Immigration and Invention: Does Language Matter?

Kirk Doran, Chungeun Yoon*

ABSTRACT

Economists have long noted that ethnolinguistically diverse immigrants might have a particularly large impact on innovation and creativity. On the other hand, if innovation depends on communication, and communication depends on a common language, then it should be immigrants who speak that language who have the largest impact on innovation. In this paper, we make use of unique features of the 1920s U.S. immigration quotas that discouraged immigration of both english and non-english speakers to cities with both relatively many and relatively few pre-existing english speakers. This variation allows us to show that the effect of immigration on innovation reported in (Doran and Yoon, 2018) is strongest when immigrants and the pre-existing population share a common language. It appears that communication is an important channel through which immigration may affect innovation.

JEL classification: J61, N32, O31, Z13

*University of Notre Dame. All errors are our own.

1 Introduction

Economists have long noted several ways that immigration could affect innovation. Highly-skilled immigrants may innovate directly, while low-skilled immigrants could affect the scale of production, encouraging labor-complementary inventions, and discouraging strongly-labor-saving inventions (Acemoglu, 2010). But the literature on immigration and innovation has failed to address the potential importance of one of the most obvious differences between immigrants and natives: language differences. On the one hand, immigration may have a larger impact on innovation when there is a language similarity between the immigrants and natives. Strongly labor-complementary inventions may be incentivized more by a large homogeneous workforce that can work easily together rather than by heterogeneous labor inputs that have trouble communicating with each other. On the other hand, immigration may have a larger impact on innovation when there is a language dissimilarity between the immigrants and natives. After all, a large literature explores the possibility that a diverse ethnolinguistic mix "brings about variety in abilities, experiences, and cultures that may be productive and may lead to innovation and creativity" (Alesina and La Ferrara, 2005).

When do immigrants have the biggest impact on innovation: when they share a language with natives, or when they do not? In order to answer this question, we need a setting in which the language of immigrants varies independently of the language mix of the people already living in the locations the immigrants are immigrating to. This is difficult, because traditional shift-share style instruments build on exactly that variation in immigration that is correlated with ethnolinguistic variation in the pre-existing population across locations. In this paper, we make use of subtle features of 1920s U.S. immigration quotas that ended up discouraging immigration of native english speakers to cities with a relatively low pre-existing population of english speakers, as well as immigration of non-english speakers to cities with a relatively high pre-existing population of english speakers. These "off-diagonal" terms allow us to estimate the effects of english-speaking and non-english speaking immigrants to cities with both relatively many and relatively few english speakers.

The results are striking. Native-born inventors in cities with a relatively high population of english speakers are most greatly affected by immigrants from english-speaking countries. At the same time, native-born inventors in cities with a relatively low population of english speakers are most affected by immigrants from non-english speaking countries.

It is important to note that, as (Doran and Yoon, 2018) explains, the effect of these low-skilled immigrants on native inventors is through a change in the scale of production that incentivizes strongly labor complementary inventions (Acemoglu, 2010). The role of communication, therefore, is happening in the context of a low-skilled workforce, not in the context of highly skilled innovators themselves. It appears that an increase in the scale of homogeneous labor inputs incentivized the strongly labor-complementary inventions of the 1920s. Indeed, for production inventions that depend on maximizing the potential for the division of labor, the benefits of a large homogeneous low-skilled workforce that can communicate well, rather than multiple heterogeneous low-skilled workforces that can not communicate with each other, are not surprising. In the low-skilled immigration context, therefore, an increase in immigrant linguistic diversity may mute the benefits to production scale and hence invention. In contrast, most of the benefits of an ethnolinguistically diverse workforce for innovation have been described in terms of new abilities, ideas, and experiences (Alesina and La Ferrara, 2005), all of which plausibly could help high skilled natives and those they directly communicate with to innovate. Therefore, the theorized benefits of immigrant linguistic diversity for innovation may be more likely in the context of highly skilled immigrants than in the context of low-skilled immigrants.

The paper is organized as follows. In Section II, we review the literature on the 1920s quotas, and explain where our results fit in the context of that literature. In Section III, we introduce the data set, referring especially to (Doran and Yoon, 2018). In Section IV, we introduce our empirical strategy and estimating equations. In Section V, we describe our results. In Section VI, we conclude.

2 Existing Economics Literature on the Quotas

In the last several years, a total of seven papers have emerged studying the economic impacts of the 1920s U.S. immigration quotas. These papers have been written almost simultaneously by separate teams of authors, with subtle differences in the implementation of the identification strategies, and without a planned consistency. Here we argue that in fact these seven papers tell a largely consistent history, in which the reported economic impacts of the quotas correspond with those predicted by models such as (Borjas, 1987), (Acemoglu, 2010), and (Tabellini, 2018). In particular, it appears that these quotas: (1) reduced immigration from some sources but not others; (2) reduced immigration to some locations but not others; (3) induced differential wage changes among natives in affected locations; (4) induced a native migration response to affected locations that was less than one for one with the immigration reductions; (5) decreased the scale and mechanization of production in affected locations; and (6) decreased natives' inventions in affected locations, especially those inventions relevant for industries that lost a large number of immigrant workers. This set of results is not only consistent with itself, but is also consistent with the new results reported here, in which the reduction in inventions was strongest when pre-existing workers and immigrants shared a common language.

In this section, we review the results of this existing literature, summarizing the results and comparing them to models such as (Borjas, 1987), (Acemoglu, 2010), and the model in Appendix B of Tabellini (2018).

One of the most important papers in this literature is "Immigration in American Economic History" (Abramitzky and Boustan (2017)). Abramitzky and Boustan (2017) review the literature on historical and contemporary immigration. They focus on three major questions in the economics of immigration. First, the paper questions whether immigrants are positively or negatively selected from their home countries over time. Second, they explore how immigrants assimilate into the US. Third, they examine the effects of immigration on the economy, especially native employment and wages. In particular, they cover the two main eras of mass immigration–the Age of Mass Migration from Europe (1850-1920), an era of unrestricted migration, and a recent period of constrained mass migration from Asia and Latin America (1965-present).

