Title
Immigrant Assimilation and Labor Market Outcomes

Author
Kim, Yoonha

Publication Date
2017

Peer reviewed|Thesis/dissertation

Immigrant Assimilation and Labor Market Outcomes

By

Yoonha Kim

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ross Levine, Co-chair
Professor John Morgan, Co-chair
Professor Noam Yuchtman
Professor David Card

Spring 2017

Abstract

Immigrant Assimilation and Labor Market Outcomes

by

Yoonha Kim

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Ross Levine, Co-chair

Professor John Morgan, Co-chair

In this dissertation, I examine immigrants' integration into host societies by examining systematic patterns in the U.S. labor market. In the first two chapters, I study how differences in language and culture explain the formation of small business by highly educated immigrants and what implications that have for firms trying to identify the most productive workers in an increasingly global labor force; in the last chapter, I study another aspect of social bias by exploring how the September 11 terrorist attacks shaped labor market outcomes for subgroups of the immigrant population.

In Chapter 1, I show that linguistic-cultural backgrounds of immigrants can lead to a systematic misallocation of scarce talent: highly educated, foreign-born workers more likely sort out of salaried work, and into self-employment, than otherwise similar U.S.-born individuals. This differential sorting can be theoretically understood as a rational, but flawed, response to the difficulties of credibly signaling capabilities—the cultural distance between the employer and the candidate generates noisy signals, when precise signaling is more critical for applicants to more demanding jobs. Using surveys representative of the U.S. population and measuring cultural mismatch with "linguistic distance", I find evidence consistent with this theory: not only the highly educated—who apply to more demanding jobs—but also the linguistically distant—who send noisier signals of ability—disproportionately sort into self-employment; immigrants who have culturally assimilated or who are surrounded by co-ethnics are less likely to exhibit such pattern. Furthermore, I show that immigrants' English language deficit, among other potential drivers, does not, in and of itself, explain the differential sorting. This suggests that the systematic sorting pattern appears to reflect inefficient allocation of talent.

In Chapter 2, I further discuss the implications of findings from Chapter 1. My empirical analyses suggest an important role for cultural frictions in the labor market and therefore, inability for firms to correctly identify the productive workers. Awareness of the phenomenon can, in principle, allow firms to better harness the untapped talent pool of highly educated immigrants sorting into self-employment—the hidden gems. I provide suggestions for how

firms should adjust their hiring practices as well as estimate for how much firms may be able to benefit by solving the hidden gems problem.

In Chapter 3, I examine how discrimination arising from preferences manifests itself in labor market outcomes across different minority groups over time. I exploit an exogenous shock in taste-based bias towards a subset of the immigrant group arising from the September 11 terrorist attacks: individuals who may appear to be Middle Eastern. By using a difference-in-differences approach based on a worker panel data, I provide suggestive evidence confirming previous studies, finding a negative effect of 9/11 on immigrants' labor market outcomes. My results further suggest that there may potentially be heterogeneous effects associated with educational attainment, where the less well educated are increasingly worse off relative to the more educated. I partly ascribe this to the heterogeneous effect of occupational discrimination, where the less well educated increasingly sort into less complex jobs over time. I discuss alternative channels that may drive my results.

While I study labor market imperfections in the context of immigrant workers in the U.S., the findings of my studies can be applied more broadly to better understand how social biases affect economic outcomes of minority groups. Given that there are particularly pronounced social divides between an immigrant and a non-immigrant, studying the experiences of immigrants provide a unique vantage point to examine the effect of labor market discrimination. I hope to contribute to expanding our understanding for how social factors such as language, culture and preferences importantly drive matching of workers to firms in the labor market, generating systematic and persistent patterns of occupational segregation.

For My Loving Family

Acknowledgements

I am indebted to my PhD advisors, Ross Levine, John Morgan and Noam Yuchtman, for their patience, guidance and support in achieving my doctorate milestone. Ross, you were the first person to teach me how to love and conduct research. You asked "Are you happy?" to help me find my dissertation topic; then you kept me focused by asking, "What exactly is your question?". Thank you, I will remember that simplicity is key. John, of all the things you taught me, you showed me that positive willpower is the most powerful characteristic a researcher, or any person, could possess. I will continue to "turn things on their heads"—as you like to say—in search of blue ocean opportunities both in research and in life. Thank you. Noam, I admire your encyclopedic knowledge of such a wide range of topics inside and outside of Economics. Your enthusiasm for research and quest for knowledge is contagious. You taught me how to engage in research out of genuine curiosity. Thank you.

Many other great scholars generously helped me along the way. I would like to especially thank David Card, Ernesto Dal Bó, Minjung Park, Yona Rubinstein, and faculty members of the Business and Public Policy program whose suggestions, mentorship and support I benefitted greatly from. I am also deeply thankful to Romain Wacziarg for generously sharing his distance data which shaped my dissertation to a great degree; to Jesse Rothstein and the IRLE staff, especially Robert Hiramoto and Terry Huwe, for their help in gaining access to the geocoded NLSY79 data; and to Maria Carkovic and the Sasakawa Foundation for generously funding my PhD years. I also thank my BPP peers—both current students and alums; I especially thank Jeffrey Kuhn and Santiago Truffa, without whose help and camaraderie I would not have been able to make it this far.

I am an exceptionally lucky person because of my family. I especially thank Kyung-Ryul—my life companion—who encouraged me to pursue a doctoral degree. Despite having been physically apart for many months, you were with me through all the emotions of my PhD life. The future I envision with you is always brighter; you are the best decision I have ever made. I am grateful to my mother-in-law and father-in-law for their encouragement throughout my PhD journey. I aspire to follow in your footsteps as an academic. I thank Kyung-Il and Sunah—my brother and my sister-in-law—for their support; you, Saejoon, and Jaejoon are close to my heart. Last but not least, I would like to thank my mom and dad—my best friends—for providing me with the best environment to pursue my dreams and for believing in me, no matter what. Your love and support is much greater and deeper than what one could imagine—you were the first teachers that shaped every little one of my habits that enable me to navigate the past, current and future chapters of my life. The love and respect I have for you only continues to grow as I mature. Thank you for raising me to be the person I am.

As part of my dissertation I study how foreigners have a harder time making themselves understood even after they become proficient with the language. I cannot express in any words in any language how deeply grateful I am to all the people that helped guide me along this journey. Thank you all very much.

Contents

List of Figures

List of Tables

Chapter 1

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

1.1 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, I argue that linguistic-cultural differences and the greater informational asymmetries that they create for high-end jobs account for the systematic sorting pattern involving highly educated immigrants. I show this in three steps: (1) I examine through a formal model how cultural differences might create frictions in matching workers to firms; (2) I test predictions that emerge from this framework empirically accounting for cultural disconnect using "linguistic distance"; (3) I verify that other factors such as linguistic deficit, which affects immigrants' productivity, does not explain the different sorting pattern. Based on my findings, I discuss the broader implications of cultural differences on the integration of foreign workers and efficient allocation of human capital in the labor market.

First, I describe a theoretical framework that predicts how linguistic-cultural differences generate differential sorting of immigrants into business ownership, differently for the high and low educated. While culture is an elusive concept, I make this problem tractable in a framework that codifies how cultural differences create frictions in a hiring setting. Predictions of business formation patterns arising from this framework enable me to isolate the impact of the greater informational asymmetries between immigrants and employers created by cultural differences on the sorting of immigrants into self-employment over salaried employment, depending on workers' human capital.

Suppose that in a hiring interview, employers receive verbal and nonverbal signals of ability from candidates that shape their beliefs about a candidate's ability to perform the job. Communication is more effective when two people share the same linguistic-cultural background; hence, immigrant candidates, who lack shared culture, are more likely to send noisier signals than their U.S.-born counterparts. This implies that the employer is less sensitive to an immigrant candidate's signals in updating her prior beliefs; as a result, the employer may make rational, but flawed, judgements, where she fails to hire sufficiently capable immigrants. Highly educated immigrants are especially subject to this error, as they apply to jobs in which the talent to perform the job is scarce, and thus it is harder to convince the employer.

Among theoretical models of discrimination, Morgan and Várdy (2009) build on this economic intuition, to solve the employer's dynamic optimization problem in a repeated interview setting. They show how employment outcomes between two groups may differ solely based on differences in the variance of signals of abilities, even when the average beliefs about the abilities are the same across populations. The difference in employment outcomes between the immigrants and U.S.-born candidates grows with employers' uncertainty, which is determined by a) the difficulty of the job and b) the relative noisiness of candidates' signals. These two factors give rise to informational frictions that drive heterogeneous selection into self-employment. Assuming that highly educated workers have higher propensities to apply to difficult jobs and that immigrants have noisier signals, highly educated immigrants are disproportionately more likely to fail to find an appropriate match and to run their own businesses.

Second, I test predictions of this framework by using individual-level surveys representative of the U.S. population—the American Community Survey (ACS; 2005 – 2012) and the Current Population Survey (CPS; 1994 – 2012) from the U.S. Integrated Public Use Microdata Series (Flood et al. 2015; Ruggles et al. 2015)—and by exploiting different levels of educational attainment and the diverse set of linguistic-cultural backgrounds that immigrants bring from their country of origin. Specifically, I account for noisiness of candidates' signals using the "linguistic distance" measure from the development economics literature (Wacziarg and Spolaore 2009), built on Fearon's (2003) approach of tracing the number of branches that separate two languages in a language tree.

While the sociological literature has long provided insights for how cultural differences hinder immigrants' assimilation in mainstream economic markets, it has been difficult to empirically identify the effect of linguistic-cultural differences. Studies have shown how immigrants' lack of communication skills affect their labor market outcomes (Ferrer et al. 2006, Peri and Sparber 2009, Imai et al. 2014), but they do not necessarily address immigrants' cultural barriers as cultural differences persist even among people who speak the same language and therefore, even when an immigrant is proficient in English. The linguistic distance measure helps overcome this challenge by enabling me to explore the frictions that linguistic-cultural differences cause beyond communication barriers that immigrants face in the typical sense. I also validate my examination with a measure of cultural distance, also constructed by Wacziarg and Spolaore (2009) but based on a completely different method using the World Value Survey, to show that linguistic distance indeed captures an immigrant's familiarity with the culture beyond mere language proficiency.

The theoretical prediction that informational friction increases with the difficulty of the job and the noisiness of the signal is supported by my data: the likelihood of immigrant self-employment systematically increases not only with higher education levels but also with greater linguistic distance. I examine linear probabilities of a worker owning a business as opposed to being a salaried worker where the main explanatory variables of interest are educational attainment and linguistic distance, as well as the interactions between the two sets of variables. My empirical results suggest that linguistically distant immigrants are, on average, 23-40% more likely to enter into self-employment than similarly qualified U.S.-born workers and that this effect is larger for the highly educated: with an additional year of education, the likelihood to self-employ increases by 3-5%. Moreover, my results qualitatively hold when I use the cultural distance measure in lieu of linguistic distance, suggesting that linguistic distance contains the degree of cultural acquaintance.

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

Third, I evaluate the predictions of the framework relative to the predictions from alternative hypotheses. Among other potential drivers, I particularly investigate whether linguistic distance 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.

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

Second, I complement the linguistic distance measure with individuals' self-reported English scores. The border between a lack of communication skills and imprecise signaling owing to cultural differences is often indistinct. However, I show that my results hold on a subsample of immigrants who report speaking English well. The fact that the theoretical predictions hold even when I 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, I show that the differential sorting between immigrants and non-immigrants in the labor market does not reflect immigrants' inability to communicate. Rather, I argue that linguistic-cultural mismatch importantly accounts for this differential sorting, representing a systematic bias: firms make false negative judgements in evaluating talented immigrants who could perform well. Hence, firms can gain competitive advantage by adjusting for such talent misallocation.

1.2 Literature review

This study relates to several strands of the literature. First, this paper contributes to the literature on the economic effects of language (Wacziarg and Spolaore 2009, Bleakley and Chin 2010). In particular, I add to this literature by investigating how imprecise signals in immigrant's job search affect their employment outcomes. More specifically, I examine how language causes friction in the discovery of immigrant talent rather than in immigrants' ability to perform a job. Prior work on immigrants' language and occupational choice mainly considers language to be a productivity input (Chiswick and Miller 1995; Davila and Mora 2004). For example, Imai et al. (2014) show an incomplete transfer of foreign skills from the source to host country in jobs that rely more heavily on communication skills, Lewis (2011) shows how language skills drive immigrant and non-immigrant substitutability, and Peri and Sparber (2009) show how immigrants sort into manual tasks, while non-immigrant workers shift to more language-intensive jobs. In contrast to these studies, I exploit variation in job and language characteristics to investigate how immigrants face barriers in the job search due to cultural differences. While English ability in and of itself is an important consideration and explains a large part of the earnings differential (Ferrer et al. 2006), it is important to explore how language affects immigrants' labor market outcomes beyond the traditional channel of their ability to communicate, given that about 70% of immigrants today consider themselves to speak English well. In this evaluation, I argue that the similarity of an immigrant's first language to English matters more than her proficiency.

Second, this paper relates to the literature on manager biases in the hiring process and, more broadly, the allocation of human resources. Petersen and Saporta (2004) note that 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. I 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); I show how such biases affect matching of workers to firms in the labor market. Based on my findings, I 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).

Finally, this study speaks to the entrepreneurship literature. As a byproduct of cultural friction, I 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; Kacperczyk 2013) or individual traits (Lazear 2005; Levine and Rubinstein 2016). 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). I build on these discussions by investigating how cultural friction, which causes false negative judgements in firms' hiring process, plays a role in sorting highly educated foreign workers into business ownership.

1.3 Theoretical framework

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

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

There are different theoretical models 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.

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 preciseness of signals and hence the variance of the distribution of talent. Thus, immigrants who have completely assimilated culturally—whose signals are just as precise as that of U.S.-born candidates—should not face this problem. This motivates the following prediction:

Prediction 4: Immigrants who send precise signals should sort into self-employment less than otherwise similar immigrants.

I 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, and hence where the employer is more likely to be from the same ethnic group, they would not suffer from this informational friction. This leads to the following testable prediction:

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

While the theoretical framework is specific to an interview setting, the implications of the model are not confined to the hiring process. First, promotion decisions could also be viewed as an organic hiring decision. I argue that a manager-level job requires a different set of skills than an entry-level job; hence, promoting a worker can be viewed as hiring her for a new role. Second, the model also has implications for employee retention. Firing decisions affect minority groups in a similar manner to hiring decisions. Hence, the long-run workforce composition that we observe in the labor market would be a more skewed version of the composition that is initially suggested by the model, which is that minorities are systematically underrepresented in jobs in which talent to perform the job is relatively scarce.

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.

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

1.4 Data description

To test the theoretical predictions outlined above, I use two distinct data sets to examine (a) differential selection into self-employment by U.S.-born and foreign born workers, (b) systematic patterns of selection into self-employment across subgroups of immigrants and (c) potential alternative explanations that may be driving this pattern. I use the ACS for the years 2005 to 2012 along with the March Supplements of the US CPS for the years 1994 to 2012. Both surveys provide baseline characteristics and occupational and productivity information on individuals. While the ACS is used to present the main results, I use the CPS to further check the robustness of the results and to conduct analyses that require metro area-level divisions. The main empirical findings hold across both datasets.

Table 1.1 provides basic demographics and labor market outcomes for the sample, where Panel A summarizes the ACS data and Panel B summarizes the CPS data. For both surveys, I include male workers aged 18 – 65 years who worked full-time in the entire year for their work year. Calculations for both samples are weighted using the population weights provided by the respective surveys. I identify first-generation immigrants as those who and whose parents were born outside the US for the CPS and those who are indicated as foreign-born for the ACS. The indicator for self-employment versus salaried employment is the main dependent variable of interest and both surveys classify all workers as either salaried or self-employed.

Details on demographics are as follows: Whites are individuals with the race code "White alone" excluding individuals identified as Hispanics. Blacks, Hispanics and Asian are those who answered yes to "Black or African American", "Spanish/Hispanic/Latino origin", and "Asian", respectively. For educational attainment, I use actual grade levels or degrees attained as well as years of education. I categorize education into 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, I take the average year of education.

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

To test the specific theoretical predictions, I need proxies for the difficulty of the job and the noisiness of the signal. I proxy for the difficulty of the job with the average education level of workers employed within jobs. Therefore, the higher the worker's educational attainment, the more likely her job will demand difficult tasks, where the employer believes that the talent to perform the job is scarce. Measuring the noisiness of signals poses a greater challenge. To overcome this challenge, I run my results using three different measures of noisy signaling: (1) an indicator for immigrant status; (2) a continuous measure of linguistic distance; and (3) a 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.

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. I 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 US 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 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, I assign a mid-value. I identify de jure English-speaking countries based on the Central Intelligence Agency's

World Factbook. Linking these data to the ACS and CPS data provides the linguistic distance for immigrants from over 150 countries.

