hence, we also report the p-value from the t-test to test the equality between coefficients B and C.

The results again are shown to hold for both measures using the ACS. The coefficients can be interpreted as follows: the rate of self-employment for individuals who are foreign born, or who have a linguistic distance of 1, with a high school degree (college education) are 2.9% (4.4%) and 3.2% (5.1%) higher respectively than their U.S.-born counterpart with a high school degree (college education). Moreover, the increase between high school degree and college education is statistically significant. The main results strongly hold using both immigrant status and linguistic distance.

We also report results using unincorporated self-employment rather than self-employment as the dependent variable in columns (3) – (4). Prior research has shown that unincorporated self-employment may proxy for necessity-based, rather than opportunity-based, entrepreneurship (Levine & Rubinstein 2017). Interpretation of the coefficients are similar as above. Notably, results for selection into unincorporated self-employment is stronger than the selection into self-employment. This suggests that the highly skilled immigrants who differentially select into self-employment are more likely do so out of necessity.

Additional robustness checks for the main results are shown in Appendix Table 1, 2 and 3. In Appendix Table 1, we replicate Table 2.2 for a subset of immigrants. In Panel A, we limit the sample of immigrants to those who immigrated after the age of 25. This limitation aims to ensure that the results are robust to an unlikely but possible reverse causality where people attain more education to change their employment outcomes. In Panel B, we limit our sample to immigrants who spent more than 10 years in the U.S. This restriction aims to check that the results are not primarily driven by other temporary factors, such as restrictions imposed by immigrants' visa status. The results qualitatively hold for both immigrant status and linguistic distance for both subsets. 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 controlling for time spent in the US in our specification—which would correlate with worker experience—we partly control for such statistical discrimination arising from the mean. The fact that there are abilities that remain uncertain to the employer even after we control for these experiences suggests that statistical discrimination on the variance also plays an important role in the labor market.

In Appendix Table 2, we replicate the results using the CPS. Again, the results directionally hold, and the respective coefficients for B and C are statistically significant. We use a conservative test using clustered standard errors at the country level, and as a result, we lose significance between coefficients B and C.

In Appendix Table 3, we show results of a more conservative test clustering standard errors at each of the education category by origin county level. This procedure is appropriate if one believes that there are correlated characteristics of particular education groups of a particular country. While we lose statistical significance at the 10% level for the between coefficient comparisons, the results qualitatively hold.

In Appendix Table 4, as previously mentioned, we replicate Table 2.2 using the cultural distance measure. Even though there is sparse representation of countries, as shown in Figure 1.2, we find quite strong results consistent with previous findings. There are two things to especially note. First, it is remarkable that the predictions of the framework hold using different measures of country level distance measures, collected based on a completely different method, across two different data sets representative of the U.S. population. Second, the fact that the measure of cultural distance and linguistic distance show qualitatively similar results, suggest that linguistic distance has the capacity to explain cultural difference, beyond mere linguistic difference.

[Insert Table 2.3 Here]

In Table 2.3 we further evaluate the differential selection by linguistic distance categories; we subset our immigrant sample approximately into quartiles based on their linguistic distance. There are two examinations to make. First, to ensure that the empirical results are not driven by threshold effects or other nonlinear effects of linguistic distance, the effects would need to hold across columns (1) through (4). Second, theory would predict that the stronger positive sorting by immigrants will be the least intense for the least noisy category, column (1), and the most intense for the noisiest category, column (4).

The results shown in columns (1) through (3) satisfy both examinations: immigrants' positive sorting with more education holds for columns (2) and (3) but not for column (1). The fact that the systematic sorting pattern appears in both columns (2) and (3) assures us that it is not a particular segment of the linguistic distance that is driving the result. Furthermore, the fact that column (1) does not exhibit the same pattern convinces us that those who have less noisier signals do not suffer from the same problem. In short, these result shows that the effect of imprecise signaling holds within immigrant groups subgroups, not just between immigrants and non-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.

While the result shown in column (4) is less consistent, this is not surprising. While the overall selection effect into self-employment is very strong, the systematic sorting pattern does not hold across education categories. The framework suggests that this is the group to which the systematic sorting effect would most strongly apply. This incongruence may be owing to the fact that column (4) lumps many countries under one category. This is shown clearly in figure 1.1 where very different countries such as Israel, Korea, Vietnam and Laos are all grouped together in the furthest linguistic distance group. More importantly, immigrants from linguistically far countries tend to be more highly educated relative to other immigrant groups and therefore we lack variation in educational attainment within the category. Hence, it is not surprising that the linguistic distance measure loses explanatory power for this subset.

In Appendix Table 5, we replicate the results for Table 2.3 using the CPS. Similarly, the results hold for the first two columns but not for columns (3) and (4), where the linguistic distance measure have less explanatory power.

There are some limitations to mapping the theoretical framework to our empirical results, however. 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. We assume that the highly educated apply to the jobs in which the talent is scarce. While the overall direction of the measure fits the framework well, some tapering effects exists when education categories are further broken down into advanced degree and college degree. The reason for this discrepancy between theory and empirics may be that educational 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.

5.2 Cultural assimilation and selection into self-employment

In this section, we test the subsequent hypotheses consistent with the framework, predictions 4: whether culturally assimilated immigrants select into self-employment less. We proxy for the degree of an immigrant's cultural assimilation in two ways: we examine (1) immigrants exposed to the U.S. education system; and (2) the subset of immigrants who immigrated at a younger age. Given that both subsets involve immigrants' age at arrival, the examination is conducted within immigrant groups in order to effectively control for immigrant cohort.

First, we assess whether the selection effect into self-employment is mitigated for immigrants with a U.S. education. Using immigrants' age of immigration, we identify immigrants who arrived in the U.S. before the age of 21 and received their high school or college education in the U.S. The results are shown in Table 3. Individuals who have been exposed to the U.S. education system will be represented in the interaction terms labelled A and B, as well as the respective education categories.

[Insert Table 3 Here]

The interpretation of the interaction terms in column (2) are as follows: the self-employment rate for immigrants who have had high school (college) education in the U.S. is 1.1% (1.5%) lower than their counterparts who received high school education outside of the U.S. In other words, for both high school and college education categories there is a mitigating effect. The mitigating effects for these two categories do not appear to be significantly different as shown by the reported p-values of the t-tests comparing coefficients.

In columns (3) and (4) of Table 3 we further breakdown college and above into some college and college degree. This enables us to assess whether the offsetting effect is driven by acquisition of a college degree: if college degree drives the mitigating effect this would suggest that the differential selection is a result of observed ability of the candidate rather than noise in their signals; on the other hand, if some college offsets the selection effect, it suggests cultural assimilation importantly accounts for the selection.

In support of the hypothesis that cultural differences push immigrants into self-employment, the magnitude of the coefficients for those with some college education in the US and those who completed their degree in the US are similar. This suggests that the offsetting effect coming from cultural adjustment is just as strong as the effect that comes from cultural adjustment and credential acquisition. 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. We argue that it is the imprecise signal owing to cultural differences that affects immigrants' employment outcomes.

Second, we examine whether immigrants who immigrate at a younger age suffer less from the noisy signaling problem and are less likely to select into self-employment. We exploit the fact that cultural assimilation naturally interacts with immigration age. While Bleakley and Chin (2010) compared social outcomes for immigrants depending on their age of immigration, we study the relation between age of immigration and employment outcomes.

We add a variable that indicates whether an immigrant came to the US before the age of 10. The results are shown in Table 4. In column (1), we add the indicator to the standard specification used in column (3) of Table 2.1, in column (2), we further interact this indicator with the linguistic distance measure. The framework would predict that the coefficient for this indicator would be negative, as immigrants who come to the US at a younger age will more likely be culturally assimilated, and that this effect should be stronger among the linguistically distant. In line with the predictions, we find strong negative coefficients for both terms in columns (1) and (2).

[Insert Table 4 Here]

There are two important things to note. First, the negative coefficient for the indicator in column (1) suggests that the immigrant subset that arrived before the age of 10 exhibit a 3% lower rate of self-employment. In other words, the selection effect differs across immigrants depending on their time spent in the U.S. Second, and more importantly, this masks heterogeneity across immigrant groups of different linguistic distance: the self-employment rate for immigrants who are part of a linguistically distant group and arrived at a young age is 5.7% lower, as shown in column (2).

We further argue that immigrants who have been culturally assimilated from a young age may develop a very nuanced but specific skill set, which are not reflected in language proficiency. Columns (3), and (4) 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. Contrary to what one would expect, the decrease in the selection effect is not monotonic in column (4). 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.

