Acculturation and the Labor Market in Mexico

Javier Cano-Urbina[*]   and   Patrick L. Mason[†]

December 20, 2016

Abstract

This paper empirically examines the relationship between the self-identity as Indigenous and earnings inequality in the Mexican labor market. Using Mexican Census data and a large set of wage covariates reveals the existence of an earnings penalty for self-identification as Indigenous. There is an additional and larger penalty for Indigenous persons who are fluent in an Indigenous language, regardless of Spanish language fluency. Further analyses using the Mexican Family Life Survey reveal that these earnings gaps persist after we also control for an individual's cognitive ability. Ethno-linguistic inequality is particularly strong in smaller cities and among self-employed workers.

[*]Corresponding author. Department of Economics, The Florida State University. 113 Collegiate Loop, 257 Bellamy Building, Tallahassee, Florida 32306. E-mail: jcanourbina@fsu.edu.

[†]Department of Economics, The Florida State University. 113 Collegiate Loop, 266 Bellamy Building, Tallahassee, Florida 32306.

1 Introduction

There is a vast literature in economics that addresses the issue of group differences in labor market outcomes (e.g. Lang, 1986; Neal and Johnson, 1996; Darity Jr. and Mason, 1998; Mason, 2004; Darity Jr. et al., 2006; Lang and Manove, 2011). This paper addresses group differences in labor market outcomes in Mexico under the light of the wealth of literature for the United States. Our study focuses on the labor market differences in outcomes between Indigenous workers and those who are not Indigenous. We find that Indigenous workers carry a significant penalty in the labor market and that this penalty increases in the level of attachment to the Indigenous group. In particular, we find an earnings penalty for those individuals who self-identify as members of an Indigenous group, but we also find that those individuals who also speak their Indigenous language carry an even higher penalty. This result persists after including a large set of controls that include, among others, the individual's education, cognitive ability and fluency in Spanish.

There are many reasons to believe that the situation in the Mexican labor market is similar to that of other Latin-American and Caribbean countries where there exists clear ethnic differences in the population. For example, Atal, Ñopo, and Winder (2009) did an interesting study that surveys gender and ethnic wage gaps in 18 Latin American countries. However, Mexico is a particularly interesting case for the analysis of group differences in the labor market given the ideological and legal frameworks of this country. First, Mexico celebrates its racial and cultural mixing through the social ideal of *mestizaje*. *Mestizaje* is characterized by the belief that out of many Indigenous, African, and European peoples with diverse historical and political economic experiences, Mexico has formed one hybrid and inclusive national identity. In 2001, Mexico's Secretary of State made modifications and additions to the Mexican Constitution to represent the unique character of Mexico and its multi-cultural composition which is based on its Indigenous groups (see SEGOB, 2008a).

Second, starting in 1999 the Criminal Code for Mexico City prohibited "labor discrimination on the basis of age, sex, pregnancy, marital status, race, language, religion, ideology, sexual orientation, skin color, nationality, social origin or position, work or profession, economic position, physical characteristics, disability or state of health" (Commission for Labor Cooperation, 2006, p. 15). In 2002, this Criminal Code further outlawed discrimination on the basis of ethnic origin (see Art. 206 in ALDF, 2002). Similarly, in 2003 Mexico's Secretary of State published the Federal Law to Prevent and Eliminate Discrimination (see SEGOB, 2008b) and in 2014, Articles 3 and 133 of the Federal Work Law of Mexico were modified to prohibit labor discrimination due to ethnic origin.

Third, Mexico has 68 officially recognized Indigenous languages described in the Catalog of National Indigenous Languages published in 2008 (see INALI, 2008). This is "the largest population of speakers of indigenous languages [···] and presents the highest cultural diversity in the Americas, with regard to the number of languages spoken" (Terborg et al., 2006, p. 417). Bilingual education, where the ability to read, write, and speak Spanish is taught through transitional programs in Indigenous languages, has been an official Mexican policy since the 20th century; hence, during 1921-2000 the proportion of bilingual Indigenous persons expanded from 38 percent to over 81 percent of the Indigenous population (Vazquez Carranza, 2009). The 2003 General Law on the Linguistic Rights of Indigenous Peoples outlaws linguistic discrimination in the work, education, and judicial environments.[1]

Therefore, the finding that Indigenous self-identification carries an earnings penalty and that the penalty is higher if the individual is also capable of speaking an Indigenous language

---

[1]In particular this law states that: (i) All Mexicans have the right to speak their language without restriction of any kind and without any kind of discrimination; (ii) Spanish and Indigenous languages have equal status and both are valid in any public or private sector and in any kind of social activity; (iii) The right of Indigenous People to bilingual and bicultural education in the compulsory levels, respecting and dignifying their cultural identity; (iv) The right to have access to the judicial system through Indigenous languages; and, (v) The State and its three governmental orders (Federation, States, and Municipalities) will protect, preserve, promote and develop the Indigenous languages through the participation of the Indigenous population and their communities (see Vazquez Carranza, 2009, p. 205).

is at odds with the ideological and legal environment of Mexico. This finding is also at odds with the theory of human capital, as "Language skills satisfy the three requirements for human capital, that it is productive, costly to produce, and embodied in the person" (Chiswick, 2008, p. 4). Understanding that Spanish is the dominant language in Mexico, if Indigenous languages have no value in the labor market, at the minimum the ability to speak such a language should be orthogonal to the labor market outcomes of the individual.

Chiswick et al. (2000) found similar results for Bolivia, and they conclude that "Bilingual speakers may be penalized in the labor market because of a poorer proficiency in Spanish" (see Chiswick et al., 2000, p. 365). We address this issue in our study by controlling for whether individuals report speaking Spanish at home. Speaking Spanish at home has a positive effect on earnings, but it does not account for the whole earnings gap.

Evidently, *Mestizaje* is not a one-dimensional ideal, but it also associates "progress" with European or white acculturation. Hence, despite the appearance of inclusiveness, *blanqueamiento* (or whitening) has been and remains an important element of mestizaje (Safa, 2005). Mexico, in that regard, is similar to the US, where dark complexion/Indian phenotype Mexican Americans have a 12 percent income penalty relative to light complexion/European phenotype Mexican Americans and where not speaking Spanish raises an individual's earnings by 13 percent, holding constant a Mexican American's English-fluency (Mason, 2004).[2]

Our study updates the literature of group differences in the labor market for Mexico in regards to the Indigenous population (e.g. Panagides, 1994; Patrinos, 1994; Hisamatsu and Ukeda, 2002). Also, this study substantially improves upon the regression analysis of this important topic. We exploit both the Mexican Census and the Mexican Family Life Survey to address group differences in the labor market. The Mexican Census allows us to study

---

[2]Another strand of the literature has focused on immigrant assimilation, for example McManus et al. (1983) or Lleras-Muney and Shertzer (2015), which can be associated with the gains from acculturation for immigrants. Here we focus on the gains from acculturation for individuals born in the same country but that belong to non-dominant ethnic groups.

the earnings differential between those self-identified as Indigenous and those who are not and the Mexican Family Life Survey allows us to control for cognitive ability and proficiency in Spanish, but at the expense of smaller sample sizes.

The most general finding of this paper is that Indigenous identification carries an earnings penalty compared to non-Indigenous workers and that workers self-identified as Indigenous who preserve their Indigenous roots, which manifests in their ability to speak an Indigenous language, have an even larger earnings penalty. While the earnings penalty for Indigenous self-identification is not a new result, there has been a limited effort on finding the mechanism behind this gap (e.g. see Borja-Vega, Lunde, and García Moreno, 2007). We discuss possible channels that can generate the earnings gap observed in the data such as differences in job search strategies and labor market discrimination. We also explore whether the earnings gap is the result of occupational segmentation and we find that both wage/salary workers and self-employed workers suffer an earnings penalty but the penalty for self-employed Indigenous workers is substantially larger. Another result from the analysis reveals that the earnings gap is mainly a small-communities phenomenon, Indigenous workers still suffer an earnings penalty in large cities, but the penalty is substantially larger in small communities. We explore other possible explanations for the earnings gap such as labor market segmentation by education or ability, the persistence of Indigenous self-identification, and whether the Indigenous language leads Indigenous workers to have different behavior towards skills investment and the labor market.[3]

The paper is organized as follows. Section 2 presents the results from the Mexican Census. Working with the Mexican Census allows us to work with a large sample drawn from all Mexican states and also allows us to perform the analysis by the type of Indigenous language. Section 2.1 allows us to address some of the concerns with the estimation results

---

[3]A recent study by Sandoval (2014) documents the lack of educational mobility among Indigenous people, and so our paper may help explain one source of the lack of mobility, that is, earnings penalties in the labor market to Indigenous people could be a disincentive to attain higher educational levels.

from the Census since we can control for cognitive ability. Section 3 discusses alternative channels that could generate the earnings gap observed in the data. Section 4 concludes.

2   Group Differences in the Mexican Labor Market

In this section, we use data from the 2010 Mexican Census to gauge group differentials in labor market outcomes between the Indigenous and non-Indigenous population. The census data are collected and distributed by the *Instituto Nacional de Estadística y Geografía* from Mexico, INEGI (for its acronym in Spanish), which is the Mexican National Statistics Bureau.[4]

The 2010 Mexican Census collects two pieces of information that are especially relevant for our study. First, it collects information on whether an individual is capable of speaking an Indigenous language as well as the individual's ability to speak Spanish, which is the dominant language in Mexico.[5] Second, it collects information on whether an individual self-identifies as a member of the Indigenous population.[6] In what follows, we call Mestizos those individuals who do not self-identify as Indigenous, and we call Indigenous those individuals who self-identify as Indigenous.

A first step towards understanding group differentials between Indigenous and Mestizo workers in the Mexican labor market is to understand the geographical location of these groups. Figure 1 shows that the Indigenous population in Mexico is concentrated in the central and southern states of Guerrero, Puebla, Hidalgo, Campeche, Chiapas, Quintana Roo, Oaxaca, and Yucatan. In the state of Guerrero, 22.69 percent of the population is Indigenous; in Puebla, is 25.31 percent; in Hidalgo, is 30.26; in Campeche, is 32.11; in

---

[4]We would like to thank the Department of Economics from the Universidad Autónoma de Nuevo León for providing the 2010 Census data.

[5]As mentioned in the Introduction, Mexico has 68 national Indigenous languages (INALI, 2008).

[6]Notice that, as in many household surveys, the individual in question might not be the respondent of the questionnaire.

Chiapas, is 32.82; in Quintana Roo, is 34.79; in Oaxaca, is 58.26; and in Yucatan, is 63.26 percent of the population. These proportions are quite high compared to the concentration of the African-American population in the United States, who represent about 21 percent of the population in those states with the highest concentrations of African-Americans, and coincidentally are also located in the south of the United States.