First, they find that migrant selection was mixed in the past (with some migrants being positively selected and others being negatively selected from their home countries), while migrants are positively selected in the present. Specifically, migrant selection during the Age of Mass Migration is consistent with a Roy model (Roy, 1951), as developed by (Borjas, 1987). The Roy model would predict positive selection from northern and western Europe and negative selection from southern and eastern Europe, with differences in productive skills of migrants and income equality across sending countries. Historical evidence on income distribution supports their argument. Income distribution in western European countries was similar with that of the US at that time while income distribution in the European periphery was less equal than that of the US. Consistent with the model, historical evidence suggests that low-skilled workers from southern and eastern Europe immigrated to the US and that they are negatively selected. The positive selection of immigrants today can be explained by both the increase in income inequality in the US (as the model would predict) and the increasing selectivity of US immigration policy, which would favor high-skilled immigration.

Second, they find that assimilation of immigrants into US economy is not consistent with the stereotypical "American Dream", whereby poor immigrants work hard and eventually become rich. During periods of mass migration, immigrants did not catch up with US natives in the past and they do not do so today, because immigrants start behind natives, and their occupational upgrading and earnings grow at a similar pace to that of US natives over time. Although immigrants experienced earnings convergence to some extent, they do not catch up with US natives in the labor market. However, these gaps diminish across generations because many children of immigrants are educated and grow up in US. The authors find some mixed results on assimilation of immigrants, thus claiming that methods and data matter, especially in the contest of a substantial heterogeneity in skills and earnings of immigrants from sending countries.

Third, they argue that immigrants do not in general have negative effects on the US economy, but they do decrease the wages of some natives to some degree during the two periods. In particular, immigrants during the Age of Mass Migration were more substitutable with natives in agriculture and manufacturing, and therefore there was some effect of immigration on native wages. They also find that immigration in the past contributed to the growth in large factories for mass production. In addition, unskilled immigrants and assembly-line machinery were complementary at that time. Whereas investments for mass production took place in immigrant-receiving areas in the past, skill-biased investments such as computerization today do not increase at higher rates.

Another key paper is "Closing Heaven's Door: Evidence from the 1920s U.S. Immigration Quota Acts" (Ager and Hansen (2017)). Ager and Hansen (2017) examine the effects of the 1920s immigration quotas on the economic outcomes in areas more likely to be exposed to the quotas. First, the change in the foreign-born share and population growth are examined before and after the policy shock. Second, they explore which groups in the labor market benefited from the policy change by increasing their earnings. Third, the effect of the quotas on productivity in the manufacturing sector is examined.

Ager and Hansen exploit several sources of variation. The quotas restricted immigration from southern and eastern Europe more than from northern and western Europe. For instance, the number of Italian immigrants decreased by over 70 percent before and after the 1921 quota while Swedish immigrants increased. The missing immigrants for each country are estimated using the different number of quotas by country and expected immigrants without the quotas predicted by previous immigrant inflows before the quotas. Another source of variation is that the number of foreign-born population by country varies across the local areas in the US. They employ the differences-in-differences strategy using three different samples: US Census at the county level (1900-1940); the one percent sample of US Census microdata (1900-1940); and US Census of Manufactures at the city level.

The first main finding is that the areas with a large decline in incoming immigrants due to the quotas experienced a decrease in the foreign-born share and lower population growth. Specifically, one additional missing immigrant per-100-inhabitants-per-year led to a decline in the foreign-born share by 1.6 percentage points and a decrease in the 10-year population growth rate by 6.7 percentage points at the county level. In addition, the corresponding decrease in marriage rates due to the quotas could contribute to the decline in population growth. Second, they show that the quotas have a significant effect on the earnings of native workers. Natives in counties exposed to the quotas were more likely to change to lower-wage occupations, though the effect varies by gender and race. In particular, white workers experienced earning losses while black workers benefited from the quotas. Earnings of white female workers were not affected, while black female workers significantly gained the benefits. These findings suggest that immigrant workers during the 1920s had a higher elasticity of substitution to black native workers. Third, they find that labor productivity in manufacturing at the city level declined under the quotas. They find no changes in the capital intensity or the capital-output ratio thus speculating that the lower productivity may be caused by agglomeration externalities and the degree of substitutability.

Overall, they find the significant effects of the immigration quotas during the 1920s on the economic consequences in the US. The affected areas experienced a decline in population growth and lower productivity in manufacturing sector. Especially, there were winners and losers caused by the immigration policy. Black workers who were more substitutable with immigrants at that time benefited from the quota restrictions, while white workers loss their earnings. However, they do not look at the labor market outcomes of immigrants who already came to the US prior to the quotas and new incoming immigrants under the quotas. It is possible that they benefited more from the quotas. Further, the effects of the quotas could vary across immigrants because the quotas restricted immigration across the sending countries. A difference in skill levels across them might affect the labor productivity in manufacturing sector.

A third important paper in this literature is Tabellini (2018). This paper makes two main additional contributions above and beyond the points already made in the literature described above. First, Tabellini (2018) introduces a notion of linguistic distance adapted from Chiswick and Miller (2005). The results show that the impact of immigration is tied closely to the linguistic distance of the source country language compared to English. The second main contribution is to introduce a model (in online Appendix B of Tabellini (2018)) that makes the following predictions: (1) (unskilled) immigration favors capital accumulation in the unskilled sector; (2) "immigration has a positive and unambiguous effect on high skilled wages"; and (3) immigration has an ambiguous effect on low skilled wages. This theoretical framework is consistent with Tabellini (2018) by construction, but it is clearly consistent with the results of Ager and Hansen (2017) as well.