As a robustness check, we replicate column (2) of Table 4 in Appendix Table 6, using different indicators cutting immigration age at age 7, 8, 9 and 11. Our results are not sensitive to how we define immigrants who arrive at a young age.

5.3 Sorting when immigrants compose the majority group

In this section, we test prediction 5: whether immigrants who compose the majority in their group select into self-employment less. While we 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.

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. Thus, the framework would predict that immigrants interviewing with another immigrant from the same ethnic group, or residing in an ethnic enclave, would be less exposed to this information problem.

While an ideal data set would identify the ethnicity of both the applicant and the recruiter, this specific information is not available. Instead, we identify immigrants surrounded by their co-ethnics, using information from the CPS about the metro area of individuals' residence. Specifically, we identify the metro area-industry-occupation cluster of workers and use it to proxy for their "workplace". To describe the measure we chart the self-employment rate of ethnic groups against the average representation in the "workplace" in Figure 2.

[Insert Figure 2 Here]

We show regression results in Table 5, which includes a variable for representation of immigrants' ethnic group in their cluster. Along with standard controls, we include fixed effects for age, immigrant cohort, state, year, industry and occupation categories.

[Insert Table 5 Here]

The framework predicts that immigrants surrounded by their co-ethnics in their workplace are less likely to select into self-employment and the regression results support this: immigrants with highest representation in their workplace are 3.7% less likely to enter into self-employment than those who are not part of the majority group. Moreover, coefficients in column (3) suggest that this mitigating effect primarily comes from the linguistically distant immigrants: the self-employment rate of a linguistically distant immigrant who is most represented in the workplace 5.9% lower relative to a linguistically close immigrant. This result is in line with a story that when immigrants compose a critical mass in their organization, they will face less informational friction. In other words, there is path dependence in hiring practices owing to cultural mismatch and indicates how a diverse workforce can beget a diverse workforce.

This result can also be generalized to understand dynamics near ethnic enclaves. Our findings resonate with 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.

6 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 fail to appropriately match with a firm. How much of this can be explained by language proficiency or other factors? In this section, we compare the predictions of the framework with the predictions of alternative hypotheses. In particular, we 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 cultural differences accounts for immigrants' self-employment decisions. We address these empirical challenges in this section.

6.1 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 our study, we 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, we address this challenge in three ways.

First, we 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. If we suppose that the systematic pattern is driven by the fact that communication between the employer and the candidate determines is an important input into production then language deficiency would damage workers in communication-intensive occupations to a greater extent than in less communication-intensive occupations. If this were the 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, we decompose occupations by their skill requirements. Specifically, we use the O*Net Skill measure to characterize occupations by their degree of communication intensity. We 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 6 reports the results. Columns (1) and (2) 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 (3) and (4) compare salaried jobs that require high levels of language skills with self-employment.

[Insert Table 6 Here]

There are two important things to note from our results. First, the selection effect holds not only in jobs that are communication intensive, but also in jobs in which communication is less intensive, as shown by the 2.8% coefficient for the linguistic distance variable in column (1). The interpretation of this coefficient is that the self-employment rate of the most linguistically distant immigrant is 2.8% higher than their U.S.-born counterpart. Considering the base rate of self-employment of 21%, this also suggests that the linguistically distant immigrants are 14% more likely to self-employ. Second, and more importantly, this selection effect is qualitatively similar across jobs with different communication intensity—the resulting selection effect from jobs that require higher levels of communication skills in column (3) is also 14%. In other words, the sorting effect does not increase as a function of the communication intensity of the job. Thus, we reject the hypothesis that linguistic distance merely measures language as a productivity input

Second, we 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 would disappear, once the sub-setting on immigrants reports that they speak English well. Table 7 presents results replicating columns (3) and

(4) of Table 2.1. Columns (1) and (2) repeat results from Table 2.1 and columns (3) and (4) replicate the analysis on a subset of immigrants who report to speak English well.

[Insert Table 7 Here]

Our results hold even when we include only immigrants who speak English well; Column (3) of Table 7 suggests that the self-employment rate of the most linguistically distant immigrants is 3.9% higher; the interaction term in column (4) suggests that the positive selection effect also holds.

One thing to note is that the selection effect from the linguistically distant immigrants is stronger from the subset of the immigrant who speak English well. Hence, results using self-reported English scores would show the opposite results of those using linguistic distance, where immigrants who do not speak English very well tend to select into self-employment less often. This finding is consistent with previous studies (Fairlie and Meyer 1996, Portes and Zhou 1996) which show that more linguistically deficient individuals are less likely to enter into self-employment.

Our 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).

We 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.

Finally, we assess the effect of cultural distance while controlling for linguistic distance. In other words, we 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.

Our results support that linguistic distance proxies for cultural differences. In Table 8, we show results controlling for both cultural and linguistic distance in using a similar specification as in Table 2.1. Our results shown in column (2), suggest that 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 while holding linguistic distance fixed. Hence, these results confirm that the linguistic distance measure proxies for the noise effect owing to differences in the discourse system, and furthermore that linguistic distance has capacity to explain beyond language proficiency.

[Insert Table 8 Here]

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 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.

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

6.3 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, we show that ethnic group factors that affect employment choices do not fully explain self-employment decisions. We address this concern in two ways.

First, we 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. We exploit the fact that English shares two more branches with German or Dutch than with French. Hence, we 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 9 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), we find strong support for the hypothesis with a coefficient of 17.2%, when we include standard controls but do not include any fixed effects. Once we include fixed effects for 22 major occupation categories, however, the result loses significance, as shown in column (2). We conjecture that the test may lose variation since immigrants from Belgium or Switzerland may be concentrated in particular occupation categories. Hence, instead, we 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.

[Insert Table 9 Here]

Overall, we 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 when ethnic group-specific factors are taken into account.

Second, we test whether the positive selection effect holds for a subset of immigrant groups that are not surrounded by their co-ethnics. In order to test this, we create an indicator for whether an immigrant is part of the most represented ethnic group in her metroarea-industry-occupation cluster, as discussed in section 5.3. The results for this test are presented in Appendix Table 7. We replicate the main specification shown in columns (1) and (2) for a subset of immigrants in columns (3) and (4).

The main results of the study hold for the subset of immigrants less likely to be influenced by their ethnic group: column (3) shows that the selection into self-employment holds just as strong from the subset of immigrants that are not surrounded by their co-ethnics; column (4) shows that the stronger positive sorting with respect to education level holds as well.

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, we 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. While there may be other factors that could generate any one of the empirical pattern, the set of empirical results we present 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. We attribute such bias to systematic bias arising from cultural mismatch in the context of hiring immigrant workers.

7 Managerial Implications

In this section, we discuss managerial implications for this study. Specifically, we first discuss how firms can use the findings of the paper and benefit by adjusting their HR practices. Second, we estimate the productivity gain for society at a partial equilibrium in which immigrant workers who are misallocated as self-employed are employed in firms. Third, we assess which firms in a given industry occupation category would realize the productivity gains. Finally, we discuss the limitations of the paper.

7.1 How and which firms should effectively screen?

How should firms improve their hiring strategies to attract the most productive workers? 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. Siegel et al. (2014) show how multinational firms can gain competitive advantages from hiring the excluded group to positions of managerial authority; we argue that firms can domestically gain competitive advantages by overcoming barriers to attracting immigrant workers.

First, our findings suggest 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 scrutinizes people of different cultural and ethnic backgrounds more carefully. 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. Our 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. We 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, we conduct our analysis based on the field of degree. We use the ACS to collapse over 150 fields of degree into 20 major categories, as listed in Appendix Table 8.

We 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, we identify fields from which society may substantially gain from having misallocated workers in salaried work, and we 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 3.

[Insert Figure 3 Here]

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. We assume a perfectly competitive labor market where workers are paid the value of their marginal product of labor. We 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 our 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:

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

Given this difference along with the number of immigrants who major in the different fields, we are able to rank order fields by the acuteness of the misallocation problem. Based on our 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 3.

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 3. In this figure, the Y-axis measures the rank ordering of the diversity of the misallocated ethnic pool depending on their field. Specifically, we 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 our 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 the following specifications:

$$Degree\,of\,Differential\,Sorting\,into\,SelfEmp_i = SelfEmp_i - Sel\widehat{fEmp}_i$$

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

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

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, we 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.

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 3 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 art 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.

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

We 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 Büttner (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. We conduct a productivity analysis to evaluate the potential social gains from correctly identifying self-employed immigrant workers.

The productivity analysis is based on the following specification, where worker wage is determined by the type of employment—salaried work or self-employment—and their human capital.

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

We 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 (2017), the unincorporated self-employed represent a non-entrepreneurial type of self-employed individuals; hence, we use this group as a lower bound for the productivity analysis.