[Insert Figure 1 here]

For the analysis we focus on males ages 16 to 64 who are not attending school and have no missing values in all relevant variables. Appendix A provides more details on the sample used in the analysis and the construction of the key variables. Table 1 presents the sample summary statistics. Sample statistics are presented separately for Mestizos and Indigenous individuals. Within the Indigenous we further distinguish between those who are bilingual (speak both Spanish and an Indigenous language) and those who are monolingual. Finally, within the Indigenous Monolingual, we distinguish between those who only speak an Indigenous language (Monolingual-Indigenous) and those who only speak Spanish (Monolingual-Spanish). As indicated by the statistics in the table, these groups have important differences in key variables.

[Insert Table 1 here]

First, note that Monolingual-Indigenous are older than both bilingual Indigenous and Monolingual-Spanish, which suggests that younger generations are more acculturated. Second, note that Mestizos have more years of education on average than Indigenous, however within Indigenous there are substantial differences in educational attainment. Monolingual-Spanish have 8.56 years of education on average, whereas Monolingual-Indigenous have 1.68 years of education on average. These differences in educational attainment are consistent with the substantial differences in the labor market. For example, among those who are

7

monolingual in Indigenous language 47.88 percent are self-employed, they have the least hours of work per week, and have the lowest monthly earnings.

In terms of the size of residential location, Table 1 reveals negative correlation between Indigenous cultural attachment and population size of residence. That is, the vast majority of Monolingual-Indigenous live in very small communities, about 43 percent of the Bilingual live in this type of communities, but only 27 percent of the Monolingual-Spanish Indigenous workers live in very small communities.

Next, in terms of occupation, Monolingual-Indigenous are mostly agricultural and fishery and crafts workers, whereas the distribution of Monolingual-Spanish is closer to that of the Mestizos. Finally, in terms of the industry of occupation, Monolingual-Indigenous are concentrated in agriculture, and once again Monolingual-Spanish are distributed within industries in a way similar to Mestizos. Overall, the numbers in Table 1 indicate that Monolingual-Spanish are the most acculturated among the Indigenous groups.

In Table 2 we explore earnings differentials between Indigenous and Mestizos conditional on some of these observable characteristics. The dependent variable is the natural logarithm of monthly earnings. All specifications control for education, age, location size, and state of residence. The results in column (1) indicate that individuals who self-identify as members of an Indigenous group have earnings that are on average 9 percent lower than Mestizos and that individuals who speak an Indigenous language have earnings that are on average 18 percent lower than those who only speak Spanish. That is, the results suggest that the lower the degree of acculturation of the individual the higher the earnings penalty. Human capital covariates are significant and have the expected signs suggesting positive returns to the investment in human capital.

[Insert Table 2 here]

Next, in columns (2) and (3) of Table 2 we also control for occupation and industry.

Typically, for regressions examining group differences in earnings, occupation and industry are considered endogenous variables, since discrimination in occupation and industry of employment might be important paths for generating earnings inequality; hence, occupation and industry are not appropriate regressors. Nevertheless, we include these variables here under the extreme assumption that intergroup differences in occupation and industry are due solely to intergroup differences in ability and employment preferences. This extreme assumption works against finding statistically significant differences in the earnings of Indigenous and Mestizos. However, the estimates on Indigenous self-identification and speaking an Indigenous language remain significant and of the same order of magnitude.

Table 3 further explores the findings of Table 2, differentiating between individuals who are bilingual or monolingual in the Indigenous and Spanish languages. The dependent variable and the controls are the same as those in the regressions of Table 2. Columns (1)-(3) include only Indigenous people and the omitted category is Monolingual-Indigenous. Columns (4)-(6) include only Mestizos and the omitted category is Bilingual. The results indicate that among Indigenous males, those who are Bilingual have monthly earnings that are on average 30 percent higher than Monolingual-Indigenous language and those who are Monolingual-Spanish have monthly earnings that are on average 45 percent higher than Monolingual-Indigenous. Accounting for occupation reduces the gains of Bilingual and Monolingual-Spanish by about 30 percent while accounting for industry barely affects the estimates. Columns (2) and (3) indicate that, adding occupation and industry to the controls, Bilinguals have monthly earnings that are on average 19 percent higher than Monolingual-Indigenous while Monolingual Spanish have monthly earnings on average 32 percent higher than Monolingual-Indigenous.

[Insert Table 3 here]

The results in columns (4)-(6) indicate that Mestizos who are Monolingual-Spanish have

a wage advantage over Bilingual Mestizos of 6 percent. We examined the observable characteristics of both groups of male Mestizos, those who are Monolingual-Spanish and those who are Bilingual, and we found no significant difference in terms of education, work position, occupation, industry, or population size. Moreover, we include these variables as controls in the regression analysis. Although, the Bilingual differential for Mestizos is statistically significant, it is fairly small compared to the male language differentials for Indigenous in column (1). One possibility is that the differential reflects finer occupation and industry differences or systematic differences in unobserved ability. Hence we remain agnostic regarding the reason for this differential. One can argue then, that we should subtract 6 percent from the wage differentials in column (1) for males; even so, the earnings differential among the indigenous population remains quite sizable.[7]

Finally, notice that the results from Table 3 indicate that the returns to education are lower for Indigenous workers than for the Mestizo workers once we include the occupation and industry controls. Borja-Vega et al. (2007) find similar results in their study.

Since linguistic ability is a skill, especially fluency in Spanish — the dominant language, it is not surprising that Monolingual-Indigenous have lower earnings than both bilingual Indigenous and Monolingual-Spanish. But, it is counter-intuitive that bilingual speakers have lower earnings than Monolingual-Spanish. All other things equal, bilingual persons have greater skill than monolingual persons.

The results in both Tables 2 and 3 suggest that workers who speak an Indigenous language have an earnings disadvantage in the Mexican labor market. Now we explore whether this disadvantage is similar for different Indigenous languages. Table 4 presents the results from log-earnings regressions where we distinguish between the different Indigenous languages. As mentioned below, in Mexico there are 68 official Indigenous languages. These 68 languages

---

[7]The sample size for this regression is 1,337,916 and from those 10,502 are Bilingual Mestizos and the rest are Monolingual-Spanish Mestizos.

belong to 11 language families as described in the Catalog of National Indigenous Languages (see INALI, 2008). Here we combine three of these families into one since the proportion of people speaking an Indigenous language in each of these families is very small. The sample only includes those individuals who self-identify as Indigenous and the omitted category is Monolingual-Spanish. Table 4 also presents the proportion of Indigenous in the sample who speak a language that is affiliated with a particular language family. The estimates indicate that for males the Huave have the largest penalty followed by the Maya, the Tarasca, and the Yuto-Nahua. Interestingly, these language families, with the exception of the Tarasca, are among the largest Indigenous groups. Notice that for the case of the Chontal de Oaxaca they have an earnings advantage compared to those who are monolingual in Spanish, however they represent a very small portion of the Indigenous groups. In terms of the geographical location of these language families: the Huave are mainly concentrated in the state of Oaxaca, the Yuto-Nahua in the states of Veracruz and Puebla, the Maya in the states of Chiapas and Yucatan, the Tarasca in the state of Michoacan, and the Chontal de Oaxaca in Oaxaca.

[Insert Table 4 here]

To conclude, the results from the Mexican Census suggest that the labor market in Mexico values acculturation. That is, self-identification as Indigenous carries a penalty in the labor market, but this penalty is higher if the individual also speaks an Indigenous language. It is possible that the education received by Indigenous workers has lower quality than that of the Mestizos. Although we have no measures of school quality we do control for state of birth, which may be correlated with differences in school quality among states. It is also possible that there is an unobserved component common to less acculturated individuals that generates the earnings differentials observed in the data, e.g. unobserved ability. In an attempt to solve this puzzle we turn, in the next section, to the analysis of the Mexican Family Life Survey, a dataset that contains additional information that can shed light on

this question.

2.1 Measuring Group Differences with the Family Life Survey

In this section we explore group differences in the labor market of Mexico using data from the Mexican Family Life Survey (MxFLS). In particular, we work with three waves of the MxFLS: 2002, 2005-2006, and 2009-2012 (see Rubalcava and Teruel, 2006a,b, 2013, for details about this survey). Our sample is limited to persons 16 - 64 years of age and who are not currently attending school.

An advantage of working with the MxFLS is that we have measures of the respondent's cognitive ability. Cognitive ability is measured by Raven Progressive Picture Matrices. For each of 12 matrices, the respondent receives a score of one if the question is answered correctly and a score of zero if it is not answered correctly. A skill index, s_{it}, is created such that $s_{it} = \sum_{j=1}^{12} R_{it}^j$, where $R_{it}^j \in \{0, 1\}$ for question $j = 1, \ldots, 12$, for individual $i = 1, \ldots, n$ at year t. It is important to note that the matrices that compose this test were designed to measure cognitive ability independently of the individual's reading and writing skills and that respondents take the test every time they are visited, and so the test score can vary for the same individual. Following Neal and Johnson (1996) we create an age- and education-adjusted standardized cognitive skill index by standardizing the residuals obtained from regressing the test score s_{it} on individual's age and education.

A disadvantage of working with the MxFLS is that it does not cover all the Mexican territory (see Appendix B) and that we will have a smaller sample compared to the Census data. Another disadvantage of working with the MxFLS data is that given the order of the questions asked in the MxFLS questionnaire and the pattern of valid skips, we cannot perfectly identify Indigenous individuals who only speak an Indigenous language (the Monolingual-Indigenous), as we did with the Census data. That is, from the MxFLS data we know whether individuals speak Spanish in their households, but we do not know if they

speak Spanish outside the household. Hence, with the MxFLS data we can only identify three groups of individuals: (i) Mestizos, (ii) Indigenous monolingual in Spanish, (iii) Indigenous who speak an Indigenous language. The latter group will include Indigenous who only speak an Indigenous Language, but we cannot separately identify them from those who are bilingual in Spanish and an Indigenous language.[8]

Table 5 presents the summary statistics for the MxFLS data. Similar to what we found in the Census data, Indigenous who only speak Spanish have observable characteristics similar to those of Mestizos. Notice that on average the sample from the MxFLS is about one year less educated than the Census sample for all groups of individuals although the ranking of average years of education by group is preserved.[9] The differences in terms of the proportion of individuals who are self employed are not as stark as they were in the Census. Even though the classification used for occupations is somewhat different in the MxFLS and the Census data, the patterns are similar in both data sets, that is, Indigenous workers who only speak Spanish have a distribution of occupations similar to that of Mestizos, and half of the Indigenous who speak an Indigenous language work in agricultural activities compared to only a quarter of the other two groups. We do not use industry where the individual is employed when working with the MxFLS data because this information is missing in the third wave of the MxFLS.[10]

---

[8]In the MsFLS, within the respondents who declare to be Indigenous and speak an Indigenous language, we can identify those who speak or do not speak Spanish at home. There are 59 respondents in our sample who are Indigenous, speak an Indigenous language, and do not speak Spanish at home. Even though we cannot tell whether they speak Spanish outside their home (so we could identify the Monolingual-Indigenous as we did in the Census) it is clear that these 59 respondents are still less acculturated than those who speak Spanish at home. We did the analysis separately identifying these individuals and the results are very similar.

