Society for the Study of Economic Inequality

Working Paper Series

Conditional occupational segregation of minorities in the U.S.

Carlos Gradín

ECINEQ WP 2010 – 185

ECINEQ 2010-185
October 2010

www.ecineq.org

Conditional occupational segregation of minorities in the U.S.*

Carlos Gradín †
Universidade de Vigo

Abstract

In this paper, we use a propensity score-based methodology to analyze the role of demographic and human capital characteristics of minorities in the U.S. in explaining their high occupational segregation with respect to whites. Thus, we measure conditional segregation based on an estimated counterfactual distribution in which minorities are given the relevant characteristics of whites. Our results show that the different levels of attained education by ethnicity and race explain a substantial share of occupational segregation of non-whites in the U.S., while English skills or immigration status are especially relevant for explaining segregation among Hispanics and Asians.

Keywords: conditional occupational segregation, race and ethnicity, United States.
JEL Classification: D63, J15, J16, J71, J82.

* I acknowledge financial support from the Spanish Ministerio de Educación y Ciencia (Grant ECO2010-21668-C03-03/ECON).

† **Address of correspondence:** Carlos Gradín, Facultade de CC. Económicas e Empresariais, Universidade de Vigo, Campus Lagoas-Marcosende, s/n 36310 Vigo, Galicia, Spain. cgradin@uvigo.es. Phone: +34 986813527; Fax. +34 986812401.

Introduction

The unequal distribution of population groups across occupations in the United States based on their race and ethnicity has received considerable attention during past years. Confinement of a demographic group to certain types of jobs undermines social cohesion in a country and may be linked to racial discrimination. It, further, may reduce people's opportunities to earn income and social recognition, especially when minorities are able to work only in low-paid jobs regardless of their skills.

The fact that minorities in the United States such as Hispanics, Asians, or blacks face high levels of segregation in their jobs has already been documented (Albelda, 1986; King, 1992; Spriggs and Williams, 1996; Queneau, 2009; Alonso-Villar, Del Río and Gradín, 2010). Less clear is to what extent their high level of segregation is explained by differences in the attributes of these groups with respect to whites. Especially important are skill-related characteristics such as attained level of education, the ability to speak English, or immigrants' labor experience in the host country. The level of human capital determines the types of jobs that a worker qualifies for, and her exclusion from certain jobs on the basis of lacking the necessary skills leads to segregation of a different nature based on pre-labor market conditions that should be treated differently than segregation that remains even when these variables are taken into account. The uneven geographical distribution of workers across the country is another factor that could affect segregation. If Asians or Hispanics are concentrated mostly in certain areas of the U.S., it is reasonable to expect that they will be overrepresented in jobs that are more available in those regions. Similarly, other factors such as workers' age or the limited transferability of skills of immigrants could partially explain observed segregation.

Little has been done so far, however, in measuring occupational segregation conditioned by these covariates in a sufficiently general way, and in quantifying the individual contribution of each of these differential characteristics in explaining the observed level of segregation. The literature has dealt, so far, with this problem in various ways, such as using decomposition by subgroups, designing specific segregation indices, or running multivariate regressions that exploit the variability of segregation across local markets or along time, among other methods.

It is important to note that a low level of conditional segregation does not preclude any form of discrimination. Discrimination could still be present because a low level of human capital in a specific group might be the result of either unequal opportunities in obtaining the demanded skills or of anticipating low returns on these skills due to prevailing discrimination in the labor market. What is clear is that both types of segregation, explained and unexplained by characteristics, are of a different nature and should be treated differently. Furthermore, the identification and quantification of which factors underlie segregation will help to better understand this phenomenon, its evolution, and what policies would be more effective in reducing it.

The aim of this paper is to adapt a methodology based on the DiNardo, Fortin, and Lemieux (1996) propensity score technique, used in decomposing wage differentials, to construct a counterfactual employment distribution in which one minority, i.e. blacks, is given the relevant characteristics of the reference group, i.e. whites. The level of occupational segregation in this counterfactual distribution is the conditional (or unexplained) segregation, while the difference between conditional and unconditional segregation represent the part that is explained by characteristics and can be further decomposed into the individual contribution of each covariate of interest. This method, which is consistent with the way in which conditional wages are usually computed to decompose wage differentials, is more general than others that have been previously proposed in the literature because it fits any kind of segregation measure and allows an analysis of segregation curves. In fact, it allows an explanation of the different proportions of workers by race in each occupation. Further, it applies to any number and type of covariates, either discrete or continuous. Using this methodology, we conduct an empirical analysis of segregation among minorities in the United States using the American Community Survey 2005-07, which will show that a significant proportion of the unconditional segregation of non-white workers is related to their specific endowments, with the magnitude and explicative factors varying by race and ethnicity.