A fourth paper in this literature is Doran and Yoon (2018). This paper addresses the question of how mass migration affects innovation. In particular, the paper questions whether low-skilled immigrants could influence innovations through labor-complementary inventions or labor-saving inventions. The results show that incumbent inventors in cities exposed to fewer low-skilled immigration inflows due to the 1920s quotas applied for fewer number of patents. To be specific, inventors living in quota-exposed cities that experienced every ten percent reduction in new immigrants reduced their patent applications by 0.5 percent per year. Further, the effect of quotas on patents is driven by fewer patent applications relevant for the quota-exposed industries that lost immigrant workers. Our findings suggest that the mechanism in Acemoglue (2010) is at work through labor-complementary inventions during the immigration quotas that reduced low-skilled immigrants from southern and eastern Europe in quota-exposed cities and industries, thus leading to fewer inventions overall.

A paper which addresses whether return migration was affected by the quotas is "Birds of passage: Return migration, self-selection and immigration quotas" (Ward (2017)). Ward (2017) explores the intentions of migrants to stay in the US or return to Europe during the 1920s. In particular, the author examines the effect of the quotas on whether migrants left the US and returned to Europe as planned or unexpectedly. This paper complements Greenwood and Ward (2015) in several ways. First, this paper mainly contributes to work on return migration in that he collects the novel data on whether migrants at arrival wanted to return home or stay in the US. Instead, Greenwood and Ward (2015) focus on actual return migration rather than migrants' return intentions. Second, this paper compares return migrants to migrants who stay in the US, thus examining the selection of return migrants. The author explores whether return migrants were positively or negatively selected while Greenwood and Ward (2015) estimate the rate of return migration. Finally, this paper complements the previous work by examining the effects of the quotas on return intentions. The findings in this paper suggest how the immigration policy can influence return intentions of immigrants.

The dataset comes from the Ellis Island records in which he uses a sample of 27,000 arrivals at the Island from Europe between 1917 and 1924. The records explicitly ask whether incoming migrants intend to return to their country, though the answer on their intentions to leave might not be truthful. He creates the unique dataset by linking the new dataset including return migrant intentions to the 1930 US Census based on the first name, last name, year of birth, and country of birth. Failing to find a migrant does not always imply return migration, but the author finds that the linking rate for planned return migrants is lower than for planned permanent migrants. Another data used in the paper on return migration is found in the Annual Report of the Commissioner General of Immigration between 1908 and 1932. Although the data does not provide individual level data on out-migration, this aggregate data allows him to compare the rates and characteristics of planned return migrants to those of actual return migrants.

He finds that the planned return rate was 15.4 percent prior to the 1921 immigration quota whereas the estimate on the actual return rate is at least 40 percent. The low rate of the planned return and the gap between the planned and actual return rates suggest that migrants did not plan to return at the time of arrival but they unexpectedly returned. Before the quotas, single males from southern and eastern Europe contributed to the high unexpected return rate. In contrast, immigrants from northern and western Europe were more likely to stay in the US. He also finds lower earnings of return migrants using data on occupational scores and argues that return migrants were negatively selected. After the quotas in 1921 and 1924, the rate of actual return migration decreased along with the fact that immigration inflows significantly dropped due to the quotas, especially from southern and eastern Europe. However, he shows that the quotas do not lower the rate of planned return migration. These findings suggest that unexpected return migration decreased after the quotas. Furthermore, immigrants might experience improved outcomes in the US and thus were less likely to return to their home countries. Migrants who already entered prior to the quotas and those who entered under the quotas are most affected by the immigration policy yielding a large shock to migration inflows. The author argues that restricted migration policy causes less competition with migrants and thus migrants are less likely to return. On the contrary, liberalized migration policy leads to more intense competition thus yielding that migrants are more likely to return.

The main finding shows that most migrants did not plan to return at arrival but they unexpectedly returned prior to the immigration quotas, but unplanned return immigration decreased after the quotas. Though the paper contributes to the existing literature on return migration in many respects, an imperfect measure of return migrants could weaken his findings. In particular, failing to link to the Census could be caused by a common name of a migrant, changing one's name or death rather than returning to their home country. It is also possible that migrants might return to other countries instead of their home country or immigrate to the US again after returning. Further, the effects of the quotas examined in this paper may be contemporary or temporary because data is not available after 1924. The paper's results, however, are once again consistent with the wage effects and selection effects suggested in the existing theoretical literature, as well as the empirical wage effects reported in Ager and Hansen (2017).

3 Data

Our analysis relies on a panel of individual inventors, a measure of how locations are exposed to quotas, and a linguistic distance between quota-exposed countries and cities in the U.S. To obtain the inventor sample, we follow the method in (Doran and Yoon, 2018). We use the European Patent Office's PATSTAT database, which provides characteristics such as inventor's full name, year of patent application, and the number of citations of each patent application granted by the U.S. Patent Office from 1899 to the present. We exploit a fuzzy matching procedure that merges patents at the individual-name level into the complete count 1920 U.S. Census with names. Given the combination of first name, middle name, and last name, 43% of the U.S. population is made up of people with a unique name in the 1920 Census. To increase the accuracy of matching, we consider the 43% of the population with a unique name, with an implied age of invention at time of patent application between the ages of 18 and 80, for patent applications between the years 1919 and 1929 for main regressions (Doran and Yoon, 2018).

In (Doran and Yoon, 2018), we digitize immigration inflows by source country and year and the exact number of quotas by country and year from administrative data from Willcox et al. (1929) and U.S. Department Commerce (1924, 1929, 1931). The complete-count 1920 U.S. Census gives us each individual's birth year, birth place, citizenship, nationality, arrival year of immigration, location, and other characteristics. We collect the following characteristics of each city: total population, foreign-born population, southern and eastern immigrant population, northern and western immigrant population, and immigrant populations by nationality and year of immigration to the U.S.. In the empirical strategy section, we explain how to measure the quota exposure by locations.

We include linguistic information about the pre-existing populations in quota-exposed and non-quota exposed locations, as well as a measure of linguistic distance between the language of immigrants from specific source countries and English. First, we compute the number of pre-existing residents who speak English in each city from the 1920 Census. We then create an indicator variable for a city having a relatively high number of pre-existing persons who speak English (above the median) versus a relatively low number of pre-existing persons who speak English (below the median). For the measure of linguistic distance from a language of immigrants from a specific source country to English, we use a language score computed by Chiswick and Miller (2004), as in Tabellini (2018). The score ranges from 1 to 3 and implies how the language is linguistically close to English. We show these scores in Table 1. For example, the score for Greek is 1.75 and for French 2.5 and thus Greek is more linguistically far away from English than is French. Norwegian and Swedish are the least distant from English. In Table 1, column 3 reports whether a main national language is linguistically far from English, meaning its linguistic distance is equal to or below the median among quota-exposed countries.