[Insert Figure 4 Here]

Estimations based on the ACS for the period from 2005 to 2012 by education group are shown in the bar charts in Figure 4. 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 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, the estimation 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). Given these other forces, there are limitations to assessing the general equilibrium effects of having the self-employed immigrant in salaried work.

7.3 When do diversity in organizations generate higher returns?

We then consider which organizations would realize this productivity gain the most. Along with the complexity of the job explored in Section 7.2, another key parameter of the model driving the differential sorting pattern is the noisiness of the signals, or in other words, the employer's ability to discern the talent of their potential employees. While we do not have a precise measure for signal noise in a dyadic employer-employee pair, one way to proxy for this is to consider worker productivity conditional on the degree of workplace diversity. One implication of the model is that for a complex job, a minority worker that is hired for the job will likely outperform a majority worker hired for the same job, as the minority worker would have sent a stronger set of signals to surpass the hiring threshold.

Being a minority or majority in the workplace is defined within the workplace, however. Even for immigrants from the same country of origin, some will likely be a majority or minority in their workplace depending on their occupation and location. Therefore, we conduct an analysis categorizing immigrants by their coethnic representation in the workplace, based on metroarea-industry-occupation clusters, defined in section 5.3. We consider a worker present in a less ethnically represented metroarea-industry-occupation cluster to be in a more “diverse” organization.

To investigate when returns to diversity is the highest, we compare the relative productivity of workers with same observable traits while varying the levels of organizational diversity. To measure relative productivity, we collect the individual residuals from the following regression specification estimated based on all full-time salaried workers who are immigrants[4]:

$$HourlyWages_i = \beta_0 + \beta_1 X_i + \beta_2 X_c + \beta_3 Industry_i + \beta_4 Occupation_i + \beta_5 Metroarea_i + \epsilon$$

Here, X_i includes the standard individual controls including four race categories, education categories, time spent in the U.S.; X_c includes origin country level controls such as linguistic distance and log GDP per capita, and then we include fixed effects for industry, occupation, metroarea as well as age and year. The residuals of this specification allow us to measure the gap between the actual and the predicted wage of the individual based on other immigrants’ productivity who work in the same metroarea-industry-occupation cluster and share the same observable characteristics. In other words, it captures a worker’s hourly wage relative to comparable immigrants.

The residuals are visually summarized in a bar chart in Figure 5. In order to fix skill requirements for the job while varying the ability of the employer to discern capacities of their potential employees, we broadly divide the workers into high (top chart) and low educated (bottom chart). We further group immigrants into those that originate from linguistically close vs far countries.

[Insert Figure 5 Here]

Interestingly, Figure 5 suggests that there exists an asymmetry for when diversity is beneficial depending not only on the complexity of the job but also on the ethnic representation of the workplace. In particular, among high skilled workers, immigrants

[4] We conduct this analysis only within immigrants as most U.S.-born are majorities in their workplace and combining them would skew the analysis.

are more productive when their ethnic representation in the workplace is lower, while among low skilled workers, immigrants are more productive when their ethnic representation in the workplace is higher.

In comparing the bar charts for linguistically distant and close immigrants, the asymmetric pattern seems especially stronger for the linguistically distant group. While this is consistent with the implication of the theory—conditional on being a minority in one's workplace, the more linguistically distant will send a signal that is harder to decode and thus, those who are hired despite their linguistic distance will likely be more productive—the difference between the respective bar charts are not statistically significant. Hence, we present the difference between linguistically distant and close immigrants only as suggestive evidence.

The more important observation from the varying degrees of immigrant worker's productivity, however, is that when the job is difficult, workers in ethnically diverse groups are more productive, while when the job is easy, workers in homogeneous groups are more productive. This is because in the case where organization lacks diversity, employers will likely overlook capable employees with different linguistic-cultrual backgrounds (i.e. make false negative decisions) in high skilled jobs, while employers may be too lenient in assessing incapable employees with different linguistic-cultural backgrounds (i.e. make false positive decisions) in low skilled jobs.

While diversity in organizations are, in general, perceived to be desirable, our study highlights that whether diversity is productivity enhancing or not is context dependent: returns to diversity vary depending on the complexity of the job and the degree to which employers can discern abilities from interviews. Our findings suggest firms and policy makers should consider factors such as the difficulty of the job and the ethnic representation of the work place, rather than uniformly implementing a 'one size fit all' diversity policies across different types of organizations.

7.4 Limitations

In this section we discuss several limitations of this study. First, we do not causally identify the effect of linguistic-cultural differences and linguistic distance may be correlated with other unobservable ethnic level factors unaddressed in this study. Instead we argue the existence of a linguistic-cultural difference effect by presenting a constellation of empirical patterns that is consistent with the lost-in-translation theory; we argue that it is difficult for other models on discrimination or immigrant entrepreneurship to generate similar set of findings.

Second, while we posit that cultural mismatch can cause friction in matching workers to firms that is orthogonal to worker productivity, we 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.

Last, while the margin of adjustments that we 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, we 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. While there are also other employment options between high skilled salaried work and self-employment, we argue that self-employment being one of many options makes it less likely for us to find an effect.

8 Conclusion

In this study, we examine how informational frictions owing to cultural differences in labor markets differentially shape the sorting of workers into either self-employment or salaried work depending on human capital. To conduct this examination, we study the experiences of immigrants, who likely face especially large labor market frictions owing to linguistic-cultural barriers. We apply a theoretical framework that presumes that shared culture lubricates communication and hence mismatch in the linguistic-cultural backgrounds between an interviewer and a candidate in hiring setting cause immigrants to be less effective in conveying their ability. The framework predicts that immigrants are less likely to find an appropriate 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. We 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, we show that immigrants are more likely to sort into self-employment, particularly when they have noisier signals and higher education. We proxy for the degree of imprecise signalling with "linguistic distance," a measure based on how many branches separate two languages in a language tree, and we 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.

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

The immigrant talent pool composes almost 18% of the working age population. We 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 positive sorting into self-employment by immigrants with education reflects inefficient allocation of talent. 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. Furthermore, our study suggests that returns to diversity is context dependent; by implementing diversity policies taking factors such as the difficulty of the job and the ethnic representation of the workplace into account, firms and policy makers could better utilize talented immigrants who are abundant but hidden in the US.

References

Alesina, A., and P. Giuliano. 2009. "Family Ties and Political Participation." NBER Working Paper No. 154115. National Bureau of Economic Research, Cambridge, MA.

Altonji, J. G., and C. R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1): 313-350.

Arcidiacono, P., P. Bayer, and A. Hizmo. 2010. "Beyond Signaling and Human Capital: Education and the Revelation of Ability." *American Economic Journal: Applied Economics* 2 (4): 76-104.

Arrow, K. J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by O. Ashenfelter and A. Rees, 3-33. Princeton: Princeton University Press.

Åstebro, T., J. Chen, and P. Thompson. 2011. “Stars and Misfits: Self-employment and Labor Market Frictions.” *Management Science* 57 (11): 1999-2017.

Åstebro, T., H. Herz, R. Nanda, and R. A. Weber. 2014. “Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics”. *Journal of Economic Perspectives* 28 (3): 49-69.

Autor, D., F. Levy, and R. Murnane. 2003. “The Skill Content of Recent Technological Change: An Empirical Exploration.” *Quarterly Journal of Economics* 118 (4): 1279-1334.

Autor D, Scarborough D (2008) “Does job testing harm minority workers? Evidence from retail establishment.” *Quarterly Journal of Economics*. 123(1):219-277.

Battisti, M., G. Peri, and A. Romiti. 2016. “Dynamic Effects of Co-ethnic Networks on Immigrants’ Economic Success.” NBER Working Paper No. 22389, National Bureau of Economic Research, Cambridge, MA.

Becker, G. 1957. *The Economics of Discrimination.* 2nd ed. Chicago: University of Chicago Press.

Bell A, Chetty R, Jaravel X, Petkova N, Van Reenen J (2016) “The lifecycle of inventors.” Working paper.

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

Bleakley, H., and A. Chin. 2010. “Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants.” *American Economic Journal: Applied Economics* 2 (1): 165-192.

Borjas, G. J. 1986. “The Self-Employment Experience of Immigrants.” *Journal of Human Resources* 21 (4): 485-506.

Borjas, G. J. 1987. “Self-Selection and the Earnings of Immigrants.” *American Economic Review* 77 (4): 531-553.

Clark, K., and S. Drinkwater. 2000. “Pushed Out or Pulled In? Self-Employment Among Ethnic Minorities in England and Wales.” *Labour Economics* 7 (5): 603-628.