The structure of the paper is as follows. The first section briefly reviews the literature. Then we introduce our methodology, to continue with the empirical analysis of segregation among minorities in the United States. The final section summarizes the main contributions of the paper.

1. The literature

Occupational segregation in the U.S. based on race or ethnicity has been the focus of several studies in the past. Albelda (1986) measured the dissimilarity index for the 1958-81 period using annual data provided by the Department of Labor, identifying a decreasing trend in occupational segregation by race, especially among women. This convergence in employment distribution by race was driven by major structural changes in the economy such as changes in educational distribution and the business cycle. Other studies (King, 1992; Spriggs and Williams, 1996) have further documented reductions in segregation by race and extended the analysis to population subgroups. Queneau (2009) showed that the reduction in racial segregation (blacks vs. non-blacks) between 1983 and 2003 was accompanied by an increase in ethnic segregation (Hispanic vs. non-Hispanics), indicating that both trends occur mainly due to a composition effect. More recently, using the 2007 American Community Survey, Alonso-Villar, Del Río, and Gradín (2010) measured multigroup segregation and showed that Latino and Asian minorities faced the highest levels of segregation by comparing their distributions across occupations with the employment structure of the economy (local segregation). These authors also remarked on the relatively lower level of segregation among female racial/ethnic groups compared with males.[1]

The extent to which the observed level of segregation can be explained by the groups' having different characteristics has already been addressed in various ways. One obvious solution involves computing segregation by specific partitions of the population by some relevant characteristics (for instance, Massey, 1979 or, more recently, Hellerstein and Neumark, 2008, or Alonso-Villar, Del Río, and Gradín, 2010], but this alternative does not solve the problem, as it does not allow the researcher to control for many attributes at the same time or to consider continuous variables. Several approaches have tried to identify factors explaining segregation by estimating multivariate regressions exploiting variability either along time (i.e. Albelda, 1986; Tomaskovic-Devey *et al.*, 2006) or across local markets (Alonso-Villar, Del Río and Gradín, 2010). Carrington and Troske (1998) estimated different OLS and ordered probit models of racial composition. Other papers have used the information of multivariate regressions to construct specific indices accounting for conditional

[1] Other studies have, rather, focused on workplace and inter-firm segregation (for example, Carrington and Troske, 1998; Tomaskovic-Devey *et al.*, 2006; Hellerstein and Neumark, 2008).

segregation. Spriggs and Williams (1996) constructed the *L*-index of segregation, which measures the extent to which race or sex affects the probability of being in an occupation, using logit estimates.[2] Measuring residential segregation, Bayer, McMillan and Rueben (2008) estimated regressions for the percentage of households of each race by census blocks and used the coefficients to compare own-race exposure predicted by the average characteristics of the model and by the average characteristics of the population as a whole. Aslund and Skans (2009) developed an approach in which they non-parametrically estimated the propensity of individuals' holding jobs with some discrete characteristics to be immigrants and used these estimates to achieve a counterfactual distribution by randomly allocating minority status to individuals within each cell resulting from crossing these characteristics, using the probability of being an immigrant as equal to the fraction of immigrants in the cell. Other statistical procedures can be found in Sethi and Somanathan (2009) and Mora and Ruiz-Castillo (2009).

2. Methodology

2.1 Measuring unconditional segregation

The measurement of occupational segregation is still a controversial issue in labor economics. In order to approach racial/ethnic segregation by occupation, most often, segregation has been measured in pair-wise comparisons between two given groups, typically a non-white minority (blacks or Hispanics) and whites.[3] With respect to measurement, several indices can be found in the literature, with the dissimilarity index (Duncan and Duncan, 1955) being, by far, the most popular in empirical analysis despite its well-known limitations. Other indices have been proposed fulfilling better properties, most of them borrowed from measurements of income inequality. Examples of these are the Gini index or the Generalized Entropy family of indices, which embraces the Theil index or the Hutchens square root as particular cases

[2] Kalter (2000) also links the dissimilarity index to a multivariate logit model.

[3] Most recently, the study of multigroup segregation has allowed the measurement of the overall segregation of all racial groups considered together. Reardon and Firebaugh (2002) surveyed several of these indices and evaluated their properties. In some cases, this overall segregation can be interpreted as the weighted sum of segregation of each group with respect to the whole economy (Alonso-Villar and Del Río, 2010).

(Duncan and Duncan, 1955, Hutchens, 1991, 2004). For this reason, in our empirical analysis, we will use a bundle of indices in order to check the sensitivity of our results.

