Labor Market Segmentation: Evidence from U.S. Janitorial Jobs Advertised in English and Spanish

By ZIJUN XU

May 8, 2020

Do jobs advertised in Spanish pay less than those advertised in English? Using a random sample of janitorial jobs from an online job board, I find that jobs advertised in Spanish face a 8.14% penalty. Evidence shows that the janitorial market segments into a primary English segment and a secondary Spanish segment. While linguistic skills have limited economic return, firms use language(s) to find workers with desired socioeconomic background. In states with a large population of unauthorized janitors, the hourly wage is 3.49% higher. Furthermore, legal janitors receive a 2.96% wage premium whereas illegal janitors in the Spanish segment face a 2.44% penalty. There is also an inverse relationship between the wage penalty and the proportion of firms verifying worker authorizations.

Key words: Wage Gap, Labor Market Segmentation, Language, Illegal Worker

I. INTRODUCTION

Based on the estimates from Bureau of Labor Statistics, there are over 2 million janitors and custodians in United States. Their interquartile hourly wage ranges from 10.57 to 16.1 dollars. In my sample, roughly 20% of the job advertisements use Spanish. This raises the question as to whether janitorial jobs posted in Spanish pay the same amount compared to those advertised in English. Furthermore, since most custodial jobs are low-skill and do not involve sophisticated communication, why do firms ask for language skills?

To answer these questions, I randomly sampled entry-level janitorial jobs from the online job board indeed.com. I find out that there is an 8.14% penalty for jobs advertised in Spanish after controlling for other covariates. After partitioning jobs into those written in English and those written in Spanish, there are different wage-setting mechanisms across segments. For example, a non-commercial driver's license increases the wage by 4.97% in the English segment but raises

[‡]I am very grateful to my senior thesis advisor Professor David Card. This article would not have been possible without his tremendous support and guidance. All mistakes belong to me alone.

the wage by 19.3% in the Spanish segment. Also, while high-school degree has a 3.93% return in the English market, its impact in the Spanish market is not significantly different from zero. Although job ads are written in different languages, coefficients for 'SPA' or 'ENG' suggest limited economic return. Through logistic regression on the language(s) used in specific job advertisement, I show the screening effects of languages.

Although it is hard to directly estimate the return of being legal and the penalty of being illegal, I make mild assumptions for the approximations. Work authorization is used as a proxy for employment eligibility and the estimated legal premium is 2.96%. This approximation should be reasonable because illegal workers can only work for firms that do not check work authorizations. Through iterated expectation, I estimate that illegal workers in the Spanish segment face 2.44% penalty. Results also suggest that as more firms verify work authorizations, the labor force of illegals in the Spanish segment faces less penalty.

My study contributes to three related literatures. First, my study suggests that there exists a primary English segment and a secondary Spanish segment. Second, it shows that the languages could encourage candidates to self-select based on their own language skills. Moreover, it can be effective for firms to screen candidates by choosing a specific language. Specifically, firms not looking for Spanish-speaking janitors could simply advertise their opening in English alone. The third related literature is illegal labor force. Evidence suggests an inverse relation between the wage penalty in the Spanish segment and the proportion of employers verifying employment eligibility. Also, if we accept that illegal workers do not receive a premium, then at most 84.4% janitorial firms are strictly verifying the employment eligibility. Furthermore, my results suggests that wages tend to be higher in states with a large proportion of unauthorized workforce.

My study has two policy implications. First, it can help governments better fund language programs that prepare new immigrants for U.S. labor market. Immigrants arrive with varying degrees of English proficiency. For immigrants working in the low-skill occupations, it is of practical importance to assess the benefit of learning English and to provide the necessary support. Second, my study is relevant to the discussion on restricting ex-offender or unlawful resident from possessing a driver's license. Because driver's license could increase the hourly pay by almost 6%, such restriction could limit low-skill workers' labor market performance.

The rest of the paper proceeds as follows. Section II describes the sampling procedure and covariates in the OLS model. Section III discusses different model specifications and justifies my preferred specification. Section IV identifies the wage gap between jobs advertised in English and Spanish and tests the labor market segmentation. Section V discusses roles of languages in wage-setting. Section VI estimates the wage premium for legal janitors and the wage penalty to illegal janitors. Section VII concludes the paper.

II. Data

(a) Sampling Procedure

My study uses five datasets. The first dataset is a random sample of janitorial job advertisements. Per the classification of the Bureau of Labor Statistics, there are three subgroups for the generic cleaning jobs. To control for occupational wage difference, my sampling targets one of the three cleaning subgroups, namely 'Janitors and Cleaners, Except Maids and Housekeeping Cleaners.'

The sampling has three steps. First, I filter the job titles by keywords in the search engine of indeed.com. The English keyword is 'cleaning' and the corresponding keyword in Spanish is 'limpieza.' Second, I sort the jobs in a reverse-chronological order with the most recent job placed at the top. Third, as I go through the search page, I exclude jobs lacking wage or city/state information. I also exclude jobs belonging to the other two subgroups of janitorial workers, i.e., 'Maids and Housekeeping Cleaners' and 'First-Line Supervisors of Housekeeping and Janitorial Workers.'

For each job advertisement, the language classification is based on Google Translate. After uploading the job description, I checked the auto language detection at the bottom of the page, in which 'English' or 'Spanish' was displayed. Then, I record the hourly wage. If annual salary or monthly salary is given instead, I divide the wage by 2080 hours per year or 173.33 hours per month to get the hourly wage. I also record years of work history or experience needed. If any info is not given in the job description, zero is entered. I then assign values for a list of covariates including ‘Spanish,’ ‘English,’ ‘ SPA,’ ‘ENG,’ ‘Work Authorization,’ ‘Background Check,’ ‘Full-time,’ ‘Part-time,’ ‘High school,’ and ‘Driver’s license.’ Details of the covariates are presented in the table below.

Because the wage of any given job advertisement depends on its local price level, I incorporate additional two datasets to control for local wage level. The second dataset is Principle Cities of Metropolitan and Micropolitan Statistical Areas, the delineation file from the Census Bureau. It enables me to map Core Based Statistical Area (CBSA) Code to each observation based on the city and state. The third dataset is Occupational Employment Statistics (OES), a cross-sectional survey from the Bureau of Labor Statistics. For each sampled job, I used OES to find the mean wage of each metropolitan area by matching the CBSA code.

The last dataset comes from Pew Research Center. It lists estimates of unauthorized immigrant population for each state as of 2016. The size of illegal immigrants varies significantly among states and many of them actively participate in the U.S. labor market. It is thus important to investigate the association between wage and size of illegal workforce. I use the last dataset to identify states with the largest illegal immigrant sizes. Since there are different measures to the population sizes, the best metrics is selected. A dummy variable is then generated to indicate if a given job is located in any of these states.

(b) Covariate Description

Table 9 in the Appendix displays the covariate names and their descriptions. Based on the job description, I manually assigned and recorded numerical values for all covariates in each sampled job.

(c) Limitation

The use of Google Translate introduces potential error to my data. Natural language processing is still a developing technology. Although Google Translate is one of the most sophisticated and reliable translation platform, it is possible that few of the observations are mislabelled.

The three-month duration of my experiment could make my dataset vulnerable to external labor market perturbations. Since the data collection process is not automated, the number of jobs I sampled varied from day to day. Thus, in this article, I am making the assumption that all the sampled jobs are drawn from the same underlying population.

Because I was excluding jobs with no wage information, I might have introduced upward bias to my wage estimates. Disclosure of wage can be self-selective and firms with less competitive offers might choose not to disclose their wage in the ad.

(d) Summary Statistics

Let us first visualize distributions of wage. Orange represents jobs written in English and blue represents jobs written in Spanish. Note that the right figure excludes job ads written in both languages. In particular, the right figure has larger entropy between orange and blue distributions. Intuitively, the wage gap is more evident if we only compare jobs written in a single language. Nevertheless, two figures are more similar than dissimilar. The distribution for English have longer right tails whereas the distribution for Spanish places more mass on the left spectrum, i.e., low-paying jobs. And the mean and median wages in the English segment are approximately one dollar higher.