[9]This could also be a consequence of not using the sampling weights to calculate the means. We do not use the MxFLS sampling weights because we did not consider them reliable. In many cases the sampling weight was missing and in some others the sampling weight was zero. Sampling weights are calculated as the inverse of the probability that a given respondent is included in the sample, hence zero sampling weights are not a sensible weight. Missing sampling weights are also a problem.

[10]While working with the MxFLS data we will not be controlling for the industry where the respondent is employed because the information on the industry is not available for the last wave of the the survey and so we will loose about a third of the sample if we want to restrict to observations with a valid industry. However, as revealed by the results in Tables 2 and 3 it is the occupation control rather than the industry

[Insert Table 5 here]

Something new in the summary statistics compared to the Census data is that we have a measure of cognitive ability in the MxFLS data. Table 5 indicates that Mestizos are on average 0.07 standard deviations above the mean, Monolingual-Spanish Indigenous are 0.11 standard deviations below the mean, and Indigenous who speak an Indigenous language are 0.30 standard deviations below the mean.

Tables 6 and 7 address the question of whether group differences in cognitive ability explain the earnings gap between these groups. Table 6 presents estimates of a specification similar to the specification estimated in Table 2. Comparing columns (1) and (3) we can see that including the standardized test score has barely an effect on the estimates, while still suggesting a positive return of cognitive ability in the labor market. Similar to the estimates with Census data, the results indicate that Indigenous workers suffer an earnings penalty compared to Mestizos, but this penalty is higher if these individuals also speak an indigenous language. Columns (2) and (4) also condition on occupational controls and even though the estimates on the effect of speaking an Indigenous language are significantly reduced when including occupational controls they are not really affected by the inclusion of the test score. We observed a similar effect on the coefficient estimates when including occupational controls in the regressions using Census data. The similarity between the estimates with and without controlling for the test score suggest that systematic differences in cognitive ability cannot explain the earnings gap observed in the Census data and confirmed with the MxFLS data.

[Insert Table 6 here]

[Insert Table 7 here]

Columns (5) and (6) address the possibility that poor Spanish proficiency leads to the earnings gap between Mestizo and Indigenous workers, as suggested by Chiswick et al. (2000).

controls that have a significant effect on the estimates.

This is done by including a control for whether the respondent speaks Spanish at home. We would expect that Indigenous workers who speak Spanish at home have better Spanish proficiency than those who only speak an Indigenous language at home. The results indicate that whether or not we include speaking Spanish at home there are large earnings penalties associated with Indigenous self-identification and speaking an Indigenous language. Including occupational controls yields similar estimates for columns (2), (4), and (6). However, speaking Spanish at home is associated with significant earnings gain. Our data are not sufficiently rich to determine if this is a Spanish fluency effect or an acculturation effect.

Next, in Table 7 we only include individuals who self-identify as Indigenous. The omitted category is Indigenous who only speak Spanish. Once again a comparison of columns (1) and (3) indicates that including our measure of cognitive ability has no effect on the estimates while still indicating a positive return to cognitive ability in the labor market. The estimates from these columns are in line with the results in Table 6 as Indigenous who also speak an indigenous language have earnings that are on average lower than Indigenous who only speak Spanish, and this earnings gap cannot be explained by systematic differences in cognitive ability between these groups. Notice that once we include occupational controls the earnings gap between these two groups is substantially reduced. However, there are no real differences between the estimates that control for the test score conditional on occupation. The last two columns of Table 7 address differences in Spanish proficiency. It is important to notice that the effect of speaking Spanish at home on the respondent's earnings is stronger when we compare only Indigenous workers.

2.2   Indigenous Identity Persistence

Throughout this study we have used Indigenous self-identification as a key variable. It is however possible that individuals self-identify as Indigenous depending on the evolution of their economic success or failure. In order to investigate whether this could be biasing

15

the results we can use the MxFLS data, since this is a panel, and see whether individuals change their responses to the question of Indigenous self-identification depending on their economic condition. To this end we identified individuals who changed the response to the Indigenous self-identification question. Then among those who changed their response we have two groups for those who changed their self-identification from: (i) Indigenous to Mestizo and (ii) Mestizo to Indigenous. We also calculated the annual earnings change. Our results indicate that in the MxFLS data there are 183 cases in which individuals changed their identification from Indigenous to Mestizo. Among these 183 cases, the average change in annual earnings is a positive change of 50.17 percent with a standard error of 17.07 percent. Similarly, there are 189 cases in which the individual changed their identification from Mestizo to Indigenous. Among these 189 cases, the average change in annual earnings is a also a positive change of 50.63 percent with a standard error of 10.99 percent. Hence, it is not the case that individuals who change their self-identification from Indigenous to Mestizo experience positive wage growth, and that those who change their self-identification from Mestizo to Indigenous experience negative wage growth. The limited change in self-identification is unlikely to have an effect on our results.

3 Discussion

The empirical evidence drawn from the analysis of the Census data suggests the existence of earnings gaps between more- and less-acculturated groups in Mexico. However, the analysis with the Census data does not allow us to rule out the possibility that an unobserved component, e.g. unobserved ability, is driving these earnings differentials. To ameliorate this problem, we use data from the MxFLS which allows us to control for individuals' cognitive ability, at the expense of a smaller and less representative sample of the Mexican labor market. The empirical evidence drawn from the MxFLS indicates that systematic differences

in cognitive ability are not enough to explain the earnings gap observed in the data. The question remains as what can explain the earnings gap. In this section, we consider a series of potential explanations.

3.1 Language and Behavior

Some behavioral economics research asserts that linguistic factors can affect cognitive processes other than intellectual ability. In particular, Chen (2013) and Sutter et al. (2015) present evidence that language affects intertemporal choice. This line of research is not without major critics; nevertheless, we wish to explore whether its core claims are consistent with the empirical results reported in this study.[11] According to this research, some languages have strong future-time reference (s-FTR) while other languages have weak future-time reference (w-FTR). Persons who speak languages with s-FTR see the future as distant and grammatically separable from the present; hence, future outcomes are strongly discounted relative to present outcomes. The grammatical capacity to reference future outcomes within the present tense raises the importance of future outcomes because they seem closer to the present and more certain to be established; with w-FTR languages the future is less distant.

Using the language of the economics of intertemporal choice, persons who speak a s-FTR language have less patience than persons who speak a w-FTR language. For the latter languages, the future is not grammatically separable from the present. Hence, language influences the extent of intertemporal patience and therefore investment decisions related to health, education, savings, etc. So, children reared within a w-FTR language family may make greater investments in reproducible skills with high time and pecuniary costs in the present but with greater payoffs in the future. Importantly for this study, the effects of a language's future-time on economic patience are independent of IQ as measured by the Raven Progressive Picture Matrices (Sutter et al., 2015, page 7).

---

[11]See Dahl (2013) for a linguistics critique of Chen (2013).

Sutter et al. find that "when parents in a household speak both languages, then the child's level of patience in experimental choices is intermediate to the cases when both parents in the household speak only one language, either Italian [the s-FTR language] or German [the w-FTR language]" (see Sutter et al., 2015, page 4). Hence, a labor market prediction is that bilingual persons will earn less than mono-lingual w-FTR speakers but more than monolingual s-FTR speakers.

Chen assumes that a person's thoughts are structured by the language spoken at home (see Chen, 2013, page 10). Sutter et al. show that bilingual German children are less future-oriented than mono-lingual German children, while bilingual Italian children are more future-oriented than monolingual Italian children. As such, the language spoken at home may affect economic behavior even if the person also speaks a language in addition to the language spoken at home. Hence, bilingual-home Indigenous should have a different income penalty than bilingual-literate Indigenous, but the same income penalty as fully bilingual Indigenous. In comparison to Monolingual-Spanish Indigenous, Bilingual-home Indigenous and fully bilingual Indigenous will have market penalties or market premia according to whether the grammar of an Indigenous language creates more or less impatience than the grammar of Spanish. Indeed, if Spanish has a relatively higher s-FTR than Indigenous languages, then bilingual Indigenous should receive a market premium.

Thus far the language and behavior research is quiet on the relationship between language and other attributes that may affect income attainment, for example, risk-taking, drive and ambition, the ability of overcome obstacles, or language group cultural pressure for supra-normal effort and achievement. In any case, an evaluation of the relationship between the relative FTRs of Spanish and Indigenous languages and their impact on income is beyond the scope of this paper. We note however that for males Table 4 shows that there are significant and substantively large negative income penalties or an insignificant effect for all Indigenous language family-groups. Hence, for the language-behavior explanation of Mestizo-Indigenous

income inequality to hold it would have to be the case that every Indigenous language family-group has a s-FTR language relative to Spanish. Further, all other language effects on cognitive skills other than intelligence must also favor Spanish speakers. Even so, the language-behavior explanation does not explain why Monolingual-Spanish Indigenous males have a 10 percent income penalty.

3.2 Discrimination

Next, there exists the possibility that less-acculturated individuals in Mexico are discriminated against in the labor market. Indigenous individuals who speak an Indigenous language have stronger ties to their Indigenous roots than Indigenous who do not speak an Indigenous language. Then, in terms of the *Mestizaje* with *blanqueamiento* ideal, Indigenous who speak an Indigenous language are less acculturated than those Indigenous who only speak Spanish.

Aguilar (2011) performs a lab experiment to examine the social and political consequences of Mexicans' different racial appearances. She finds that Mexicans tend to socially evaluate more positively European-looking Mexicans. She also finds that people are cognizant of more negative traits attached to Indigenous-looking individuals than to Mestizo or European-looking persons, while attaching more positive traits to the latter. Similarly, Arceo-Gomez and Campos-Vazquez (2014) perform a correspondence study to identify racial discrimination in Mexico throughout the range of phenotypes generated by racial mixing since the Spanish colonization. The authors find that firms in Mexico tend to discriminate against applicants of Indigenous appearance. And so there seems to be some evidence of discrimination against indigenous in Mexico.

The results from Sections 2 and 2.1 indicate that less-acculturated Indigenous workers have a significant penalty compared to Indigenous workers who speak only Spanish or those who are Mestizos. It is clear that Indigenous workers who speak only an Indigenous language have an earnings disadvantage because the dominant language in the Mexican labor market

is Spanish. But, it is not clear why Indigenous workers who speak only Spanish have an earnings advantage compared to Indigenous workers who are bilingual. Indigenous bilingual individuals speak both Spanish and an Indigenous language; hence at a minimum, there should be no significant difference in their earnings compared to the earnings of Indigenous workers who speak only Spanish. However, the results from the Census data indicate that Indigenous who speak only Spanish have a significant earnings advantage over Indigenous who speak both Spanish and an Indigenous language (see Tables 2 and 3). The results from the MxFLS are less precise but also suggest an earnings advantage for more acculturated individuals (see Tables 6 and 7).