In the next section, we explain how unique features of the implementation of the quotas allow us to identify how the impact of low-skilled immigration on American innovation varies by both the linguistic distance of the immigrants and the degree of English ability of the pre-existing population.

4 Empirical Strategy

Typically, a shift-share instrument for immigration relies on variation in the national origin of the pre-existing population across locations, and assumes that the new immigrants will have a tendency to locate in locations where people of their ethnicity or nationality already live. In most cases, this would also imply linguistic sorting, in which immigrants who speak English (or a language close to English) end up sorting to locations full of people who already speak English (or a language close to English). Given such linguistic sorting it would be difficult to use such an instrument to determine the differential impact of immigrants who speak a relatively common language among the pre-existing population from that of immigrants who speak a relatively rare language among the pre-existing population. We would need a natural experiment in which immigrants who speak English, for example, are often attracted to locations with relatively few English speakers, and immigrants who do not speak English are often attracted to locations with relatively many English speakers. These "off-diagonal" sortings would enable us to determine whether immigrants have a differential impact when they are located in areas with relatively many or relatively few people speaking their language.

In this paper, we exploit unique features of the 1920s U.S. immigration quotas that attracted English speakers to locations with both relatively many and relatively few English speakers, and attracted non-English speakers to locations with both relatively many and relatively few English speakers. In particular, the quota for the United Kingdom was set based on the portion of the U.S. population in 1890 that was born in the United Kingdom. This massively underrepresented the portion of the U.S. population with English ethnicity, since most English settlers had arrived generations earlier. As a result, after the quotas there were many "missing immigrants" from the United Kingdom.

The UK immigrants followed immediate relations and family to cities with many recent UK immigrants in them. Crucially, some of these cities also had relatively high populations of English speakers overall, while others of these cities had relatively low populations of English speakers overall. As a result, we can can use the quotas to compare the effects of English-speaking immigrants in locations with many English speakers to the effects of English-speaking immigrants in locations with few English speakers. Furthermore, we can determine whether the effects are due to the language similarity in general, or to something specific to English itself: the quotas also affected the immigration of some non-English-speaking immigrants who had tended to locate in cities with a relatively high number of

English speakers as well as the immigration of non-English-speaking immigrants who had tended to locate in cities with a relatively low number of English speakers. By comparing the treatment effects associated with all four groups, we can determine whether the effect of immigrants on innovation is highest when the immigrants speak a relatively common language in their destination location, or when they do not.

We follow the method in (Doran and Yoon, 2018) and (Ager and Hansen, 2018) to expand on the analysis in (Doran and Yoon, 2018), in which we examined the effects of immigration on American innovation. To identify the differential effects of quotas on innovation through a linguistic distance between American inventors and immigrants, we exploit the fact that some countries were linguistically far to English while others were linguistically close, among those countries whose immigration was restricted due to the 1921 and 1924 quotas, and that some cities had a large proportion of their residents who speak English while other cities had a relatively small proportion of English speaking residents before the quotas were enacted.

The identification strategy is a difference-in-differences strategy that depends on variation across locations and years. To expand on the strategy in (Doran and Yoon, 2018), we calculate the quota exposure for each location through the following equation:

$$Quota_c^{Lang,Eng} = \frac{100}{P_{c,1920}} \sum_{j=1}^{J} \left(\widehat{Immig}_{j,22-30} - Quota_{j,22-30} \right) \frac{FB_{jc,1920}}{FB_{j,1920}} \times Lang_j \times Eng_c \quad (1)$$

where $P_{c,1920}$ is the population of city c from the 1920 census. $\widehat{Immig}_{j,22-30}$ estimates the average immigration inflows per year from country j during the post-quota years between 1922 and 1930 if the quota acts had not been enacted. $Quota_{j,22-30}$ is the average quota limit per year from country j and thus the difference between $\widehat{Immig}_{j,22-30}$ and $Quota_{j,22-30}$ estimates the "missing" immigrants per year from country j due to the quotas. This difference is set to zero if its value is negative in which the number of estimated immigrants without the quotas is smaller than the quota.

$Lang_j$ represents two different indicators, ENG_j and NON_j, which respectively determine whether the main national language in percentages in country j is English or not. Likewise, Eng_c are two different indicators, $HIGH_c$ and LOW_c, which represent the degree of English-speaking corresponding to the percentages of English speakers in city c from the 1920 census. We define a city as high (respectively low) if the proportion of English speakers in a city is above (respectively below) the median. Finally, we exploit four treatment variables $Quota_c^{Lang,Eng}$ across a language distance of sending country j from English, ENG and NON, and the degree of English speaking in city c, $HIGH$ and LOW. Specifically, $Quota_c^{ENG,HIGH}$ represents the average annual number of missing immigrants due to quotas from English speaking countries per-100-inhabitants in city c where a proportion of residents who speak English is higher than that in other cities, while $Quota_c^{NON,LOW}$ is the missing immigrants from non-English speaking countries per-100-inhabitants in low English speaking city c.

We can also repeat the above analysis replacing the English and non-English categories with categories for languages close to English and languages far from English, estimating difference-in-differences specifications of the following forms:

$$\begin{aligned} Y_{ict} = \alpha &+ \beta_{NL}(Quota_c^{NON,LOW} \times Post_t) + \beta_{NH}(Quota_c^{NON,HIGH} \times Post_t) \\ &+ \beta_{EL}(Quota_c^{ENG,LOW} \times Post_t) + \beta_{EH}(Quota_c^{ENG,HIGH} \times Post_t) + \theta X_{it} + \tau_t + \gamma_i + \epsilon_{ict} \end{aligned} \tag{2}$$

where Y_{ict} is the number of patents or citations of incumbent inventor i who already had at least one patent in 1910 or 1919 before the quotas in city c and year t. The quartic of age of inventor i in year t, the individual fixed effect, and the year fixed effect are included. Specifically, we test whether β_{NL} and β_{EH} are statistically significant. β_{NL} reports the effect of quotas on innovation of American inventors in a low English speaking city which experienced the decline in immigration inflows from non-English speaking countries, while β_{EH} is the estimate in high English speaking cities where immigrants from English speaking countries decreased. Thus both estimates show the effect of quotas on innovation of native inventors who lost access to more linguistically close immigrants who can communicate better with the pre-existing population given the linguistic environment.