Cornell, B., and I. Welch. 1996. “Culture, Information, and Screening Discrimination.” *Journal of Political Economy* 104 (3): 542-571.

Dávila, A., and M. T. Mora. 2004. "English-Language Skills and the Earnings of Self-Employed Immigrants in the United States: A Note." *Industrial Relations* 43 (2): 386-391.

Fairlie, R.W., and M. Lofstrom. 2015. "Immigration and Entrepreneurship." In *Handbook on the Economics of International Migration, Vol. 1B*, edited by B. Chiswick and P. Miller, 877-912. Amsterdam: Elsevier.

Fairlie, R. W., and B. Meyer. 1996. "Ethnic and Racial Self-Employment Differences and Possible Explanations." *Journal of Human Resources* 31 (4): 757-793.

Fearon, J. D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8 (2): 195-222.

Fernández, R. 2011. "Does Culture Matter?" In *Handbook of Social Economics, Vol. 1A*, edited by J. Benhabib, M. O. Jackson, and A. Bisin, 481-510. Amsterdam: North-Holland.

Fernández, R. 2007. "Women, Work and Culture." *Journal of the European Economic Association* 5 (2-3): 305-332.

Fernández, R., and A. Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1 (1): 146-177.

Fernández, R., and A. Fogli. 2006. "Fertility: The Role of Culture and Family Experience." *Journal of the European Economic Association* 4 (2-3): 552-561.

Ferrer, A., D. A. Green, and W. C. Riddell. 2006. "The Effect of Literacy on Immigrant Earnings." *Journal of Human Resources* 41 (2): 380-410.

Fossen, F. and T. Buttner. (2013) "The returns to education for opportunity entrepreneurs, necessity entrepreneurs, and paid employees." Economics of Education Review. 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. and C. Rouse. (2000) "Orchestrating impartiality: The impact of "blind" auditions on female musicians." *American Economics Review* 90(4):715-741.

Hamilton, B. 2000. “Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment.” *Journal of Political Economy* 108 (3): 604-631.

Hegde, D., and J. Tumlinson. 2014. “Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity’s Role in US Venture Capital.” *Management Science* 60 (9): 2355-2380.

Hegde, D., and J. Tumlinson. 2015. “Unobserved Ability and Entrepreneurship.” Working Paper. April 20, 2015.

Hoffman M, Kahn L, Li D (2015) “Discretion in hiring”. *The Quarterly Journal of Economics* 133 (2): 765-800.

Hsieh C, Hurst E, Jones C, Klenow P (2016) “The allocation of talent and U.S. economic growth.” NBER Working Paper No. 18693, National Bureau of Economic Research, Cambridge.

Hurst, E., and B. W. Pugsley. 2015. *Wealth, Tastes, and Entrepreneurial Choice. Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*. Chicago: University of Chicago Press.

Imai, S., D. Stacey, and C. Warman. 2014. “From Engineer to Taxi Driver? Language Proficiency and the Occupational Skills of Immigrants.” Working Paper No. 1275. Queen’s Economics Department, Kingston, Ontario.

Joshi, A.M., Lahiri, N. (2014) Language friction and partner selection in cross-border R&D Alliance Formation. *Journal of International Business Studies* 46(2):123-152.

Kerr, S.P., Kerr, W. R., and W.F. Lincoln (2015) “Skilled immigration and the employment structures of US firms.” *Journal of Labor Economics* 33(S1):S147-S186.

Kerr, W., and Lincoln, W. (2010) “The supply side of innovation: H-1B Visa reforms and U.S. ethnic invention.” *Journal of Labor Economics* 28(3):473-508.

Kerr, W., and M. Mandroff. 2015. “Social Networks, Ethnicity, and Entrepreneurship.” NBER Working Paper No. 21597. National Bureau of Economic Research, Cambridge, MA.

Kulchina, E. (2016) “A path to value creation for foreign entrepreneurs.” *Strategic Management Journal*. 37(7):1240-1262.

Lang, K. 1986. “A Language Theory of Discrimination.” *Quarterly Journal of Economics* 101 (2): 363-382.

Lazear, E. 2005. “Entrepreneurship.” *Journal of Labor Economics* 23 (4): 649-680.

Levine, R., and Y. Rubinstein. 2017. “Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?” *Quarterly Journal of Economics* 132 (2): 963-1018.

Lewis, E. 2011. “Immigrant-Native Substitutability: The Role of Language Ability.” In *Immigration, Poverty, and Socioeconomic Inequality*, edited by D. Card and S. Raphael, 60-97. New York: Russell Sage Foundation.

Light, I. 1972. *Ethnic Enterprise in America*. Berkeley: University of California Press.

Light, I. 1979. “Disadvantaged Minorities in Self-Employment.” *International Journal of Comparative Sociology* 20 (1-2): 31-45.

Lundberg, S. J., and R. Startz. 1983. “Private Discrimination and Social Intervention in Competitive Labor Markets.” *American Economic Review* 73 (3): 340-47.

Luttmer, E. F. P., and M. Singhal. 2011. “Culture, Context, and the Taste for Redistribution.” *American Economic Journal: Economic Policy* 3 (1): 157-179.

Montalvo, J. G., and M. Reynal-Querol. 2005. “Ethnic Polarization, Potential Conflict, and Civil Wars.” *American Economic Review* 95 (3): 796-816.

Morgan, J., and F. Várdy. 2009. “Diversity in the Workplace.” *American Economic Review* 99 (1): 472-485.

Nanda, R., and J. B. Sørensen. 2010. “Workplace Peers and Entrepreneurship.” *Management Science* 56 (7): 1116-1126.

Nanda, R., and T. Khanna, (2010) “Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software industry.” Journal of Economics and Management Strategy 19(4):991-1012.

Oreopoulos P (2011) “Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes”. *American Economics Journal: Economic Policy* 3(4):148-171.

Peri, G., and C. Sparber. 2009. “Task Specialization, Immigration, and Wages.” *American Economic Journal: Applied Economics* 1 (3): 135-169.

Phelps, E. 1972. “The Statistical Theory of Racism and Sexism.” *American Economic Review* 62(4):659-661.

Phillips, K. W., Liljenquist, K. A., and M.A. Neale. (2009) "Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers." *Personality and Social Psychology Bulletin* 35(3):336-350.

Portes, A., and M. Zhou. 1996. "Self-Employment and the Earnings of Immigrants." *American Sociological Review* 61 (2): 219-230.

Ridlon, R., and J. Shin (2013) "Favoring the winner or loser in repeated contests". *Marketing Science* 32(5):768-785.

Rivera, L. A. 2012. "Hiring as Cultural Matching: The Case of Elite Professional Service Firms." *American Sociological Review* 77 (6): 999-1022.

Sauermann, H. and W. M. Cohen (2010) "What makes them tick? Employee motives and firm innovation." *Management Science*. 56(12):2134-2153.

Siegel J. I., Pyun, L. and B. Y. Cheon (2018) "Multinational firms, labor market discrimination, and the capture of competitive advantage by exploiting the social divide". *Administrative Science Quarterly 1-28*.

Slezkine, Y. 2004. *The Jewish Century*. Princeton: Princeton University Press.

Spolaore, E., and R. Wacziarg. 2009. "The Diffusion of Development." *Quarterly Journal of Economics* 124 (2): 469-529.

Stenard, B. S., and H. Sauermann. 2016. "Educational Mismatch, Work Outcomes, and Entry into Entrepreneurship." *Organization Science* 27 (4): 801-824.