Notice that, in this study, we use the individual's capacity to speak an Indigenous language as a metric for being more or less acculturated. However, one can argue that this difference in the level of acculturation maps into systematic differences in the way that less-acculturated individuals interact in the labor market, the way they dress, or the way they speak Spanish (i.e. their accent), their phenotype, among other things. As a result, even without explicitly knowing that an Indigenous individual is bilingual it is possible that labor market participants infer that some individuals are less acculturated than others and so if less-acculturated individuals are discriminated against this results in the earnings gap observed in the data.

It is also important to notice that we found that speaking Spanish at home is associated with an earnings gain. While we expect that those respondents who speak Spanish at home should have better Spanish proficiency, it is also true that respondents who speak Spanish at home are more acculturated than respondents who speak an Indigenous language at home.

Finally, notice the effect that occupational controls (i.e. occupation and industry) have on the earnings gap. The results from Sections 2 and 2.1, indicate that without conditioning on occupational outcomes the earnings gap is even wider. It is possible that labor market discrimination against less-acculturated individuals is also reflected in restricted ac-

cess to certain occupations or industries. The fact that the earnings gap persists even after controlling for occupation and industry suggests that less-acculturated individuals are both discriminated in terms of their job choices and in terms of the return to their work.

3.3   Job Search and Networks

Another mechanism that could generate the earnings gaps observed in the data are differences in job search and labor market networks. Consider the search and matching framework of Mortensen and Pissarides (1994). In this framework the wages are determined through Nash bargaining and the resulting wage is positively related to the labor market tightness, which is defined as the ratio of the number of vacancies to the number of unemployed workers. If Mestizo workers have larger networks in the labor market than Indigenous workers, this will result in a more favorable labor market tightness for Mestizo workers, resulting in better bargained wages than Indigenous workers. Also, the descriptive statistics in Tables 1 and 5 reveal that more-acculturated Indigenous workers, that is those who only speak Spanish, have labor market outcomes more similar to the Mestizos than less-acculturated Indigenous workers. Hence, if more-acculturated Indigenous workers have better networks than less-acculturated Indigenous workers, the more-acculturated group will face more favorable labor market tightness and bargain better wages than the less-acculturated Indigenous workers. A model that can be used to explain these earnings gaps in the Mexican labor market is developed in Mortensen and Pissarides (2003) which assumes segmented labor markets for different groups of workers in the labor market.

Alternatively, we can consider the equilibrium search framework of Burdett and Mortensen (1998) to understand the earnings gaps observed in the data. Once again if Mestizos have larger networks than Indigenous workers, Mestizo workers will face higher job-offer arrival rates than Indigenous. Similarly, if within Indigenous workers, those who are more acculturated would have higher job-offer arrival rates than less-acculturated. In recent work,

Bowlus, Miyairi, and Robinson (2014) develop an equilibrium search model to study immigrant job search assimilation in Canada where immigrants have smaller networks and face other difficulties such as language and knowledge of the Canadian labor market which translate into lower arrival rates of job offers both while employed and while unemployed. The authors show that under certain conditions the distribution of wages of native Canadians first-stochastically dominates the distribution of wages for immigrants. A similar model could be used to describe the labor market of Mexico. The two graphs in Figure 2 present kernel density estimations for the distributions of log-earnings for different groups in the Mexican labor market. Panel (a) of Figure 2 compares the distribution of log-earnings for Mestizos and Indigenous and the graph suggests that the distribution of log-earnings of Mestizos first-stochastically dominates that of Indigenous. A Kolmogorov-Smirnov test for equality of the two distributions confirms this result.[12] Similarly, the graph in Panel (b) of Figure 2 compares the distribution of log-earnings for Indigenous who speak only Spanish and those who speak both Spanish and and Indigenous language. The graph suggest that the distribution of log-earnings for more-acculturated Indigenous workers first-stochastically dominates the distribution of log-earnings for less-acculturated Indigenous workers. Once again, a Kolmogorov-Smirnov test confirms this result.[13]

[Insert Figure 2 here]

---

[12]Under the assumption that the cumulative distribution function (c.d.f.) of log-earnings for Indigenous is above the c.d.f. of log-earnings for Mestizos, the largest difference between the two c.d.f.'s is 0.2353 which is significantly different from zero (p-value of 0.000). However, under the assumption that the c.d.f. of log-earnings for Mestizos is above the c.d.f. of log-earnings for Indigenous, the largest difference is -0.0004 which is not significantly different from zero (p-value of 0.915). This is consistent with the distribution of log-earnings for Mestizos first-stochastically dominating that of Indigenous.

[13]Under the assumption that the cumulative distribution function (c.d.f.) of log-earnings for Bilingual Indigenous is above the c.d.f. of log-earnings for Monolingual-Spanish Indigenous, the largest difference between the two c.d.f.'s is 0.2336 which is significantly different from zero (p-value of 0.000). However, under the assumption that the c.d.f. of log-earnings for Monolingual-Spanish is above the c.d.f. of log-earnings for Bilingual, the largest difference is 0.0000 which is not significantly different from zero (p-value of 1.000). This is consistent with the distribution of log-earnings for Monolingual-Spanish first-stochastically dominating that of Bilingual.

Borja-Vega et al. (2007) test for the effect of social networks based on group-identity or ethnic affinity on employment outcomes among Indigenous individuals in Mexico. The authors find that social networks have a positive and significant effect on males being employed as unskilled workers or being self-employed in both rural and urban areas and they also explore the effect of social networks in other outcomes such as migration and participation in public programs. Similarly, Skoufias, Lunde, and Patrinos (2010) examine the extent to which social networks among Indigenous individuals in Mexico have an effect on investment in human capital, school attendance, teenage work, migration, welfare participation, employment, occupation and employment sector. Their empirical analysis confirms that social networks plan an important role in these economic decisions, particularly in rural areas. And so it seems reasonable to think that differences in the size of networks between Indigenous and Mestizo workers could be an important source of the earnings gap observed in the data.[14]

3.4 Segmentation by Education or Cognitive Ability

Another possibility is that the Mexican labor market may be segmented by years of education or cognitive ability. If so, there may be non-linear and negative income effects associated with Indigenous identity. Theories of labor market segmentation tend to suggest limited mobility between a high-wage primary sector and a low-wage secondary sector, where the secondary sector is also characterized by limited on-the-job training and extremely few career ladders (Lewis, 1954; Gordon et al., 1982). Given the relative absence of on-the-job training and career ladders, secondary jobs require fewer years of education and less ability than primary sector jobs. If Indigenous workers are more likely than Mestizos to be employed in the secondary sector, then we will observe ethno-linguistic income differentials. In this

---

[14]Also recent study by Pérez Velasco Pavón (2014) finds that Indigenous workers tend to work in small scale productive units which are known to be less productive.

case ethno-linguistic income differentials are the result of ethnic differences in mobility from the secondary to the primary sector and the superior wages, on-the-job training, and career ladders within the primary sector.

To investigate this possibility we estimated a series of specifications for alternative years of education and cognitive ability quartiles using the MxFLS data. The quartiles are based on Mestizos years of education and cognitive ability and are presented in Tables 8 and 9 which present the estimates from specifications similar to those in Table 6. The results from Table 8 indicate that the earnings penalty for Indigenous identification only occurs in the lowest education quartile, whereas the penalty for speaking an Indigenous language only occurs at the third and fourth education quartiles. The results from Table 9 indicate that the earnings penalty for Indigenous identification only occurs at the second ability quartile, whereas the penalty for speaking an indigenous language only occurs at the first and third ability quartiles. The specifications in Tables 8 and 9 control for occupation. The results without these controls are very similar and do not change any of the conclusions.

[Insert Table 8 here]

[Insert Table 9 here]

3.5   Employment Status

The statistical analysis has included self-employed persons in the samples used to estimate the log-earnings regressions. The MxFLS data has a very uniform distribution of self-employed among the Mestizo and Indigenous workers. But the Census data, which is more representative of the Mexican population, reveals that about 30 percent of bilingual Indigenous workers and 50 percent of Monolingual-Indigenous workers are self-employed (see Table 1). As it has been documented for the case of the Unites States, there are significant differences in earnings between self-employed and wage/salary workers (see Hamilton, 2000).

If the earnings process differs by occupational segmentation and Mestizo and Indigenous workers have different occupational distributions, then the penalties for Indigenous self-identification and bilingual linguistic ability may represent spurious correlation. Hence, we use the Census data to estimate specifications similar to those in Tables 2 and 3 separately for self-employed and wage/salary workers.[15]

The results are presented in Table 10 where all specifications control for occupation and industry, the results without these controls are very similar and do not change any of the conclusions. Columns (1) and (2) present the estimates for a specification identical to that of Table 2, whereas Columns (3) and (4) present the estimates for a specification identical to that of Table 3. The results from all specifications are in line with the findings using the whole sample, and so we conclude that segmentation into self-employment is not the cause of ethno-linguistic identity wage penalties. Notice however, that the earnings differentials are higher for self-employed workers than for wage/salary workers. This is particularly important as the compensation of wage/salary workers must be somewhat restricted by labor regulations or even the status quo. However, the compensation of self-employed workers will be mostly determined by bargaining between the self-employed worker and the buyer of the worker services, and so this compensation could carry some discount for hiring a less-acculturated worker.

[Insert Table 10 here]

3.6 Location Size

Similarly, even when we control for the size of the location of residence of the respondents in the sample, it is possible that the earnings gap is driven by the fact that there is a larger

---

[15]Recall that the Census does not allow us to control for differences in cognitive ability, but as seen before, systematic differences in the measure of cognitive ability available in the MxFLS are not enough to explain the gap (see Tables 6 and 7). On the other hand, working with the Census allows us to control for a large set of observable characteristics, has larger samples and it is more representative of the Mexican population.

proportion of indigenous workers living in small communities whereas there is a larger proportion of Mestizos and Monolingual-Spanish Indigenous workers living in large cities. Then, similar to what we did in Table 10 we estimate separate regressions for small communities and large cities. We define small communities as locations with population less than 100,000; and large cities as locations with population larger than 100,000. The results are shown in Table 11. The results from Columns (1) and (2) indicate that the earnings gap between Indigenous and Mestizos is mainly a small-communities phenomenon. In large cities, those individuals who self-identify as indigenous still suffer an earnings gap but it is only of five percentage points, whereas in these locations speaking an Indigenous language is not associated with an earnings penalty; on the contrary, Indigenous workers who speak an Indigenous language have an earnings advantage of about one percent, although this estimate is borderline significant. On the other hand, the results in Column (1) indicate that it is in small communities that self-identification as Indigenous has a large earnings gap and the penalty is higher if the individual speaks and Indigenous language. Similarly, when comparing among Indigenous workers in Columns (3) and (4) we find that the gains for acculturation is completely a phenomenon that occurs in small communities. In large cities there is no earnings gain for acculturation.[16] The specifications in Table 11 control for occupation and industry, the results without these controls are very similar and do not change any of the conclusions.

[Insert Table 11 here]

3.7   Policy Implications of the Alternative Explanations

This paper finds a large ethnic identity earnings gap. Our best estimates indicate that Indigenous workers earn about 9.5 percent less than otherwise identical Mestizo workers.