In the next section, we determine whether the quota-induced changed in immigration had differential impacts on innovation depending on whether the immigrants spoke a relatively common local language or not.

5 The Effect of Quota and Linguistic Distance

We report the results for a patents outcome variable in Table 2. In most of specifications that vary across the sample restrictions, years covered, and cutoff year for the post-quota period, we find larger declines in the number of patents applied for per year by incumbent inventors living in quota-exposed cities when the immigrants speak a relatively common language among the pre-existing population.

To be specific, the decline in immigrants from non-English speaking countries decreases the number of patents of incumbent inventors living in low English speaking cities shown in the first row in Panel A and B. For example, a one-unit increase in the treatment variable of Quota^{NON,LOW}, one less immigrant per-100-inhabitants-per-year, decreases patent applications per year by 5%. Further, we find the decrease in innovation of incumbent inventors in high English speaking cities caused by the decline of immigration inflows from English speaking countries in the fourth row in Panel A and B. On the other hand, the number of patents does not decrease when immigration inflows declined from countries linguistically far from the pre-existing population to quota-exposed cities.

Figure 2 and Figure 3 plot the estimated coefficients from the results in Table 2. The large error bar shows the statistically insignificant results for either the quota effect on innovation in high English speaking cities which decreased non-English speaking immigrants, or in low English speaking cities which experienced the decline in English speaking immigrants.

The outcome variable means reported in the table demonstrate that the variation in the size of the coefficients across the four main combinations (English v. non-English X high v. low cities) is not an artifact of different outcome variable means. Rather, the outcome variable means are similar across all four combinations, while the coefficients are dramatically different. The evidence clearly suggests that the effect of immigration on patenting reported in (Doran and Yoon, 2018) is driven by two groups: UK immigrants to cities that had a relatively high number of English speakers, and non-UK immigrants to cities that had a relatively low number of English speakers. The other two combinations do not significantly affect innovation.

While the results are less significant for citations or for the alternative measures of linguistic distance in Chistwick and Miller (2004), the results have qualitiative similarities. Future work should explore alternative outcome variables and measures of linguistic distance in more detail. Further work should also address which classes of patents are affected by language, as in (Doran and Yoon, 2018).

6 Conclusion

In this paper, we explore the mediating role of language in the effect of immigrants on innovation. We find, as in (Doran and Yoon, 2018), that immigrants affect innovation of pre-existing native inventors. But we uncover that this effect is driven by immigrants whose language corresponds to the relatively common language of their destination location. English-speaking immigrants to locations with relatively high levels of English speaking among the pre-existing population, as well as non-English-speaking immigrants to locations with relatively low levels of English speaking among the pre-existing population, are collectively responsible for the significant innovation effects of immigrants. The ”off-diagonal” combinations of mismatched languages show smaller coefficients and insignificant effects.

This strongly suggests that language may have a mediating role in the effect of low-skilled immigrants on innovation. In (Doran and Yoon, 2018), we find that low-skilled immigrants can affect innovation through changing the scale of the workforce. The results in this paper are consistent with a setting in which only low-skilled immigrants who can speak the same language as the pre-existing workforce increase the scale of production in a way that is relevant for incentivizing new inventions. In contrast, those low-skilled immigrants whose arrival produces a heterogeneous workforce whose different groups cannot communicate easily together do not increase the scale of production in the same way.

This result does not imply that highly skilled immigrants do not provide greater impacts on innovation when they are ethnolinguistically diverse. Instead, it suggests that for the low-skilled, their ability to impact innovation depends crucially on their ability to communicate easily. Future research should determine whether the benefits of new ideas, abilities, and experiences from a linguistically diverse highly skilled immigrant pool outweigh any communication barriers they bring.

References

Abramitzky, R. and L. Boustan (2017). Immigration in American economic history. *Journal of Economic Literature 55 (4)*, 1311-4.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122(3)*, 467-506.

Acemoglu, D. (2010). When does labor scarcity encourage innovation? *Journal of Political Economy 118(6)*, 1037-1078.

Ager, P. and C. W. Hansen (2018). Closing heaven's door: Evidence from the 1920s us immigration quota acts.

Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature, 43(3)*, 762-800.

Chiswick, B. R. and Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development, 26(1)*, 1-11.

Doran and Yoon (2018). Immigration and Invention: Evidence from the Quota Acts.

Greenwood, M. J. and Z. Ward (2015). Immigration quotas, world war i, and emigrant flows from the united states in the early 20th century. *Explorations in Economic History 55*, 76-96.

Tabellini, M. (2018). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration.