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. While I could use cultural distance as the main explanatory variable throughout the study, it has less variation as the measure covers only 74 countries. 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.

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

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

I also consider institutional factors that shape immigrants' employment choices. In particular, I account for H-1B visa holders, whose career trajectory would likely differ from others because their immigration status ties them to a specific employer, and exclude them from my 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, I identify immigrant subgroups that would compose ~70% of the H-1B holders based on USCIS' FY2012 Annual Report to Congress. Specifically, I exclude Indian, Chinese and Canadian immigrants with a college degree working in universities or in computer- or engineering-related occupations.

Finally, given that immigrants are not proportionately distributed across space, I construct two additional measures. First, to determine the different dynamics in ethnic enclaves, I create a proximate indicator for whether an immigrant resides in an enclave. Specifically, for each ethnic group, I rank metro areas by the size of the ethnic group population and identify the metro areas that are above the 95th percentile and 99th percentile of the distribution. This measure captures slightly over half and one third of the immigrants residing in the US. I assign 1 if an immigrant resides in these metro areas and 0 otherwise. Second, I construct an indicator denoting whether an immigrant composes the majority of their organization. I proxy for an organization by using an occupation category within an industry in a metro area and assess the proportion each ethnic group represented in each metro area-industry-occupation cluster.

1.5 Empirical methodology

In this section, I discuss the empirical methodology employed to test the main predictions of the framework. I use linear probability models with an indicator for self-employment as the dependent variable and individual- or origin country-level characteristics as explanatory variables. I use the following specification to test 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 I use three different measures: (1) an indicator for first-generation immigrant status; (2) a continuous measure of linguistic distance; and (3) a continuous measure of cultural 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, I 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 $\beta 1$ 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 $\beta 2$ and $\beta 3$ 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, I examine significant differences in the sorting effect with higher education by categorizing $\beta 2$ and $\beta 3$ into three education categories and conducting t-test between coefficients. I further categorize $\beta 1$ and $\beta 3$ 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.

To test for subsequent hypotheses concerning whether the selection effect is mitigated for immigrants who send more precise signals I use the following specification:

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

This specification is used to test subsequent hypotheses within immigrants, where I 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. I 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 whether the immigrant is part of a majority. I assign 1 if an immigrant is considered a majority and 0 otherwise.

The main coefficient of interest here is $\beta 2$ and $\beta 3$. The $\beta 2$ 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. $\beta 3$ will help further us to assess whether there is heterogeneity across immigrant groups with different degrees of noisy signals.

1.6 Key empirical findings

In this section, I test the predictions delineated in section 3 by using the empirical setting and methods discussed in sections 4 and 5. First, I test the predicted differential sorting pattern across educational attainment and noisy signal. Second, I examine sorting patterns of immigrants who have culturally assimilated. Third, I investigate immigrants' propensities to self-employ when they are surrounded by their co-ethnics.

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

In this section, I 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 1.1 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.

In Table 1.2, I 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; I show results using three different measures of noisiness of signals: immigrant status, linguistic distance and cultural distance.

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.

The results in Table 1.2 provide evidence that monotonicity directionally holds. In line with prediction 2, the coefficients for the noisy signal measures in columns (1), (3) and (5) are all 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, having a linguistic distance of 1 or having a cultural distance of 1—is 2.4%, 3.0% and 10.6% 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[1].

Furthermore, in support of predictions 1 and 3, which suggests that there is a stronger sorting effect for the linguistically distant and the more educated, the interaction term between the noisy signal measure and years of education added in columns (2), (4) and (6) is 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 educational attainment. With an additional year of education, individuals with a noisy signal have a 0.4% - 0.6% higher rate of self-employment. Considering the base rate of self-employment of 13%, this translates into a selection effect of 3 – 4%.

There are two things to especially note. First, it is remarkable that the predictions of the framework hold using three different 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.

In Table 1.3 I 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, I 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 all three measures using the ACS. The coefficients can be interpreted as follows: the rate of self-employment for individuals who are

[1] For cultural distance, since most countries have cultural distance of 0.5 rather than 1, it is less meaningful to calculate the selection effect.

foreign born, or who have a linguistic distance of 1, or who have a cultural distance of 1, with a high school degree (college education) are 2.9% (4.4%), 3.2% (5.1%), and 4.8% (7.0%) 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 P-value for cultural distance is not significant at the 10% level, however, it is very likely that this lack of signifying is due to a lack of variation owing to fewer coverage of countries. However, the results qualitatively hold for all measures. For the remainder of the analysis I use linguistic distance as the main measure.

Additional robustness checks for this main result are shown in Appendix in Tables A.1, A.2 and A.3. In Table A.1, I replicate Table 1.3 for a subset of immigrants. In Panel A, I 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, I limit my 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 illegal immigrants or other short term factors. The results qualitatively hold for all three measures for both subsets. The statistical significance between coefficients B and C is somewhat weaker for the cultural distance measure; however, the respective coefficients for B and C are statistically significantly positive. In Table A.2, I replicate the results using the CPS. Again, the results directionally hold, and the respective coefficients for B and C are statistically significant. I use a conservative test using clustered standard errors at the country level, and as a result, I lose significance between coefficients B and C. In Table A.3, I 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 I lose statistical significance at the 10% level for the between coefficient comparisons, the results qualitatively hold.

In Table 1.4 I further evaluate the differential selection by linguistic distance categories; I subset my 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 or cultural 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.

The result shown in column (4) is less favorable. 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. Hence, it is not surprising that the linguistic distance measure loses explanatory power for this subset.

In Appendix Table A.4 I replicate the analysis using cultural distance. Here, I do not find strong results in line with the prediction. Again, this is likely a result of the sparse representation of countries for the cultural distance measure. In Appendix Table A.5 I replicate the results 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.

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

There are some limitations to mapping the theoretical framework to my 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. I 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 break 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.

1.6.2 Cultural assimilation and selection into self-employment

In this section I test the subsequent hypotheses consistent with the framework, predictions 4: whether culturally assimilated immigrants select into self-employment less. I proxy for the degree of an immigrant's cultural assimilation in two ways: I 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, I assess whether the selection effect into self-employment is mitigated for immigrants with a U.S. education. Using immigrants' age of immigration, I 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 1.5. 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.

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 1.5, I further breakdown college and above into some college and college degree. This enables me 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. I argue that it is the imprecise signal owing to cultural differences that affects immigrants' employment outcomes.

Second, I 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. I 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, I study the relation between age of immigration and employment outcomes.

I add a variable that indicates whether an immigrant came to the US before the age of 10. The results are shown in Table 1.6. In column (1), I add the indicator to the standard specification used in column (3) of Table 1.2, in column (2), I 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, I find strong negative coefficients for both terms in columns (1) and (2).

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.

I 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, I replicate column (2) of Table 1.6 in Appendix TableA.6, using different indicators cutting immigration age at age 7, 8, 9 and 11. My results are not sensitive to how I define immigrants who arrive at a young age.

1.6.3 Sorting when immigrants compose the majority group

In this section, I test prediction 5: whether immigrants who compose the majority in their group select into self-employment less. While I apply theory to the setting of immigrant workers in the U.S., the model can be interpreted more generally as a mismatch between two individuals with a different cultural background.

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, I identify immigrants surrounded by their co-ethnics, using information from the CPS about the metro area of individuals' residence. Specifically, I identify the metro area-industry-occupation cluster of workers. I chart the self-employment rate of ethnic groups against the average representation in the "workplace" in Figure 1.3.

I show regression results in Table 1.7, which includes a variable for representation of immigrants' ethnic group in their cluster. Along with standard controls, I include fixed effects for age, immigrant cohort, state, year, industry and occupation categories.

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 higher representation in their workplace are 3.4% less likely to enter into self-employment than those who are not part of the majority group. Moreover, although the coefficients lose statistical significance at the 10% level, the magnitude of the 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 represent a majority is 3.8% lower. This result suggests 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. My 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. This result may also be understood in conjunction with Borjas et al. (2017), who show how the influx of Chinese graduate students increased the productivity of Chinese Math professors.

1.7 Potential alternative explanations

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

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

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

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

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

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

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

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

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

This expression suggests that the more important communication skills are for a job, the larger the agent's marginal product. Hence, if communication skills are an important productivity input, employers have much to lose from hiring someone that does not speak English well. Hence, language deficiency would damage workers in communication-intensive occupations to a greater extent. In such a case, we should empirically observe stronger sorting into self-employment for the subset of workers in jobs that require more communication skills.

To test this hypothesis, I decompose occupations by their skill requirements. Specifically, I use the O*Net Skill measure to characterize occupations by their degree of communication intensity. I take the average scores of reading comprehension, speaking and writing skills required for the job, and divide salaried occupations into jobs that require above and below median language skills in order to compare their effects regarding sorting into self-employment. Table 1.8 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.

There are two important things to note from my 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, I reject the hypothesis that linguistic distance merely measures language as a productivity input

Second, I complement the linguistic distance measure with individuals' self-reported English scores. If linguistic distance measures technical language skills rather than cultural distance, the sorting effect would disappear, once the sub-setting on immigrants reports that they speak English well. Table 1.9 presents results replicating columns (3) and (4) of Table 1.2. Columns (1) and (2) repeat results from Table 1.2 and columns (3) and (4) replicate the analysis on a subset of immigrants who report to speak English well.

My results hold even when I include only immigrants who speak English well; Column (3) of Table 1.9 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.

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

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

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

My results support that linguistic distance proxies for cultural differences. In Table 1.10, I show results controlling for both cultural and linguistic distance in using a similar specification as in Table 1.2. My 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.

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.

1.7.2 Linguistic distance as a measure for distaste for unfamiliarity

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

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

One way to tackle this question is to, again, exploit how language proficiency naturally interacts with acquisition age, as shown in Table 1.6. Those who immigrate at a young age share the same observable characteristics as those who immigrate at a later age, except they do not suffer from linguistic-cultural barriers. If it were taste-based discrimination, we should see the same effect for this subgroup of immigrants. My results showing that coming before age 10 mitigates selection into self-employment for the linguistically distant is in line with the imprecise signaling hypothesis. This result demonstrates that linguistic distance does not simply measure distaste for immigrant group-specific attributes.

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

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

1.7.3 Linguistic distance as a proxy for ethnic group specific factors

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

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

The results reported in Table 1,11 show weak support for this hypothesis. The sample includes immigrants who were born in either Belgium or Switzerland and those who speak Dutch, German or French at home. In column (1), I find strong support for the hypothesis with a coefficient of 17.2%, when I include standard controls but do not include any fixed effects. Once I include fixed effects for 22 major occupation categories, however, the result loses significance, as shown in column (2). I conjecture that the test may lose variation since immigrants from Belgium or Switzerland may be concentrated in particular occupation categories. Hence, instead, I create 6 categories of occupations constructed based on the complex problem solving skill measure from the O*Net Skill scores, to include them as fixed effects. As shown in column (3), the results regain significance.

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

Second, I 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, I 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 6.3. The results for this test are presented in Appendix Table A.7. I 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, I rule out potential alternative explanations, including language as an input to production, taste-based discrimination and ethnic group-specific factors, that may explain why linguistic distance predicts immigrants sorting into business ownership. The series of empirical results suggest that there is systematic bias in the context of hiring immigrant workers that is not fully explained by conventional factors noted in previous work. I attribute such bias to systematic bias arising from cultural mismatch in the context of hiring immigrant workers.

1.7.4 Limitations

In this section I discuss several limitations of this study. First, while I posit that cultural mismatch can cause friction in matching workers to firms, I do not examine how cultural fit may shape organizations' productivity. Prior studies have discussed how cultural fit may facilitate coordination (Van den Steen 2005) and how ethnic ties help generate business leads and meet financing needs (Nanda and Khanna 2010). Conversely, studies have also suggested how firms may benefit from diverse teams, as such teams are more likely to make decisions more carefully and become more open to new ideas (Phillips et al. 2009); more generally, Collier (2001) finds that fractionalized societies perform better in the private sector. 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.

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

1.8 Conclusion

In this study, I 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, I study the experiences of immigrants, who likely face especially large labor market frictions owing to linguistic-cultural barriers. I apply a theoretical framework that presumes that shared culture lubricates communication and hence mismatch in the linguistic-cultural backgrounds between an interviewer and a candidate in a 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. I empirically test these predictions by investigating whether there exist differential patterns of sorting out of salaried work and into self-employment between immigrants and non-immigrants across subsets of education categories.

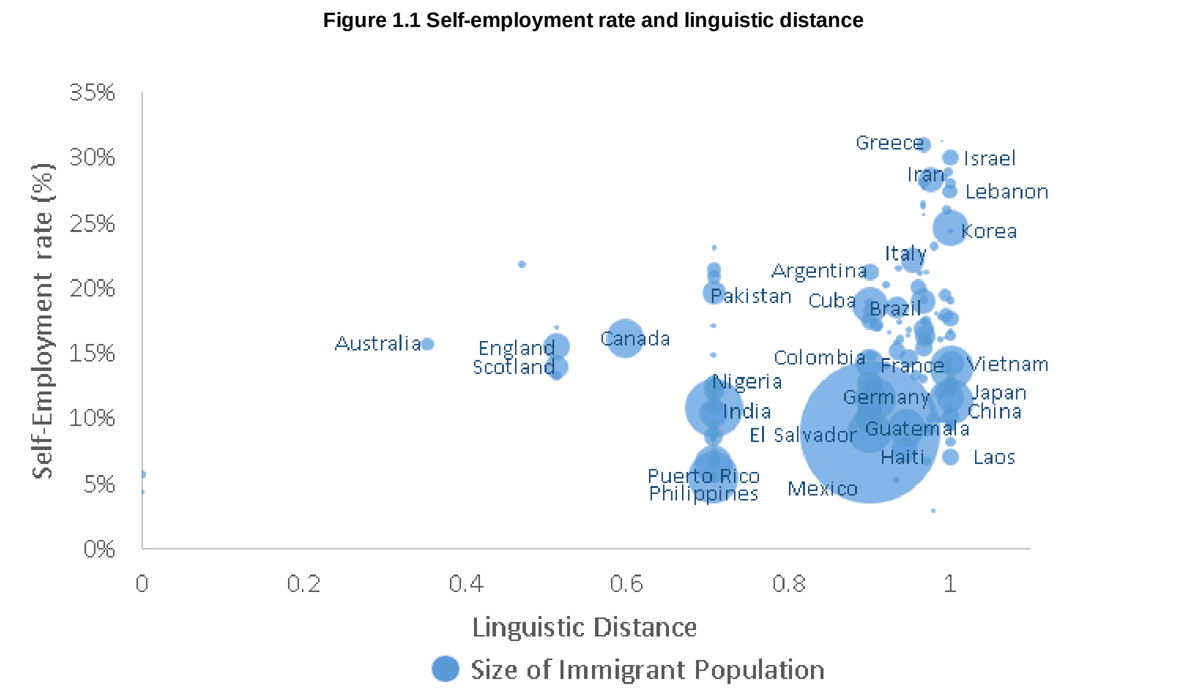
Consistent with the theoretical framework, I show that immigrants are more likely to sort into self-employment, particularly when they have noisier signals and higher education. I proxy for the degree of imprecise signaling with "linguistic distance," a measure based on how many branches separate two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

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

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

1.9 Figures

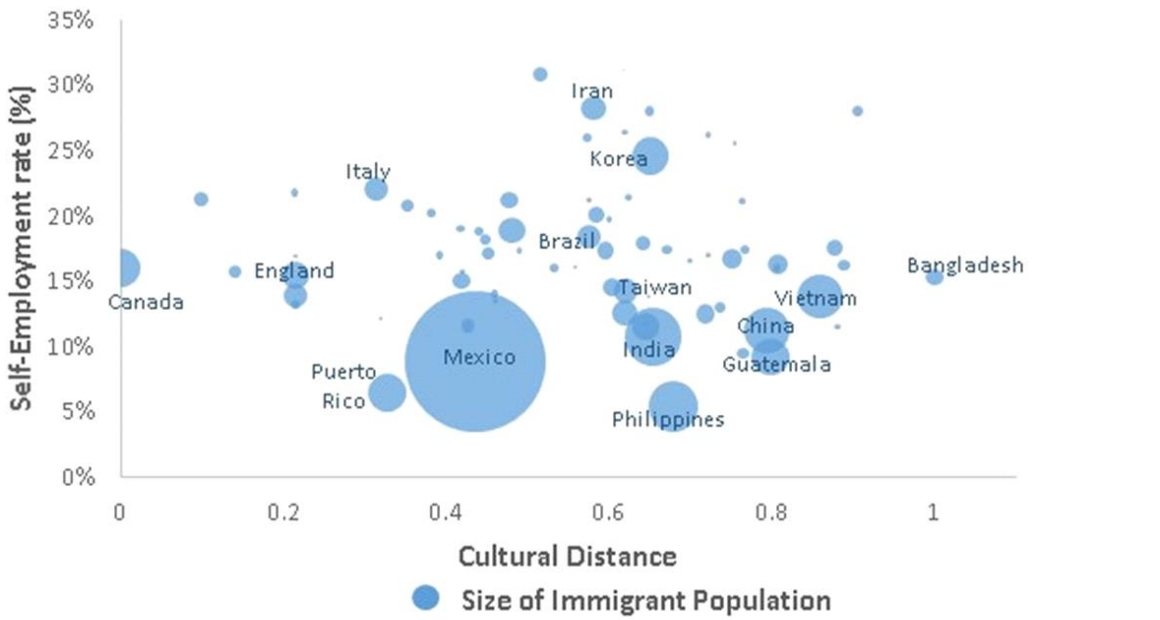
Figure 1.1 Self-employment rate and linguistic distance by ethnic groups



Source: American Community Survey, 2005 - 2012; Linguistic 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.