___

[16]Although this could be a result of a very small number of Monolingual-Indigenous living in large cities since our sample only has 19 individuals who are Monolingual-Indigenous living in large cities. However, we ran the same specification in Column (4) of Table 11 excluding Monolingual-Indigenous, so that only Bilingual and Monolingual-Spanish individuals were included and the estimates indicate no significant gain for Monolingual-Spanish.

These estimates control for cognitive ability, Spanish fluency, and other productivity-linked attributes. There is an additional 22 percent penalty for Indigenous workers who also speak an Indigenous language. Additionally, our findings suggest that the social and labor market mechanisms generating the earnings gap are structural, rather than behavioral. Specifically, our results are inconsistent with the notion that Indigenous language culture discourages skill accumulation and thereby contributes to ethnic inequality in the Mexican labor market. These results are consistent with several mechanisms: 1) earnings discrimination in the most narrow sense, that is, lower pay for Indigenous workers who are as productive as Mestizo workers; 2) larger or more effective job search networks for Mestizos relative to Indigenous people; 3) segmentation in the labor market with Indigenous workers disproportionately located in what are most likely secondary sector jobs; 4) Indigenous identity and language penalties are substantively higher among self-employed workers; 5) earnings penalties are strongly concentrated among workers located in small communities, that is, locations with population less 100,000 persons.

Tying all these outcomes together suggests Indigenous workers lack sufficient job networks to obtain the most desirable employment in larger urban locations, that is, jobs within the primary employment sector and where there is the least earnings inequality. Accordingly, Indigenous workers especially those with the strongest attachment to Indigenous identity are disproportionately sorted into self-employment occupations and jobs located in smaller municipalities. All of these results suggest that the earnings gap between Indigenous and Mestizo workers can be reduced through policies that: (i) improve Indigenous worker access to waged jobs in order to reduce their concentration in self-employment, and (ii) develop mechanisms to improve the development of Indigenous communities in small communities.

In regards to improving access to waged jobs, Mexico has already started moving in the right direction. In the last decade, Mexico has established standards for labor equality and non-discrimination (see the labor standard NMX-R-025-SCFI-2009 and more recently

27

the labor standard NMX-R-025-SCFI-2015). Although most of these standards focus on the male-female earnings gap (especially in the earlier versions), rather than the earnings gaps due to ethnic differences, the labor standards for equality exist. Moreover, there have been additions and modifications to the labor laws and the Mexican constitution to prevent discrimination in employment. It suggests that more work needs to be done in guaranteeing that these labor standards and labor regulations are enforced, especially for Indigenous workers.

For example, firms can obtain certification in the most recent Mexican Standards for Labor Equality and Non Discrimination NMX-R-025-SCFI-2015. Certification allows the firm to be recognized as an employer that satisfies employment equality, which can be used by the firm to improve its image and perhaps its position in consumers' preferences. However, certification is voluntary so the labor standards for equality are not really binding. While we believe that making these standards mandatory for every firm would not be viable, the authorities might try to induce firms' certification in a variety of ways.

One possibility is to make certification mandatory for firms that have significant economic transactions with the Federal Government of Mexico. This would be similar to the United States Executive Order 11246 which prohibits discrimination in employment for firms which do over $10,000 in Government businesses in one year and impose penalties if firms do not comply with these regulations. Mandatory certification could also be based on firm size. Other possibilities include offering tax incentives to firms that are certified in these labor standards. The feasibility and the implementation of these incentives are beyond the scope of this paper.

Similarly, as the results suggest, the earnings gap between Indigenous and Mestizo workers seem to be higher in smaller communities. Hence the economic development of Indigenous people in small communities could help in reducing the earnings gap we observe between Indigenous and Mestizo workers. In Mexico, the Government agency in charge of this is

28

the National Commission for the Development of Indigenous Peoples (CDI for its acronym in Spanish) which has at its objectives "the integral and sustainable development of the indigenous peoples and communities." This agency has developed a series of programs that are aimed at developing Indigenous communities, such as programs to improve production and productivity in Indigenous communities, foster education and improve infrastructure in Indigenous communities. Once again, the Mexican authorities seem to be moving in the right direction, but much more work needs to be done to reduce or eliminate the Indigenous-Mestizo earnings gap.

4   Conclusion

The results herein reveal that the Mexican labor market penalizes Indigenous identity, that is, even highly acculturated Monolingual-Spanish Indigenous will receive a wage penalty relative to Mestizos. Moreover, the penalty will increase if Indigenous workers also speak an Indigenous language — regardless of the degree of Spanish fluency.

Cognitive ability does have a large, positive, and statistically significant effect on wages. For a one standard deviation increase in the age- and education-adjusted test score, male wages will rise by 6-7 percent. However, including a cognitive ability measure into the wage equation has little or no effect on the ethno-linguistic identity coefficients. Hence, it is possible to obtain consistent coefficient estimates of ethno-linguistic inequality from regressions estimated on data drawn from the Mexican Census, a very large dataset which includes information on earnings, hours, and individual characteristics but which does not include a measure of ability.

It is important to note that the Mexican labor market is characterized for having a sizable informal sector (e.g. see Maloney, 1999; Cano-Urbina, 2015, 2016). However, the present study abstracts from that feature of the labor market. The main reason for such abstraction

is because the informal sector is an urban phenomenon, and concentrating on urban areas would leave a very important group out of our sample: the monolingual Indigenous as it was shown in Table 1 that the vast majority of these individuals live in rural areas. It is left for future research to explore the labor market experience of Indigenous and Mestizos in the urban labor market and address directly the issue of informality of the job.

The discussion in Section 3 provides many possible channels that could explain the earnings gap between Indigenous and Mestizo workers. In all likelihood, the earnings gap is a result of a combination of multiple mechanisms. But it is left for future research to explore the empirical relevance of each of these.

References

Aguilar, R. (2011, August). Social and political consequences of stereotypes related to racial phenotypes in Mexico. Working Paper 230, CIDE, Mexico City, Mexico.

ALDF (2002). Asamblea legislativa del distrito federal: Código Penal para el Distrito Federal. *Gaceta Oficial del Distrito Federal*, 16 July 2002.

Arceo-Gomez, E. O. and R. M. Campos-Vazquez (2014). Race and marriage in the labor market: A discrimination correspondence study in a developing country. *American Economic Review 104*(5), 376 – 380.

Atal, J. P., H. Ñopo, and N. Winder (2009, October). New century, old disparities: gender and ethnic wage gaps in Latin America. IDB Working Paper IDB-WP-109, Inter-American Development Bank, Washington, DC.

Borja-Vega, C., T. Lunde, and V. García Moreno (2007, February). Economic opportunities for indigenous people in Mexico. Case study, The World Bank, Washington, DC. Conference Edition.

Bowlus, A. J., M. Miyairi, and C. Robinson (2014, August). Immigrant job search assimilation in canada. Working paper, The University of Western Ontario.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*(2), 257 – 273.

Cano-Urbina, J. (2015). The role of the informal sector in the early careers of less-educated workers. *Journal of Development Economics 112*, 33 – 55.

Cano-Urbina, J. (2016, January). Informal labor markets and on-the-job training: Evidence from wage data. *Economic Inquiry 54*(1), 25–43.

Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review 103*(2), 690–731.

Chiswick, B. R. (2008). The economics of language: an introduction and overview. IZA Discussion Paper No. 3568, Institute for the Study of Labor (IZA).

Chiswick, B. R., H. A. Patrinos, and M. E. Hurst (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change 48*(2), 349–367.

Commission for Labor Cooperation (2006, June). Mexican Chapter. Principles 7 and 8: Elimination of Employment Discrimination and Equal Pay for Women and Men.

Dahl, O. (2013). Stuck in the futureless zone. http://dlc.hypotheses.org/author/oestendahl. Posted on 2013/03/09.

Darity Jr., W. and P. L. Mason (1998, Spring). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives 12*(2), 63–90.

Darity Jr., W., P. L. Mason, and J. Stewart (2006, July). The Economics of Identity: The Origin and Persistence of Racial Norms. *Journal of Economic Behavior and Organizations 60*(3), 283–305.

Gordon, D. M., R. Edwards, and M. Reich (1982). *Segmented Work, Divided Workers: The historical transformation of labor in the United States.* New York: Cambridge University Press.

Hamilton, B. H. (2000). Does entrepreneurship pay? an empirical analysis of the returns to self-employment. *Journal of Political Economy 108*(3), 604–631.

Hisamatsu, Y. and H. Ukeda (2002). Economic disadvantage of being Indigenous: an analysis of Mexico's first National Survey of Employment in Indigenous Areas. Working Paper.

INALI (2008). Instituto Nacional de Lenguas Indígenas: Catálogo de las Lenguas Indígenas Nacionales. *Diario Oficial de la Federación*, 14 January 2008. Vol. DCLII, No. 9.

Lang, K. (1986, May). A language theory of discrimination. *Quarterly Journal of Economics 101*(2), 363–382.

Lang, K. and M. Manove (2011, June). Education and labor market discrimination. *American Economic Review 101*(4), 1467–1496.

Lewis, W. A. (1954, May). Economic development with unlimited supplies of labour. *The Manchester School 22*(2), 139–191.

Lleras-Muney, A. and A. Shertzer (2015). Did the Americanization movement succeed? An evaluation of the effect of english-only and compulsory schooling laws on immigrants. *American Economic Journal: Economic Policy 7*(3), 258 – 90.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review 13*(2), 275–302.

Mason, P. L. (2004, October). Annual income, hourly wages, and identity formation among mexican americans and other latinos. *Industrial Relations 43*(4), 817–834.

McManus, W., W. Gould, and F. Welch (1983). Earnings of Hispanic men: The role of english language proficiency. *Journal of Labor Economics 1*(2), 101 – 130.

Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies 61*(3), 397 – 415.

Mortensen, D. T. and C. A. Pissarides (2003). Taxes, subsidies and equilibrium labor market outcomes. In E. S. Phelps (Ed.), *Designing inclusion: tools to raise low-end pay and employment in private enterprise*, pp. 44 – 73. Cambridge; New York: Cambridge University Press.

Neal, D. A. and W. R. Johnson (1996, October). The role of premarket factors in black-white wage differences. *Journal of Political Economy 104*(5), 869–895.

Panagides, A. (1994). Mexico. In G. Psacharopoulos and H. A. Patrinos (Eds.), *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington: World Bank.

Patrinos, H. A. (1994). The cost of discrimination in Latin America. Human Resources Development and Operational Policy Working Papers No. 45, World Bank.

Pérez Velasco Pavón, J. C. (2014, August). Economic behavior of indigenous peoples: the Mexican case. *Latin American Economic Review 23*(1), 1–58.

Rubalcava, L. and G. Teruel (2006a). Guía del usuario para la primera encuesta nacional sobre niveles de vida de los hogares. www.ennvih-mxfls.org.

Rubalcava, L. and G. Teruel (2006b). Mexican family life survey, second round. Working Paper, www.ennvih-mxfls.org.