Van den Steen, E. (2005) Organizational beliefs and managerial vision. *Journal of Law, Economics, and Organizations* 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 2.1 Selection into self-employment

Measure of noisy signal:	Self-employment (vs Salaried)			
	Immigrant Status 1gImm		Linguistic Distance LD	
	(1)	(2)	(3)	(4)
Panel A: American Community Survey (ACS), 2005 - 2012				
Noisy signal	0.024**	-0.035*	0.030*	-0.037
	0.012	0.02	0.016	0.024
Years of education	0.002***	0	0.002***	0
	0	0.001	0.001	0.001
Noisy signal x Yrs of education		0.004***		0.005***
		0.001		0.001
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Constant	-0.009	0.027	-0.016	0.021
	0.051	0.046	0.058	0.052
Number of Observations	5280414		5280414	
Base rate of self-employment	13%		13%	
Selection effect				
The noisiest signal relative to U.S.-born	18%		23%	
With an additional year of education	3%		4%	
Panel B: Current Population Survey (CPS), 1994 - 2012				
Noisy signal	0.039***	-0.026	0.051**	-0.029
	0.012	0.018	0.02	0.028
Years of education	0.001	-0.001*	0.001	-0.001*
	0.001	0	0.001	0
Noisy signal x Yrs of education		0.005***		0.006***
		0.001		0.001
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Constant	-0.159***	-0.124**	-0.163**	-0.127*
	0.06	0.06	0.07	0.068
Number of Observations	583189		583189	
Base rate of self-employment	13%		13%	
Selection effect				
The noisiest signal relative to U.S.-born	31%		40%	
With an additional year of education	4%		5%	

Source: Panel A uses the American Community Survey and Panel B uses the March Supplements of the Current Population Survey.
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. Reports results using two measures of noisy signal: columns (1) and (2) use immigrant status, columns (3) and (4) use linguistic distance.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and time spent in US for which U.S.-born are assigned their age.
Fixed effects include age, year, state, industry and occupation categories.
Selection effect divides the coefficient of noisy signals by the base rate of self-employment.
Reported Standard Errors are clustered at origin country level; *, **, 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 2.2 Differential selection into self-employment (and unincorporated self-employment) by education categories

	Self-employment (vs Salaried)		Uninc. Self-employment (vs Salaried)	
Measure of noisy signal:	**Immig. Status**	**Ling. Dist**	**Immig. Status**	**Ling. Dist**
	1gImm	**LD**	**1gImm**	**LD**
	(1)	(2)	(3)	(4)
Education (vs Grade School)				
High School	0.000	0.000	-0.006**	-0.006**
	0.003	0.003	0.002	0.002
College	0.004	0.004	-0.007	-0.007
	0.005	0.005	0.004	0.004
A. Noisy signal (1gImm / LD)	-0.009	-0.008	-0.020**	-0.020*
	0.014	0.017	0.009	0.011
B. (1gImm / LD) x High School	0.029***	0.032***	0.026***	0.029***
	0.007	0.008	0.005	0.005
C. (1gImm / LD) x College	0.044***	0.051***	0.039***	0.045***
	0.009	0.010	0.006	0.007
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	5280414	5280414	5005239	5005239
P-values comparing coefficients				
B = C	0.055	0.041	0.000	0.021

Source: American Community Survey, 2005 - 2012
Notes: Table reports linear estimates of the probability of a worker to be (unincorporated) self-employed over salaried work.
Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants.
Reports results for self-employment in columns (1) through (2) and for unincorporated self-employment in columns (3) to (4) using two measures o
noisy signal: columns (1) and (3) use immigrant status, columns (2) and (4) use linguistic distance.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
and time spent in US for which U.S.-born are assigned their age.
Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level; *, **, and *** indicate significance at 10%, 5% and 1%, respectively.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Table 2.3 Selection into self-employment by education x linguistic distance categories

	Self-employment (vs Salaried)			
Linguistic Distance category:	**<0.8**	**0.8 - 0.9**	**0.9 - 0.95**	**0.95 - 1**
	(1)	(2)	(3)	(4)
Education (vs Grade School)				
High School	0.004	0.003	0.001	0.004
	0.000	0.001	0.002	0.000
College	0.010	0.010	0.006	0.011
	0.000	0.001	0.005	0.000
A. Linguistic Distance	0.008	0.159	-0.027	0.096
	0.027	0.027	0.009	0.019
B. Linguistic Distance x High School	0.007	0.019	0.022	0.004
	0.009	0.007	0.009	0.007
C. Linguistic Distance x College	0.006	0.045	0.050	-0.015
	0.018	0.007	0.008	0.008
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	4641040	4545618	4858984	4655217
P-values comparing coefficients				
B = C	0.904	0.001	0.000	0.005

Source: American Community Survey, 2005 - 2012
Notes: Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and time spent in US for which U.S.-born are assigned their age.
Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Table 3 Selection into Self-Employment by Education Categories (Immigrants only)

	Self-employment (vs Salaried)			
	(1)	(2)	(3)	(4)
Education (vs Grade School)				
High School	0.013	0.018	0.012	0.017
	0.004	0.005	0.004	0.005
College	0.018	0.026		
	0.004	0.005		
Some college			0.001	0.001
			0.004	0.005
College degree and above			0.014	0.021
			0.004	0.005
Immigrate before 21		-0.033		-0.033
		0.006		0.006
Exposure to U.S. education				
A. In US x High School		-0.011		-0.011
		0.004		0.004
B. In US x College		-0.015		
		0.004		
C. In US x Some College				-0.018
				0.004
D. In US x College degree and above				-0.014
				0.005
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of observations	806845	806845	806845	806845
P-values comparing coefficients				
A = B		0.357		
C = D				0.450

Source: American Community Survey, 2005 - 2012
Notes: Tests whether exposure to the U.S. education system has a mitigating effect for entering into self-employment
Results ran only for working age, male immigrants in the sample; identified immigration age as well as exposure to U.S. education based on immigrants' reported year of entry.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Columns (3) and (4) reports results that further breakdown College into some college and college degree and above.
Reported Standard Errors are clustered at origin country level.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

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

	Self-employment (vs Salaried)			
	(1)	(2)	(3)	(4)
Linguistic Distance (LD)	0.069	0.078	0.069	0.079
	0.031	0.033	0.031	0.034
Age of immigration				
Before 10	-0.030	0.021	-0.037	0.014
	0.007	0.026	0.011	0.029
Between 10 to 15			-0.01	0.005
			0.007	0.022
Between 15 to 20			-0.007	-0.012
			0.003	0.023
LD x Age of immigration				
LD x Before 10		-0.057		-0.058
		0.028		0.029
LD x Between 10 to 15				-0.017
				0.022
LD x Between 15 to 20				0.006
				0.024
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	806845	806845	806845	806845

Source: American Community Survey, 2005 - 2012
Notes: Tests whether immigrating at a younger age has a mitigating effect for entering into self-employment
Results ran only for working age, male immigrants in the sample; identified immigration age based on immigrants' reported year of entry.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and years of education.
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Columns (3) and (4) reports results that further breakdown immigrants who come between 10 to 15 and between 15 to 20.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Table 5 Workplace representation and selection into uninc. self-employment (among immigrants)

	Self-employment (vs Salaried)		
	(1)	(2)	(3)
Linguistic Distance (LD)	0.073***	0.071***	0.087***
	0.024	0.024	0.029
"Workplace" representation		-0.037***	0.007
		0.008	0.018
"Workplace" rep. x Ling. Dist			-0.059**
			0.028
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	68689	68689	68689

Source: Current Population Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ by immigrants surrounded by their co-ethnics.
Results ran only for working age, male immigrants in the sample, who lived more than 10 years in the U.S.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and years of education.
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table 6 Language as input to productivity

Communication intensitiy of salaried jobs:	Self-employment (vs Salaried)			
	Low		High	
	(1)	(2)	(3)	(4)
Linguistic Distance (LD)	0.028	0.012	0.036	0.05
	0.015	0.018	0.009	0.039
Years of education	0.004	0.003	-0.005	-0.005
	0.001	0.001	0.001	0.001
LD x Years of education		0.001		-0.001
		0.001		0.002
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	3323900	3323900	2622300	2622300
Base rate of self-employment	21%		26%	
Selection effect	14%		14%	

Source: American Community Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ depending on communication intensity of the salaried job; sample includes working age, male, foreign-born and U.S.-born workers who worked full time full year in the survey year.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
Fixed effects include age, immigrant cohort, year, state, major industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Table 7 Assesing selection into self-employment from a subset of immigrants proficient in English

	Self-employment (vs Salaried)			
	All Immigrants		Proficient immigrants	
	(1)	(2)	(3)	(4)
Linguistic Distance (LingD)	0.030	-0.037	0.039	-0.006
	0.016	0.024	0.021	0.033
Years of education	0.002	0.000	0.001	0.000
	0.001	0.001	0.000	0.000
LingD x Yrs of education		0.005		0.003
		0.001		0.001
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	5280414		4929141	
Base rate of self-employment	13%		13%	
Selection effect				
The most LingD relative to U.S.-born	23%		30%	
With an additional year of education	4%		2%	

Source: American Community Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ ; sample includes working age, male, foreign-born and U.S.-born workers who worked full time full year in the survey year.
Columns (3) and (4) subset immigrants who report to speak English either Well or Very Well.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Table 8 Selection into Self-Employment

	Self-employment (vs Salaried)		
	(1)	(2)	(3)
Cultural Distance	0.106	0.149	0.055
	0.039	0.061	0.053
Linguistic Distance		-0.027	-0.020
		0.027	0.019
Cultural Dist x Yrs of educ.			0.006
			0.002
Other controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	5069458	5069458	5069458

Source: American Community Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ ; sample includes working age, male, foreign-born and U.S.-born workers
who worked full time full year in the survey year.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
years of education, time spent in US
Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Table 9 Sorting of immigrants from multilinguial countries (Belgium & Switzerland)

	Self-employment (vs Salaried)		
	(1)	(2)	(3)
Distance of Language spoken at home	0.172	0.056	0.184
	0.095	0.09	0.096
Controls	✓	✓	✓
Fixed effects			
Major occupation category		✓	
Occupation complexity category			✓
Number of Observations	996	983	983

Source: American Community Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ based on linguistic distance of language spoken at home;
sample includes working age, male, immigrants from Belgium or Switzerland, who worked full time full year in the survey year.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
and five education categorical variables
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Calcualtions weighted using the population weights provided.