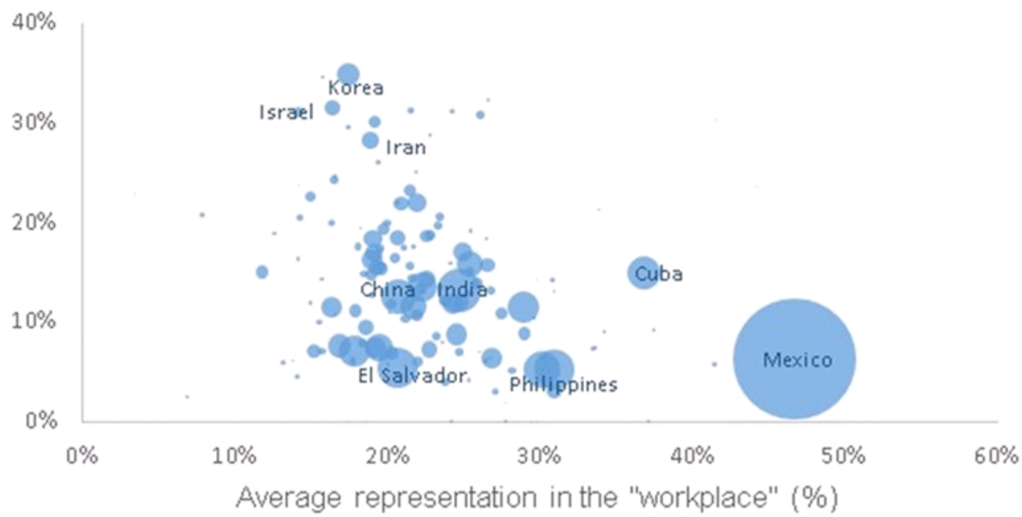
Figure 1.2 Self-employment rate and cultural distance by ethnic groups

Figure 1.2 Self-employment rate and cultural distance



Source: American Community Survey, 2005 - 2012; 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 1.3 Self-employment rate and 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.

1.10 Tables

Table 1.1 Sample description

	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 1.2 Selection into self-employment

	Self-employment (vs Salaried)					
Measure of noisy signal:	**Immigrant Status 1gImm**		**Linguistic Distance LD**		**Cultural Distance CD**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: American Community Survey (ACS), 2005 - 2012						
Noisy signal	0.024**	-0.035*	0.030*	-0.037	0.106***	0.02
	0.012	0.02	0.016	0.024	0.039	0.047
Years of education	0.002***	0	0.002***	0	0.002***	0
	0	0.001	0.001	0.001	0	0
Noisy signal x Yrs of education		0.004***		0.005***		0.006**
		0.001		0.001		0.002
Controls	✓	✓	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓	✓	✓
Constant	-0.009	0.027	-0.016	0.021	-0.209	-0.173
	0.051	0.046	0.058	0.052	0.129	0.12
Number of Observations	5280414		5280414		5069458	
Base rate of self-employment	13%		13%		13%	
Selection effect						
The noisiest signal relative to U.S.-born	18%		23%		n/a	
With an additional year of education	3%		4%		4%	
Panel B: Current Population Survey (CPS), 1994 - 2012						
Noisy signal	0.039***	-0.026	0.051**	-0.029	0.116**	0.02
	0.012	0.018	0.02	0.028	0.046	0.059
Years of education	0.001	-0.001*	0.001	-0.001*	0.001	0
	0.001	0	0.001	0	0.001	0
Noisy signal x Yrs of education		0.005***		0.006***		0.007***
		0.001		0.001		0.003
Controls	✓	✓	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓	✓	✓
Constant	-0.159***	-0.124**	-0.163**	-0.127*	-0.306*	-0.270*
	0.06	0.06	0.07	0.068	0.154	0.152
Number of Observations	583189		583189		562874	
Base rate of self-employment	13%		13%		13%	
Selection effect						
The noisiest signal relative to U.S.-born	31%		40%		n/a	
With an additional year of education	4%		5%		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 three measure of noisy signal: columns (1) and (2) use immigrant status, columns (3) and (4) use linguistic distance, columns (5) and (6) use cultural 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 1.3 Differential selection by education categories

	Self-employment (vs Salaried)		
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**	**Cultural Dist CD**
	(1)	(2)	(3)
Education (vs Grade School)			
High School	0.000	0.000	0.002
	0.003	0.003	0.002
College	0.004	0.004	0.008**
	0.005	0.005	0.003
A. Noisy signal (1gImm / LD / CD)	-0.009	-0.008	0.047
	0.014	0.017	0.042
B. (1gImm / LD / CD) x High School	0.029***	0.032***	0.048***
	0.007	0.008	0.014
C. (1gImm / LD / CD) x College	0.044***	0.051***	0.070***
	0.009	0.010	0.022
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	5280414	5280414	5069458
P-values comparing coefficients			
B = C	0.055	0.041	0.185

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 measure of noisy signal: columns (1) and (2) use immigrant status, columns (3) and (4) use linguistic distance, columns (5) and (6) use cultural 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.
Reports p-values from t-tests testing equality between coefficients of the interaction terms.
Calcualtions weighted using the population weights provided.

Table 1.4 Differential selection by education and 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; *, **, 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.

Table 1.5 Immigrants' selection into self-employment, some college vs college degree

	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			0.020***	0.030***
			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; *, **, 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.

Table 1.6 Immigrants' selection into self-employment by age of arrival

	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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.

Calcualtions weighted using the population weights provided.

Table 1.7 Immigrants' workplace representation and selection into self-employment

	Self-employment (vs Salaried)		
	(1)	(2)	(3)
Linguistic Distance (LD)	0.062**	0.061**	0.070**
	0.025	0.025	0.029
"Workplace" representation		-0.034***	-0.006
		0.006	0.015
"Workplace" rep. x Ling. Dist			-0.038
			0.023
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	94059	94059	94059

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.
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 1.8 Assessing language skills as input to productivity

	Self-employment (vs Salaried)			
Communication intensitiy of salaried jobs:	**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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table 1.9 Assessing 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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table 1.10 Effect of cultural distance controlling for linguistic distance

	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	-0.02
		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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table 1.11 Sorting by immigrants from multilingual 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.
*, **, and *** indicate significant at 10%, 5% and 1%, respectively. Calcualtions weighted using the population weights provided.

Chapter 2

Finding the Hidden Gems: An Arbitrage Opportunity for Firms

2.1 Introduction

Mr. Lee[§] had originally planned to become a Certified Public Accountant when he first moved to the United States. After several years and several attempts looking for employment, however, he ultimately became disillusioned about making it in corporate America. Instead, he decided to open up a sushi restaurant. Stories of immigrant entrepreneurs such as Mr. Lee are not uncommon. Asian immigrants often form their own businesses out of necessity, as they struggle to find and maintain jobs. Although Mr. Lee had hoped to become a successful worker in a major accounting firm, events did not unfold as he planned.

As much as Mr. Lee's story describes immigrants facing barriers in their job search, it also illustrates a fundamental challenge for corporations looking to hire. Identifying the most talented and productive immigrants in an increasingly diverse labor force, represents both a challenge and an opportunity for firms. In part, the difficulty stems from the fact that those screening the workers often have markedly different cultural backgrounds than the immigrant population. This imposes a hurdle not only for immigrants trying to assimilate in mainstream economic markets, but also for firms hoping to recruit talented workers. So then, how can firms successfully identify and recruit from a diverse workforce in the face of such challenge?

In this chapter, I discuss how firms may exploit the arbitrage opportunities arising from screening frictions owing to immigrant's different linguistic-cultural backgrounds, discussed in Chapter 1. First, I provide some stylized facts consistent with the framework on the information asymmetry problem foreign workers face and discuss implications for worker productivity. Second, I highlight the importance of solving this information problem for firms by quantifying the potential benefits. Last, I discuss how firms may develop more effective hiring strategies.

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

2.2 Implications of the hidden gems problem

Immigrants select into self-employment differently from the non-immigrant population. My previous chapter discusses how the different selection pattern into self-employment reflects imprecise signalling of immigrants who come from a different linguistic-cultural background. In this section, I examine the outcomes of the predictions of the model through stylized facts from surveys representative of the U.S. population.

The main implication of the framework is that there are differential sorting depending on the difficulty of the job and the relative precisions of the two signals that the candidates send. Considering that immigrants send a noisier signal relative to U.S.-born workers, I provide two simple stylized facts that is consistent with the framework.

First, I compare the rate of self-employment by educational attainment by immigrants and their U.S.-born counterparts. The differential selection predicted by the framework is summarized well in the bar chart shown in Figure 2.1. The bar chart represents the self-employment rate of an immigrant and her propensities to self-employ were she to be assigned the propensities to self-employ of an American, who has similar observable traits.

Operationally, the bar chart compares immigrants' rate of self-employment to the fitted rate of self-employment, $\widehat{SelfEmp}$, for immigrants using coefficients from running the following regression only for U.S.-born workers:

$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon \qquad (2.1)$$

X includes the standard controls including four race categories, time spent in the U.S. and fixed effects for age, industry, occupation, state and year. The underlying idea is that by assigning U.S.-born workers' coefficients to immigrants, I estimate the likelihood that an immigrant would have entered into self-employment, were it not for her immigrant status.

Interestingly, while the U.S. counterfactuals are less likely to enter into self-employment with higher education, immigrants have a higher tendency to enter into self-employment. As a result, the gap between the self-employment rates conditional on educational attainment increases with educational attainment.

Second, I then consider some stylized facts from a firms' perspective by considering the distribution of talent for occupations requiring different levels of skills. The opportunity cost of having these immigrants in self-employment in our society means that they are not in salaried employment. In particular, I look at the distribution of talent among salaried workers that are sorted into different types of occupations. Figure 2.2 shows a box plot of educational attainment by occupation categories, the top figure using the March CPS and the bottom figure using the American Community Survey. Occupations are categorized by the complexity of their jobs using the Occupational Information Network (O*Net) skill scores, a hedonic measure of skills generated by the U.S. Department of Labor.

There are two important observations to note: First the relative spread of the boxes, indicating the 25th and 75th percentile, changes across the type of occupations. Specifically, immigrants have a more compressed education range for complex jobs than for routine jobs. Immigrants with a wide range of education is hired in routine jobs, whereas the opposite is true for complex jobs. Second the mean and median of the two groups systematically differ depending on type of occupations—the red line in the chart indicates the median, the mean numbers and standard errors are shown in the table below. These numbers suggest that in complex jobs, immigrants on average are more likely to have higher educational attainment while in routine jobs, the opposite is true—immigrants are likely to have lower educational attainment.

The theory predicts that immigrants who send noisier signals have a harder time convincing the employers in complex jobs, where employers are more sensitive to the precision of the signals, while in routine jobs that is not the case. The observations based on the differential rate of self-employment and talent distribution across different type of salaried work fit the theoretical framework quite well.

2.3 Are high skilled salaried immigrants more productive?

The framework then has further implications about the productivity of salaried workers. If it is the case that screening frictions hinder immigrants from joining as a salaried employee in complex jobs, it would also imply that immigrants that are hired as salaried workers outperform non-immigrant salaried workers. I test whether this is the case by examining changes in hourly wages for a salaried individual i using the following specification:

$$Log(HourlyWage)_{i,FE} = \beta_0+\beta_1\mathbb{1}(Immigrant)_i+\beta_2 Education_i+\beta_3\mathbb{1}(Immigrant)_i\times Education_i$$
$$+\beta_4 X_i+\lambda_{FE}+\epsilon_{i,FE}$$

Educational attainment is used to proxy for the difficulty of the job. For *Education*, I use either years of education or education categories, including less than a high school degree, high school degree and some college and above. X_i includes individual specific controls, such as race, time spent in the US. The specification also includes fixed effects for age, year, state, industry, and occupation. Standard errors are clustered at the origin country level.

The coefficient for the indicator term for immigrant status, β_1, suggests the relative wage performance of the least educated immigrants relative to the least educated non-immigrants; the coefficient for the interaction term, β_3, indicates the differential wage gap a more educated immigrant face relative to their native counterpart and also relative to the low educated immigrant. I run this specification using the Current Population Survey and the American Community Survey. Results are shown in Table 2.1.

Assuming that the more highly educated are more likely to apply to and work in complex jobs, the theoretical framework predicts that immigrants with higher education, will more likely outperform their native counterparts. However, the negative interaction terms in columns (1) and (3) suggest that this does not hold true. Columns (2) and (4) further suggest that this negative effect is mostly driven by immigrants with a high school education. The statistically insignificant coefficient for the interaction term *1gImm x College* suggests that immigrants with a college education earn a statistically comparable wage relative to their non-immigrant counterparts.

The result that highly educated immigrants in salaried employment do not significantly outperform the U.S.-born is inconsistent with the framework. If it were the case that cultural mismatch cause a particularly large friction for highly educated immigrant workers who can perfectly perform the job well, and if such screening friction were the only force affecting immigrant workers, the hired immigrants should significantly outperform the non-immigrant workers as such friction would draw upon immigrant workers from the very top tail of the distribution. The insignificant difference in the productivity of workers of college educated workers may suggest that for the same reasons immigrants face screening frictions in hiring settings they may also be less productive in firms.

However, there may be some alternate channels that give rise to this outcome. Specifically, salaried immigrants may face discrimination even after the hiring stage. For example, frictions may take place anytime over one's career—promotion decisions could be understood as organic hiring decisions and each staffing opportunity provided to workers could be understood as a screening process. This is in line with Altonji & Pierret (2001), who show how training opportunities may increasingly be given to more productive workers. In other words, screening frictions can affect immigrant workers more generally beyond the hiring setting, and the labor market pattern we observe in the population may reflect an even more skewed version of the populous shaped by the initial hiring decision.

2.4 How much better off can firms be?

Then, provided that it is better to have the highly skilled immigrants in salaried work, what is the degree of magnitude of the benefits that firms can gain by alleviating this misallocation problem? I estimate the productivity gain for society at a partial equilibrium in which immigrant workers who are misallocated as self-employed are employed in firms. Throughout this analysis, I assume an elastic demand for labor.

I estimate the potential economic gains 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. Table 2.2 shows the top 10 occupations that the highly educated immigrants and non-immigrants self-employ in. Interestingly, restaurant and food store managers are among the most pursued businesses by immigrants but not among the U.S.-born.

These suggest that there may be social losses associated with how immigrants sort in the labor market and that society can better leverage their skills. I conduct a productivity analysis to evaluate the potential social gains from correctly identifying self-employed immigrant workers. The productivity analysis is based on the following specification:

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

Here, worker wage is determined by the type of employment—salaried work or self-employment—and their human capital. X includes standard observable characteristics such as race, age and time spent in U.S. as well as year and state fixed effects. However, it does not include industry or occupation fixed effects, accounting for the fact that people may self-employ in different industries and occupations.

I assess the difference between salaried employment and self-employment as well as the difference between salaried employment and unincorporated self-employment. According to Levine and Rubinstein (2016), the unincorporated self-employed represent a non-entrepreneurial type of self-employed individuals; hence, I use this group as an upper bound for the productivity difference between those who are in salaried work and self-employment.

Estimations based on the American Community Survey for the period from 2005 to 2012 by education group are shown in Figure 2.3. In summary, the estimates suggest that the potential gains from hiring each talented immigrant who is misallocated in the market may be ~\$3,000-\$9,000 for an average worker and ~\$6,000-\$18,000 for a highly-educated immigrant worker, annually. The equivalent for U.S.-born workers in self-employment are ~\$3,000-\$12,000 for an average worker and \$4,000-\$15,000 for the highly educated.

Two observations are worth noting. First, for both immigrants and U.S.-born, the potential social gain is disproportionately large for firms that require higher skilled labor. Second, for those firms looking for skill level that require less than a college education, there is more to gain from hiring a self-employed U.S. born, while for firms in search for high skilled labor, there is more to gain from hiring a self-employed immigrant. In other words, firms requiring college educated workers may gain ~\$3,000 more by hiring a highly educated self-employed immigrant rather than a highly educated self-employed U.S.-born.