Rubalcava, L. and G. Teruel (2013). Mexican family life survey, third round. Working Paper, www.ennvih-mxfls.org.

Safa, H. I. (2005). Challenging Mestizaje: A Gender Perspective on Indigenous and Afrodescendant Movements in Latin America. *Critique of Anthropology 25*(3), 307–330.

Sandoval, M. (2014). Movilidad social de la población indígena de México. Master's thesis, CIDE, Mexico City, Mexico.

SEGOB (2008a). Secretaría de Gobernación: Decreto de Reforma de la Constitución Política de los Estados Unidos Mexicanos. *Diario Oficial de la Federación*, 14 August 2001. Vol. DLXXV, No. 10.

SEGOB (2008b). Secretaría de Gobernación: Ley Federal para Prevenir y Eliminar la Discriminación. *Diario Oficial de la Federación*, 11 June 2003. Vol. DXCVII, No. 8.

Skoufias, E., T. Lunde, and H. A. Patrinos (2010). Social networks among indigenous peoples in Mexico. *Latin American Research Review 45*(2), 49–67.

Sutter, M., S. Angerer, D. Gltzle-Rtzler, and P. Lergetporer (2015). The effects of language on children's intertemporal choices. Discussion Paper 9383, The Institute for the Study of Labor (IZA).

Terborg, R., L. García Landa, and P. Moore (2006, December). The language situation in Mexico. *Current Issues in Language Planning 7*(4), 415–518.

Vazquez Carranza, A. (2009). Linguistic rights in mexico. *Revista Electronica de Linguistica Aplicada 8*, 199–210.

Table 1: Census Descriptive Statistics: Mean

	Mestizo	Indigenous			
		All	Bilingual	Monolingual	
				Indigenous	Spanish
Age	36.76	36.28	36.87	41.63	35.85
Years of education	9.45	7.74	6.49	1.68	8.56
Self-employed (%)[†]	20.73	25.06	29.93	47.88	21.94
Working hours per week[†]	49.70	48.46	47.48	42.83	49.11
Monthly earnings[‡]	6,530.96	4,631.91	3,752.44	1,757.88	5,172.97
Location size (%)					
1 - 2,499	16.13	33.21	43.62	76.18	26.63
2,500 - 14,999	13.06	21.99	23.67	20.11	21.10
15,000 - 99,999	15.92	15.84	11.82	2.50	18.31
99,999 or more	54.89	28.97	20.89	1.21	33.96
Occupation (%)					
Legislators / officials / managers	3.87	1.77	0.91	0.08	2.28
Professionals	2.64	1.43	0.68	0.09	1.87
Technicians / associate professionals	15.00	9.60	6.60	0.33	11.43
Clerks	3.85	2.53	1.60	0.11	3.08
Service workers / sales	18.96	15.86	14.24	2.78	17.01
Skilled agricultural / fishery	8.07	17.49	24.58	67.72	12.60
Crafts / trades	17.55	19.92	20.77	10.45	19.63
Plant-machine operators / assemblers	15.23	11.01	7.20	0.99	13.31
Elementary occupations	14.52	19.95	22.88	17.45	18.38
Armed forces	0.31	0.44	0.53	0.00	0.41
Industry (%)					
Agriculture, fishing, and forestry	11.29	23.34	32.38	76.50	17.32
Mining	0.82	0.69	0.45	0.28	0.83
Electricity, gas and water	0.69	0.53	0.31	0.07	0.67
Construction	13.86	17.76	19.49	10.55	16.94
Manufacturing	17.86	14.31	11.67	5.54	15.94
Wholesale and retail trade	16.81	14.21	12.45	3.53	15.39
Transportation and communications	8.56	6.07	4.07	0.98	7.27
Financial services and insurance	0.90	0.37	0.12	0.00	0.52
Real estate and business services	6.52	3.86	2.81	0.42	4.50
Educational services	3.42	3.39	3.84	0.09	3.21
Health care and social work services	1.69	0.89	0.44	0.00	1.15
Amusement and recreational services	0.98	0.83	0.58	0.17	0.98
Hotel and restaurant services	4.26	4.15	4.20	0.57	4.20
Other services except government	6.93	4.69	2.75	0.60	5.84
Private household services	0.57	0.73	0.84	0.18	0.69
Public administration and defense	4.84	4.18	3.62	0.51	4.55
Sample Size	1,331,883	429,543	211,086	16,892	201,565

Notes: Means using Census sampling weights. Sample includes males ages 16-64 not attending school. [†] The descriptive statistics for this variable was obtained from a smaller sample due to missing values. [‡] Monthly earnings is in Mexican Pesos of 2nd half of December 2010.

Table 2: Effect of Ethnic Identification on Earnings

	(1)	(2)	(3)
Identify as indigenous	-0.0887	-0.0886	-0.0898
	(0.0016)	(0.0016)	(0.0015)
Indigenous language	-0.1832	-0.1599	-0.1633
	(0.0020)	(0.0019)	(0.0019)
Years of education	0.0587	0.0406	0.0385
	(0.0001)	(0.0001)	(0.0001)
Age	0.0479	0.0407	0.0396
	(0.0003)	(0.0002)	(0.0002)
Age^2	-0.0005	-0.0005	-0.0004
	(0.0000)	(0.0000)	(0.0000)
Additional controls:			
Occupation		y	y
Industry			y
R^2	0.2995	0.3591	0.3737
No of obs	1,761,426	1,761,426	1,761,426

Notes: Standard errors in parenthesis. Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of monthly earnings. All specifications control for size of location, state of residence, and state of birth.

Table 3: Effect of Acculturation on Earnings

	Indigenous			Mestizo		
	(1)	(2)	(3)	(4)	(5)	(6)
Monolingual Spanish	0.4486	0.3186	0.3195	0.0542	0.0532	0.0577
	(0.0063)	(0.0059)	(0.0058)	(0.0064)	(0.0062)	(0.0061)
Bilingual	0.2993	0.1957	0.1893			
	(0.0060)	(0.0056)	(0.0055)			
Years of Education	0.0524	0.0279	0.0254	0.0597	0.0435	0.0415
	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0002)	(0.0002)
Age	0.0414	0.0331	0.0310	0.0500	0.0431	0.0422
	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0003)
Age^2	-0.0005	-0.0004	-0.0004	-0.0006	-0.0005	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Additional controls:						
Occupation		y	y		y	y
Industry			y			y
R^2	0.2560	0.3580	0.3827	0.2613	0.3123	0.3251
No of obs	429,543	429,543	429,543	1,331,883	1,331,883	1,331,883

Notes: Standard errors in parenthesis. Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of monthly earnings. All specifications control for size of location, state of residence, and state of birth. The regressions in columns (1)-(3) only include individuals who are Indigenous. The regressions in columns (4)-(6) only include individuals who are Mestizos. The omitted category in columns (1)-(3) is Monolingual Indigenous, and the omitted category in columns (4)-(6) is Bilingual.

Table 4: Effect of Ethnic Identification on Earnings from Various Ethnic Groups

	Proportion	Regressions		
	(%)	(1)	(2)	(3)
Indigenous Language:				
Yuto-Nahua	29.67	-0.1739	-0.1534	-0.1522
		(0.0040)	(0.0037)	(0.0036)
Oto-Mangue	28.13	-0.1489	-0.1029	-0.1111
		(0.0040)	(0.0037)	(0.0036)
Maya	32.38	-0.1711	-0.1538	-0.1663
		(0.0048)	(0.0044)	(0.0043)
Totonaco-Tepehua	4.05	-0.1386	-0.0772	-0.1018
		(0.0073)	(0.0067)	(0.0066)
Tarasca	2.40	-0.1875	-0.1958	-0.1659
		(0.0143)	(0.0133)	(0.0131)
Mixe-Zoque	2.83	-0.0396	-0.0105	-0.0348
		(0.0079)	(0.0073)	(0.0072)
Chontal de Oaxaca	0.04	0.1953	0.1714	0.1405
		(0.0650)	(0.0603)	(0.0591)
Huave	0.35	-0.4076	-0.2394	-0.2580
		(0.0258)	(0.0239)	(0.0234)
Other	0.13	0.2661	0.2447	0.2197
		(0.1570)	(0.1456)	(0.1428)
Years of Education		0.0550	0.0297	0.0272
		(0.0003)	(0.0003)	(0.0003)
Age		0.0416	0.0331	0.0311
		(0.0005)	(0.0005)	(0.0005)
Age^2		-0.0005	-0.0004	-0.0004
		(0.0000)	(0.0000)	(0.0000)
Additional controls:				
Occupation			y	y
Industry				y
R^2		0.2523	0.3570	0.3815
No of obs		432,517	432,517	432,517

Notes: Standard errors in parenthesis. Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of monthly earnings. All specifications control for size of location, state of residence, and state of birth. The regression only include individuals who are Indigenous. The omitted category is the group of Indigenous who are monolingual in Spanish. The family "Other" is composed of three families: Algica, Cochimi-Yumana, and Seri

Table 5: MxFLS Descriptive Statistics: Mean

	Mestizo	Indigenous		
		All	Speak Indigenous Language	Only Speak Spanish
Age	36.70	37.73	38.60	36.07
Years of education	7.79	6.08	5.37	7.42
Standardized test score	0.07	-0.24	-0.30	-0.11
Self-employed (%)†	13.20	13.04	13.48	12.21
Working hours per week†	45.83	43.44	42.99	44.29
Annual earnings‡	51,022.15	33,711.93	29,315.03	42,107.83
Occupation (%)				
Professionals	3.87	1.63	0.59	3.62
Technicians	7.86	6.22	4.50	9.50
Education	2.41	3.42	4.27	1.81
Arts, performance and sports	0.60	0.93	0.83	1.13
Directors public/private	0.67	0.39	0.36	0.45
Agricultural/foresting/fishing	21.19	43.08	52.96	24.21
Chiefs artisan/industrial	0.63	0.31	0.36	0.23
Workers artisan/industrial	16.89	13.37	13.15	13.80
Operators fixed machinery	3.94	1.94	1.18	3.39
Assistants artisan/industrial	6.46	4.35	3.55	5.88
Conductors of machinery/transport	6.81	3.81	2.61	6.11
Chiefs administrative/services	1.58	0.78	0.59	1.13
Workers administrative	4.65	3.19	2.01	5.43
Retailers and sales agents	10.02	7.54	5.92	10.63
Street sales and services	1.10	1.32	1.07	1.81
Workers personal establishments	4.84	3.42	2.84	4.52
Workers domestic services	0.68	0.54	0.47	0.68
Protection and armed forces	3.93	2.41	1.66	3.85
Occupations Non-classified	1.88	1.32	1.07	1.81
Sample Size	9,382	1,286	844	442

Notes: Unweighted means. Data from the 2002, 2005, and 2009 MxFLS. Sample includes males ages 16-64 not attending school. † The descriptive statistics for this variable was obtained from a smaller sample due to missing values. ‡ Annual earnings is in Mexican Pesos of 2nd half of December 2010.