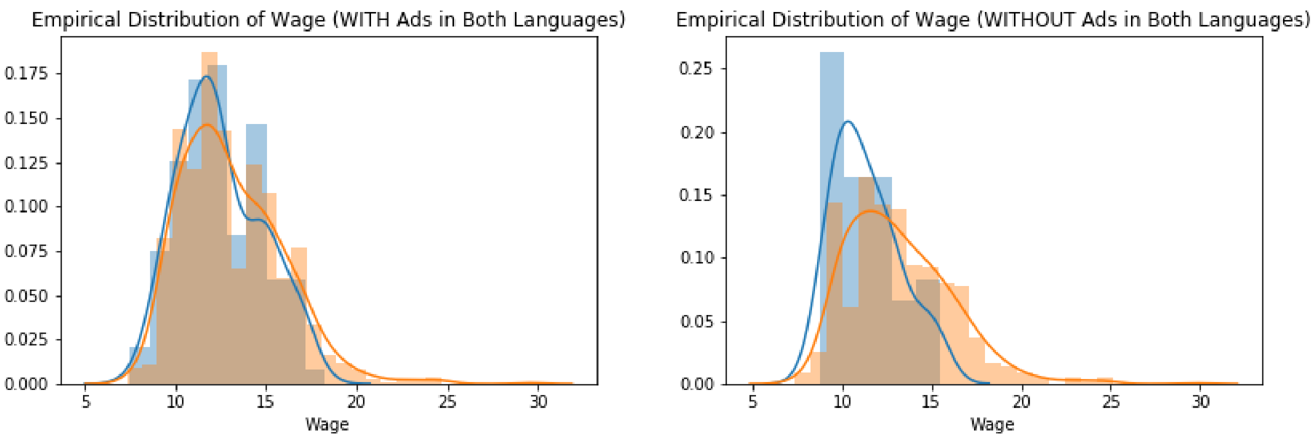

FIGURE 1
DISTRIBUTIONS OF WAGES (BLUE: SPANISH; ORANGE: ENGLISH)

Table 10 in the Appendix displays the summary statistics computed after grouping by the language(s) in which the job ads are written. More than 75% of the sampled jobs are written in English alone, less than 5% in Spanish alone, and almost %20 in both languages. Jobs written in English alone have the highest average wage and exceed jobs written in Spanish along by almost 2 dollars per hour. Jobs written in both languages have an average wage near the former. Because those jobs were sampled nationally, differences in local prices partially contributes to the wage gap. The higher the mean wage for janitors in a city, the higher the observed average wage in this city. But do all jobs respond to the local mean in the same way?

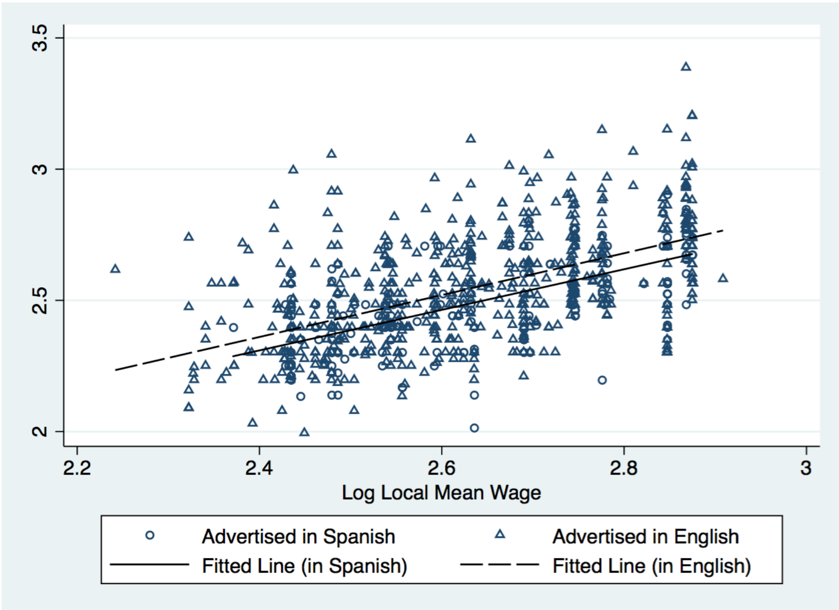

FIGURE 2
SCATTER PLOT OF LOG WAGE AGAINST LOG LOCAL MEAN WAGE

The above figure visualizes the association between log wage of a given job and its log local mean wage. We discover that the fitted line for Spanish (solid line) and the fitted line for English (dashed line) are parallel to one another. This resemblance in slopes indicates that local wage levels have nearly identical impacts on wage. However, it is clear that these lines have distinct intercepts. Hence, there must be other factors causing the wage gap. The next section uses the ordinary least squares to study a more comprehensive set of explanatory variables.

III. OLS MODEL

(a) Specification: Wage

To find the preferred specification, I propose six OLS models. All six models have the same set of fixed covariates, but some models have additional controls or slightly different output variables, i.e., log-transformed or linearly transformed. The first model is the base model where we regress the wage in absolute scale on the full set of covariates.

TABLE 1
SPECIFICATIONS ON WAGE

	(1) Wage	(2) Log Wage	(3) Wage Deviation	(4) Log Ratio	(5) Wage	(6) Log Wage
Spanish	-1.764^{***}	-0.137^{***}	-0.942^{**}	-0.0785^{**}	-1.191^{***}	-0.0932^{***}
	(0.404)	(0.0297)	(0.347)	(0.0248)	(0.334)	(0.0242)
English and Spanish	1.211^{**}	0.100^{**}	0.0713	0.0197	0.417	0.0399
	(0.435)	(0.0320)	(0.374)	(0.0267)	(0.360)	(0.0261)
SPA	-0.884	-0.0563	-0.510	-0.0314	-0.624	-0.0376
	(0.618)	(0.0455)	(0.531)	(0.0380)	(0.510)	(0.0370)
ENG	1.152^{***}	0.0848^{***}	-0.0616	0.00298	0.306	0.0235
	(0.241)	(0.0177)	(0.207)	(0.0148)	(0.203)	(0.0147)
Work Authorization	0.716^{***}	0.0508^{**}	0.376^{*}	0.0280^{*}	0.479^{**}	0.0338^{**}
	(0.214)	(0.0158)	(0.184)	(0.0132)	(0.177)	(0.0128)
Background Check	-0.266	-0.0190	0.270	0.0170	0.108	0.00799
	(0.193)	(0.0142)	(0.166)	(0.0118)	(0.160)	(0.0116)
Full-time	0.252	0.0190	0.0832	0.00669	0.134	0.00977
	(0.237)	(0.0175)	(0.204)	(0.0146)	(0.196)	(0.0142)
Part-time	-1.139^{***}	-0.0889^{***}	-1.059^{***}	-0.0837^{***}	-1.083^{***}	-0.0850^{***}
	(0.237)	(0.0175)	(0.204)	(0.0146)	(0.195)	(0.0142)
Driver License	0.904^{***}	0.0624^{***}	0.842^{***}	0.0579^{***}	0.861^{***}	0.0591^{***}
	(0.224)	(0.0165)	(0.192)	(0.0137)	(0.184)	(0.0134)
High School	0.174	0.0153	0.629^{**}	0.0466^{**}	0.491^{*}	0.0387^{*}
	(0.251)	(0.0185)	(0.216)	(0.0154)	(0.208)	(0.0151)
Years of Relevant Work	0.345^{*}	0.0229^{*}	0.226	0.0143	0.262^{*}	0.0165^{*}
	(0.135)	(0.00992)	(0.116)	(0.00829)	(0.111)	(0.00807)
Local Mean Wage					0.697^{***}	
					(0.0331)	
Log Local Mean Wage						0.749^{***}
						(0.0341)
Constant	13.21^{***}	2.562^{***}	-0.977^{***}	-0.0782^{***}	3.324^{***}	0.585^{***}
	(0.265)	(0.0195)	(0.228)	(0.0163)	(0.518)	(0.0913)
Observations	948	948	948	948	948	948
Adjusted R^2	0.135	0.139	0.116	0.126	0.413	0.432

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As expected, the log-transformed wage leads to better goodness of fits. This can be seen by pair-wise comparisons between the first two and the middle two models. Although taking the log wage does not change the significance levels of covariates, it does improve adjusted R-squares by a small margin. We also need to decide whether to include the local mean wage, because wage variations could be partly attributed to different local price levels.