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

Previous studies have discussed the positive role of small businesses, however. First, self-employment is a channel for immigrants to assimilate in host societies better—Lofstrom (2002) shows how immigrants in self-employment reach earnings parity faster than immigrants that are in salaried employment. Second, Ottaviano and Peri (2006) show that immigrants provide a variety ethnically diverse local goods and services. Third, Glaeser et al. (2010) show that many small, entrepreneurial employers are highly correlated with higher regional economic growth, although in a more recent work, Haltiwanger et. al (2013) show that there is no systematic relationship between firm size and employment growth once a firm's age is controlled for, and furthermore that small mature businesses negatively affect net job creation.

While I acknowledge these ongoing debates and that small businesses may positively contribute to the creation of jobs and ethnically diverse products, the purpose of this study is to highlight that there are economic tradeoffs from having an immigrant capable enough to work in a more complex salaried job in a lower task self-employed job. Perhaps the society would benefit from having the competent immigrant laundromat owners as salaried workers in accounting firms and competent U.S.-born construction workers as self-employed as laundromat owners. The systematic pattern we observe across population where immigrants are more educated in higher skilled jobs and less educated in lower skilled jobs suggest that there is room to better allocate talent.

2.5 How should firms effectively screen?

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. There are two directions a social planner may take in alleviating this misallocation problem. One is to minimize screening frictions by implementing training programs for immigrants. Another, perhaps a more realistic adjustment, would be to have firms implement more effective hiring policies. Siegel et al. (2014) show how multinational firms can gain competitive advantages from hiring the excluded group to positions of managerial authority; I argue that firms can domestically gain competitive advantages by overcoming barriers to attracting immigrant workers. In this section, I discuss how firms should improve their hiring strategies to attract the most productive workers and which firms may be able to benefit by investing in their HR practices.

First, my 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. My suggestions contrast this common belief; however, they resonate with the handicapping principle in the contest literature: Ridlon and Shin (2013) indicate that giving a boost to those with a disadvantage yields better outcomes in competitions when there is severe heterogeneity.

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

One factor to consider is that for most firms that hire highly educated workers, employers make very specific searches by conducting campus recruiting at top tier schools rather than searching the local labor market. Hence, I conduct my analysis based on the field of degree. I use the ACS to collapse over 150 fields of degree into 20 major categories, as listed in Appendix Table B.1.

I conduct a cost-benefit analysis to identify when it is worthwhile for firms to make the investment to hire an HR representative who speaks the candidate's language. For this purpose, I identify fields from which society may substantially gain from having misallocated workers in salaried work, and I then suggest how costly it would be for firms if a more diverse set of ethnic groups were to pursue their particular fields. The results are visually summarized in Figure 2.4.

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

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

Given this difference along with the number of immigrants who major in the different fields, I am able to rank order fields by the acuteness of the misallocation problem. Based on my sample, Engineering and Business majors presents the largest social benefits, while Psychology, Biology and Health Services majors provide the lowest benefits. The solid line demarcates the point where the social gain for having a worker in salaried work becomes positive and hence where immigrants with a Psychology or a Biology major are likely to earn more through self-employment. This rank ordering is plotted along the X-axis of Figure 2.4.

Furthermore, the problem is more widespread when a more diverse set of ethnic groups pursue those particular fields, as shown by the Y-axis of Figure 2.4. In this figure, the Y-axis measures the rank ordering of the diversity of the misallocated ethnic pool depending on their field. Specifically, I count the number of misallocated immigrants by 12 country-of-origin categories from Lofstrom (2002) and then count the number of "peaks" in the distribution. Peaks are defined to have more than one misallocated immigrant in my sample in an origin category. The misallocated immigrants among engineering majors are heavily focused in a few ethnic categories, primarily in the Middle East or Latin America, while misallocated immigrants among business majors occur for a diverse set of ethnic groups. The misallocated measure is a metric for differential sorting into self-employment using the following specifications:

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

where $\widehat{SelfEmp}$ is the fitted value for immigrants using coefficients from running the specification in equation (2.1).

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 2.4 suggests that if a firm is in search of a worker in the first quadrant, with an engineering or a computer science degree, it should conduct a targeted search, as the ethnic category span of misallocated workers for those majors is quite narrow, while the productivity gains from having those workers in a firm can be large. Conversely, if a firm is in search of someone in the third quadrant, with a liberal 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.

2.6 Conclusion

In this chapter, I argue that firms should view the labor market imperfection arising from cultural differences as a source of competitive advantage and suggest how firms may adjust their hiring practices to better recruit hidden foreign talent.

The immigrant talent pool composes almost 18% of the working age population, and the potential economic gains for society as a whole from correctly identifying immigrants, especially those who are highly educated, are thus meaningful. Indeed, partial equilibrium estimates suggest that the annual potential gains from hiring each talented immigrant who is misallocated in the market is ~ $3,000-$9,000 for an average worker and ~$6,000-$18,000 for a highly-educated worker.

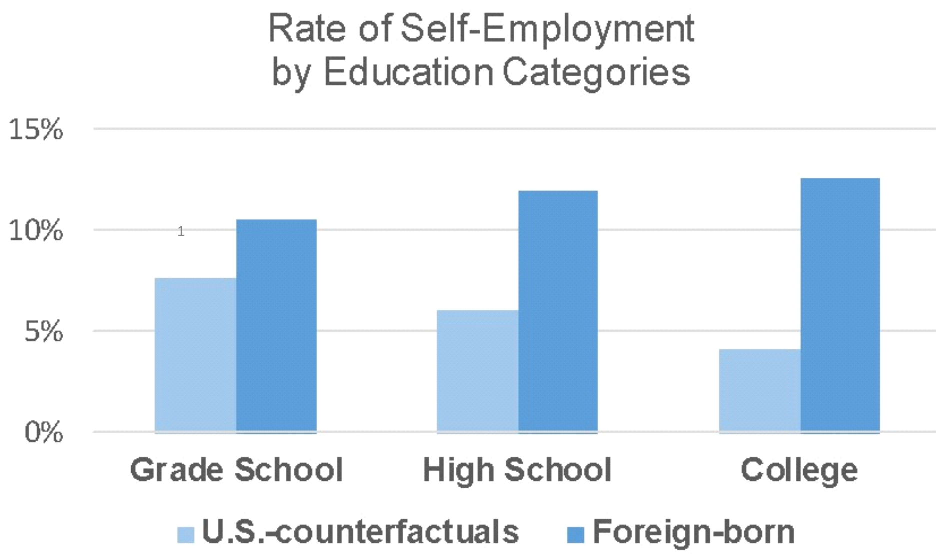
My findings suggest that for immigrant policy towards high skilled immigrant workers to work its full extent, firms should invest in hiring technologies. In contrast to the common belief that more objective measures in hiring will overcome social biases, in some cases, it is better to implement hiring practices that examine people of different cultural and ethnic backgrounds more carefully. I also suggest firms may benefit from hiring people who can decipher immigrants' signal better depending on which fields of degree they are targeting.

While policy discussions have focused on seeking talented workers from outside the US, I hope to inform policy makers and firms about the potential gains from searching inside the US. A number of studies have examined the impact of the H-1B program on US innovation (Kerr and Lincoln 2010, Kerr et al. 2015)[2]. While understanding the effect of that particular segment of the immigrant labor force is crucial, only 650,000 of the 41 million immigrants are estimated to be H-1B visa holders[3]. I argue that by adjusting their managerial practices, firms can better harness the untapped talent pool of high skilled immigrant workers, who are abundant but hidden within the U.S.

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

2.7 Figures

Figure 2.1. Propensity to self-employ between immigrants and their U.S.-counterfactuals

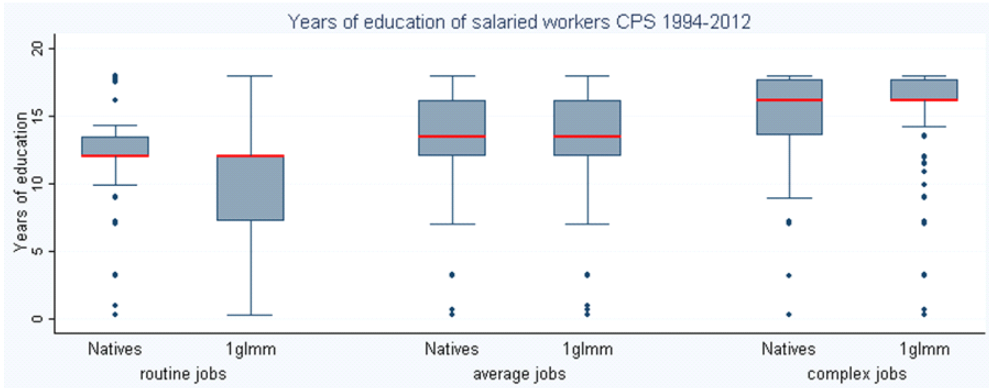


Source: American Community Survey 2005 -2012
Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year.
Differential sorting based on an imputed measure comparing immigrant's likelihood to self-emply to U.S.-born workers.
This is calculated as (Self-Employed) - (Estimated Self-Employed), where (Estimated Self-Employed) is the fitted value for immigrants using the coefficients from running the following regression for U.S. born workers:

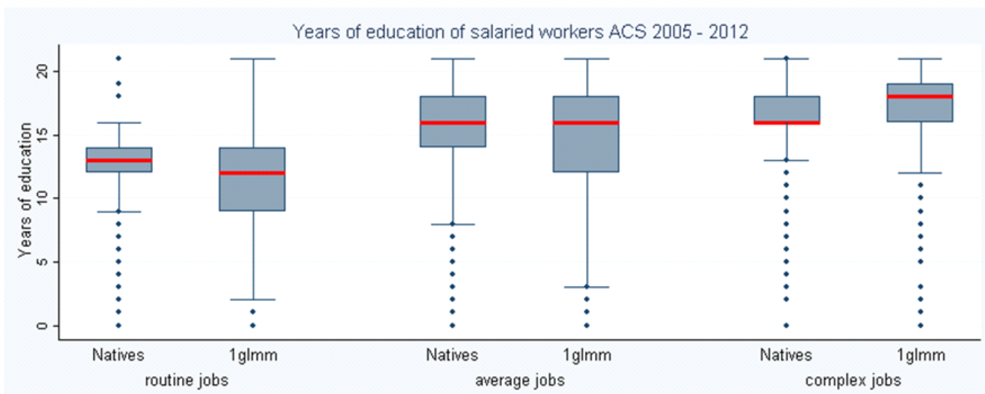
$$SelfEmp_i = \beta_0 + \beta_1 EducCategories_i + \beta_2 X_i + \epsilon$$

All education categories and linguistic categories defined exclusively.
Immigrants compose 65% of 'Grade Scjpp;' category, suggesting that the imputation is driven by a smaller U.S.-born sample.
Calcualtions weighted using the population weights provided by the survey.

Figure 2.2. Talent Distribution of salaried workers by types of occupation



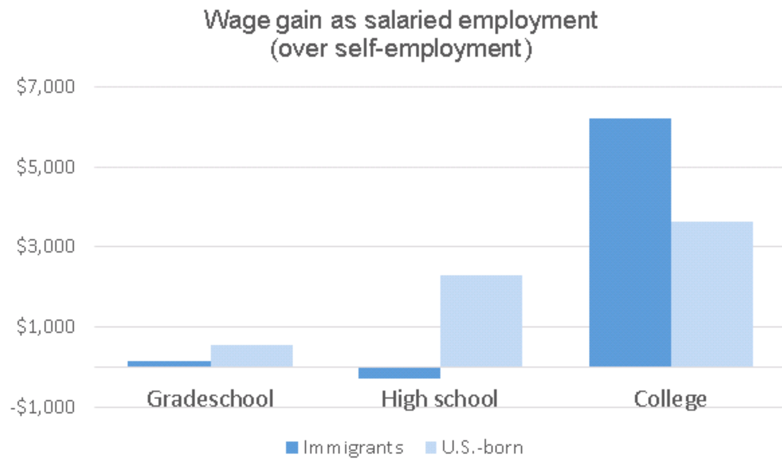
Mean:	12.55	10.55	13.69	12.56	15.52	16.16
Standard Error:	0.00	0.02	0.01	0.02	0.01	0.01

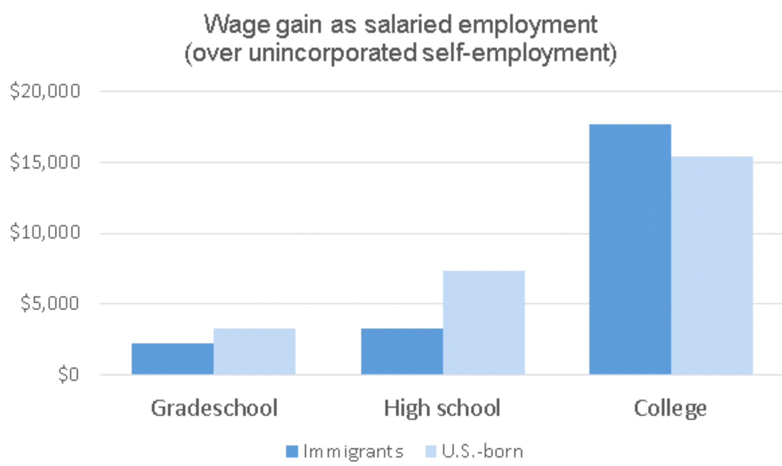


Mean:	13.41	11.46	15.58	14.65	16.59	17.75
Standard Error:	0.00	0.01	0.00	0.01	0.01	0.01

Source: Current Population Survey 1994 - 2012 ; American Community Survey 2005 - 2012
Notes: Sample includes salaried immigrant and non-immigrant workers; males between 18 and 65, who worked full-time full-year. Calcualtions weighted using the population weights provided.

Figure 2.3. Potential annual gain per worker by education categories



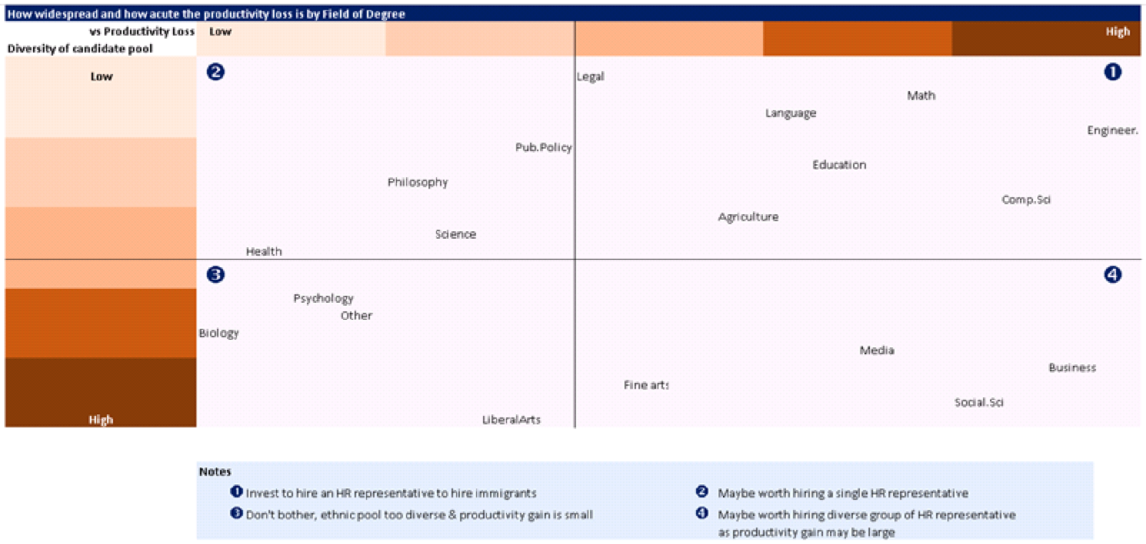


Source: American Community Survey 2005 -2012
Notes: Sample includes male immigrant workers, between 18 - 65 old in the survey year, who worked full-time for the entire year. Potential wage gain evaluated as the sum of coefficients beta 2 and beta 3 from the following specification:

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

All education categories and linguistic categories defined exclusively; uses winsorized earnings at 5% level. Calcualtions weighted using the population weights provided by the survey.