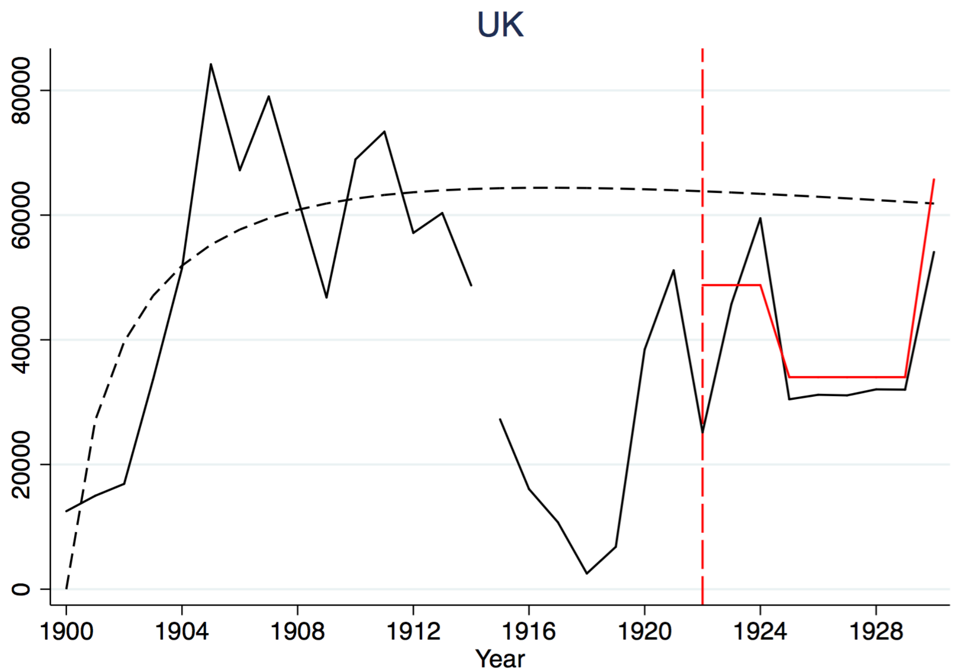
U.S. Department of Commerce (1924, 1929, 1931). Statistical Abstract of the United States.

Ward, Z. (2017). Birds of passage: Return migration, self-selection and immigration quotas. *Explorations in Economic History 64*, 37-52.

Willcox, W. F. et al. (1929). International migrations, volume i: Statistics. *NBER Books.*

Xie, B. (2017). The effects of immigration quotas on wages, the great black migration, and industrial development.

Figure 1: **Immigration Inflows under Quota**



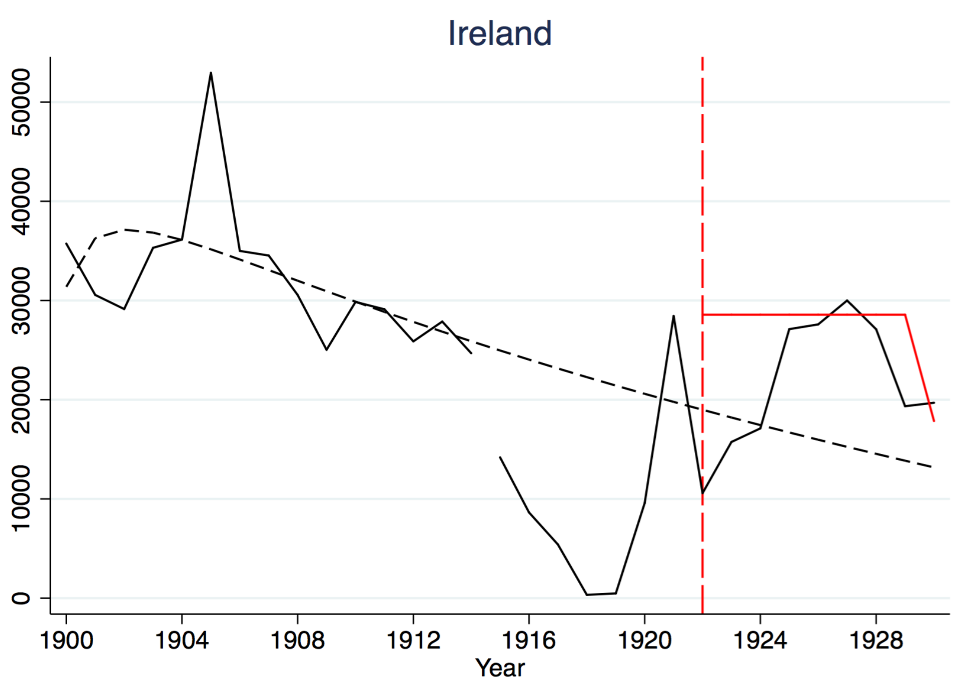
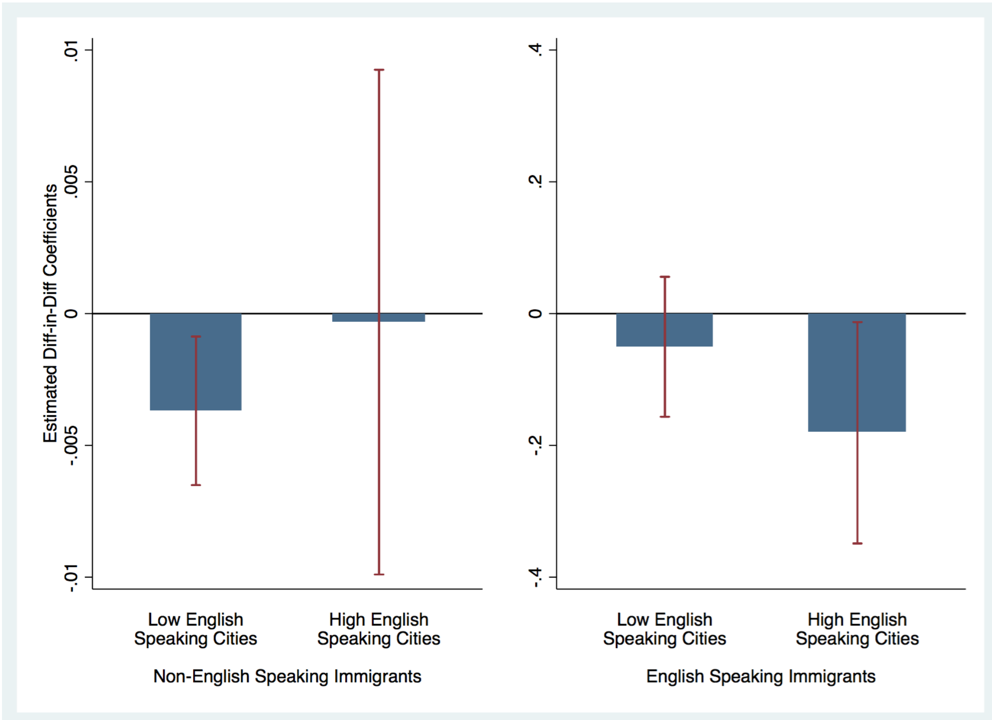
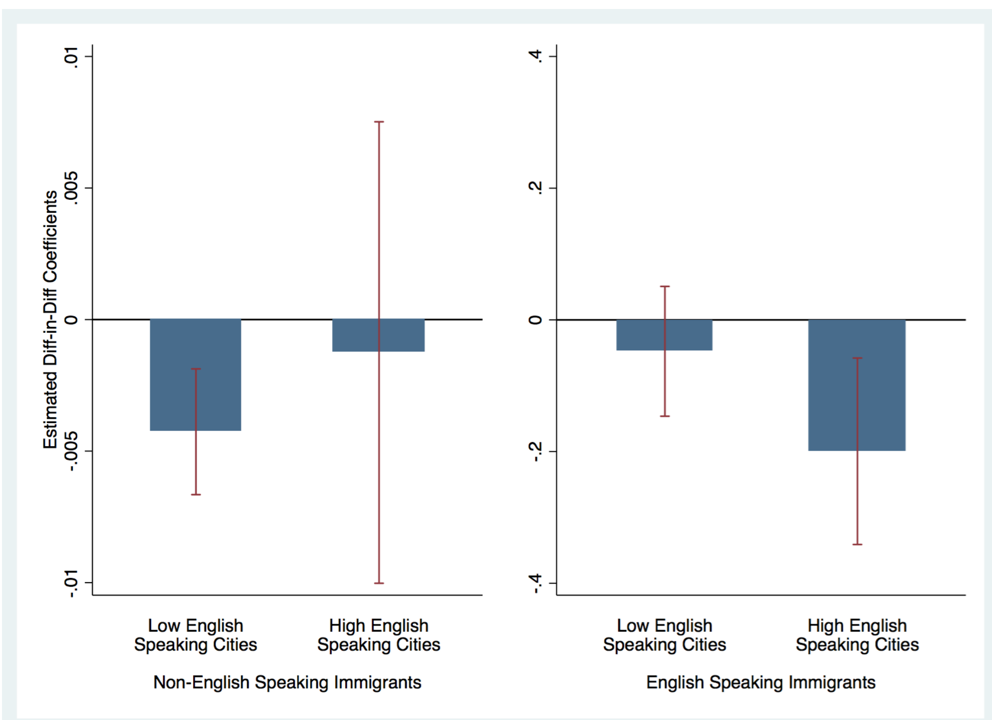