The adjusted R-squared of the basic model is 0.135. If we add the mean wage as an additional control , the goodness of fit increases tremendously from 0.135 to 0.413. Alternatively, we could subtract the mean wage from the output variable and regress the same set of covariates on this wage deviation. However, the goodness of fit reduces from 0.135 to 0.116. This is because we are implicitly making an unreasonable assumption that the regression coefficient for the local mean wage is 1. However, the coefficient from the fifth model becomes 0.7, which means that the previous assumption overestimates the true coefficient. Therefore, we should include the local mean wage as a free variable. The fifth and the last models compares the effect of log-transforming the local mean wage. The log-transformed mean wage leads to an approximately 0.2 increase in adjusted R-squared. Log transformation works because it makes our model less impacted by outliers from the right-tailed wage distribution. By accounting for the geographical wage disparities and fully utilizing the logarithms in wage analysis, I choose the last model as my preferred specification.

Among my proposed specifications, the indicator on 'Spanish' is uniformly robust, showing a consistent wage penalty to jobs advertised in Spanish. Besides, both indicators for 'Work Authorization' and 'Driver's License' are uniformly robust as well, which shows the premium for being a legal worker or a driver. Additionally, 'Years of Relevant Work' and 'Part-time' are also robust to specifications. However, 'English and Spanish' losses its significance level after controlling the log local mean wage. The same applies to 'ENG.' On the other hand, 'High School' gains significance level in my preferred specification.

(b) Additional Control: Labor Supply

As of 2017, the Pew Research Center estimated that California had the largest population of illegal immigrants at around 2 million. Texas was ranked the second at around 1.6 million and Florida followed at around 0.8 million. An alternative ranking system sorts states by their relative sizes of illegals. In this case, the top three states are California, Texas and Nevada.

Does size of unauthorized workers influence the wage-setting in a given state? To better understand this question, I plot state-level average wages against the proportion of illegals.

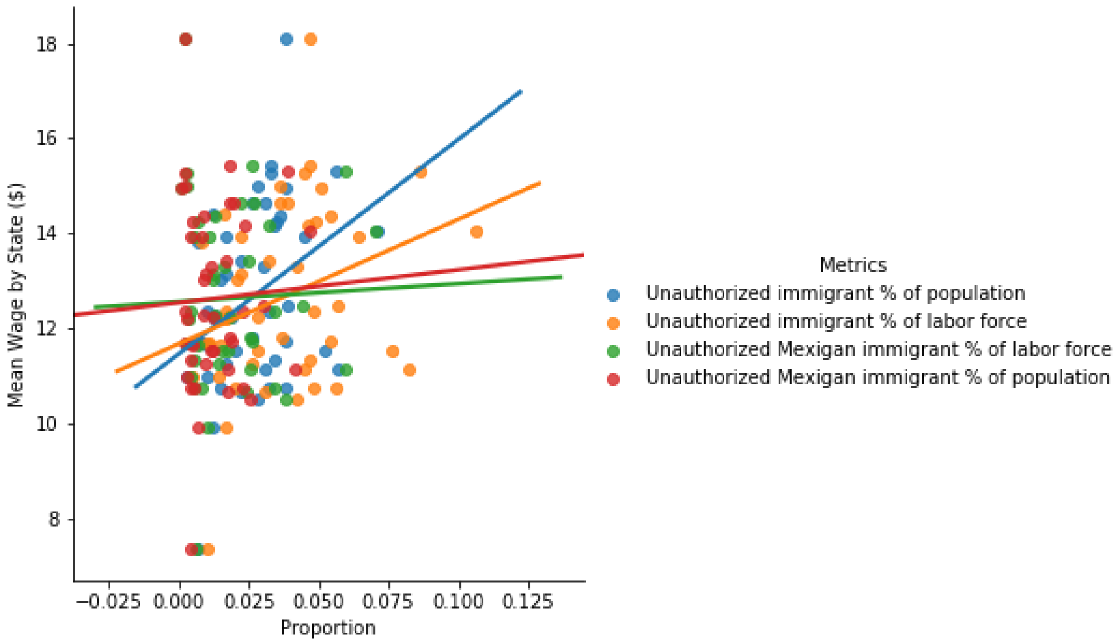

FIGURE 3
SIZE OF ILLEGAL IMMIGRATION VERSUS STATE MEAN WAGE

In Figure 3, regardless of the metric chosen, I find a positive slope and hence a positive association between proportion of unauthorized workers and state-level mean. To study the effect of illegals' size on an individual job, I generate a new dummy variable that turns on when a given job is located in any of the states with the largest illegal population. The next table compares the original regression model and another two models with added controls.

The relative measure 'Highest Immigration (%)' clearly outperforms the absolute measure 'Highest Immigration (#).' First, relative-measure model improves the adjusted R-squared of the original model by by 0.05 while the absolute-measure model makes zero improvement. Furthermore, the coefficient for 'Highest Immigration (%)' is statistically different from zero. Therefore, I will include 'Highest Immigration (%)' to improve my model.

TABLE 2
CONTROL THE LOCAL UNAUTHORIZED LABOR SUPPLY

	(1) Log Wage	(2) Log Wage	(3) Log Wage
Log Local Mean Wage	0.757***	0.753***	0.730***
	(0.0338)	(0.0340)	(0.0348)
SPA	−0.0386	−0.0391	−0.0401
	(0.0367)	(0.0366)	(0.0365)
ENG	0.0225	0.0222	0.0224
	(0.0145)	(0.0145)	(0.0145)
Work Authorization	0.0331**	0.0315*	0.0296*
	(0.0127)	(0.0128)	(0.0127)
Background Check	0.00774	0.00883	0.00936
	(0.0115)	(0.0115)	(0.0114)
Full-time	0.00968	0.00968	0.0102
	(0.0141)	(0.0141)	(0.0140)
Part-time	−0.0867***	−0.0858***	−0.0841***
	(0.0141)	(0.0141)	(0.0140)
Driver License	0.0586***	0.0583***	0.0585***
	(0.0132)	(0.0132)	(0.0132)
High School	0.0383*	0.0384*	0.0366*
	(0.0149)	(0.0149)	(0.0149)
Years of Relevant Work	0.0171*	0.0167*	0.0157*
	(0.00799)	(0.00800)	(0.00798)
Spanish	−0.0783**	−0.0816***	−0.0814***
	(0.0243)	(0.0244)	(0.0242)
English and Spanish	0.0245	0.0263	0.0265
	(0.0261)	(0.0262)	(0.0260)
Highest Immigration (#)		0.0127	
		(0.0109)	
Highest Immigration (%)			0.0349**
			(0.0116)
Constant	0.565***	0.572***	0.625***
	(0.0906)	(0.0908)	(0.0924)
Observations	947	947	947
Adjusted R^2	0.438	0.438	0.443

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

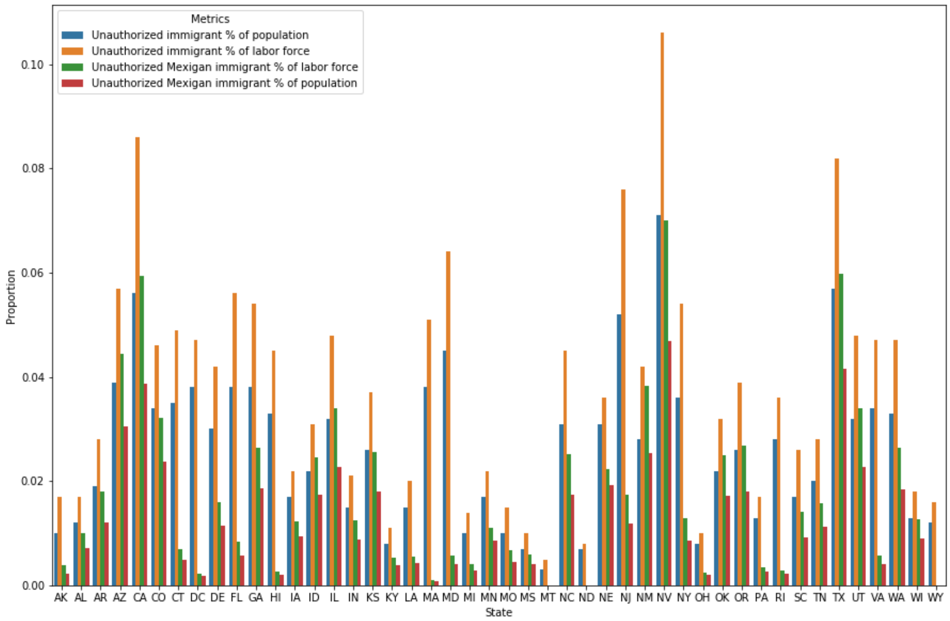

FIGURE 4
BAR PLOTS FOR MEASURES OF ILLEGAL IMMIGRANTS

Moreover, by taking a close look at Figure 4, we see that 'Highest Immigration (%)' actually has a richer interpretation than we previously defined. In fact, CA, TX and NV are exactly the same top three states if we rank by 'proportion of illegals of the labor force,' 'proportion of Mexican immigrants of the population,' and 'proportion of Mexican immigrants in the labor force.'