For simplicity, let us consider a population of size N divided into two groups: N^1, whites, and N^0, non-whites. We are interested in measuring the segregation of this population across T occupations in the economy. Let us denote by $n^i = \left(n_1^i, \ldots, n_T^i\right)$ the distribution for one group across occupations, such that $N^i = \sum_{j=1}^{T} n_j^i$, i={0,1}. Then, based on the proportions of whites and non-whites in each occupation, we define the following segregation indices[4]:

$$D(n^0, n^1) = \frac{1}{2}\sum_{j=1}^{T}\left|\frac{n_j^1}{N^1} - \frac{n_j^0}{N^0}\right|$$

$$H(n^0, n^1) = 1 - \sum_{j=1}^{T}\sqrt{\frac{n_j^0}{N^0}\frac{n_j^1}{N^1}} \tag{1}$$

$$T(n^0, n^1) = \sum_{j=1}^{T}\frac{n_j^0}{N^0}\ln\left(\frac{n_j^0 / N^0}{n_j^1 / N^1}\right)$$

D is the dissimilarity index proposed by Duncan and Duncan (1955). H is the Hutchens square root index, whose appealing properties are well-described in Hutchens (1991, 2004).[5] Additionally, if occupations are sorted by the increasing magnitude of the ratio $e_j^1 / \left(e_j^0 + e_j^1\right)$, then we can write the Gini index as:

$$G(n^0, n^1) = 1 - \sum_{j=1}^{T}\frac{n_j^1}{N^1}\left(\frac{n_j^0}{N^0} + 2\sum_{h=j+1}^{T}\frac{n_h^0}{N^0}\right) \tag{2}$$

Note that D, G, and H are bounded between 0, when there is no segregation because whites and non-whites have the same distribution across occupations, and 1, when segregation is at its maximum because there is no overlap between both distributions (whites and non-whites work in different occupations).

[4] For simplicity, we will focus the analysis on pair-wise comparisons, but note that our methodology could be adapted to the multigroup case.

[5] Note that H (multiplied by 4) is a member of the family of Generalized Entropy $GE(c)$ measures when $c= 0.5$, while T corresponds to $c=1$. T is not defined if $n_j^1 = 0$ for any j.

An alternative and more robust approach to rank two distributions according to their level of segregation is to directly compute the segregation curves (Duncan and Duncan, 1955; see a formalization in Hutchens, 1991) that represent the cumulative proportion of whites on the ordinate and the cumulative proportion of non-whites on the abscissa when occupations are ordered in increasing values of $e_j^1 / \left(e_j^0 + e_j^1\right)$. The 45º line indicates the case of no segregation and, thus, nonintersecting segregation curves for two distributions indicate a lower level of segregation for the one with the curve lying closer to the 45º line. A variety of segregation indices will be consistent with this partial ordering, including those discussed above.[6]

2.2 Measuring conditional segregation

In this section, we adapt the approach of DiNardo, Fortin, and Lemieux (1996) to the measurement of occupational segregation.[7] This propensity score technique was initially proposed in the context of decomposing the wage differential between two given distributions across the entire distribution. In presenting the procedure, we first need to reformulate the notation. Each individual observation belongs to a joint distribution $F(e, z, W)$ of occupations $e \in \{1, 2..., T\}$, (continuous or discrete) individual characteristics $z = (z_1, z_2, ..., z_k, ..., z_K)$ defined over the domain Ω_z, and a dummy W indicating group membership. The joint distribution of occupations and attributes of each group is the conditional distribution $F(e, z \mid W)$. The discrete density function of occupations for each group, $f^i(e)$, can be expressed as the product of two conditional distributions:

$$f^i(e) \equiv f(e \mid W = i) = \int_z dF(e, z \mid W = i)dz = \int_z f(e \mid z, W = i) \cdot f(z \mid W = i)dz \quad \textbf{(3)}$$

where i=1 for whites and 0 for non-whites.

Then, under the general assumption that the structure of occupations of non-whites, represented by the conditional density $f(e \mid z, W = 0)$, does not depend on the

[6] Note that D indicates the maximum vertical distance between the 45º line and the segregation curve, while G is twice the area between the 45º line and the segregation curve.
[7] The approach presented here can, obviously, be applied to segregation across other types of units (such as workplaces or schools, for example).

distribution of attributes, we can define the hypothetical counterfactual distribution $f_Z(e)$:

$$f_Z(e) = \int_z f(e \mid z, W=0) \cdot f(z \mid W=1) dz = \int_z f(e \mid z, W=0) \cdot \psi_Z \cdot f(z \mid W=0) dz = \int_z \psi_Z f(e, z \mid W=0) dz \quad \textbf{(4)}$$

as the density that would prevail if the population of non-whites kept their own conditional probability of being in a given occupation, $f(e \mid z, W=0)$, but had the same characteristics of whites given by their marginal distribution $f(z \mid W=1)$. Expression (4) shows that this counterfactual distribution can be produced by properly reweighting the original distribution of the target group. The reweighting scheme ψ_Z can be obtained, after using Bayes' theorem, as the product of two probability ratios:

$$\psi_Z = \frac{f(z \mid W=1)}{f(z \mid W=0)} = \frac{\Pr(W=0)}{\Pr(W=1)} \frac{\Pr(W=1 \mid z)}{\Pr(W=0 \mid z)}. \quad \textbf{(5)}$$