Figure 1 Self-employment rate and linguistic / cultural distance by ethnic groups

Figure 1.1 Self-employment rate and linguistic distance

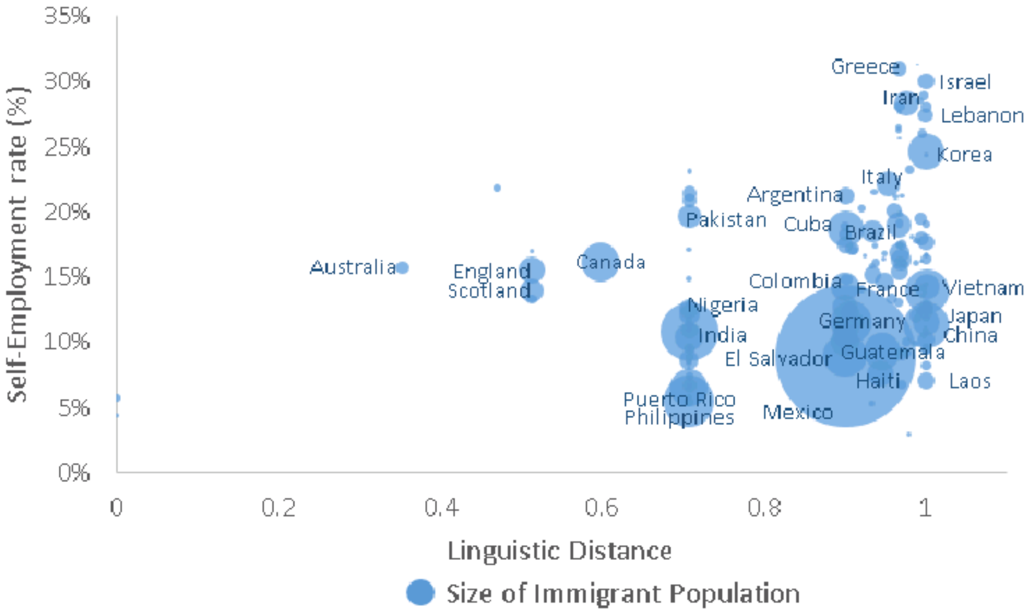
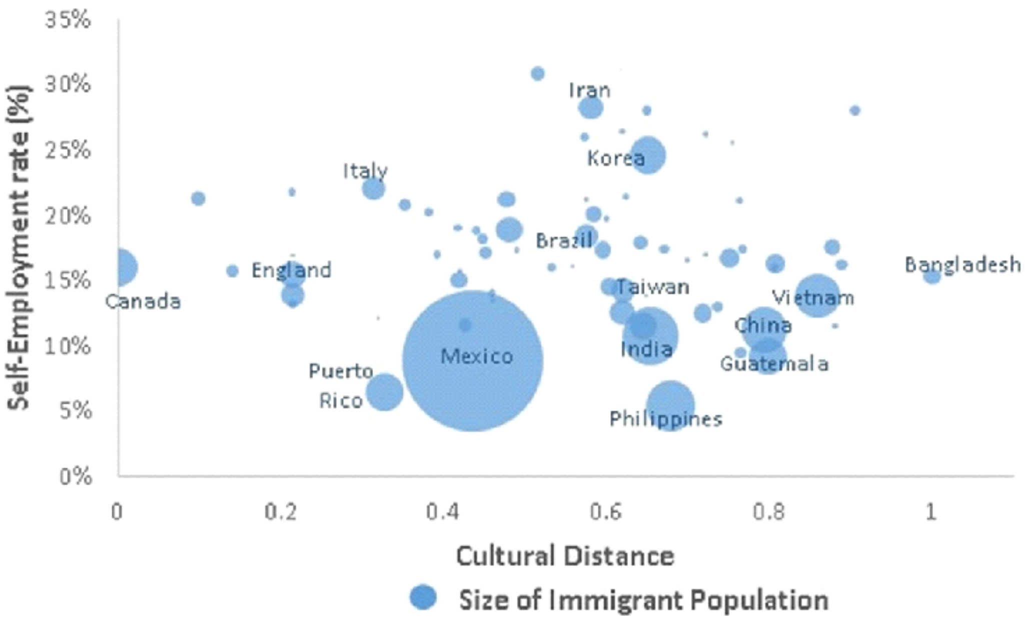
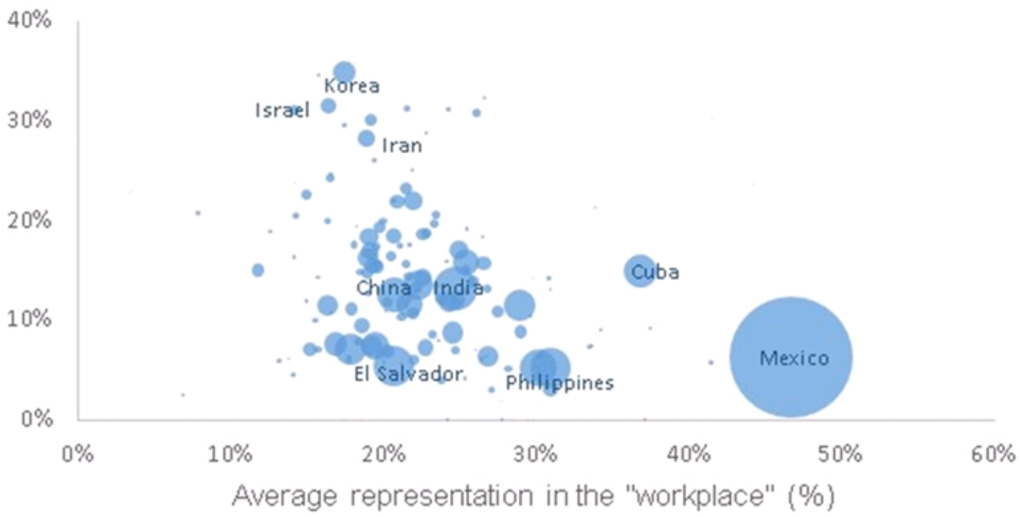


Figure 1.2 Self-employment rate and cultural distance



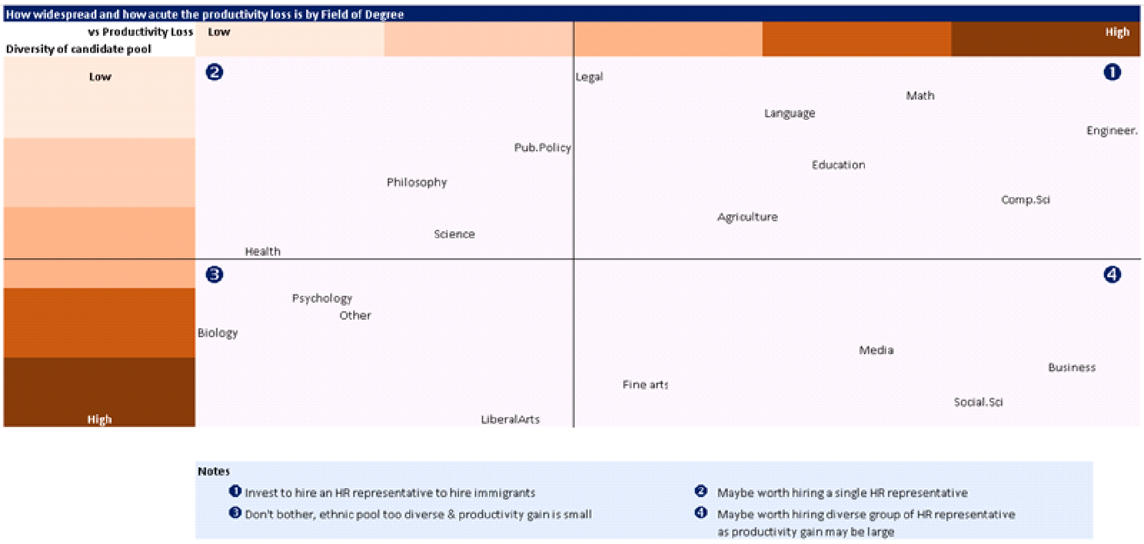
Source: American Community Survey, 2005 - 2012; Linguistic and Cultural distance measures based on Wacziarg and Spolaore (2009)
Notes: Standardized distance measures between 0 and 1; For linguistic distance, assigned mid-value for de jure English speaking countries, based on the Central Intelligence Agency's World Factbook.
Calcualtions weighted using the population weights provided.