The third interpretation is most relevant to my study, because unauthorized Mexican immigrants primarily use Spanish at work and they also actively participate in the U.S. labor market. Given a fixed labor demand, an increase in the labor supply usually leads to a decrease in the wage. However, there is a 3.5% increase in the hourly wage if the job is located in CA, TX or NV. In other words, when a janitor position is located in a state with many immigrants, the hourly wage tends to be higher. Empirically, immigrant inflow must be more complicated than a simple shift in the labor supply. Card and Shleifer (2009) point out that 'immigrants and natives appear to be imperfect substitutes within broad education classes.' Using the Spanish Labor Force Survey, Alcobendas and Rodriguez-Planas (2009) argue that 'low-skilled immigrants assimilate better in a dual labor market.' Suppose there does exist a primary market and a secondary market. When the unauthorized population grows, many of them settle down in the secondary market in which there is less legal regulation and hence easier to enter. As a result, the labor market becomes more segmented in size. However, some native workers might still be employed in the secondary segment due to factors such as their inability to pass a background check. When more unauthorized workers assimilate with natives who have higher wages, the average wage in the secondary segment could raise and hence explain the positivity of the coefficient for 'Highest Immigration.'

(c) Full Model

Combining the discussion in the previous section leads to the OLS in Table 3 as my final model specification with a full set of controls.

First of all, jobs posted in Spanish face an 8.14% wage penalty on average. With the regression coefficient of 0.73, local mean wage is highly predictive of a job's compensation. This is not surprising as sample mean often serves as a good estimate and it controls for the varying price levels across states. While we fail to reject the null that coefficient of 'full-time' is zero, 'part-time' jobs receive an 8.41% penalty in the OLS model. Simpson (1986) estimates a penalty around 10% as well, but Rodgers (2004) argues the penalty would become statistically insignificant from zero once we adjust for the selection effect of workers into full-time or part-time jobs.

Employment eligibility and background check are two variables related to the legal background. Jobs that explicitly require work authorization receive a 2.96% wage premium. However, this does not mean that the return of being legal is 2.96%, because some firms may 'forget' to mention it in the ad but still verify the legal documents during the hiring. Aside from the employment eligibility, firms may also choose not to recruit applicants with criminal history. A clean criminal history is usually needed to pass a background check. While the estimated coefficient for jobs asking for background check is 0.9%, it is not statistically significant to reject the null hypothesis that 'background check' has significant impact on wage.

Firms also value on-the-job training. For each additional year of work history requested by the firm, the hourly wage increases by 1.57%. Corcoran and Duncan (1979) finds out the yearly

TABLE 3
FULL MODEL

	(1) Log Wage
Log Local Mean Wage	0.730***
	(0.0348)
SPA	−0.0401
	(0.0365)
ENG	0.0224
	(0.0145)
Work Authorization	0.0296*
	(0.0127)
Background Check	0.00936
	(0.0114)
Full-time	0.0102
	(0.0140)
Part-time	−0.0841***
	(0.0140)
Driver License	0.0585***
	(0.0132)
High School	0.0366*
	(0.0149)
Years of Relevant Work	0.0157*
	(0.00798)
Spanish	−0.0814***
	(0.0242)
English and Spanish	0.0265
	(0.0260)
Highest Immigration (%)	0.0349**
	(0.0116)
Constant	0.625***
	(0.0924)
Observations	947
Adjusted R^2	0.443

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

returns to employment history to be 1.3%, 2.7%, 1%, and 1.4% respectively for White male, African American male, White female, and African American female. Although my estimate is close to Corcoran and Duncan's, it is uncertain if the marginal return of the work experience will apply workers with many years of experience. This is my OLS model is fitted on a sample with few jobs asking for more than 2 years of work history. Furthermore, entry-level janitor would typically need to become a supervisor to gain a sizable raise.

While none of the sampled jobs ask for college degrees, the return of having high-school diploma is 3.66%. Nevertheless, it is modest compared to the 9% estimate by Jaeger and Page (1996), which more than doubles my estimate. One might ask what the value of high school diploma is to janitorial firms. If we accept the finding of Clark and Martorell (2014) that there is very limited evidence of the signaling value of high school diploma, the estimated 3.66% return might be indeed capturing the knowledge premium gained from schooling. Since janitors are sometimes expected to read instructions on chemicals, literacy signaled by a high school diploma could be the only motivation for employers who are not interested in other skills one from high school.

Although most of the job duties do not include driving or transporting tools using vehicles, sometimes recruiters do require a valid driver's license since some job sites do not have easy access to public transportation. This is particularly common among night-shift jobs in which public transportation is less accessible. Among all skills or qualifications in my model, driver's license has the highest return at 5.85%. Kleiner and Krueger (2010), however, finds an OLS coefficient of 15% for acquiring occupational licensing. This approximately 10% difference may be attributed to the non-occupational use of an occupational license. In other words, job duties of a janitor rarely involve driving.

It is worth taking a look at the four language-related covariates—'SPA,' 'ENG,' 'Spanish' and 'English.' While 'Spanish' indicates a substantial usage of Spanish in the advertisement, 'SPA' indicates the language proficiency in Spanish. Their coefficients are -8.14% and -4.01%, which are quite different in magnitudes. Further details on languages are discussed in details in Section V.

IV. Test of Dual Labor Market

In my dataset, the majority of jobs are advertised in English, but jobs advertised in Spanish pay less on average. Reich, Gordon, and Edwards (1973) suggest that the labor market segmentation leads to the separation of primary segment and secondary segment. In my study, the English segment is primary and the Spanish segment is secondary. To test the separation between these segments, I use Dickens and Lang's approach. They propose two criterions for testing the existence of dual labor market. Both markets must have 'different wage setting mechanism' and there must exist 'barriers to mobility' between two markets (Dickens and Lang 1983).

(a) Different Wage-setting Mechanism

To compare and contrast the wage setting mechanisms of the English segment and the Spanish segment, I subset the dataset by language(s) used in the job description. Note that I classify all jobs into one of 'English-only,' 'Spanish-only', and 'English and Spanish.' Then, I regress the log wage on my full set of controls except the above language variables. First, the coefficient for 'Log Local Mean Wage' is 8.8% lower in the Spanish segment. Hence, in the English segment, individual wages are more responsive to changes in the local means.

From Table 4, we find that education and work experience have significantly less return in the Spanish segment. While a high-school degree on average increases the wage by 3.93% in the English segment, the coefficient in the Spanish segment is almost zero if not negative. As for the work history, the English segment rewards a 2.14% increase for each additional year of experience. However, the coefficient in Spanish segment is not significant enough to reject the null. The largest difference in the regression coefficients occurs for the driver's license. The economic return of a driver's license more than triples in the Spanish segment. Holding other factors constant, possessing a driver's license increases the wage by 4.97% in the English market whereas the return surges to 19.3% in the Spanish market.

Clearly, the English segment and the Spanish segment have two distinct wage equations. Moreover, the 'English and Spanish' segment also differs from both of them. For example, the return of having a driver's license is 8.21%, sandwiched between 4.97% of the English segment and 19.3% of the Spanish market. While holding a high-school degree in the English segment has a positive return, the coefficient is not statistically different from zero in the third regression. Similarly, the estimated return for another year of work experience is not statistically different from zero as well.