Table 6: Effect of Ethnic Identification on Earnings Conditional on Cognitive Ability

	(1)	(2)	(3)	(4)	(5)	(6)
Identify as indigenous	-0.1033	-0.0918	-0.0956	-0.0859	-0.0949	-0.0854
	(0.0392)	(0.0374)	(0.0392)	(0.0374)	(0.0391)	(0.0374)
Indigenous language	-0.2312	-0.1613	-0.2316	-0.1627	-0.2183	-0.1527
	(0.0510)	(0.0487)	(0.0507)	(0.0485)	(0.0505)	(0.0484)
Years of education	0.0723	0.0455	0.0662	0.0412	0.0660	0.0411
	(0.0024)	(0.0026)	(0.0025)	(0.0027)	(0.0025)	(0.0027)
Age	0.0726	0.0660	0.0713	0.0651	0.0714	0.0652
	(0.0043)	(0.0042)	(0.0043)	(0.0042)	(0.0043)	(0.0042)
Age^2	-0.0008	-0.0007	-0.0008	-0.0007	-0.0008	-0.0007
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Test Score			0.0686	0.0542	0.0683	0.0540
			(0.0092)	(0.0089)	(0.0091)	(0.0089)
Speak Spanish at home					0.2266	0.1765
					(0.0904)	(0.0886)
Additional controls:						
Occupation		y		y		y
R^2	0.2119	0.2580	0.2163	0.2607	0.2170	0.2611
No of obs	10,668	10,668	10,668	10,668	10,668	10,668

Notes: Standard errors in parenthesis. Data from the 2002, 2005, and 2009 MxFLS. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of annual earnings. All specifications control for year, state of residence, and state of birth.

Table 7: Effect of Acculturation on Earnings Conditional on Cognitive Ability

	(1)	(2)	(3)	(4)	(5)	(6)
Indigenous language	-0.1855	-0.1107	-0.1889	-0.1149	-0.1714	-0.1031
	(0.0633)	(0.0621)	(0.0626)	(0.0617)	(0.0622)	(0.0616)
Years of Education	0.0745	0.0372	0.0686	0.0338	0.0663	0.0323
	(0.0077)	(0.0084)	(0.0080)	(0.0086)	(0.0080)	(0.0085)
Age	0.0888	0.0813	0.0860	0.0797	0.0874	0.0808
	(0.0131)	(0.0127)	(0.0131)	(0.0127)	(0.0131)	(0.0127)
Age^2	-0.0010	-0.0010	-0.0010	-0.0009	-0.0010	-0.0010
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Test Score			0.0869	0.0592	0.0822	0.0558
			(0.0267)	(0.0256)	(0.0264)	(0.0253)
Speak Spanish at home					0.3701	0.2964
					(0.1260)	(0.1268)
Additional controls:						
Occupation		y		y		y
R^2	0.2387	0.3128	0.2451	0.3158	0.2515	0.3198
No of obs	1,286	1,286	1,286	1,286	1,286	1,286

Notes: Standard errors in parenthesis. Data from the 2002, 2005, and 2009 MxFLS. Sample includes males ages 16-64 not attending school. All regressions only include individuals who are Indigenous. The dependent variable is the natural logarithm of annual earnings. All specifications control for year, state of residence, and state of birth.

Table 8: Log-Earnings Regressions by Education Quartiles: MxFLS Data

	Education Quartiles			
	Q1	Q2	Q3	Q4
Identify as indigenous	-0.1230	-0.0762	-0.1705	-0.0247
	(0.0568)	(0.0677)	(0.1920)	(0.0861)
Indigenous language	-0.0941	-0.1025	-0.6173	-0.2570
	(0.0693)	(0.0899)	(0.3163)	(0.1236)
Years of education	-	-	-	-
Age	0.0536	0.0652	0.0762	0.0733
	(0.0062)	(0.0084)	(0.0315)	(0.0110)
Age^2	-0.0006	-0.0007	-0.0009	-0.0007
	(0.0001)	(0.0001)	(0.0005)	(0.0001)
Test Score	0.0402	0.0604	0.0944	0.1007
	(0.0130)	(0.0160)	(0.0582)	(0.0202)
Additional controls:				
Occupation	y	y	y	y
R^2	0.2083	0.1589	0.2890	0.2302
No of obs	4,813	3,209	270	2,376

Notes: Standard errors in parenthesis. Data from the 2002, 2005, and 2009 MxFLS. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of annual earnings. All specifications control for year, state of residence, and state of birth. The cut-offs for the education quartiles are: 6, 9, and 10 years of education.

Table 9: Log-Earnings Regressions by Cognitive Ability Quartiles: MxFLS Data

	Cognitive Ability Quartiles			
	Q1	Q2	Q3	Q4
Identify as indigenous	-0.0220	-0.1552	-0.0672	-0.0635
	(0.0690)	(0.0790)	(0.0757)	(0.0830)
Indigenous language	-0.2703	0.0066	-0.2057	-0.1519
	(0.0859)	(0.1004)	(0.1003)	(0.1060)
Years of education	0.0315	0.0326	0.0460	0.0549
	(0.0051)	(0.0053)	(0.0055)	(0.0053)
Age	0.0460	0.0642	0.0740	0.0781
	(0.0082)	(0.0078)	(0.0084)	(0.0095)
Age^2	-0.0005	-0.0007	-0.0008	-0.0009
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Test Score	-	-	-	-
Additional controls:				
Occupation	y	y	y	y
R^2	0.2174	0.2218	0.2459	0.3109
No of obs	2,852	2,649	2,620	2,547

Notes: Standard errors in parenthesis. Data from the 2002, 2005, and 2009 MxFLS. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of annual earnings. All specifications control for year, state of residence, and state of birth. The cut-offs for the standardized test-score quartiles are: -0.67304718, 0.0924423, and 0.7972154.

Table 10: Log-Earnings Regressions by Employment Status: Census Data

|	Indigenous and Mestizos		Indigenous only	
	Wage Worker	Self-Employed	Wage Worker	Self-Employed
	(1)	(2)	(3)	(4)
Identify as indigenous	-0.0689	-0.1293		
	(0.0016)	(0.0039)		
Indigenous language	-0.1117	-0.2167		
	(0.0020)	(0.0046)		
Monolingual Spanish			0.2687	0.3440
			(0.0064)	(0.0115)
Bilingual			0.1710	0.1540
			(0.0062)	(0.0105)
Years of education	0.0355	0.0402	0.0260	0.0253
	(0.0002)	(0.0004)	(0.0003)	(0.0007)
Age	0.0374	0.0422	0.0304	0.0356
	(0.0003)	(0.0007)	(0.0005)	(0.0012)
Age^2	-0.0004	-0.0005	-0.0004	-0.0004
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Additional controls:				
Occupation	y	y	y	y
Industry	y	y	y	y
R^2	0.3772	0.3776	0.3893	0.3571
No of obs	1,274,626	416,760	287,889	127,589

Notes: Standard errors in parenthesis. Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of monthly earnings. All specifications control for size of location, state of residence, and state of birth. The regressions in columns (1)-(2) include Indigenous and Mestizo workers. The regressions in columns (3)-(4) only include individuals who are Indigenous, and the omitted category in these columns is Monolingual Indigenous.

Table 11: Log-Earnings Regressions by Location Size: Census Data

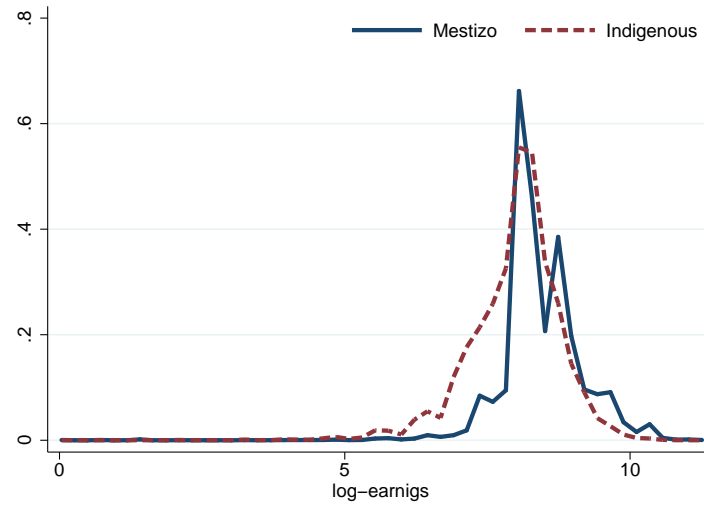|	Indigenous and Mestizos		Indigenous only	
	(1)	(2)	(3)	(4)
	Small Community	Large City	Small Community	Large City
Identify as indigenous	-0.1005	-0.0478		
	(0.0016)	(0.0045)		
Indigenous language	-0.1788	0.0130		
	(0.0020)	(0.0078)		
Monolingual Spanish			0.3389	0.1513
			(0.0059)	(0.1270)
Bilingual			0.1929	0.1406
			(0.0056)	(0.1270)
Years of education	0.0346	0.0534	0.0249	0.0390
	(0.0002)	(0.0003)	(0.0003)	(0.0010)
Age	0.0383	0.0466	0.0307	0.0455
	(0.0003)	(0.0006)	(0.0005)	(0.0018)
Age^2	-0.0004	-0.0005	-0.0004	-0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Additional controls:				
Occupation	y	y	y	y
Industry	y	y	y	y
R^2	0.3476	0.2565	0.3639	0.2617
No of obs	1,394,909	366,517	400,853	28,690

Notes: Standard errors in parenthesis. Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. The dependent variable is the natural logarithm of monthly earnings. All specifications control for state of residence, and state of birth. The regressions in columns (1)-(2) include Indigenous and Mestizo workers. The regressions in columns (3)-(4) only include individuals who are Indigenous, and the omitted category in these columns is Monolingual Indigenous.