Figure 2.4. Diversity vs Importance of potential candidate pool by field of degree



Source: American Community Survey, 2010 -2012
Notes: Over 150 field of degree categories grouped into 20 categories.
X-axis is the rank ordreing of the productivity loss associated with having a worker as self-employed rather tnan salaried work
High indicates high productivity loss, Low indicates low productivity loss
Productivity loss evaluated using the sum of coefficients beta2 and beta 3 from the following equation:

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

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

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

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

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

2.8 Tables

Table 2.1 Productivity of salaried workers by education categories

	Log hourly earnings			
	Current Population Survey		American Community Survey	
	(1)	(2)	(3)	(4)
First generation immigrants (1gImm)	0.373***	0.123***	0.342***	0.089**
	0.086	0.037	0.065	0.035
Average treatment effect				
Years of education (educ)	0.062***		0.048***	
	0.003		0.003	
1gImm x educ	-0.019***		-0.018***	
	0.006		0.004	
With education categories (base: grade school)				
High School		0.156***		0.092***
		0.001		0.001
College		0.299***		0.202***
		0.004		0.004
1gImm x High School		-0.042***		-0.007
		0.01		0.009
1gImm x College		0.008		0.023
		0.027		0.026
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓

Source: Current Population Survey 1994 - 2012 ; American Community Survey 2005 - 2012
Notes: Sample includes salaried immigrant and non-immigrant workers; males between 18 and 65, who worked full-time full-year. Column (1) and (2) show results based on the Current Population Survey, columns (3) and (4) use the American Community Survey. Controls include four race categories including (Non-hispanic) White, Black, Hispanic and Asian; and time spent in U.S. for which U.S.-born are assigned their wage. Fixed effects include age, year, state, industry and occupation categories.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table 2.2 Top 10 occupations of highly educated self-employed workers

	Immigrants		Natives
1	Physicians and surgeons	1	Lawyers and judges
2	Restaurant managers	2	Construction managers
3	Business services workers	3	Physicians and surgeons
4	Construction managers	4	Business services managers
5	Business services managers	5	Real estate agents and brokers
6	Food store managers	6	Dentists
7	Real estate agents and brokers	7	Accountants and auditors
8	Wholesale trade managers	8	Insurance agents and brokers
9	Dentists	9	Retail store managers
10	Lawyers and judges	10	Building managers

Source: Current Population Survey 1994 - 2012 ; American Community Survey 2005 - 2012
Notes: Sample includes self-employed workers; males between 18 and 65, who worked full-time full-year; with a college degree

Chapter 3

Persistence of Discrimination: The Effects of 9/11

3.1 Introduction

According to the Economist, the annual incidence of hate crimes in 2016 reported to the FBI increased by 35% over 2015. Historically, such a rise was short-lived—the article notes that while there was a sharp increase in the number of hate crimes following the September 11 terrorist attacks in 2001, the number fell back quickly. It is quite unique and remarkable that hatred towards a certain group could be stimulated and dampened so quickly. Given that hatred is a concept that is tightly linked with one well-established theory of discrimination (Becker 1957), this poses some questions about what this means for the discussion on preference-based discrimination that minority groups face. Was the effect of discrimination also short-lived as suggested by the number of hate crime incidents? Were there any long-term consequences? I explore these questions by examining how the anti-Islamic sentiment arising from the September 11 terrorist attacks shaped labor market outcomes in the long run for Arabic immigrants in the US.

Whether the degree of discrimination vary across time or not has been an important defining characteristic of two long standing explanations for minority wage gaps—Becker's taste-based discrimination and Arrow and Phelps' statistical discrimination. Statistical discrimination has been understood to attenuate over time as employers learn about workers' productivity, while taste-based bias is assumed to be fixed—Altonji & Pierret (2001) use these characteristics in their seminal article to identify the existence of statistical discrimination based on education, and show that educational status becomes less important as employers learn about unobserved ability.

While it is not unreasonable to think that taste and preference towards a group of people may change quickly, the time-invariability assumption of taste-based biases has received surprisingly little empirical challenge. There are two potential reasons for this. First, it is hard to distinguish among taste-based biases from other factors. In order to identify workers' labor market outcomes reflecting preference-based bias at the population level, there needs to be a large enough negative shock affecting a particular group of people. Second, even in the presence of such a shock, most data sets representative of the population are cross-sectional, making it hard for examiners to disentangle the long-term effects directly associated with taste-based discrimination from other types of negative selection biases (Abramitzky et al. 2014). To address this problem, I use a panel data to examine how the tragic incident of the September 11 terrorist attacks shaped labor market outcomes for Arabic immigrants. While there are existing studies such as Kaushal et al. (2007) and Davila and Mora (2005) that examine this setting, I believe this is the first study to address long-term labor market consequences following the incident.

In this paper, I provide suggestive evidence that there are heterogeneous effects of 9/11 across time associated with educational attainment. I use a difference-in-differences approach to examine how the taste-based discrimination arising from 9/11 manifests itself in immigrants' labor market outcomes for different subgroups. I then discuss underlying drivers of the heterogeneous effects I find.

I use the 1979 National Longitudinal Survey of Youth (NLSY79) to assess biennial labor market outcomes before and after the September 11 terrorist attacks in 2001, from 1994 to 2012. My sample only includes a small number of Arabic or Middle Eastern immigrants, however. Hence, I only provide suggestive results by identifying immigrants who are exposed to discrimination two ways. First, I use a genetic distance measure constructed by Wacziarg and Spolaore (2009), which quantifies the genetic similarity between countries; I use the distance between immigrants' origin countries to Middle Eastern countries to categorize immigrants into those that are more or less affected by the terrorist attacks. Second, I more narrowly define the treatment group using geographic distance which include immigrants from countries in the Middle East or North Africa.

The main findings of this paper are as follows: First, I confirm previous studies that the terrorist attacks have negative effects on labor market outcomes of immigrants from countries that are genetically or geographically close to the Middle East. Immigrants who are more genetically close to the Middle East face a 24% discount in log hourly wage, on average, relative to the less genetically similar group post 9/11.

Second, I show that there are heterogeneous effects across time based on educational attainment where the negative effects are mostly driven by the less well educated immigrants. Moreover, the differential effect between the high and low educated among the most genetically similar group increases over time. In particular, a one standard deviation decrease in educational attainment lowers log hourly wages by 18-21%, on average, for immigrants in the treatment group. By 2011, a one standard deviation difference in education corresponds to an additional 34-41% drop in wages by the low educated relative to the highly educated. This suggests that educational attainment may be more valuable for minority groups, as it would allow them to work in high-skilled jobs, which would better shield workers from discrimination in the long run.

These findings are interesting because the standard way distaste is formalized in Becker (1957) is as a fixed cost to an employer's utility. The theoretical framework does not provide an explanation for why such cost would systematically vary over time and diverge based on observable characteristics.

Third, I examine drivers of the heterogeneous outcome associated with educational attainment. I consider whether there are systematic patterns of occupation switching by education groups. I provide suggestive evidence to partly ascribe the heterogeneous effects to occupation discrimination as seen when workers sort into different occupation categories. In particular, I use occupation characteristics such as complex problem solving skills required to show that the low educated sort into less complex jobs relative to the highly educated after the 9/11 incident. This finding suggests that immigrants would increasingly overcrowd workers in low skilled jobs in the presence of discrimination.

I further test alternative mechanisms that may be driving my results; specifically, other sources of discrimination, other channels for how the highly educated respond systematically differently or other explanations spuriously driving the wage decrease. First, I use measures of linguistic, genetic, and religious distance from the US to show that mere difference in background does not generate bias against the affected immigrants. Second, I rule out the possibility that the difference in unobserved inherent ability between the high and low educated workers is driving the heterogeneous outcome. Third, I test whether different distributions of workers in the treatment group across industries drive changes in wages. I provide evidence against these proposed hypotheses.

In this study, I provide suggestive evidence that the effects of preference-based discrimination may be amplified over time differentially for different subgroups of the discriminated population. In particular, I show that educational attainment become more important in circumventing discrimination over time. The existing literature does not explain these findings. I attribute part of the reason to how members of the affected group respond by differentially sorting into occupations. My findings suggest how discrimination cause systematic divergence in the occupational distribution among populations and how education becomes increasingly more valuable for the discriminated group over time.

3.2 Literature review

This paper contributes to the literature testing theories of taste-based discrimination. Previous studies have tested Becker's prediction in several ways. One stream of literature tests how increased competition should make employers less likely to forego profit to indulge in their discriminatory taste; therefore, more competitive markets should decrease wage gaps for minorities. Black and Brainerd (2004) test how change in import share between 1976 and 1993 in the manufacturing sector narrowed gender wage gap more in non-competitive industries. Levine et al. (2012) show how increased labor market competition from state-level banking deregulation reduced Black-White wage gaps. Another stream includes Charles and Guryan (2008) who use measures of employer prejudices and fraction of black workforce to test Becker's prediction about how the marginal discriminator determines wage gaps. I take a different approach, testing how the degree of taste-based discrimination change labor market outcomes. I exploit how the anti-Islamic sentiment following the terrorist attacks vary depending on genetic and geographic proximity of an immigrant to the Middle East. Rather than testing equilibrium predictions, I trace workers' labor market outcomes over time to show that taste-based discrimination manifest itself to give rise to occupational discrimination. This is loosely related to the discussion by Hsieh et al. (2016), who show that convergence in the occupational distribution among workers account for a quarter of US productivity growth over the past decades. The fact that occupational distribution diverges for the discriminated groups over time suggest inefficient matching in the labor market.

In my setting, I specifically question the theoretical assumption that preferences affect members of the same minority groups in a similar manner, and that this prejudice remains constant over time. This assumption in particular has been used to identify statistical discrimination—Altonji and Pierret (2001) find evidence for statistical discrimination, under the assumption that statistical discrimination attenuates over time as employers learn about worker ability while taste-based discrimination does not. In this study, I question whether taste-based discrimination also vary over time, differently depending on worker's observable characteristics. My suggestive finding may potentially help explain why Altonji and Pierret (2001) find evidence of statistical discrimination based on education but not based on race—taste-based discrimination based on race or ethnic groups could give rise to differences in labor market outcomes based on education.

Finally, this study complements other papers that discuss the effect of terrorist activity on labor market outcomes of ethnic minorities. Previous studies have looked at how the September 11 terrorist attacks affected the employment outcomes of Arabic and Muslim immigrant in the U.S. In particular, Kaushal et al. (2007) and Davila and Mora (2005) use cross-sectional data to show that there is a temporary drop in earnings of Arabic immigrants following the attacks. While their sample is representative of the population, cross-sectional data is not free of selection biases; hence, not suited for studying labor market changes in the long-term. In this study, I use a panel data set to investigate the persistence of the negative sentiment towards Arabic immigrant workers and how they shape labor market outcomes over time. While previous studies such as Gould and Klor (2016) have shown how 9/11 attacks hinder social assimilation of Muslim immigrants, I question whether there are direct effects of discrimination causing occupational segregation in the labor force.

3.3 Empirical context

3.3.1 Institutional background

On September 11th, 2001 four passenger airplanes operated by United Airlines and American Airlines were hijacked by 19 al-Qaeda terrorists, before two of them crashed into the World Trade Center complex and one into the Pentagon. This unexpected tragic perpetration caused 2,977 direct victims. Following the attacks, Arabic immigrants residing in the US have also been severely affected by the anti-Islamic sentiments caused by the terrorist attacks; according to Human Rights Watch, USA (2002), Arab and Muslims reported more than two thousand September 11-related backlash incidents.

The terrorist attacks also marked a start of a new era for the US immigration regulation. President Bush established the Department of Homeland Security, representing the largest government reorganization since the creation of the Defense Department after World War II (Mittelstadt et al. 2011). Among other changes, new visa policies were enforced and US customs and border protections were greatly enhanced.

3.3.2 Data description

The main data set of the study is the National Longitudinal Survey of Youth 1979 (NLSY79), a survey representative of the US population. I conduct my analyses only among the 985 immigrants in the sample. I further limit my sample to those who worked as salaried employees and had nonnegative earnings for the observed time periods. I also exclude immigrants from Central America per Kaushal et al (2007). In order to know the origin country of the immigrants I use the restricted access geocoded version of the NLSY79, which provides information about immigrants' country of origin.

The individuals in the NLSY79 sample were first surveyed in 1979, when they were 15 – 22 years old; they were surveyed annually up to 1994, and biennially from 1994, through the age of 48 – 55. Survey questions include workers' labor market outcomes as well as workers' cognitive and non-cognitive traits. In this study, I use labor market outcomes during the 10 biennial time periods between from 1994 and 2012.

I emphasize that a panel data set allows me to best estimate the degree of labor market changes occurring from the anti-Arabic sentiment following the terrorist attacks given profound changes in immigration policies after the terrorist attacks as described above. Cross sectional data may suffer from selection bias—there may be return migration of Arabic immigrants owing to the change in the social environment, as well as different selection relating to the entrance of new migrants. This is the reason why existing studies such as Kaushal et al. (2007) and Davila and Mora (2005) only examined labor market outcomes for a short period of time following the terrorist attacks. However, the sample size of the immigrant population is small and this study only provides suggestive evidence.

Given that immigrants are typically oversampled, the empirical results may be sensitive to the weighting of the individuals. In particular, given that I analyze the data in a longitudinal manner; in the case that a respondent misses an interview for some of the waves, the empirical results using the provided generic annual sample weights are not representative of my sample. To improve accuracy in analyzing multiple years of data, I use custom weights for the immigrant groups in the sample, separately requested from the Bureau of Labor Statistics, rather than the annual sample weights provided by the survey. The total weighted sample size is about 500 working immigrants and about 200 immigrants when I further restrict it to those who reported nonnegative earnings and exclude immigrants from central Amerca. My analytical results are not sensitive to this restriction.

This NLSY79 data set provides two major advantages to my study. First, as an extensive panel data set it enables me to make empirical improvements to existing studies studying the labor market outcomes of immigrants post the terrorist attacks (Davila and Mora 2005; Kaushal et al. 2007). Also, a panel data set allows for more selection than I typically would with a cross-sectional data. Specifically, by using individual fixed effects, I can allow individual units to differ from each other in unobserved ways that affect their outcomes in a manner that is constant over time. This helps overcome the selection issue the study may face from having a particular subset of immigrants in the NLSY79 sample.

Second, the NLSY79 data allows me to measure both traits that are observable and those that are only observable to the statistician. For observable characteristics, I focus on education levels and genetic similarity. For education levels, given that I only use data from 1994, each individual's educational acquisition would have been completed by then in most cases. I use genetic proximity to the Middle East, to categorize the treatment and control group for this study, this is designated based on the individual's country of origin. I use the country level genetic distance measure constructed by Wacziarg and Spolaore (2009), which I refer to below. For unobservable characteristics, I use the Armed Forces Qualification Test (AFQT) score, surveyed in 1980. This score is based on arithmetic ability and vertical expression scores and the raw scores are converted to an AFQT percentile score, ranging from 1 to 99. Previous studies, including Lang and Manove (2011) have shown that this measure has the capacity to capture unobserved ability not measured with education levels.

3.3.3 Treatment and control groups

To categorize the treatment and control group, I need to identify immigrants who would have become more negatively perceived after the September 11 terrorist attacks. Given that taste-based discrimination is most often associated with racism, I use a measure of genetic distance to proxy for the degree of taste-based discrimination individuals face. Specifically, I employ the country level genetic distance measure constructed by Spolaore and Wacziarg (2009) based on Cavalli-Sforza et al. (1994) who report bilateral genetic distance among 42 world populations, computed from 120 alleles. This measure captures how genealogically related two populations are by measuring the time elapsed since the last time there were common ancestors between two populations. The shorter the time elapsed the less likely the populations would differ in a wide range of traits and characteristics. Hence, I identify immigrants that are more likely to be biologically perceived as Middle Eastern from those that do not by using the genetic distance between Saudi Arabia and an immigrant's country of origin. While I use Saudi Arabia to calculate the distance, my results are not sensitive to which Middle Eastern country I use. Previous studies have shown how anti-Islamic sentiment vary with appearance. Davila and Mora (2005) show that there are stronger declines in labor market outcomes by Middle Eastern Arab men than African Arab men following the September 11 terrorist attacks in the US.

In one of the empirical specifications, I also use measures of genetic distance, linguistic distance and religious distance from the U.S., also constructed by Spolaore and Wacziarg (2009). Linguistic distance measures language similarity between two countries, based on Fearon (2003)'s approach which counts the number of branches that separate two languages in a language tree. Similarly, religious proximity is based on Mecham et al. (2006) who categorize religious family trees.

I categorize immigrants by their genetic proximity to Saudi Arabia. The top half is categorized as the treatment group, the bottom half as the control group. My empirical results are not sensitive to whether I define the treatment group as the top quartile, the top 33rd percentile or the top 50th percentile. Defining the treatment group based on appearance is supported by findings of previous research—Ratcliffe and Von Hinke Kessler Scholder (2015) show that the 2005 London Bombing incident increased hate crimes against all Asians, not just those that are Muslims.

As a robustness check, I also run the same analysis using a more narrowly defined treatment group using geographic distance to the Middle East. This group includes countries such as Israel, Iraq, Iran, Nigeria India, Greece, Italy and Portugal. My results qualitatively hold in the same direction only using these samples.