TABLE 4
WAGE-SETTING MECHANISM

	(English-only Segment) Log Wage	(Spanish-only Segment) Log Wage	(Intersection: English and Spanish) Log Wage
SPA	−0.0596	0.0952	−0.0254
	(0.0529)	(0.0950)	(0.0548)
ENG	0.0276	0.0241	−0.0171
	(0.0166)	(0.0606)	(0.0366)
Work Authorization	0.0373*	−0.0625	0.00703
	(0.0155)	(0.0520)	(0.0236)
Background Check	0.00627	0.0281	0.0361
	(0.0134)	(0.0518)	(0.0246)
Full-time	0.00377	0.0544	0.0216
	(0.0172)	(0.0422)	(0.0266)
Part-time	−0.0910***	−0.0577	−0.0591*
	(0.0175)	(0.0434)	(0.0254)
Driver License	0.0497**	0.193***	0.0821**
	(0.0155)	(0.0528)	(0.0286)
High School	0.0393*	−0.0447	−0.0146
	(0.0162)	(0.0798)	(0.0600)
Years of Relevant Work	0.0214*	0.0271	−0.0222
	(0.00917)	(0.0473)	(0.0186)
Log Local Mean Wage	0.743***	0.831***	0.711***
	(0.0416)	(0.132)	(0.0754)
Highest Immigration (%)	0.0284*	0.0349	0.0457
	(0.0142)	(0.0405)	(0.0241)
Constant	0.599***	0.229	0.609**
	(0.111)	(0.339)	(0.201)
Observations	725	45	177
Adjusted R^2	0.416	0.544	0.501

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(b) Barrier of Entry

In a competitive labor market, workers are able to freely move from one job to another, seeking opportunities to move from low-paying jobs to high-paying jobs. In this article, the relevant question is whether custodians in the Spanish segment are capable of moving to the

English segment. In many studies on market segmentation, one of the biggest challenge is to neatly separate a market into two segments. In my study, however, the linguistic barrier produces a natural and effective partition.

In general, the first step for any worker to successfully apply for a job is to read and understand the job advertisement. In the entry-level janitorial industry, it is not uncommon that workers only know a single language, either English or Spanish. Thus, even when Spanish-speaking janitors discover a English job ad that pays well, they will find it hard to understand the job description and to apply to the opening. One might say that workers can submit an application with the help of modern translation softwares. It is true. For firms looking for workers with a specific language skills, however, candidates' real language skill can be easily examined in a face-to-face interview. The same dilemma applies to English-speaking workers interested in jobs written in Spanish. Therefore, for monolingual workers, language skill is a barrier to entry for either direction.

(c) Dual Market Structure

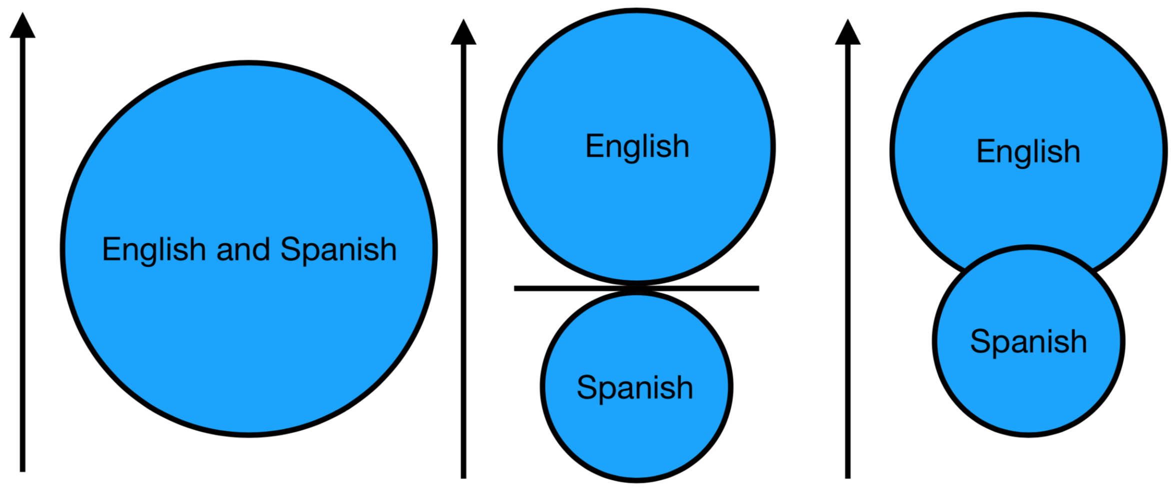

FIGURE 5
STRUCTURES OF ENTRY-LEVEL JANITORIAL LABOR MARKET

From a traditional perspective, all janitorial jobs form a single coherent labor market, as illustrated in the leftmost graph. Alternatively, dual market theory suggests a separation of primary segment and secondary segment, as shown in the middle graph in which some noneconomic barriers separate the two segments. This section provides empirical evidence to the segmentation hypothesis. Specifically, language is the noneconomic barrier dividing the English segment and the Spanish segments that have distinct wage-setting mechanisms.

However, more than 15% of sampled jobs are advertised simultaneously in English and Spanish. This group of jobs is illustrated in the rightmost graph. As the intersection of two segments, these jobs are part of the English segment and part of the Spanish segment. Moreover, language is no longer an active barrier. However, they do not fit in the wage equation of neither segment. The intriguing question is why workers in the Spanish segment do not move to the bilingual market and why workers in the bilingual market do not move to the English segment. By doing so, they could have landed a job with more attractive compensation. After all, in either case, language would not be a barrier for them to seeker better job offers. One explanation for this paradoxical finding is the existence of multiple barriers of entry. While linguistic barrier is not at work for one person, he or she faces legal barriers such as work authorization to prevent him or her to leave the Spanish or the bilingual segment for better-pay jobs.

We can classify any janitorial worker into one of the six types below.

TABLE 5
TYPES OF WORKERS BY LEGAL STATUS AND LINGUISTIC SKILL

	Authorized to work	Unauthorized to work
Speak English only	Type 1	Type 4
Speak Spanish only	Type 2	Type 5
Speak both English and Spanish	Type 3	Type 6

Assume that a small number firms in both segments do hire illegal workers. Let us consider a Type 5 worker who currently has an illegal job. He or she is unable to find a legal job in the Spanish-only segment or to find a better-paying illegal job in the English-only segment. Even for a Type 6 worker who masters two languages, chances are that he or she will still have a hard time finding a job that pays better, because illegal jobs generally pay less than legal jobs. Even for Type 2 people with valid work authorizations, they still cannot land a job in the bilingual market because they do not speak English in addition to Spanish.

One demand-side explanation rests on the fact that there exist lawbreaking firms that take advantage of unauthorized workers by paying them less than the state minimum. In my 1000 sampled jobs, I have found two jobs that advertise wages below their respective state minimum wages. If these recruiters did not unintentionally post the wrong wage, they must confidently believe in their ability to fill the illegal vacancies. In other words, they believe that workers of Type 4, Type 5, or Type 6 would not have any other choices and have to accept a low pay rate just to make a living.

V. Roles of Language

(a) Self-Selection among Janitors

Previous researches have investigated the economic return of acquiring a second language. In the literature, language is commonly seen as a skill. For example, Dustmann and Soest estimate around a 5% economic return for speaking German as an immigrants' second language. In my full model (Table 3), however, the returns of speaking English is 2.24% and that of speaking Spanish is -4.01% (0.016). Why does Spanish-speaking have negative wage return?

Let us also take a close look at the first two OLS models in Table 4. Within the English-only segment, the return to English-speaking is 2.76% (0.0166) but that of speaking Spanish is -5.96% (0.0529). Within the Spanish-only segment, the return to English-speaking is 2.41% (0.0606) but that of speaking Spanish is 9.52% (0.0529). Although these estimates are not statistically significant, we could still interpret their signs. English-speaking is always value-adding whereas Spanish-speaking is only value-adding in the Spanish segment. In other words, both the primary and the secondary markets award the primary language—English. On the other hand, the primary market punishes the secondary language—Spanish.

One hypothesis is that languages send signals to workers if they are welcomed as applicants. DiNardo and Pischke (1997) study the return of using office tools like pencils and hand tools like hammers. They suggest that the high return of computers may be attributed to the selection effect. Similarly, in our case, a janitor may be more encouraged to apply to openings that are advertised in his or her mother tongue. If Spanish-speaking janitors and English-speaking janitors have heterogeneous socioeconomic characteristics, the same job would draw a different pool of applicants if a different language were used. Because English-speaking workers might have more desirable characteristics as a group, deviations in the coefficients may only reflect the result of sorting among workers.