Figure 2 Self-employment rate and average representation in the workplace by ethnic groups



Source: Current Population Survey, 1994 - 2012
Notes: Average representation in the workplace imputed using metro area - industry - occupation cluster of workers.
Size of the bubbles indicate relative size of the ethnic population.
Calcualtions weighted using the population weights provided.

Figure 3: 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:

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

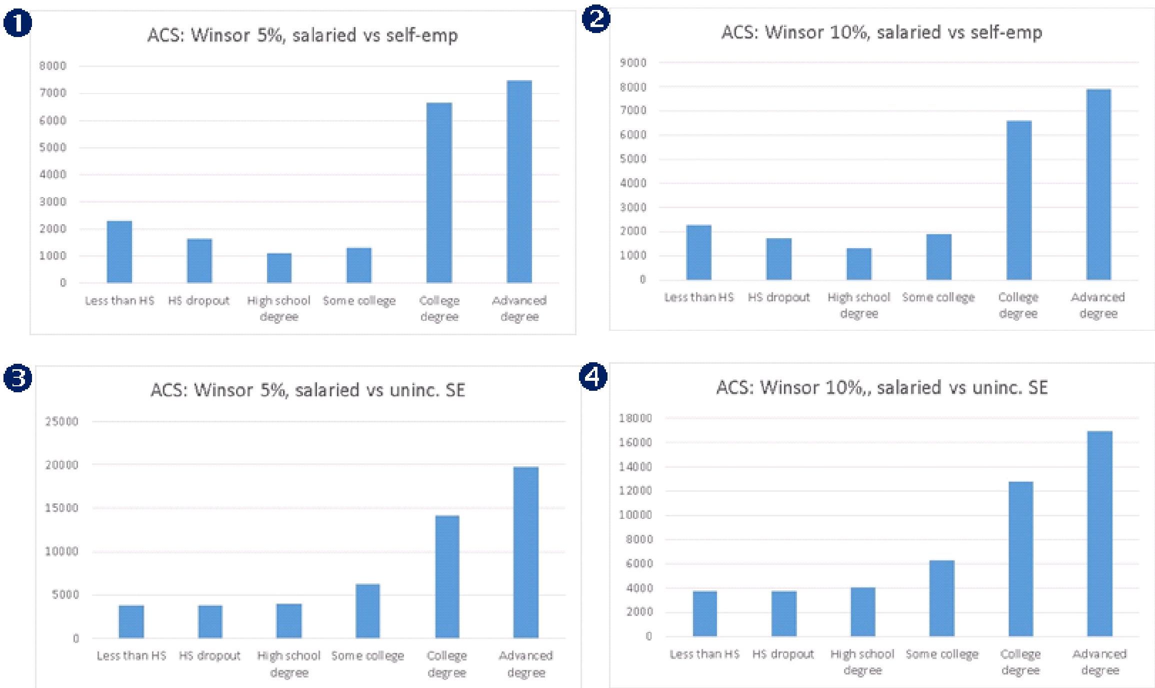
Y-axis measures the rank ordering of how diverse the misallocated ethnic candidate pool is depending on their field of degree.
Specifically, I count number of misallocated immigrants by 12 country of origin categories and then count the number of "peaks" in the distribution. Peaks defined as having more 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 4: Potential annual social gain per worker by education categories



Source: American Community Survey 2005 -2012

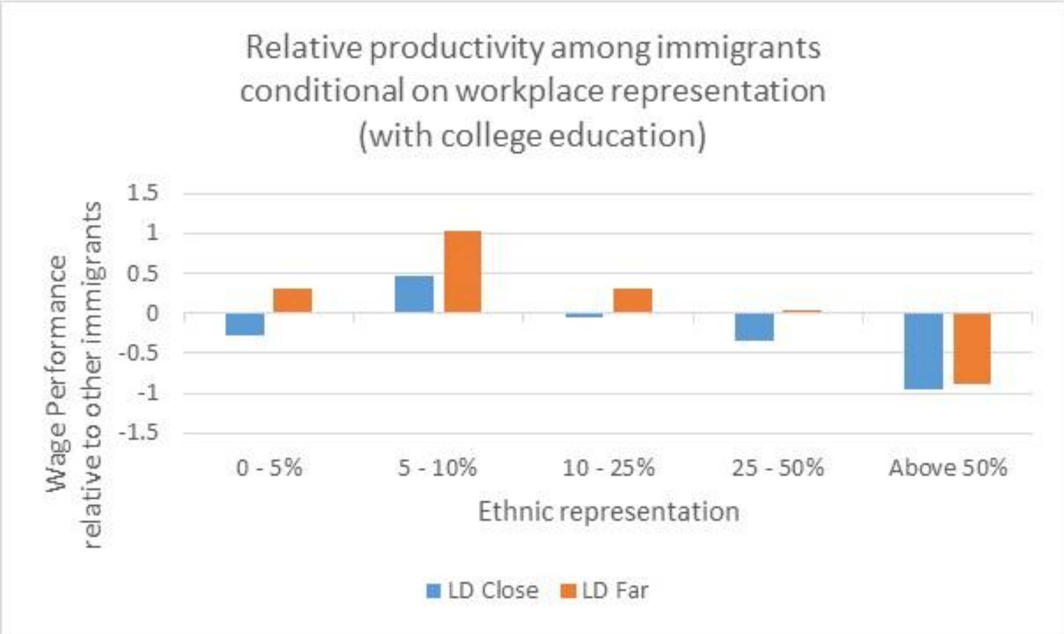
Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year Productivity loss evaluated the sum of coefficcients beta2 and beta 3 from the following equation:

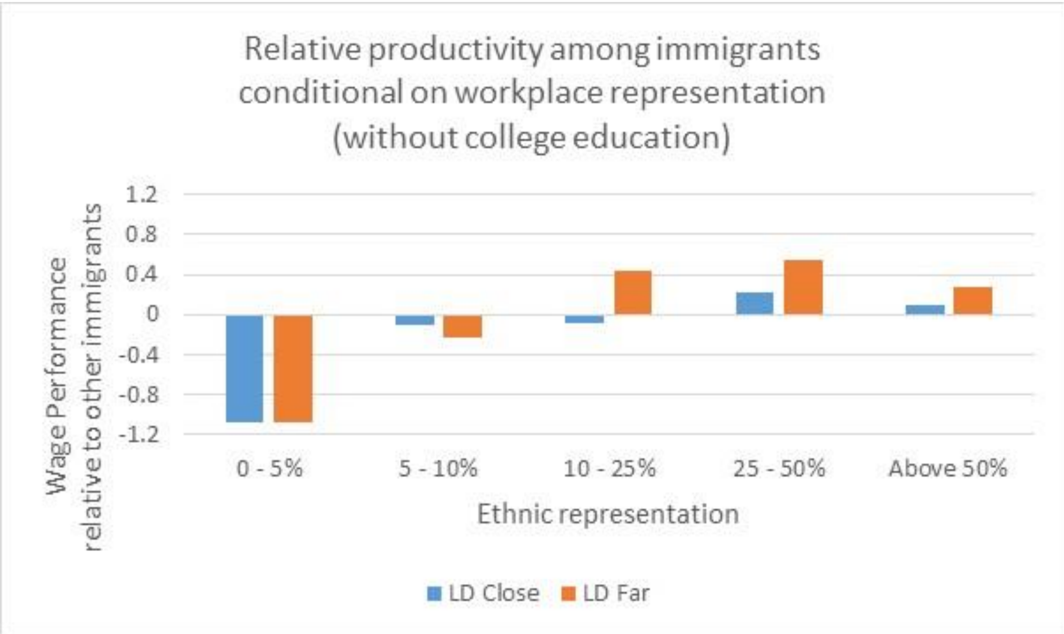
$$Wages_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 Salaried_i + \beta_3 Salaried \times EducCategories_i + \beta_4 X_i + \epsilon$$

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.

Figure 5 Relative productivity of immigrants by ethnic representation





Source: Current Population Survey, 1994 - 2012
Notes: Ethnic representation in the workplace imputed using % of coethnics in metro area - industry - occupation cluster of workers.
Bars represent average relative wage performance of workers in that ethnic representation group category.
Relative wage performance calculated as the actual - predicted hourly wage earnings.
Calculations weighted using the population weights provided.