Figure 2: **Estimated Coefficients on Patents and English Speaking**



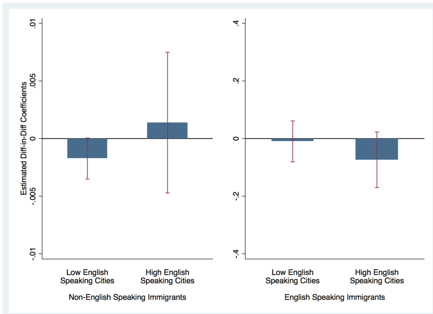
(a) Average estimates



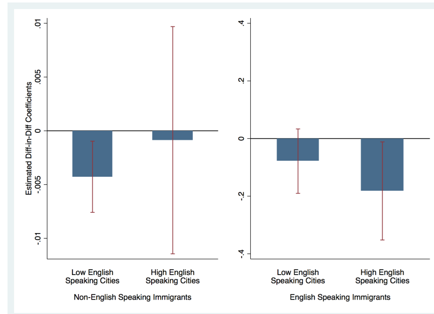
(b) Inventor in 1919, Years: 1919-1929, Post-treatment: 1924

Figure 3: **Estimated Coefficients on Patents and English Speaking, All specifications**

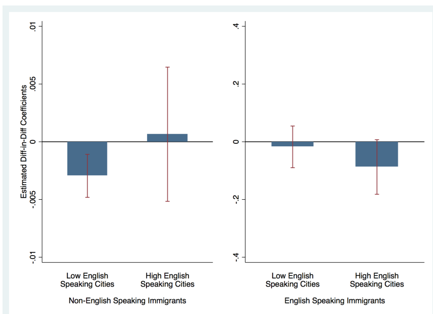
Incumbent Inventors in 1919 | **Incumbent Inventors in 1910**