At the same time, firms may respond to this self-sorting with an increase or decrease of the wage, which could contribute to the wage gap. The study on gender-targeted job ads by Kuhn, Shen, and Shuo (2018) finds that gender profiling discourages prospective applicants from applying to targeted ads. Recruiters sometimes introduce mismatch penalty such as low call-back rates to those not belonging to the targeted gender. Likewise, janitorial firms advertising in English alone may only look for English-speaking custodians and those advertising in Spanish may only seek Spanish-speaking custodians. As a result, the 5.96% wage penalty to Spanish-speaking in the English segment may be a mismatch penalty. However, this does not explain the 2.41% premium to English-speaking in the Spanish segment. One explanation is that English is always desirable as a primary language. Another interesting observation is that the reward to Spanish-speaking far exceeds the reward for English-speaking by 7%, which could represent a matching prize in the Spanish segment.

(b) Screening Effect

One advantage of advertising jobs in Spanish is that only Spanish-speakers can read and hence apply to these jobs. Do recruiters use language(s) in job ads to screen workers? To study the probability of firms explicitly asking for language skills, we fit a logistic regression on 'SPA,' 'ENG,' and 'Bilingual.'

Although the coefficients for 'Spanish Only' are not statistically different from zero, the coefficient for 'English Only' has a fairly large critical values. In the first regression, the negative coefficient of -1.109 means that jobs only advertised in English are less likely to specify Spanish-speaking skills. Similarly, the coefficient of -1.236 in the third regression means that jobs advertised in English alone are less likely to ask for both Spanish-speaking and English-speaking skills. It may suggest that workers applying to jobs advertised in English have various English proficiency. Consequently, 'English' may not be an effective screening device for

TABLE 6
LOGISTIC REGRESSION: SIGNALING AND SCREENING EFFECTS OF LANGUAGES

	(1) SPA	(2) ENG	(3) Bilingual
main			
Work Authorization	0.830	0.545**	0.327
	(0.500)	(0.209)	(0.589)
Background Check	−0.662	0.390	−0.625
	(0.574)	(0.199)	(0.641)
Full-time	0.334	−0.625*	−0.0497
	(0.607)	(0.258)	(0.687)
Part-time	0.492	−0.868***	−0.134
	(0.603)	(0.256)	(0.672)
Driver License	1.213*	0.700***	1.381*
	(0.491)	(0.210)	(0.536)
High School	−0.772	0.364	−0.571
	(1.054)	(0.251)	(1.068)
Years of Relevant Work	0.293	−0.0647	0.360
	(0.280)	(0.146)	(0.294)
Log Local Mean Wage	1.305	3.602***	1.586
	(1.599)	(0.669)	(1.840)
Highest Immigration (%)	0.110	0.0267	0.256
	(0.512)	(0.205)	(0.568)
Spanish Only	0.225	0.153	−0.283
	(0.843)	(0.589)	(1.118)
English Only	−1.109*	0.480	−1.236*
	(0.521)	(0.264)	(0.568)
Constant	−7.562	−11.36***	−7.893
	(4.375)	(1.841)	(5.032)
Observations	947	947	947

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

English-speaking skills and employers hence adapt by explicitly stating the necessity of English skill in their job description.

To summarize, language has a limited return as a skill. However, languages used in the advertisements play two roles—self-selection and screening. Job-seekers might interpret the language as signals as to whether they are welcome to apply, which then leads to them to self-select into different jobs. Moreover, if firms intentionally use languages to profile workers, mismatched workers may be penalized. For example, Spanish-speaking janitors who apply to jobs exclusively written in English are penalized. Another role of languages is screening. Because one must understand the job description before applying, firms can intentionally exclude those who do not speak the language selected for advertising.

VI. Legal and Illegal Workers

By the federal law, employers are required to verify the employment eligibility of their employees both before hiring and during employment. Firms knowingly hire unauthorized workers could face monetary penalty up to 20,130 dollars per violation or criminal prosecution for serious violation, according to U.S. Immigration and Customs Enforcement (2020). Although it is against the law to hire unlawful workers, some employers still do so. Because of the difficulty in collecting data on illegal workers, the estimation of wage penalty that they face has been challenging. Borjas and Cassidy (2019) find that the average wage gap between undocumented and legal immigrants is over 35 percent but this wage penalty shrinks to 4 percent after controlling for observable socioeconomic characteristics.

In this section, I estimate the legal premium—the return of being a legal worker, and the illegal penalty—the penalty of being an illegal worker in Spanish segment. These two estimates need not be identical. Let us assume firms bear legal risks when recruiting illegal workers. In a industry where everyone has a work authorization, being a legal worker would not be rewarded and being an illegal would find it hard to get a job anywhere because of the abundant supply of legal workers. However, in the case when the majority of workers are illegal, legal workers would receive legal premium because firms bear zero risk when hiring them.

(a) Legal Premium

When a job ad asks for a valid work authorization, it is possible that some unlawful workers still get hired due to negligence of recruiters. When a job ad does not explicitly require a work authorization, recruiters might still be law-abiding and no illegal workers get hired. It is thus unrealistic to assume that firms ask for work authorization if and only if all their workers are legal. Nevertheless, this approach can be a decent start. In Table 7, the regression coefficient for 'work authorization' is my estimate of the legal premium. As shown in the first column of the regression table, the average return of having work authorization is 6%. However, after adding the full set of controls, the estimated premium dropped to 2.96%.

(b) Illegal Penalty

Because my dataset does not indicate if any firm hires unlawful workers, I use the law of iterated expectation to estimate the illegal penalty. Let us assume that firms explicitly require work authorization if and only if they do not hire unauthorized workers. Denote the log wage by W. Let $I_{Spanish}$ be the dummy variable indicating the use of Spanish in job ad and let X be the vector of full control variables excluding $I_{Spanish}$. Also, let β and γ respectively be the slope coefficients for $I_{Spanish}$ and X. Specifically,

$$X = (1, X_{English}, X_{ENG}, X_{SPA}, X_{BkgdCheck}, X_{FullTime}, X_{PartTime}, X_{DriverLicense}, X_{HighSchool}, X_{YearsExp}, X_{LocalMean}).$$

Write the population model as

$$W = I_{Spanish}\beta + X\gamma + \varepsilon, \text{ in which } \varepsilon \text{ is the error.}$$

First, let us condition on the event that work authorization is stated in the job description. Next, condition on the event that work authorization is stated in the job description plus that jobs asking for work authorization only hire legal workers. Using the law of iterated expectation, $\hat{\beta}$, the estimated coefficient for applying to jobs written in Spanish, can be decomposed as

$$\hat{\beta} = \hat{\mathbb{P}}(\text{WorkAuth} = 1)\hat{\beta}_{\text{WorkAuth=1}} + \hat{\mathbb{P}}(\text{WorkAuth} = 0)\hat{\beta}_{\text{WorkAuth=0}},$$

$$\hat{\beta} = \hat{\mathbb{P}}(\text{Legal} = 1)\hat{\beta}_{\text{Legal=1}} + \hat{\mathbb{P}}(\text{Legal} = 0)\hat{\beta}_{\text{Legal=0}}.$$

TABLE 7
LEGAL PREMIUM

	(Short Regression) Log Wage	(Long Regression with Full Controls) Log Wage
Work Authorization	0.0599***	0.0296*
	(0.0161)	(0.0127)
Log Local Mean Wage		0.730***
		(0.0348)
SPA		−0.0401
		(0.0365)
ENG		0.0224
		(0.0145)
Background Check		0.00936
		(0.0114)
Full-time		0.0102
		(0.0140)
Part-time		−0.0841***
		(0.0140)
Driver License		0.0585***
		(0.0132)
High School		0.0366*
		(0.0149)
Years of Relevant Work		0.0157*
		(0.00798)
Spanish		−0.0814***
		(0.0242)
English and Spanish		0.0265
		(0.0260)
Highest Immigration (%)		0.0349**
		(0.0116)
Constant	2.533***	0.625***
	(0.00756)	(0.0924)
Observations	947	947
Adjusted R^2	0.013	0.443