Appendix

Appendix Table 1: Selection into self-employment by education categories (immigrant subgroups)

	Self-employment (vs Salaried)	
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**
	(1)	(2)
Panel A: Immigrants who came after age 25		
Education (vs Grade School)		
High School	0.002	0.002
	0.002	0.002
College	0.007**	0.007**
	0.003	0.003
A. 1gImm / LD	0.036	0.027
	0.030	0.032
B. (1gImm / LD) x High School	0.042***	0.047***
	0.007	0.009
C. (1gImm / LD) x College	0.062***	0.074***
	0.011	0.011
Controls	✓	✓
Fixed effects	✓	✓
Number of Observations	4776685	4776685
P-values comparing coefficients		
B = C	0.083	0.031
Panel B: Immigrants who spent more than 10 years in the U.S.		
Education (vs Grade School)		
High School	0.001	0.001
	0.002	0.002
College	0.006	0.006
	0.004	0.004
A. 1gImm / LD	-0.020	-0.018
	0.013	0.017
B. (1gImm / LD) x High School	0.034***	0.038***
	0.007	0.008
C. (1gImm / LD) x College	0.046***	0.052***
	0.009	0.010
Controls	✓	✓
Fixed effects	✓	✓
Number of Observations	5075506	5075506
P-values comparing coefficients		
B = C	0.171	0.15

Source: American Community Survey, 2005 - 2012

Notes: Replicate columns (1) and (2) of Table 2.2 only including subgroup of immigrants: those who immigrated after age 25 (Panel A) ; and immigrnats who spent more than 10 years in the U.S. (Panel B)

Reported Standard Errors are clustered at origin country level; *, **, and *** indicate significant at 10%, 5% and 1% respectively.

Reports p-values from t-tests testing equality between coefficients of the interaction terms.

Calcualtions weighted using the population weights provided.

Appendix Table 2 Differential selection into self-employment (using the Current Population Survey)

	Self-employment (vs Salaried)	
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**
	(1)	(2)
Education (vs Grade School)		
High School	0.013	0.000
	0.013	0.000
College	0.005**	0.005*
	0.002	0.003
A. Noisy signal (1gImm / LD)	0.006	0.017
	0.004	0.020
B. (1gImm / LD) x High School	0.027***	0.035***
	0.006	0.009
C. (1gImm / LD) x College	0.033***	0.044***
	0.007	0.009
Controls	✓	✓
Fixed effects	✓	✓
Number of Observations	583189	583189
P-values comparing coefficients		
B = C	0.362	0.362

Source: Current Population Survey, 1994 - 2012
Notes: Table replicates Table 2.2 using the CPS rather than the ACS.
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.
Reports results using three measures of noisy signal. Column (1) uses immigrant status and column (2) uses linguistic distance, as measure of noisy signal, respectively; Reported Standard Errors are clustered at origin country level;
*, **, and *** indicate significant at 10%, 5% and 1% respectively.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Appendix Table 3 Differential selection into self-employment (cluster standard errors at educ x country categories)

	Self-employment (vs Salaried)	
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**
	(1)	(2)
Education (vs Grade School)		
High School	0.000	0.000
	0.003	0.003
College	0.004	0.004
	0.007	0.007
A. Noisy signal (1gImm / LD)	-0.009	-0.008
	0.011	0.013
B. (1gImm / LD) x High School	0.029**	0.032**
	0.014	0.016
C. (1gImm / LD) x College	0.044***	0.051***
	0.010	0.011
Controls	✓	✓
Fixed effects	✓	✓
Number of Observations	5280414	5280414
P-values comparing coefficients		
B = C	0.206	0.189

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, either U.S.-born or first generation immigrants.
Reports results using three measures of noisy signal. Column (1) uses immigrant status and column (2) uses linguistic distance, as measures of noisy signal, respectively. Reported Standard Errors are clustered at three education categories by origin country; *, **, and *** indicate significant at 10%, 5% and 1% respectively.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Appendix Table 4 Differential selection into (unincorporated) self-employment using cultural distance

	Self-employment (vs Salaried)	Uninc. Self-employment (vs Salaried)
	(1)	(2)
Education (vs Grade School)		
High School	0.002	-0.004***
	0.002	0.001
College	0.008**	-0.004
	0.003	0.002
A. Noisy signal (1gImm / LD)	0.047	0.019
	0.042	0.024
B. (1gImm / LD) x High School	0.048***	0.042***
	0.014	0.010
C. (1gImm / LD) x College	0.070***	0.062***
	0.022	0.017
Controls	✓	✓
Fixed effects	✓	✓
Number of Observations	5069458	4806102
P-values comparing coefficients		
B = C	0.185	0.109

Source: American Community Survey, 2005 - 2012
Notes: Table replicates Table 2.2 using cultural distance.
Sample includes males between 18 and 65, who worked full-time full-year, either U.S.-born or first generation immigrants.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and time spent in US for which U.S.-born are assigned their age.
Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level; *, **, and *** indicate significance at 10%, 5% and 1%, respectively.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calculations weighted using the population weights provided.

Appendix Table 5 Selection into self-employment by education x noisy signal categories (CPS)

	Self-employment (vs Salaried)			
Linguistic Distance category:	**<0.8**	**0.8 - 0.9**	**0.9 - 0.95**	**0.95 - 1**
	(1)	(2)	(3)	(4)
Education (vs Grade School)				
High School	0.007	0.006	0.008	0.008
	0.001	0.002	0.000	0.000
College	0.010	0.008	0.011	0.011
	0.002	0.003	0.000	0.000
A. Linguistic Distance	-0.047	0.005	0.139	0.133
	0.023	0.008	0.040	0.027
B. Linguistic Distance x High School	0.035	0.022	-0.013	0.026
	0.014	0.008	0.027	0.022
C. Linguistic Distance x College	0.056	0.044	-0.049	-0.001
	0.015	0.005	0.044	0.017
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	516244	539689	493548	501164
P-values comparing coefficients				
B = C	0.025	0.003	0.224	0.012

Source: Current Population Survey, 1994 - 2012
Notes: Replicates Table 2.3 using the CPS.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and time spent in US for which U.S.-born are assigned their age.
Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Appendix Table 6 Age of immigration and selection into self-employment (Among immigrants)

	Self-employment (vs Salaried)			
Immigrate before:	**Age 7**	**Age 8**	**Age 9**	**Age 11**
	(1)	(2)	(3)	(4)
Linguistic Distance (LD)	0.079	0.079	0.080	0.081
	0.035	0.035	0.035	0.036
Immigrate at a young age	0.024	0.018	0.021	0.021
	0.03	0.029	0.029	0.028
LD x Immigrate ate a young age	-0.064	-0.056	-0.057	-0.053
	0.036	0.034	0.034	0.033
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Constant	0.003	0.005	0.004	0.001
	0.073	0.073	0.074	0.074
Number of Observations	806845	806845	806845	806845

Source: American Community Survey, 2005 - 2012
Notes: Replicates column (2) of Table 4, using different indicators for coming at a young age.
Results ran only for working age, male immigrants in the sample; identified immigration age based on immigrants' reported year of entry.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and years of education.
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Appendix Table 7 Selection into self-employment

	Self-employment (vs Salaried)			
	All Immigrants & U.S. born		Minority immigrants & U.S.-born	
	(1)	(2)	(3)	(4)
Linguistic Distance (LingD)	0.051	-0.029	0.057	-0.013
	0.020	0.028	0.019	0.028
Years of education	0.001	-0.001	0.001	-0.001
	0.001	0.000	0.001	0.000
LingD x Yrs of education		0.006		0.005
		0.001		0.001
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	583189	583189	559069	559069

Source: Current Population Survey, 2005 - 2012
Notes: Tests linear propensities to self-employ; Columns (3) and (4) limit immigrants to minority immigrants.
Minority immigrants represent those not part of the most represented ethnic group in their metroarea - industry - occupation cluster.
Results ran only for working age, male immigrants in the sample.
Controls include log GDP per capita of origin country; four race categories including (Non-hispanic) White, Black, Hispanic and Asian;
Fixed effects include age, immigrant cohort, year, state, industry and occupation categories.
Reported Standard Errors are clustered at origin country level.
Calcualtions weighted using the population weights provided.

Appendix Table 8 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
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 TECH
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 ADMINI
- TREATMENT THERAPY PROFESSIONS
- COMMUNITY AND PUBLIC HEALTH
- MISCELLANEOUS HEALTH MEDICAL PROFESSIONS

Language
- LINGUISTICS AND COMPARATIVE LANGUAGE AND LITERATURE
- FRENCH GERMAN LATIN
- 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

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
- TRANSPORTATION SCIENCES AND TECHNOLOGIES