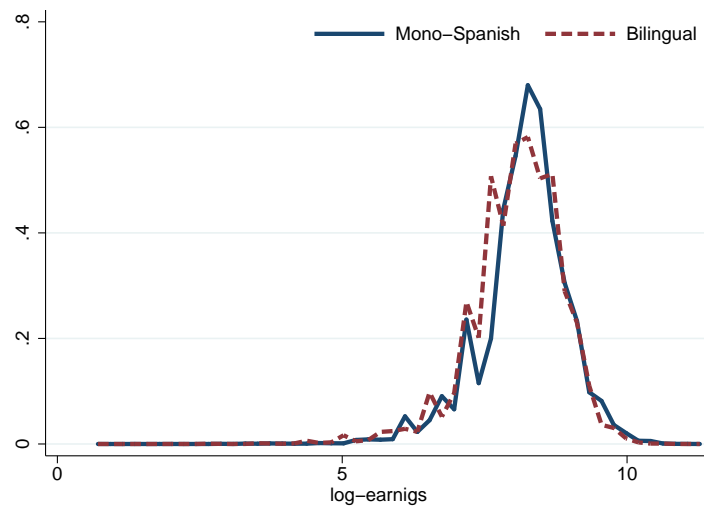Figure 1: Concentration of Indigenous Population in Mexico

21.5 – 63
11.5 – 21.5
5 – 11.5
2 – 5
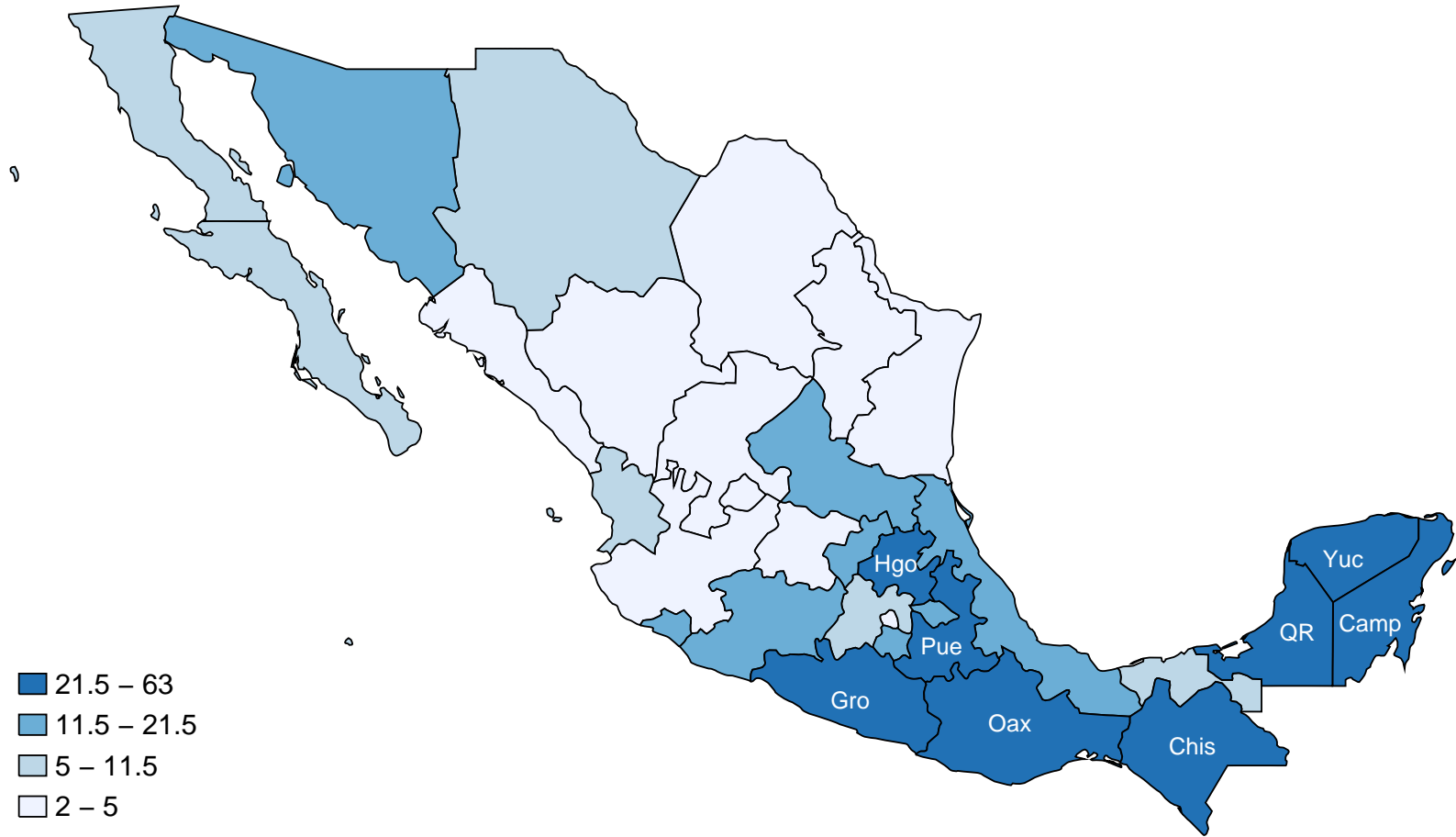
Notes: Authors' own calculations using data from the 2010 Mexican Census. The state abbreviations are: Yuc = Yucatan, Oax = Oaxaca, QR = Quintana Roo, Chis = Chiapas, Camp = Campeche, Hgo = Hidalgo, Pue = Puebla, and Gro = Guerrero.

Figure 2: Distributions of Log Earnings

(a) Indigenous Vs Mestizos

(b) Indigenous: Bilingual Vs Monolingual-Spanish

Notes: Data from the 2010 Mexican Census. Sample includes males ages 16-64 not attending school. Only observations used in the estimations presented in Table 2 are used.

APPENDIX

A   Mexican Census Data

The 2010 Mexican Census data is the 10 percent sample obtained from the *Instituto Nacional de Estadística y Geografía* from Mexico, INEGI. The sample includes 11,938,402 observations, but we only work with males and females ages 16 to 64. Respondents in this age group represent 58.89 percent of the sample. In an effort to focus on respondents who are committed to the labor market, we restrict the sample to respondents who are not attending school. We find that 89.96 percent of the respondents ages 16 to 64 are not attending school. The two main questions from the questionnaire for our analysis are: (i) the ability to speak an indigenous language, and (ii) ethnic self-identification. We find that 21.15 percent of individuals ages 16 to 64 not attending school speak an indigenous language. Similarly we find that 31.77 percent of individuals ages 16 to 64 not attending school self-identify as member of an indigenous group. We use the answer to these questions to create the indicator variables for Table 2 and the categories of Indigenous and Mestizo used in Table 3.

For individuals who indicate being able to speak an indigenous language, the questionnaire also asks them if they have the ability to speak Spanish. We find that 79.35 percent of the individuals ages 16 to 64 not attending school and that report being able to speak an indigenous language also can speak Spanish. We use this information to build the indicator variables for Table 3, which are: monolingual-indigenous, monolingual Spanish, and bilingual. We drop from the sample individuals who are monolingual indigenous but do not self-identify as members of an indigenous group. A comparison of the summary statistics reveal that these individuals are very similar to monolingual-indigenous who self-identify as members of an indigenous group and not to those who do not self-identify as members of an indigenous group. Hence, we believe that these individuals did not understand the question

correctly. Only 1,990 individuals are in this situation, and from them only 678 have valid observations so they could have been used in the estimation sample. The results are basically identical if we include or exclude these individuals.

We also drop observations with missing information in variables used in the analysis such as age, education, income, state of residence, state of birth, occupation, industry and population size (see Tables 1, 2, and 3). The income variable refers to the person's earned income and does not include any other sources of income such as income from welfare programs or transfers from other family members. The earned income is the monthly income from the previous month and it is reported as the income at the time of the census in Mexican Pesos. Since the census date was June of 2010 we use the CPI for May 2010 to deflate earned income.

The Census uses the official catalog published in the *Diario Oficial de la Federacion* (Mexico's Federal Register) to classify the indigenous language spoken by the respondent. This catalog identifies 68 different languages, and we used the the catalog published in the *Diario Oficial de la Federacion* to group these 68 languages in the 11 linguistic families used in Table 4 (see INALI, 2008). However, three of these families have very low proportions and so we group them in one category, these are: Algica, Cochimi-Yumana, and Seri. No observations are dropped from the main sample used in Tables 1, 2, and 3 for missing identification of the language.

The summary statistics presented in Table 1 are weighted using the person weight variable available in the Census, `wtper`, which indicates the number of persons in the actual population represented by the person in the sample.

B   Mexican Family Life Survey Data

We use the three waves from the Mexican Family Life Survey (MxFLS) which are 2002, 2005 and 2009 (see Rubalcava and Teruel, 2006a,b, 2013). The information provided by these surveys is organized in separate data files or "books" but the structure of these books remained constant over the three waves of the survey for the books that are relevant for our study. The following description refers to the particular books used to obtain the variables for the analysis.

First, we use the control book, Book C, to obtain geographical data such as state, city, county, and city size. We also obtain from Book C the month of the visit to the household which is used to locate the months that are relevant to deflate annual income measures, this is obtained in the back-cover section of Book C. Finally, a third section of Book C is used to obtain some individual data such as: respondent's age, gender, and annual earnings from the previous 12 months. This is the measure of earnings used in the regression analysis in Tables 6 and 7. A small number of observations (21) were deleted because of excessively high annual earnings values.

Individual language, ethnic self-identification, and education data was obtained from Book IIIA. The language and ethnic self-identification variables ask if the respondent: (i) speaks Spanish at household, (ii) reads and/or writes in Spanish, (iii) self-identify as indigenous, and (iv) speaks an indigenous language. This information is used to build the language variables in Tables 6 and 7.

The education information is very detailed for the highest grade completed up to grade 12 which is high school. We can also identify if the respondent attended the "open system" for middle school (grades 7 to 9) and for high school (grades 10 to 12) and use this information when necessary.[17] The education information for college and graduate degrees is not very

---

[17]We assign a highest grade completed of 9 if the respondent completed open middle school (6 grades for elementary school and 3 grades for middle school), and a highest grade completed of 6 if the respondent did

detailed and we only know whether the respondent attended and graduated from these degrees, and so we follow a similar strategy as with the open system schools.[18]

We use Book IIIA also to obtain the respondent's occupation or position in his/her main job. The categories for the respondent's occupation are: non-agricultural worker, agricultural worker, self-employed, and boss of his/her own company.

The measures of cognitive ability are obtained from Book EA. The measures of cognitive ability are those calculated with the Raven's progressive matrices. These matrices are designed to measure the cognitive ability of the respondent independently of the respondent's capability of reading and writing. The test of Raven's progressive matrices is taken by the respondent in all three visits to the household.

The MxFLS allows following respondents from the first wave even if they left the original household. If respondents in the original sample leave the original household and form a new household the members of the new household become part of the sample. The MXFLS follows respondents even when they change cities or states. Hence in the second and third waves we have new members in the sample which helps in reducing the problem of attrition in the sample. Merging the three waves is done following the instructions in Rubalcava and Teruel (2013).

The MxFLS does not cover all Mexican states. In the first wave, only 16 out of the 32 Mexican states are covered. These states are: Baja California Sur, Coahuila, Distrito Federal, Durango, Guanajuato, Jalisco, Estado de Mexico, Michoacan, Morelos, Nuevo Leon, Oaxaca, Puebla, Sinaloa, Sonora, Veracruz, and Yucatan. New states appear in the second and third waves as the MxFLS follows respondents who were in the original sample and then migrated.

---

not complete open middle school. Similarly, we assign a highest grade completed of 12 if the respondent finished open high school (9 grades for elementary and middle schools and 3 grades for high school) and a highest grade completed of 9 if the respondent did not complete open high school.

[18]We assign a highest grade completed of 16 if the respondent completed college or teacher's college, and 12 if they attended these levels of education but did not get the degree. Similarly we assign 18 years of education if the respondent completed a graduate degree and 16 if they attended a graduated degree but did not get the degree.

Then, in the regression analysis, there will be less than 32 dummies for state of residence, but there are 32 dummies for state of birth since we have observations where individuals were born in a different state from the state of residence.