Table 3.1 shows the summary statistics of the sample by treatment and control groups using the two different distances. I report observable characteristics including race, gender, years of education and labor market outcomes; unobservable characteristics including mother's and father's education and AFQT scores; labor market outcomes such as hourly income and which major industry categories the workers are in. I also show averages of the Spolaore and Wacziarg (2009) distance measures.

Given that groups are categorized based on genetic and geographic distances from the Middle East, it is unlikely that the groups are balanced across all observable traits. For example, the race compositions are quite different and are differences in hourly wages. Such difference is not too much of a concern, however, as I include individual fixed effects in all of my specifications to measure average percentile changes in individual wages. I also assess whether earning trends pre-9/11 are similar across different groups.

Small sample sizes inflate standard errors making it less likely to find statistically significant results. To minimize biases from outliers, I use winsorized earnings of log hourly wages at the 5% level. My empirical results do not qualitatively differ with or without making this adjustment. In the following sections, I first perform my empirical analysis comparing the treatment and control group and then I show that there is a heterogeneous effect within the treatment group.

3.4 How 9/11 shaped immigrants' labor market outcomes

In this section, I first discuss the empirical methodologies used to identify the effect of the September 11 terrorist attacks on the labor market outcomes of immigrants who are more likely to be erroneously associated with the terrorist attacks. Then I test whether there is a wage shock for those that are more genetically close to the Middle East.

3.4.1 Difference-in-differences estimation

I begin by investigating the degree of negative shock on immigrants erroneously associated with the terrorist attack, and hence use the following standard difference-in-differences specification:

$$Log(HourlyWage)_{i,s,t} = \beta_0 + \beta_1 \mathbb{1}(Post\,911)_t + \beta_2 \mathbb{1}(Treatment_s \times Post\,911_t) + \delta_i + \epsilon_{i,s,t}$$

$Log(HourlyWage)_{i,s,t}$ is the log hourly wage of individual i, of country category s, in year t. The main explanatory variable of interest is $\mathbb{1}(Treatment_s \times Post\,911_t)$, which has value 1 after the September 11 terrorist attacks for immigrants categorized as the treatment group based on the two distance measures. The variable $\mathbb{1}(Post\,911)_t$ hold fixed the time varying trends that may influence both the treatment and control groups. In a subsequent specification, I also include time fixed effects for each survey year, in lieu of $\mathbb{1}(Post\,911)_t$, to control for time specific factors affecting the entire sample in a given year. The regression includes individual fixed effects, δ_i, to control for unobserved individual characteristics that do not vary over time. Given individual fixed effects, the coefficient of the interaction term does not change regardless of whether the standalone term $\mathbb{1}(Treatment)_s$ is included in the regression. $\varepsilon_{i,s,t}$ is the error term. Given that observations pertaining to the same individual at different points in time are very likely to be correlated with each other, I cluster standard errors at the individual level.

The β_2 coefficient on the term $\mathbb{1}(Treatment_s \times Post\text{-}911_t)$ provides an estimate for the average percent changes in hourly earnings of immigrants that fall under the treatment group, categorized by either genetic similarity or geographic distance to the Middle East, after the terrorist attacks.

3.4.2 Empirical findings

I find that there is a sharp decline in labor market outcomes for the affected immigrants after the terrorist attacks. The results for average treatment effects and the results including year fixed effects are shown in Table 3.2. As shown in column (1) the affected immigrants based on genetic similarity to the Middle East face a 24% discount in log hourly wage; column (3), which uses geographic distance to define the treatment group, show that there is a 12% discount, although the result is not statistically significant.

My empirical results add to the discussion of existing studies in several ways. First, while I confirm that there is a negative effect of 9/11, I find a larger magnitude of the decrease in labor market outcomes over a longer time horizon. Specifically, Kaushal et al. (2007), who studies labor market outcomes from 1997 to 2005, find that wages declined by 9 – 11% for Arab and Muslim men post 911. Second, I show that the negative effect is strongest for the group that are most genetically similar to the Middle East. This resonates with the discussion on the existence of skin shade effect on minority wages (Goldsmith et al, 2006). Last, by showing that labor market outcomes differ with genetic distance from Saudi Arabia, rather than genetic distance from the U.S., I show that the mechanism that generates prejudice is specifically through the anti-Arabic sentiment. For example, a country such as Jamaica is genetically more distant from the U.S. than Israel is but is genetically further from Saudi Arabia and therefore is less affected. This is important to note because it highlights that the mechanism driving discriminatory behavior is based on taste against a specific physical feature associated with the terrorist attacks.

In columns (2) and (4), I show my results with year fixed effects with the base year as 1999. Hence, the coefficients of the interaction terms, *Year t x Treatment*, suggest decline relative to right before the terrorist incident. Although not statistically significant at the 10% level, the consistently negative coefficients suggest a decline in hourly wages post-9/11.

For my estimations to be unbiased, I need to address two challenges: (1) whether there is endogeneity in the rise of the exogenous shock, and more importantly, (2) that in the absence of terrorist attacks, the labor market outcomes of the treatment and control groups would have followed a similar trajectory.

First, the September 11 terrorist attacks happened unexpectedly; therefore, reverse causality is less of a concern. In other words, it is unlikely that the expected changes in labor market productivity triggered the terrorist attack.

Second, I investigate whether the parallel trends assumption holds in two ways: in a regression framework and through visual inspection. In the regression framework, I assess this using the interaction terms, *Year t x Treatment* of columns (2) and (4) of Table 3.2 for years prior to 911. If the treatment group were randomly assigned, I would expect the coefficients for these terms to be a statistical zero. Relative to the base year, 1999, the coefficients are positive but not statistically significant in column (2) and are negative but not statistically significant in column (4). In other words, while there are differences in wages between the treatment and control group prior to 911, the differences are statistically not meaningful across the years prior to the terrorist attacks. As for the visual inspection, I plot the changes in hourly earnings in Figure 3.1 grouping immigrants using the genetic and geographic distance measures. The dotted line marks the terrorist attacks. As can be seen visually, leading up to the dotted line, there are no clear difference in earnings trend before the incident. I use hourly earnings winsorized at 5%, but the visual trend is qualitatively similar even without this adjustment. Hence, both regression and visual inspections suggest that the parallel trend assumption holds.

By analyzing through a panel that covers a longer time horizon, my results importantly suggest how the terrorist attacks shape labor market outcomes for the affected immigrants overtime. This gives me an opportunity to examine not only the existence and persistence of discrimination but also how taste-based discrimination manifests itself over time for different subgroups of the immigrant population. Given this unique vantage point, I further assess heterogeneous effects across subgroups of the affected immigrants.

3.4.3 Heterogeneous effects across education groups

I further investigate whether taste based discrimination arising from 9/11 dynamically affected different immigrant subgroups overtime. I test whether there are varying effects of discrimination depending on educational attainment. The difference-in-differences specification is as follows:

$$Log(HourlyWage)_{i,s,t} = \beta_0 + \beta_1 \mathbb{1}(Treat_s \times Post-911_t) + \beta_2 \mathbb{1}(Years\,of\,educ_i \times Post-911_t)$$
$$+\beta_3 \mathbb{1}(Treat_s \times Years\,of\,educ_i \times Post-911_t) + \delta_i + \epsilon_{i,s,t}$$

As before, this specification includes individual fixed effects, δ_i. Educational attainment is completed for all individuals in the sample, therefore this specification tests whether the terrorist attacks have a differential effect for immigrants with different levels of educational attainment. Results are shown in Table 3.3: the average treatment effect is shown in columns (1) and (3) and results with year fixed effects is shown in columns (2) and (4).

My results show that there are heterogeneous effects of 9/11 depending on immigrants' educational attainment. In column (1), two coefficients are worth noting. First, the negative coefficient for the interaction term between 9/11 and the treatment group suggests that there is a large negative effect of 9/11 for the treatment group. Second, the positive coefficient for the triple interaction term, *9/11 × Treatment × Years of education*, suggests that with a college education—16 years of education—the negative effect from the terrorist attacks can be recouped. This suggests that the wage decrease following the terrorist attacks is mostly driven by those with low education among the treated immigrants. In particular, the coefficients suggest that a one standard deviation decrease in educational attainment (2.7 years) in the treatment groups leads to a 18% (0.068 × 2.7) to 21% (0.079 × 2.7) additional loss in hourly wage, on average, relative to the highly educated. The statistically insignificant coefficient for the term *9/11 × Years of education*, suggests that the trend between the high and low educated does not systematically differ before and after the terrorist attacks.

Moreover, I find that there exists a systematic time-varying effect. In particular, as shown in columns (2) and (4) with year fixed effects, the differential labor market outcome for the high and low educated increases over time, where the low educated are increasingly worse off. Not only are all the interaction terms statistically significant, but moreover, t-test comparing coefficients of the triple interaction terms 2007 x Treatment x educ and 2011x Treatment x educ suggest that the increase in gap is statistically significant. Specifically, by 2011, a one standard deviation difference in education corresponds to an additional 34% (0.128 × 2.7) to 41%(0.153 × 2.7) drop in wages relative to the highly educated. This suggests that the highly-educated cope with discrimination better than the less well educated, especially over time.

My findings contrast the discussion by Kaushal et al. (2007) who show that the magnitude of taste-based biases from the September 11 terrorist attack more or less similarly affect immigrants with different education levels, nativity status, and residential location. We show that the effect of taste-based bias differs for immigrants with different education levels in the long run.

Although my results are suggestive, I show that there potentially may be time-varying, dynamic effects of taste-based discrimination on labor market outcomes. This has an important implication for the existing literature as it questions the commonly made assumption that taste-based discrimination remains fixed overtime. For example, Altonji & Pierret (2001) use such time unvarying features to identify statistical discrimination—statistical discrimination, but not taste-based discrimination, is assumed to diminish with experience. Furthermore, my findings inform how taste-based discrimination, as well as statistical discrimination, can endogenously shape investment in human capital by forward-looking minority groups. Previously, acquisition of human capital in response to discrimination has mostly been understood on the basis of statistical discrimination models, where discrimination based on imperfect information negatively shape investment in human capital by forward looking workers who face lower expected returns[4]. In contrast, I show that the long-term consequences of taste-based discrimination may also endogenously shape investment choices of immigrant workers, but in the positive direction. What I find resonates with findings by Lang and Manove (2011), that conditional on cognitive skills, minority groups have higher levels of education. While Lang and Manove (2011) posit noisy signaling as the main driver, however, I show how acquisition of human capital may arise to better circumvent taste-based discrimination.

3.4.4 Response to discrimination

The fact that the highly-educated increasingly fare better over time suggests that the highly educated and the low educated may respond systematically differently to discrimination. I further explore the drivers of my previous result which suggests that the differential effect of discrimination for the high and low educated immigrants grows over time for the treated group. Specifically, I conjecture (1) whether discrimination cause workers to switch occupations more and (2) whether discrimination further shape the type of occupations workers transition into.

[4] For example, Coate and Loury (1993) show how misbeliefs about lower productivity of minority groups can generate a self-fulfilling equilibrium where the minority groups endogenously choose lower levels of human capital.

First, I consider the likelihood of occupation transition. I create an indicator for whether there was any change in a worker's occupation category relative to the previous year and simply compare the differential likelihood of transitioning between the high and low educated. I run the following specification for the years 2001, onwards:

$$\mathbb{1}(OccTransition)_{i,s,t} = \beta_0 + \beta_1 \mathbb{1}(Year_t) + \beta_2 \mathbb{1}(Year_t \times Years\,of\,educ_i) + \beta_3 \mathbb{1}(Treat_s \times Year_t) + \beta_4 \mathbb{1}(Treat_s \times Years\,of\,educ_i \times Year_t) + \delta_i + \epsilon_{i,s,t}$$

The results examining occupation transition since the terrorist attacks are shown in Table 3.4. As shown in columns (2) and (4), among the treatment group, the highly educated in general were less likely to switch occupation after the terrorist attacks, relative to the low educated in the treatment group. Most of the differential transition between the high and low educated happened in between 2005 and 2007.

Second, I examine what type of occupation these workers transitioned into. To test this, I assess the characteristics of occupations workers are in using the O*Net Skill scores, a hedonic measure of skills created by the Department of Labor. I use the same specification as above but with the occupation complexity scores as the dependent variable in lieu of the likelihood of occupation transition. Results are shown in Table 3.5.

Results in column (1), although not statistically significant, suggest that the treatment group is more likely to sort into less complex jobs. In column (2), however, the positive coefficients for the triple interaction terms suggest that the decrease in occupation complexity score is driven by the low educated. Specifically, with one standard deviation increase in the years of education (~2.7 years), the complexity score of the occupation that the individual is working in increases by 9.38 (2.7 ×3.500) scores after the terrorist attacks in between 2003 and 2005. The results are qualitatively in the same direction when defining the treatment group using geographic distance although not statistically significant.

In other words, my results suggest that bias against immigrants who are genetically similar to the Middle East manifests itself in labor market outcomes by sorting workers into different education categories, whereby the low educated are more likely to sort into less complex jobs.

3.5 Discussion

In this section, I explore other plausible explanations for my findings. Specifically, I discuss whether there are alternative channels of discrimination, other potential responses to discrimination, and any sectoral changes that particularly affect immigrants in the treated group.

3.5.1 Alternative sources for differential effects by genetic similarity

One may question whether the empirical results that I find are only feelings against immigrants that are perceived as Middle Eastern, or whether the terrorist attacks trigger other sources of discrimination that drive the results. For example, the terrorist attacks may give rise to negative sentiment towards people with different genetic proximity, linguistic-cultural background or different religions. I test whether the treatment effects I observe are a result of distaste towards Arab looking immigrants or whether they are driven by general differences in language, culture or belief.

In order to test this, I use measures of genetic distance, linguistic distance and religious distance from the US, also constructed by Spolaore and Wacziarg (2009). Linguistic distance measures language similarity between two countries, based on Fearon (2003)'s approach of counting the number of branches that separate two languages in a language tree. Similarly, religious proximity is based on Mecham et al. (2006) who categorized religious family trees. I test whether I find similar effects using these measures in lieu of genetic proximity to Saudi Arabia.

Results testing the alternative sources of discrimination are shown in Appendix Table C.1. The regression results using the distance measures from the U.S. are shown in columns (2), (3) and (4) respectively; column (1) shows the baseline empirical result using genetic distance from the Middle East. Discrimination would again be suggested by the coefficient of the interaction terms between 911 and the distance measures. While the measure for genetic distance from Saudi Arabia shows a statistically significant negative coefficient, none of the other measures show such result. The interaction terms are either statistically insignificant or slightly positive, suggesting that wages do not change differentially, or only slightly increase, for immigrants who are distant from the U.S. in terms of these other measures.

These results suggest that the wage changes I find in my main empirical results are due to taste against immigrants who are genetically close to the Middle East, rather than mere foreignness of an immigrant owing to their appearance, cultural or religious background.

3.5.2 Alternative sources for differential effects by education

One of the long run debates in the discrimination literature has been whether labor market wage gaps reflect discrimination or differences in skill. Neal and Johnson (1996) had argued that the black-white wage gap reflects differences in skill, by showing that the wage differentials are not driven by lower returns to cognitive skills, measured by AFQT scores, in expectations. However, Lang and Manove (2011) later revisit this argument that it is inappropriate to only control for AFQT scores considering that black workers acquire higher levels of education conditional on AFQT scores.

I conjecture whether the systematically different behavior by educational attainment reflects discrimination, as I argue, or differences in inherent skill. Perhaps, the highly educated may differ in their ability to perceive discrimination and circumvent discrimination more effectively. As in the existing studies on this debate, I use AFQT scores to proxy for such unobserved ability.

I test whether the highly educated have different levels of inherent ability other than higher educational attainment by replicating Table 3.3 using AFQT performance scores instead of years of education. The results are shown in Appendix Table C.2. The main result in Table 3.3 shows that the differential effect between the high and low educated increases over time. Relative to this finding, the heterogeneous effect does not hold with AFQT scores as shown by the statistically insignificant coefficients for the interaction terms between each year and AFQT scores.

The fact that I do not find the same heterogeneous effects of education with AFQT scores suggests that immigrants with lower educational attainment, but not necessarily those with lower unobserved ability, face a longer lasting effect of taste-based discrimination. Had I found that AFQT scores present qualitatively the same results as when I use education levels, it would have suggested that the heterogeneous effect could be ascribed to unobserved ability that affects productivity, as well as the institutional constraints that the low educated have to face. Hence, my finding helps clarify that the highly educated being in different occupations, rather than having different inherent ability, cause taste-based discrimination to shape different patterns of selection for the high and low educated immigrants in the long run.