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 8
WAGE PENALTY TO ILLEGAL WORKERS IN SPANISH SEGMENT

	(All Jobs) Log Wage	(Jobs Stating Work Auth) Log Wage	(Jobs Not Stating Work Auth) Log Wage
Spanish	-0.0583^{***}	-0.0706^{**}	-0.0554^{***}
	(0.0125)	(0.0251)	(0.0145)
SPA	−0.0375	−0.0758	−0.00726
	(0.0366)	(0.0564)	(0.0481)
ENG	0.0250	0.0369	0.0108
	(0.0145)	(0.0258)	(0.0179)
Background Check	0.0157	0.0142	0.00815
	(0.0112)	(0.0213)	(0.0136)
Full-time	0.0104	0.0138	0.0101
	(0.0140)	(0.0327)	(0.0157)
Part-time	-0.0836^{***}	-0.0697^{*}	-0.0873^{***}
	(0.0140)	(0.0317)	(0.0158)
Driver License	0.0594^{***}	0.0686^{**}	0.0577^{***}
	(0.0132)	(0.0247)	(0.0157)
High School	0.0359^{*}	0.0223	0.0416^{*}
	(0.0149)	(0.0329)	(0.0169)
Years of Relevant Work	0.0170^{*}	0.0217	0.0120
	(0.00798)	(0.0154)	(0.00950)
Log Local Mean Wage	0.738^{***}	0.835^{***}	0.705^{***}
	(0.0346)	(0.0743)	(0.0396)
Highest Immigration (%)	0.0372^{**}	0.0230	0.0393^{**}
	(0.0116)	(0.0237)	(0.0135)
Constant	0.608^{***}	0.365	0.694^{***}
	(0.0920)	(0.201)	(0.105)
Observations	947	208	739
Adjusted R^2	0.440	0.532	0.400

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We are interested in estimating $\beta_{\text{Legal=0}}$, which is the wage penalty for unlawful workers seeking employment in the Spanish segment. Since $\hat{\beta}$ is know and $\hat{\mathbb{P}}(\text{Legal} = 0) = 1 - \hat{\mathbb{P}}(\text{Legal} = 1)$, it suffices to find $\hat{\beta}_{\text{Legal=1}}$ and $\hat{\mathbb{P}}(\text{Legal} = 1)$.

For $\hat{\beta}_{\text{Legal=1}}$, I assume that $\hat{\beta}_{\text{Legal=1}} = \hat{\beta}_{\text{WorkAuth=1}}$, of which the latter is known. In words, regardless if recruiters "remember" to explicitly state work authorization, they treat legal workers consistently. For $\hat{\mathbb{P}}(\text{Legal} = 1)$, I introduce α to account for the unobserved proportion of firms that do not ask for work authorization but do check the employment eligibility during hiring. Because $\widehat{\mathbb{P}}(\text{WorkAuth = 1}) = 0.2194$, we can write

$$\hat{\mathbb{P}}(\text{Legal} = 1) = 0.2194 + \alpha \text{ , where } 0 \leq \alpha \leq 1 - 0.2194 = 0.7806.$$

By rearranging terms, we have the equation

$$\begin{aligned}
\hat{\beta}_{\text{Legal=0}} &= \frac{\widehat{\beta} - \hat{\mathbb{P}}(\text{Legal} = 1)\hat{\beta}_{\text{Legal=1}}}{\hat{\mathbb{P}}(\text{Legal} = 0)} = \frac{\widehat{\beta} - \hat{\mathbb{P}}(\text{Legal} = 1)\hat{\beta}_{\text{Legal=1}}}{1 - \hat{\mathbb{P}}(\text{Legal} = 1)}, \\
&= \frac{(-0.0583) - (0.2194 + \alpha) \cdot (-0.0706)}{1 - (0.2194 + \alpha)} \text{ where } 0 \leq \alpha \leq 0.7806, \\
&\geq -0.05422.
\end{aligned}$$

This bound shows that illegal workers seeking jobs in Spanish janitorial market would face at most 5.42% wage penalty. By treating $\alpha \in [0, 0.7806]$ as a continuous random variable, we can plot the illegal penalty in Spanish segment as a curve.

I then assume that wage penalty for illegal workers is always nonnegative, i.e., workers do not get paid more for being illegal. This means that we can discard the rightmost part where illegal workers are rewarded. Hence, the proportion of jobs actually verifying work authorization can be at most 82.6%. Furthermore, the possible values of α becomes $[0, 0.606]$. If an 'average' estimate is desired, let us assume $\alpha \sim \text{Uniform}[0, 0.606]$ and $f_\alpha(x) = \frac{1}{0.624}$ for $x \in [0, 0.624]$.

The expected value for $\hat{\beta}_{\text{Legal=0}}$ is therefore

$$\mathbb{E}\left[\hat{\beta}_{\text{Legal=0}}\right] = \int_0^{0.606} \frac{1}{0.0606} \cdot \frac{(-0.0583) - (0.2194 + \alpha) \cdot (-0.0706)}{1 - (0.2194 + \alpha)} \, d\alpha = -2.44\%,$$

which is lower but reasonably close to the estimate of 4% obtained by Borjas and Cassidy.

The curve demonstrates an inverse relation between the proportion of firms verifying employment eligibility and the wage penalty faced by illegal workers in the Spanish segment. At the

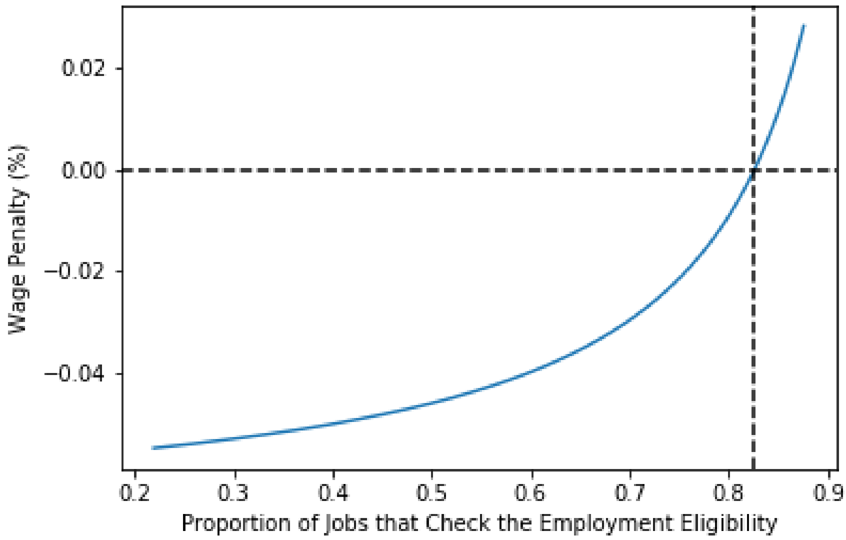

FIGURE 6
WAGE PENALTY TO ILLEGAL WORKERS IN SPANISH SEGMENT

far left of the plot in which only a 21.9% jobs check for legal working status, the industry is the least law-abiding and illegal workers are penalized the most. As the whole industry becomes more law-abiding, however, the wage penalty decreases and approaches zero.

VII. CONCLUSIONS

Using a random sample of janitorial jobs from indeed.com, my paper investigates the wage gap between jobs advertised in Spanish and jobs advertised in English. There are barriers of entry and distinct wage-setting mechanisms between the English segment and the Spanish segment. They provide evidence for the segmentation of U.S. janitorial labor market. While linguistic skills have limited return, language used in the ad encourages sorting among workers and helps firms seek workers with preferred qualification. If firms ask for work authorization if and only if all of their workers are legal, the estimated legal premium is 2.96%. Assuming workers are not rewarded for being illegal, the illegal penalty in the Spanish segment is 2.44%. It is worth noting that the legal premium and the illegal penalty are fairly close in magnitudes. Furthermore, the more firms checking work authorization, the smaller wage penalty faced by illegal workers.

Current language programs aim to prepare immigrants for the labor market by training those with limited or no English skill to acquire working knowledge of English. While I do not observe significant return to the language skill itself, acquiring basic English skill could nevertheless enable low-skilled workers to understand job ads written in English and remove the linguistic barrier of entry. Because the primary English segment has better returns to work experience and education, those workers could benefit from an increase in wage.

As an increasing number of firms uses E-verify to check work authorizations, unauthorized workers may be denied access to better-paying jobs or any legal jobs at all. If they gained authorization to work, they would gain a 2.96% legal premium. Interestingly, wage tends to be 3.49% higher in states with the largest populations of illegal immigrants. The greatest return to wage comes from a driver's license for which the pooled estimate is 5.85%. The estimate for English market alone is 4.97%, but the estimated return for the Spanish market is as high as 19.3%, more than triple the former estimate. Since acquiring a driver's license also allows job-seekers to apply to jobs inaccessible by public transportation, driver's license is one of the most rewarding investments for unlawful Spanish workers. Moreover, Holzer, Raphael, and Stoll (2003) demonstrate employment barriers facing ex-offenders and note that some state laws prohibit ex convicts from obtaining driver's license. The importance of driver's license to the U.S. workforce should therefore not be underestimated in the policy-making process.