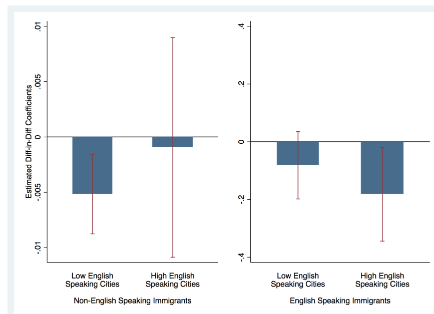
(a) 1900-1950, Post1922



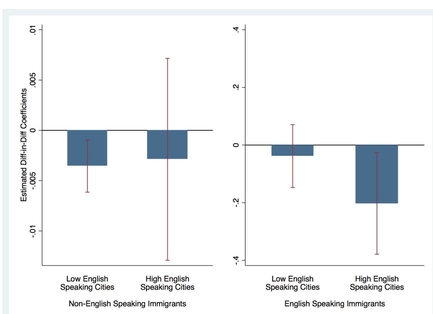
(b) 1900-1950, Post1922



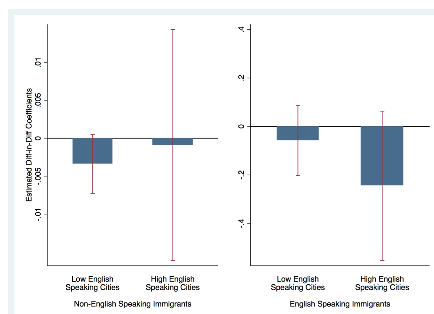
(c) 1900-1950, Post1924



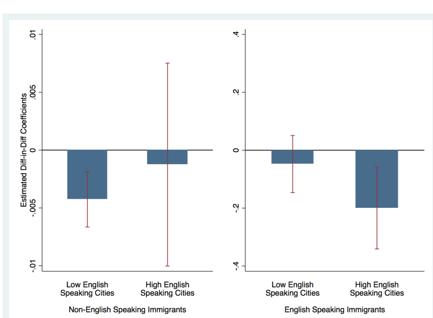
(d) 1900-1950, Post1924



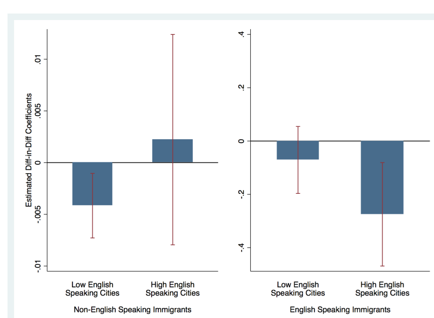
(e) 1919-1929, Post1922



(f) 1919-1929, Post1922



(g) 1919-1929, Post1924



(h) 1919-1929, Post1924

Table 1: LINGUISTIC DISTANCE BY COUNTRY

Country	Language (1)	Language Score (2)	Linguistic Distance (3)	Non-English (4)
A. Southern and Eastern Europe				
Austria	German	2.25	×	×
Bulgaria	Bulgarian	2.0	×	×
Czechoslovakia	Czech	2.0	×	×
Greece	Greek	1.75	×	×
Hungary	Hungarian	2.0	×	×
Italy	Italian	2.5		×
Poland	Polish	2.0	×	×
Portugal	Portuguese	2.5		×
Romania	Romanian	2.25	×	×
Russia	Russian	2.25	×	×
Spain	Spanish	2.25	×	×
Turkey	Turkish	2.0	×	×
Yugoslavia	Serbo-Croatian	2.0	×	×
B. Northern and Western Europe				
Belgium	Dutch	2.75		×
Denmark	Danish	2.25	×	×
Finland	Finnish	2.0	×	×
France	French	2.5		×
Germany	German	2.25	×	×
Ireland	English			
Netherlands	Dutch	2.75		×
Norway	Norwegian	3.0		×
Sweden	Swedish	3.0		×
Switzerland	German	2.25	×	×
UK	English			

Notes:
This table presents a linguistic distance from English corresponding to the main national language among quota affected countries. Column 2 reports a language score that ranges from 1 to 3 as computed by Chiswick and Miller (2004). For example, the score for Greek is 1.75 and for French 2.5 and thus Greek is more distant from English than French. Norwegian and Swedish are the least distant from English. Column 3 determines whether a main language is linguistically far from English where its linguistic distance is equal to or below the median among countries, and column 4 indicates a non-English speaking country.

Table 2: EFFECT OF QUOTA ON PATENT AND ENGLISH SPEAKING

	Year of Patent Application			
	1900-1950		1919-1929	
	Post-Treatment Year			
	1922	1924	1922	1924
	(1)	(2)	(3)	(4)
A. Dependent Variable: Patents by Incumbent Inventors in 1919				
$\text{Quota}^{NON,LOW} \times$ Post	-0.0017	-0.0030***	-0.0036**	-0.0043***
	(0.0011)	(0.0011)	(0.0016)	(0.0015)
$\text{Quota}^{NON,HIGH} \times$ Post	0.0014	0.0007	-0.0029	-0.0013
	(0.0037)	(0.0035)	(0.0061)	(0.0053)
$\text{Quota}^{ENG,LOW} \times$ Post	-0.0099	-0.0176	-0.0382	-0.0477
	(0.0432)	(0.0439)	(0.0663)	(0.0599)
$\text{Quota}^{ENG,HIGH} \times$ Post	-0.0737	-0.0874	-0.2027*	-0.1995**
	(0.0588)	(0.0575)	(0.1072)	(0.0860)
Dependent Variable Mean	0.1252	0.1206	0.1060	0.0936
- Low English Speaking City	0.1274	0.1230	0.1097	0.0977
- High English Speaking City	0.1219	0.1171	0.1003	0.0876
Number of Observations	6572144	6572144	1572335	1572335
Number of Inventors	145722	145722	145722	145722
Number of Cities	3306	3306	3306	3306
R-squared	0.2328	0.2328	0.4004	0.4004
B. Dependent Variable: Patents by Incumbent Inventors in 1910				
$\text{Quota}^{NON,LOW} \times$ Post	-0.0043**	-0.0052**	-0.0034	-0.0042**
	(0.0020)	(0.0022)	(0.0024)	(0.0019)
$\text{Quota}^{NON,HIGH} \times$ Post	-0.0009	-0.0009	-0.0009	0.0022
	(0.0064)	(0.0060)	(0.0092)	(0.0062)
$\text{Quota}^{ENG,LOW} \times$ Post	-0.0784	-0.0814	-0.0587	-0.0709
	(0.0679)	(0.0709)	(0.0878)	(0.0764)
$\text{Quota}^{ENG,HIGH} \times$ Post	-0.1818*	-0.1823*	-0.2451	-0.2748**
	(0.1034)	(0.0985)	(0.1874)	(0.1179)
Dependent Variable Mean	0.1448	0.1389	0.0808	0.0784
- Low English Speaking City	0.1477	0.1420	0.0851	0.0831
- High English Speaking City	0.1405	0.1344	0.0746	0.0717
Number of Observations	3697635	3697635	870860	870860
Number of Inventors	81245	81245	81245	81245
Number of Cities	3270	3270	3269	3269
R-squared	0.2655	0.2656	0.4426	0.4426

Notes:
This table shows the results for patents of incumbent native inventors and languages of countries affected by quotas. The estimates report the differential effect of quotas on innovation through the linguistic distance. Panel A uses the dependent variable of individual patent data from incumbent inventors who already had at least one patent in 1919 before the quota and pre-treatment periods. In Panel B, the sample is restricted to incumbent inventors who patented before the year 1910. The number of patents is winsorized at 10. Standard errors are clustered by city and asterisks denote statistical significance, where *** significant at 1% level; ** significant at 5% level; * significant at 10% level.