3.5.3 Alternative sources for decrease in wages

The individual fixed effects model depends on the assumption that the error term that varies across an individual is fixed over time. What if this is not the case? The terrorist attacks caused major changes in immigration regulations, arguably most severely affecting Arabic immigrants. While changes in immigrant policies may cause different selection of immigrants or changes in wage structure of industries heavily concentrated by immigrants, the difference-in-differences empirical design comparing Arab looking immigrants with immigrants that look less Arabic, gives me comfort that such macro changes would likely affect all immigrants in a similar manner. However, there could be other sectoral shocks that coincide with the terrorist attacks that spuriously drive my results. In particular, the burst of the Dot Com Bubble coincided with the terrorist attacks. To the extent that Arabic immigrants are overrepresented in affected industries systematic macroeconomic changes may also differentially affect changes in wages between the treatment and control groups.

I address this problem using industry fixed effects. In my baseline empirical analysis, I only include individual fixed effects, rather than including industry or occupation fixed effects, in order not to draw inferences from sparse cells. I include fixed effects for major industry categories in 2001 and then with industry fixed effects interacted with September 11 in my specification. The results including these fixed effects are shown in Appendix Table C.3. Columns (1) through (3) shows the general treatment effect, columns (4) through (6) show the heterogeneous effect by education groups within the treatment group. Columns (2) and (4) include industry fixed effects and columns (3) and (6) include industry fixed effects interacted with 9/11.

The statistically significant negative coefficients for the interaction terms *9/11 x Treatment* suggest that the discriminatory effect I find, especially for the immigrants most genetically close to the Middle East, is not owing to changes in industry level factors. The triple interaction terms including years of education, especially in column (6) suggest the same holds for the heterogeneous effects across education groups. This exercise suggests that macro level changes are not spuriously driving my results by differentially affecting the industries that the members of the treatment group are concentrated in.

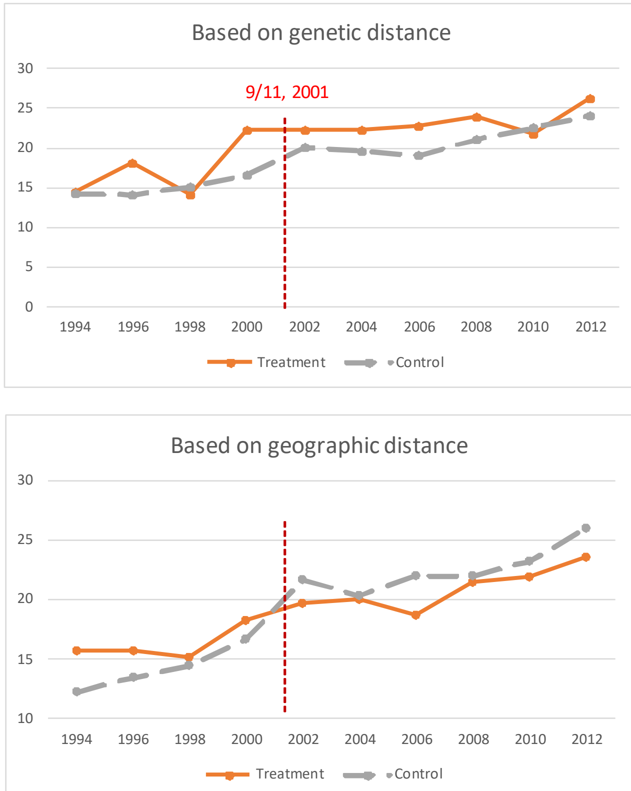
3.6 Conclusion

This study exploits a tragic institutional shock caused by the September 11 terrorist attacks to test the persistence of preference-based discrimination and assess how it manifests itself. This provides a unique setting where I follow Arabic immigrant workers overtime to observe changes in hourly wages and occupation switching behavior. I confirm previous studies that find evidence of labor market discrimination towards Arabic immigrants. More interestingly, I show that the discriminatory effects vary across subgroups of the affected immigrants, where those who look the most Arabic and who are less well educated face an increasingly large drop in wages. In other words, the value of observable traits, such as genetic similarity or educational attainment, increase over time for minority groups. I suggest that occupational discrimination contributes to such heterogeneous effects. While I only provide suggestive evidence in this study owing to the limited sample size, future studies may examine the same topic using existing panel data sets for countries other than the U.S.

Pro- or anti-immigration policies are at the center of heated political campaigns. I hope to inform people of the long-term consequences of anti-immigration policies; in particular, distaste against particular groups may cause a more segregated workforce and where immigrants may increasingly comprise the majority in less complex jobs. This suggests that negative sentiments towards minority groups or anti-immigration policies can overcrowd less complex jobs and increase competition for low-skilled, native workers. Better understanding of the long-term effects of discrimination will help us reduce its negative effects, for members of the majority group as well as the minority groups.

3.7 Figures

Figure 3.1 Hourly earnings by immigrant subgroups



Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 ;
Notes: Sample includes salaried immigrant workers with positive hourly earnings, for all time periods.
Uses winsorized earnings at 5% for hourly earnings.
Top figure groups Immigrants by genetic proximity to the Middle East. Genetic distance measure from Spolaore & Waczairg (2009)
Bottom figure groups immigrants by geographic distance to the Middle East.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

3.8 Tables

Table 3.1 Sample description

	Genetic Distance		Geographic Distance	
	Treatment	Control	Treatment	Control
# of obs	58	78	17	119
Weighted # of obs	90	118	26	183
% White	78.5%	60.8%	67.9%	72.1%
% Black	11.9%	8.7%	3.7%	12.3%
% Hispanic	0.3%	3.5%	0.0%	2.0%
% Asian	9.2%	27.0%	28.4%	13.6%
% Female	44.6%	56.0%	46.4%	49.9%
Years of education (years)	13.2	14.5	13.7	13.7
Mother's education (years)	10.8	11.2	8.8	11.5
Father's eudcation (years)	10.7	12.6	9.3	12.0
AFQT scores	48.6	50.6	33.4	53.1
Mean hourly income ($)	19.6	22.2	22.2	20.3
Median hourly income ($)	16.4	18.2	22.5	18.2
Distance measure from Spolaore & Wacziarg (2009)				
Average Genetic Distance from Middle East	256.7	689.6	272.5	428.4
From the U.S.				
Average Genetic Distance	375.4	817.6	523.7	501.5
Average Linguistic Distance	0.5	0.9	1.0	0.5
Average Religious Distance	0.7	0.7	0.8	0.7
Major industry categories				
Construction	13.1%	8.1%	22.0%	8.4%
Manufacturing	11.8%	3.3%	7.5%	8.5%
Transportatoin, Communication and Public Utilities	10.5%	9.6%	14.9%	9.0%
Wholesale and Retail trade	22.6%	5.6%	14.9%	15.8%
Finance, Insurance and Real Estate	6.9%	7.4%	7.2%	7.0%
Business and Repair services	10.3%	25.3%	7.3%	18.7%
Personal services	0.0%	3.4%	0.0%	1.7%
Entertainment and recreation services	0.0%	3.7%	0.0%	1.9%
Professional and related services	17.3%	31.3%	26.3%	22.3%
Public Administration	7.5%	2.2%	0.0%	6.6%

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012
Notes: Sample includes salaried immigrant workers with positive hourly earnings for all time periods
Excludes immigrants from Central America
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table 3.2 Change in log hourly wages post 9/11 by immigrant subgroups

	Log hourly earnings			
	Genetic Dist		Geographic Dist	
	(1)	(2)	(3)	(4)
Treatment group	-1.104***	-1.228***	-1.166***	-1.020***
	0.093	0.156	0.167	0.211
Average treatment effect				
September 11, 2001 terrorist attacks (9/11)	0.473***		0.348***	
	0.107		0.071	
9/11 x Treatment	-0.243*		-0.118	
	0.131		0.178	
Including year fixed effects (base year: 1999)				
Year 1993 x Treatment		0.420**		-0.097
		0.18		0.171
Year 1995 x Treatment		0.269		-0.004
		0.197		0.193
Year 1997 x Treatment		0.037		-0.408
		0.242		0.315
Year 2001 x Treatment		-0.11		-0.282
		0.176		0.223
Year 2003 x Treatment		-0.086		-0.275
		0.221		0.279
Year 2005 x Treatment		-0.234		-0.142
		0.183		0.206
Year 2007 x Treatment		-0.081		-0.228
		0.189		0.228
Year 2009 x Treatment		-0.14		-0.472
		0.218		0.325
Year 2011 x Treatment		-0.1		-0.237
		0.199		0.316
Individual fixed effects	✓	✓	✓	✓
Year fixed effects		✓		✓
Constant	2.454***	2.577***	2.517***	2.539***
Number of individuals	157	157	157	157

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes in log hourly wages post 9/11; use winsorized hourly earnings at the 5% level.
Treatment group defined using genetic distance for columns (1) and (2); geographic distance for columns (3) and (4)
Columns (1) and (3) show average treatment effects; Columns (2) and (4) include year fixed effects.
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table 3.3 Heterogeneous effects in changes in log hourly wages post 9/11

	Log hourly earnings			
	Genetic Dist		Geographic Dist	
	(1)	(2)	(3)	(4)
Average treatment effect				
September 11 terrorist attacks (9/11)	0.393		0.064	
	0.568		0.484	
9/11 x Treatment group	-1.084		-1.217*	
	0.689		0.717	
9/11 x Years of education	0.003		0.020	
	0.041		0.036	
9/11 x Treatment group x Years of education	0.068		0.079*	
	0.049		0.048	
Including year fixed effects (base year: 1999)				
Year 1993 x Treatment group		-0.156		-0.874
		0.878		0.811
Year 1995 x Treatment group		-1.258		-1.71
		1.149		1.065
Year 1997 x Treatment group		0.081		0.236
		1.07		1.214
Year 2001 x Treatment group		-1.332		-2.073*
		0.894		1.167
Year 2003 x Treatment group		-2.071**		-2.897**
		1.021		1.41
Year 2005 x Treatment group		-1.125		-1.675**
		0.83		0.736
Year 2007 x Treatment group		-1.491*		-1.863***
		0.832		0.692
Year 2009 x Treatment group		-1.814**		-2.338**
		0.897		0.952
Year 2011 x Treatment group		-1.796		-2.384*
		1.105		1.244
Year 1993 x Treatment group x Years of education		0.038		0.056
		0.063		0.058
Year 1995 x Treatment group x Years of education		0.106		0.124
		0.084		0.079
Year 1997 x Treatment group x Years of education		-0.01		-0.058
		0.076		0.1
Year 2001 x Treatment group x Years of education		0.091*		0.129*
		0.055		0.072
Year 2003 x Treatment group x Years of education		0.144**		0.187**
		0.07		0.091
Year 2005 x Treatment group x Years of education		0.067		0.109**
		0.056		0.048
Year 2007 x Treatment group x Years of education		0.107**		0.117***
		0.052		0.044
Year 2009 x Treatment group x Years of education		0.127**		0.134**
		0.054		0.056
Year 2011 x Treatment group x Years of education		0.128*		0.153*
		0.073		0.078
Individual fixed effects	✓	✓	✓	✓
Year fixed effects		✓		✓
Year fixed effects x education		✓		✓
Constant	0.001	1.632	2.345	2.859**
Number of individuals	157	157	157	157

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes in log hourly wages post 9/11; use winsorized hourly earnings at the 5% level.
Treatment group defined using genetic distance for columns (1) and (2); geographic distance for columns (3) and (4)
Columns (1) and (3) show average treatment effects; Columns (2) and (4) include year fixed effects.
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table 3.4 Likelihood of occupation transition from 2001

	Occupation transition			
	Genetic Dist		Geographic Dist	
	(1)	(2)	(3)	(4)
Including year fixed effects (base year: 2001)				
Year 2003 x Treatment group	0.058	-0.204	0.196	0.555
	0.129	0.594	0.183	0.792
Year 2005 x Treatment group	0.039	0.476	-0.008	-0.199
	0.131	0.615	0.157	0.703
Year 2007 x Treatment group	-0.064	1.118*	0.011	1.307**
	0.133	0.599	0.16	0.566
Year 2009 x Treatment group	-0.014	-0.039	0.117	-0.276
	0.12	0.606	0.165	0.66
Year 2011 x Treatment group	0.064	-0.492	-0.041	-0.697
	0.13	0.614	0.182	0.7
Year 2003 x Treatment group x Years of education		0.018		-0.027
		0.04		0.052
Year 2005 x Treatment group x Years of education		-0.033		0.014
		0.042		0.049
Year 2007 x Treatment group x Years of education		-0.087**		-0.095**
		0.042		0.038
Year 2009 x Treatment group x Years of education		0.001		0.03
		0.044		0.046
Year 2011 x Treatment group x Years of education		0.04		0.049
		0.043		0.049
Individual fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Year fixed effects x education		✓		✓
Constant	-0.091	-1.785	-0.090***	-2.086
Number of individuals	149	149	149	149

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine occupation transition patterns post 9/11
Treatment group defined using genetic distance for columns (1) and (2); geographic distance for columns (3) and (4)
Columns (1) and (3) show average treatment effects; Columns (2) and (4) include year fixed effects.
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table 3.5 Occupation characteristics

	Occupation complexity scores			
	Genetic Dist		Geographic Dist	
	(1)	(2)	(3)	(4)
Including year fixed effects (base year: 2001)				
Year 2003 x Treatment group	-2.78	-41.989**	6.248	-5.539
	4.314	18.766	4.956	18.93
Year 2005 x Treatment group	-4.524	-52.025**	4.411	-23.36
	4.877	20.331	5.366	20.057
Year 2007 x Treatment group	-4.124	-53.291***	3.99	-16.648
	4.69	19.812	5.103	20.186
Year 2009 x Treatment group	-7.112	-45.196**	-1.344	-16.552
	4.86	22.085	5.737	21.224
Year 2011 x Treatment group	-3.743	-44.136**	3.898	-12.744
	5.084	22.216	5.051	21.137
Year 2003 x Treatment group x Years of education		2.861**		0.505
		1.215		1.301
Year 2005 x Treatment group x Years of education		3.500***		1.626
		1.28		1.335
Year 2007 x Treatment group x Years of education		3.617***		1.101
		1.243		1.338
Year 2009 x Treatment group x Years of education		2.800*		0.702
		1.432		1.371
Year 2011 x Treatment group x Years of education		2.974**		0.814
		1.386		1.388
Individual fixed effects	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Year fixed effects x education		✓		✓
Constant	44.994***	43.936	48.697***	55.066***
Number of individuals	147	147	147	147

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes occupation characteristics post 9/11.
Treatment group defined using genetic distance for columns (1) and (2); geographic distance for columns (3) and (4).
Columns (1) amd (3) include year fixed effects; Columns (2) and (4) include year x education fixed effects as well.
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Bibliography

Abramitzky R, Boustan LP, Eriksson K (2014) A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. J. Polit. Econ. 122(3):467-506.

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

Arcidiacono P, Bayer P, Hizmo A (2010) Beyond signaling and human capital: education and the revelation of ability. Am. Econ. J. Appl. Econ. 2(4):76-104.

Arrow KJ (1973) The theory of discrimination. Ashenfelter O, Rees A, eds, Discrimination in Labor Markets (Princeton University Press, Princeton), 3-33.

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

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

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

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

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

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

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. Am. Econ. Rev. 94(4):991-1013.

Black S, Brainerd E (2004) Importing Equality? The Effects of Globalization on Gender Discrimination. Ind. and Lab. Relat. Rev. 57(4): 540 - 559.

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

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

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

Borjas GJ, Doran KB, Shen Y (2017) Ethnic complementarities after the opening of china: how Chinese graduate students affected the productivity of their advisors. J. Hum. Resour. forthcoming.

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The History and Geography of Human Genes. (Princeton University Press, Princeton, NJ).

Charles KK, Guryan J (2008) Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination. J. Polit. Econ. 116(5): 773-809.

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

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

Coate S, Loury G (1993) Will Affirmative-Action Policies Eliminate Negative Stereotypes? Am. Econ. Rev. 83(5): 1220-1240.

Collier P (2001) Implications of ethnic diversity. Econ. Pol.: a European Forum. 0:127-155.

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

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

Fairlie RW, Lofstrom M (2014) Forthcoming Chiswick B, and Miller P, eds. Handbook on the Economics of International Migration (Elsevier, Amsterdam).

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

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

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

Flood S, King M, Ruggles S, Warren R (2015) Integrated Public Use Microdata Series, Current Population Survey: Version 4.0 [Machine-readable dataset]. Minneapolis: University of Minnesota.

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

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

Glaeser E, Kerr WR, Ponzetto GAM (2010) Clusters of entrepreneurship. J. Urb. Econ. 67(1): 150-168.

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

Goldsmith A, Hamilton D, Darity W (2006) Shades of Discrimination: Skin Tone and Wages. Am. Econ. Rev. 96(2): 242-245.

Gould ED, Klor EF (2016) The Long-run Effects of 9/11: Terrorism, Backlash, and the Assimilation of Muslim Immigrants in the West. Econ. J. 126 (November): 2064-2114.

Haltiwanger J, Jarmin RS, Miranda J (2013) Who creates jobs? Small versus large versus young. Rev. Econ. Stat. 95(2): 347-361.