Because my dataset solely focuses on entry-level janitorial jobs, it is unclear if my results hold for a wider range of occupations. As suggested by Boyd and Cao (2009), relative advantage of English or French reduces in the high-skilled sector. It is reasonable to expect that English and Spanish play a diminished role in high-skilled occupations. Moreover, the type of labor market segmentation studied in this article is likely to be absent for these occupations, since almost all skilled workers in the U.S. speak English so that linguistic barrier does not exist. Nevertheless, as online job board becomes an increasingly important platform for both recruiters and job-seekers, it has great potential to provide rich and insightful first-hand data.

VIII. Appendix

TABLE 9
COVARIATE CODEBOOK

Covariate Name	Description
Spanish	At least one sentence in the job description is written in Spanish
English	At least one sentence in the job description is written in English
English and Spanish	Fulfill the above two criterions, i.e., 'Spanish' and 'English'
Spanish Only	The job description is entirely written in Spanish.
English Only	The job description is entirely written in English.
SPA	Spanish language skill is explicitly required
ENG	Spanish language skill is explicitly required
Bilingual	Fulfill the above two criterions, i.e., 'SPA' and 'ENG'
Work Authorization	Work Authorization or SSN or U.S. citizenship required
WA*Spanish	Product of 'Work Authorization' and 'Spanish'
WA*English and Spanish	Product of 'Work Authorization' and 'English and Spanish'
Background Check	Legal or criminal background check/screening required
Full-time	Label of full-time or word description of at least 35 working hours
Part-time	Label of part-time or word description of less than 35 working hours
PT*Spanish	Product of 'Part-time' and 'Spanish'
PT*English and Spanish	Product of 'Part-time' and 'English and Spanish'
Driver's License	Valid driver's License, either Class B or commercial license, is required
DL*Spanish	Product of 'Driver's License' and 'Spanish'
DL*English and Spanish	Product of 'Driver's License' and 'English and Spanish'
High School	High-school degree or GED required
HS*Spanish	Product of 'High School' and 'Spanish'
HS*English and Spanish	Product of 'High School' and 'English and Spanish'
Years of Relevant Work	Number of years of relevant work experience
Wage	Mean of the min and max wage as shown in a given job advertisement
Log Wage	Logged Mean of the min/max wage displayed in the job advertisement
Local Mean Wage	Mean wage of entry-level janitors of a given metropolitan area
Highest Illegals (#)	Top three states with the largest counts of illegal immigrants
Highest Illegals (%)	Top three states with the largest proportions of illegal immigrants

TABLE 10
SUMMARY STATISTICS

	count	mean	sd	min	p25	p50	p75	max
English Only								
Wage	725	13.20957	2.928432	7.35	11	12.84	15	29.59
SPA	725	.0137931	.1167118	0	0	0	0	1
ENG	725	.1834483	.3873008	0	0	0	0	1
Bilingual	725	.0110345	.1045361	0	0	0	0	1
Work Authorization	725	.2068966	.4053603	0	0	0	0	1
Background Check	725	.3310345	.4709103	0	0	0	1	1
Full-time	725	.417931	.4935592	0	0	0	1	1
Part-time	725	.5903448	.4921096	0	0	1	1	1
Driver License	725	.1972414	.3981908	0	0	0	0	1
High School	725	.1751724	.3803769	0	0	0	0	1
Years of Relevant Work	725	.2871255	.6712773	0	0	0	0	5
Spanish Only								
Wage	46	11.28565	1.896641	7.25	10	11	12.5	15.5
SPA	46	.0434783	.2061846	0	0	0	0	1
ENG	46	.0869565	.2848849	0	0	0	0	1
Bilingual	46	.0217391	.147442	0	0	0	0	1
Work Authorization	46	.1521739	.3631584	0	0	0	0	1
Background Check	46	.1956522	.4010855	0	0	0	0	1
Full-time	46	.6304348	.4880207	0	0	1	1	1
Part-time	46	.5434783	.5036102	0	0	1	1	1
Driver License	46	.1304348	.3405026	0	0	0	0	1
High School	46	.0434783	.2061846	0	0	0	0	1
Years of Relevant Work	46	.1630435	.3659743	0	0	0	0	1
Spanish and English								
Wage	177	12.77698	2.337413	7.5	11	12.25	14.5	18.25
SPA	177	.0451977	.2083269	0	0	0	0	1
ENG	177	.1299435	.3371951	0	0	0	0	1
Bilingual	177	.039548	.1954477	0	0	0	0	1
Work Authorization	177	.2881356	.4541794	0	0	0	1	1
Background Check	177	.2655367	.4428714	0	0	0	1	1
Full-time	177	.6666667	.4727418	0	0	1	1	1
Part-time	177	.4576271	.4996146	0	0	0	1	1
Driver License	177	.1694915	.3762498	0	0	0	0	1
High School	177	.0282486	.1661523	0	0	0	0	1
Years of Relevant Work	177	.2033898	.5548237	0	0	0	0	3
Total								
Wage	948	13.03544	2.815502	7.25	11	12.5	15	29.59
SPA	948	.021097	.1437838	0	0	0	0	1
ENG	948	.1687764	.374752	0	0	0	0	1
Bilingual	948	.0168776	.128881	0	0	0	0	1
Work Authorization	948	.2194093	.4140649	0	0	0	0	1
Background Check	948	.3122363	.4636502	0	0	0	1	1
Full-time	948	.4746835	.4996222	0	0	0	1	1
Part-time	948	.5632911	.4962399	0	0	1	1	1
Driver License	948	.1888186	.391571	0	0	0	0	1
High School	948	.1413502	.3485663	0	0	0	0	1
Years of Relevant Work	948	.2654705	.6400517	0	0	0	0	5
Observations	948							

IX. References

Alcobendas, Miguel A., and Rodriguez-Planas, Nuria. 2009. Immigrants' Assimilation Process in a Segmented Labor Market. IZA Discussion Papers no. 4394: 31

Boyd, Monica, and Cao, Xingshan. 2009. Immigrant Language Proficiency, Earnings, and Language Policies. Canadian Studies in Population 36 no.1: 63-86

Borjas, George J., and Cassidy, Hugh. 2019. The Wage Penalty to Undocumented Immigration. Working Paper

Card D. 2009. Immigration and Inequality. American Economic Review 99 no. 2:1-21

Clark, Damon, and Martorell, Paco. 2014. The Signaling Value of a High School Diploma. Journal of Political Economy 122 no. 2: 282-318

Corcoran, Mary, and Duncan, Greg J. 1979. Work History, Labor Force Attachment, and Earnings Differences between the Races and Sexes. The Journal of Human Resources 14 no. 1:3-20

Dickens, William, and Kevin Lang. 1984. A Test of Dual Labor Market Theory. NBER Working Paper no. 1314

DiNardo, John E., and Pischke, Jorn-Steffen. 1997. The Returns to Computer Use Revisited: have Pencils Changed the Wage Structure Too? The Quarterly Journal of Economics 112 no. 1:291-303

Dustmann, Christian, and Van Soest, Arthur. 2002. Language and the Earnings of Immigrants. ILR Review 55 no. 3:473-492

Form I-9 Inspection Overview. 2019. Retrieved from https://www.ice.gov/factsheets/i9-inspection.

Holzer, Harry J., Raphael, Steven, and Stoll, Michael A. 2003. Employment Barriers Facing Ex-offenders. Urban Institute Reentry

Jaeger, David A., and Page, Marianne E. 1996. Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education. The Review of Economics and Statistics 78 no. 4:733-740

Kleiner, Morris M., and Krueger, Alan B. 2010. The Prevalence and Effects of Occupational Licensing. British Journal of Industrial Relations 48 no. 4:676-687

Kuhn, Peter, Shen, Kailin, and Zhang Shuo. 2018. Gender-targeted Job Ads in the Recruitment Process: Evidence from China. NBER Working Paper no. 25365

Reich, Michael, Gordon, David, and Edwards, Richard C. 1973. A Theory of Labor Market Segmentation. The American Economic Review 63 no. 2:359-365

Rodgers, Joan R. 2002. Hourly Wages of Full-Time and Part-Time Employees in Australia. Australian Journal of Labour Economics 7 215-238

Simpson, Wayne. 1986. Analysis of Part-Time Pay in Canada. The Canadian Journal of Economics 19 no. 4: 798-807