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

Hanson GH, Slaughter M (2016) High-skilled immigration and the rise of STEM occupations in U.S. employment. NBER Working Paper No. 22623, National Bureau of Economic Research, Cambridge.

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

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

Hoffman M, Kahn L, Li D (2015) Discretion in hiring. Working Paper 16-055, Harvard Business School, Boston.

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.

Human Rights Watch, USA (2002) 'We Are Not the Enemy': Hate Crimes Against Arabs, Muslims, and those Perceived to be Arab or Muslim after September 11, 14 November 2002, G1406.

Hurst E, Pugsley BW (2015) Wealth, tastes, and entrepreneurial choice. Measuring Entrepreneurial Businesses: Current Knowledge and Challenges (University of Chicago Press, Chicago).

Imai S, Stacey D, Warman C (2014) From engineer to taxi driver? Language proficiency and the occupational skills of immigrants. Working Paper 1275, Queen's Economics Department.

Joshi AM, Lahiri N (2014) Language friction and partner selection in cross-border R&D Alliance Formation. J. Int. Bus. Stud. 46(2):123-152.

Kacperczyk AJ (2013) Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry. Organ. Sci. 24(3):664-683.

Kausha N, Kaetsner R and Reimer C (2007) Labor Market Effects of September 11th on Arab and Muslim Residents of the United States. J. Hum. Resour. 42(2): 275-308.

Kerr SP, Kerr WR, Lincoln WF (2015) Skilled immigration and the employment structures of US firms. J. Lab. Econ. 33(S1):S147-S186.

Kerr W, Lincoln W (2010) The supply side of innovation: H-1B Visa reforms and U.S. ethnic invention. J. Lab. Econ. 28(3):473-508.

Kerr W, Mandroff M (2015) Social networks, ethnicity, and entrepreneurship. NBER Working Paper No. 21597, National Bureau of Economic Research, Cambridge.

Kulchina E (2016) A path to value creation for foreign entrepreneurs. Strateg. Manag. J. 37(7):1240-1262.

Lang K (1986) A language theory of discrimination. Q.J. Econ. 101(2):363-382.

Lang K, Manove M (2011) Education and Labor Market Discrimination. Am. Econ. Rev. 101(4): 1467-1496.

Lazear E (2005) Entrepreneurship. J. Lab. Econ. 23(4):649-680.

Levine R, Levkov A, Rubinstein Y (2012) Bank Deregulation and Racial Inequality in America. Crit. Fin. Rev. 3(1): 1-48.

Levine R, Rubinstein Y (2016) Smart and illicit: who becomes an entrepreneur and do they earn more? Q. J. Econ. (in press).

Lewis E (2011) Immigrant-native substitutability: The role of language ability. Card D, Raphael S, eds. Immigration, Poverty, and Socioeconomic Inequality (Russell Sage Foundation, New York), 60-97.

Light I (1972) Ethnic Enterprise in America (University of California Press, Berkeley).

Light I (1979) Disadvantaged minorities in self-employment. Int. J. Comp. Sociol. 20(1-2):31-45.

Lofstrom M (2002) Labor market assimilation and the self-employment decision of immigrant entrepreneurs. J. Pop. Econ. 15(1):83-114.

Lundberg SJ, Startz R (1983) Private discrimination and social intervention in competitive labor markets. Am. Econ. Rev. 73(3): 340-47.

Mecham RQ, Fearon J, Laitin D (2006) Religious Classification and Data on Shares of Major World Religions (Stanford University).

Mittelstadt M, Speaker B, Meissner D, Chishti M (2011) Through the Prism of National Security: Major Immigration Policy and Program Changes in the Decade since 9/11. Migration Policy Institute.

Montalvo JG, Reynal-Querol M (2005) Ethnic polarization, potential conflict, and civil wars. Am. Econ. Rev. 95(3):796-816.

Morgan J, Várdy F (2009) Diversity in the workplace. Am. Econ. Rev. 99(1):472-485.

Nanda R, Sørensen JB (2010) Workplace peers and entrepreneurship. Manag. Sci. 56(7):1116-1126.

Nanda R, Khanna T (2010) Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software industry. J. Econ. Manag. Strateg. 19(4):991-1012.

Neal D, Johnson W (1996) The Role of Premarket Factors in Black-White Wage Differences. J. Polit. Econ. 104(5): 869-895.

Oreopoulos P (2011) Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. Am. Econ. J. Econ. Policy 3(4):148-171.

Ottaviano GP, Peri G. (2006) The economic value of cultural diversity: evidence from U.S. cities. J. Econ. Geo. 6(1): 9-44.

Peri G, Sparber C (2009) Task specialization, immigration, and wages. Am. Econ. J. Appl. Econ. 1(3):135-169.

Petersen T, Saporta I (2004) The opportunity structure for discrimination. Am. J. Sociol. 109(4):852-901.

Phelps, E (1972) The statistical theory of racism and sexism. Am. Econ. Rev. 62(4):659-661.

Phillips KW, Liljenquist KA, Neale MA (2009) Is the pain worth the gain? The advantages and liabilities of agreeing with socially distinct newcomers. Personal. Soc. Psychol. Bull. 35(3):336-350.

Portes A, Zhou M (1996). Self-employment and the earnings of immigrants. Am. Soc. Rev. 61(2):219-230.

Ratcliffe A, Von Hinke Kessler Scholder S (2015) The London Bombings and Racial Prejudice: Evidence from the Housing and Labor Market. Econ. Inq. 53(1): 276-293.

Ridlon R, Shin J (2013) Favoring the winner or loser in repeated contests. Mark. Sci. 32(5):768-785.

Rivera LA (2012) Hiring as cultural matching: The case of elite professional service firms. Am. Sociol. Rev. 77(6):999-1022.

Ruggles S, Genadek K, Goeken R, Grover J, Sobek M (2015) Integrated Public Use Microdata Series: Version 6.0 [Machine-readable dataset]. Minneapolis: University of Minnesota

Sauermann H, Cohen WM (2010) What makes them tick? Employee motives and firm innovation. Manag. Sci. 56(12):2134-2153.

Siegel JI, Pyun L, Cheon BY (2014) Multinational firms, labor market discrimination, and the capture of competitive advantage by exploiting the social divide. Working Paper 11-011, Harvard Business School, Boston.

Slezkine Y (2004) The Jewish Century (Princeton University Press, Princeton)

Spolaore E, Wacziarg R (2009) The diffusion of development. Q. J. Econ. 124(2):469-529.

Stenard BS, Sauermann H (2016) Educational mismatch, work outcomes, and entry into entrepreneurship. Organ. Sci. 27(4):801-824.

Van den Steen E (2005) Organizational beliefs and managerial vision. J. Law Econ. Organ. 21(1):256-283.

Appendix for Chapter 1

Table A.1 Selection into self-employment for immigrant subgroups

	Self-employment (vs Salaried)		
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**	**Cultural Dist CD**
	(1)	(2)	(3)
Panel A: Immigrants who came after age 25			
Education (vs Grade School)			
High School	0.002	0.002	0.003***
	0.002	0.002	0.001
College	0.007**	0.007**	0.009***
	0.003	0.003	0.002
A. 1gImm / LD / CD	0.036	0.027	0.101
	0.030	0.032	0.065
B. (1gImm / LD / CD) x High School	0.042***	0.047***	0.065***
	0.007	0.009	0.016
C. (1gImm / LD / CD) x College	0.062***	0.074***	0.092***
	0.011	0.011	0.022
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	4776685	4776685	4695102
P-values comparing coefficients			
B = C	0.083	0.031	0.150
Panel B: Immigrants who spent more than 10 years in the U.S.			
Education (vs Grade School)			
High School	0.001	0.001	0.003**
	0.002	0.002	0.001
College	0.006	0.006	0.009***
	0.004	0.004	0.002
A. 1gImm / LD / CD	-0.020	-0.018	0.040
	0.013	0.017	0.047
B. (1gImm / LD / CD) x High School	0.034***	0.038***	0.059***
	0.007	0.008	0.017
C. (1gImm / LD / CD) x College	0.046***	0.052***	0.073***
	0.009	0.010	0.025
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	5075506	5075506	4915746
P-values comparing coefficients			
B = C	0.171	0.15	0.482

Source: American Community Survey, 2005 - 2012
Notes: Replicate 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.

Table A.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	Cultural Dist CD
	(1)	(2)	(3)
Education (vs Grade School)			
High School	0.013	0.000	0.000
	0.013	0.000	0.000
College	0.005**	0.005*	0.006***
	0.002	0.003	0.002
A. Noisy signal (1gImm / LD / CD)	0.006	0.017	0.062
	0.004	0.020	0.050
B. (1gImm / LD / CD) x High School	0.027***	0.035***	0.057***
	0.006	0.009	0.015
C. (1gImm / LD / CD) x College	0.033***	0.044***	0.059***
	0.007	0.009	0.016
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	583189	583189	562874
P-values comparing coefficients			
B = C	0.362	0.362	0.879

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, column (2) uses linguistic distance, column (3) uses cultural 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.

Table A.3 Selection using standard error clustered at education x country level

	Self-employment (vs Salaried)		
Measure of noisy signal:	**Immig. Status 1gImm**	**Linguistic Dist LD**	**Cultural Dist CD**
	(1)	(2)	(3)
Education (vs Grade School)			
High School	0.000	0.000	0.002
	0.003	0.003	0.004
College	0.004	0.004	0.008
	0.007	0.007	0.008
A. Noisy signal (1gImm / LD / CD)	-0.009	-0.008	0.047
	0.011	0.013	0.033
B. (1gImm / LD / CD) x High School	0.029**	0.032**	0.048*
	0.014	0.016	0.026
C. (1gImm / LD / CD) x College	0.044***	0.051***	0.070***
	0.010	0.011	0.020
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Number of Observations	5280414	5280414	5069458
P-values comparing coefficients			
B = C	0.206	0.189	0.344

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, column (2) uses linguistic distance, column (3) uses cultural distance as measure 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.

Table A.4 Selection into self-employment by education × cultural distance categories

	Self-employment (vs Salaried)			
Cultural Distance category:	**<0.5**	**0.5 - 0.65**	**0.65 - 0.75**	**0.75 - 1**
	(1)	(2)	(3)	(4)
Education (vs Grade School)				
High School	0.002	0.004***	0.004***	0.003***
	0.003	0.000	0.000	0.001
College	0.007	0.011***	0.011***	0.010***
	0.005	0.000	0.000	0.001
A. Cultural Distance	0.075	0.040	0.192***	0.079
	0.052	0.066	0.012	0.053
B. Cultural Distance x High School	0.033***	0.013	-0.003	0.036**
	0.009	0.041	0.017	0.016
C. Cultural Distance x College	0.090***	0.003	-0.039	0.042
	0.012	0.060	0.032	0.030
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Number of Observations	4814248	4531171	4562292	4565873
P-values comparing coefficients				
B = C	0.000	0.341	0.240	0.530

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; *, **, 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.

Table A.5 Selection by education × 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; *, **, 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.

Table A.6 Selection at different age of immigration

	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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Table A.7 Selection into self-employment among immigrants not residing in enclaves

	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
	0.02	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; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using the population weights provided.

Appendix for Chapter 2

Table B.1 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

Appendix for Chapter 3

Table C.1 Changes in log hourly wages post 9/11, using different distance measures

	Log hourly earnings			
	Proximity to ME	Distance from U.S.		
	Genetic	Genetic	Linguistic	Religious
	(1)	(2)	(3)	(4)
Respective distance measures				
(Placebo) Treatment group	-1.104***	-1.202***	-0.048	-1.172***
	0.093	0.099	0.098	0.147
September 11 terrorist attacks (911)	0.473***	0.390***	0.324***	0.346***
	0.107	0.099	0.081	0.073
911 x Treatment	-0.243*	-0.103	0.005	-0.111
	0.131	0.131	0.137	0.160
Individual fixed effects	✓	✓	✓	✓
Constant	2.454***	2.496***	1.256***	2.517***
Number of individuals	157	157	157	157

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes in log hourly wages post 9/11; use winsorized hourly earnings at the 5% level.
Distance measures from Spolaore & Wacziarg (2009); Columns (2), (3) and (4) represent respective distance from the US
Sample includes salaried immigrant workers in the treated group with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table C.2 Heterogeneous effects by AFQT scores

	Log hourly earnings			
	Genetic Dist		Geographic Dist	
	(1)	(2)	(3)	(4)
Average treatment effect				
September 11 terrorist attacks (9/11)	0.593***		0.585***	
	0.152		0.113	
9/11 x Treatment group	-0.331		-0.572**	
	0.223		0.257	
9/11 x AFQT scores	-0.003		-0.004	
	0.004		0.003	
9/11 x Treatment group x AFQT scores	0.002		0.010*	
	0.005		0.005	
Including year fixed effects (base year: 1999)				
Year 1993 x Treatment group		-0.109		-0.063
		0.269		0.248
Year 1995 x Treatment group		-0.208		-0.182
		0.269		0.276
Year 1997 x Treatment group		0.386		-0.368
		0.403		0.343
Year 2001 x Treatment group		-0.27		-0.541
		0.341		0.403
Year 2003 x Treatment group		-0.711*		-1.070**
		0.413		0.477
Year 2005 x Treatment group		-0.247		-0.728**
		0.298		0.308
Year 2007 x Treatment group		-0.198		-0.710*
		0.333		0.363
Year 2009 x Treatment group		-0.405		-1.055**
		0.434		0.532
Year 2011 x Treatment group		-0.28		-0.761
		0.426		0.553
Year 1993 x Treatment group x AFQT scores		0.010*		-0.002
		0.005		0.005
Year 1995 x Treatment group x AFQT scores		0.009		0.006
		0.006		0.006
Year 1997 x Treatment group x AFQT scores		-0.008		-0.016
		0.007		0.011
Year 2001 x Treatment group x AFQT scores		0.004		0.006
		0.004		0.006
Year 2003 x Treatment group x AFQT scores		0.013		0.017**
		0.008		0.009
Year 2005 x Treatment group x AFQT scores		0		0.011
		0.005		0.007
Year 2007 x Treatment group x AFQT scores		0.003		0.01
		0.006		0.008
Year 2009 x Treatment group x AFQT scores		0.006		0.013
		0.007		0.008
Year 2011 x Treatment group x AFQT scores		0.004		0.012
		0.007		0.009
Individual fixed effects	✓	✓	✓	✓
Year fixed effects		✓		✓
Year fixed effects x education		✓		✓
Constant	2.420***	2.400***	2.423***	2.373***
Number of individuals	157	157	157	157

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes in log hourly wages post 9/11; use winsorized hourly earnings at the 5% level.
Treatment group defined using genetic distance for columns (1) and (2); geographic distance for columns (3) and (4)
Columns (1) and (3) show average treatment effects; Columns (2) and (4) include year fixed effects.
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.

Table C.3 Heterogeneous outcomes with industry fixed effects

	Log hourly earnings					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment group	-1.104***	0.837***	0.834***	2.307*	3.795***	3.295**
	0.093	0.079	0.171	1.322	1.420	1.419
September 11 terrorist attacks (9/11)	0.473***	0.600***	0.261	0.393	0.580	0.034
	0.107	0.123	0.412	0.568	0.708	0.890
9/11 x Treatment	-0.243*	-0.362**	-0.363***	-1.084	-1.505*	-1.483**
	0.131	0.146	0.123	0.689	0.787	0.710
Years of education (educ)				0.154**	0.117	0.106*
				0.066	0.075	0.064
9/11 x educ				0.003	-0.001	0.022
				0.041	0.052	0.052
Treatment x educ				-0.228**	-0.205**	-0.164
				0.093	0.099	0.107
9/11 x Treatment x educ				0.068	0.090	0.086*
				0.049	0.057	0.051
Individual fixed effects	✓	✓	✓	✓	✓	✓
Major industry fixed effects		✓	✓		✓	✓
Major industry x 9/11 fixed effects			✓			✓
Constant	2.454***	2.375***	2.300***	0.001	0.524	0.604
Number of individuals	157	157	157	157	157	157

Source: National Longitudinal Survey of Youth 1979 (NLSY79), 1994 - 2012 (Biennial labor market outcomes for 1993 - 2011)
Notes: Examine changes in log hourly wages post 9/11; use winsorized hourly earnings at the 5% level.
Columns (2) and (4) include industry fixed effects, columns (3) and (6) include industry fixed effects interacted with 9/11 as well.
Major industry categories in year 2001 used for industry fixed effects.
Grouping based on Genetic distance from the Middle East based on Spolaore and Wacizarg (2009).
Sample includes salaried immigrant workers with positive hourly earnings; includes individual fixed effects.
Reported Standard Errors are clustered at the individual level; *, **, and *** indicate significant at 10%, 5% and 1%, respectively.
Calcualtions weighted using customized weights provided by the Bureau of Labor Statistics.