The first ratio is given by the unconditional probabilities of group membership and is a constant. The second ratio is given by conditional probabilities and can be obtained by pooling the samples for whites and non-whites and estimating a logit (or probit) model for the probability of being white conditional on z. We will estimate the following logit model

$$\Pr(W=1 \mid z) = \frac{\exp(z\hat{\beta})}{1+\exp(z\hat{\beta})}, \quad \textbf{(6)}$$

where $\hat{\beta}$ is the associated vector of estimated coefficients.

For any given segregation index S, we can measure unconditional segregation defined over the distributions of occupations for whites and non-whites, $S(e) \equiv S\big(f(e \mid W=1), f(e \mid W=0)\big)$, and define segregation conditional on z to be the same index computed after replacing the density of non-whites by the counterfactual: $S(e \mid z) \equiv S\big(f(e \mid W=1), f_Z(e)\big)$. This is the amount of (unexplained) segregation that remains after controlling for characteristics. The difference between unconditional and conditional segregation (counterfactual) provides a measure of segregation that is actually explained by our covariates z. This is in line with how wage differentials are usually decomposed into their characteristics (explained) and coefficients (unexplained) effects. Then, unconditional segregation can be divided into its explained and unexplained parts:

$$S(e) = [S(e) - S(e \mid z)] + S(e \mid z) \ . \qquad \textbf{(7)}$$

One advantage of this method is that it permits the gathering of additional information using the counterfactual distribution. For example, we can also construct the conditional segregation curve or estimate the conditional density for any occupation in which we are interested.

Further, the explained term can be additionally disaggregated into the detailed contribution of each covariate (or subset of covariates) z_k in order to identify which factors are more explicative. With $s(z_k)$ being the relative contribution of covariate k,

$$S(e) - S(e \mid z) = \sum_k s(z_k)[S(e) - S(e \mid z)] \, . \qquad \textbf{(8)}$$

In order to obtain this detailed decomposition, we need to compute a new counterfactual distribution $f_{z_k}(e)$ in which the corresponding reweighting factor ψ_{zk} is obtained, setting all of the other logit coefficients but this one to zero (Lemieux, 2002). Alternatively, we can shift all of the coefficients in a specific sequence, computing the contribution of each factor as the result of changing its associated coefficients. This recalls the well-known path-dependency problem in inequality decomposition because the contribution of a factor to the overall differential in income will depend on the order in which we consider them. This difficulty will be overcome in the empirical analysis by computing the Shapley decomposition that results from averaging over all possible sequences (Chantreuil and Trannoy, 1999; Shorrocks, 1999).[8]

3. Empirical analysis

3.1 Data

The data used in the empirical analysis comes from the 2005-07 release of the Public Use Microdata Sample (PUMS) file of the American Community Survey (ACS)

[8] See Gradín (2010) for an application of a similar procedure for the decomposition of income distribution differentials across racial groups. Statistical inference for both the aggregate and the detail decompositions can be executed using bootstrapping. However, in the empirical analysis shown in the next section, no inference was made because the large dimension of the dataset makes bootstrapping too time-consuming a task.

conducted by the U.S. Census Bureau, thus reflecting the pre-recession situation in the U.S. labor market. This survey is the result of pooling a series of monthly samples jointly accounting for 3 percent of the overall population living in the U.S. in housing units during the period (and 2 percent of those living in group quarters during 2006-7). The sample amounts to 4,123,320 observations[9] of employed workers for which a variety of information is provided about their socio-demographic characteristics and labor market performance. In particular, we will analyze five different possible explicative factors: i) education, 16 groups defined by the census according to the level attained, going from no schooling to doctorate degree); ii) the ability to speak English, chosen among five categories (speaks only English, speaks English very well, well, not well, not at all); iii) immigrant status, distinguishing between those born in the U.S. and immigrants of different periods after arrival in the U.S. (less than a year, 1 to 5 years, 5 to 15 years, more than 15 years); iv) geographical location, defined as 158 metropolitan/nonmetropolitan areas of work[10]; and v) age measured in years and age squared.

Regarding race and ethnicity, people are asked in the survey to choose the race(s) with which they most closely identify and to answer whether they have or not Spanish/Hispanic/Latino origin. Based on self-reported identity, we identified the following mutually exclusive groups of workers: i) the four major single-race non-Hispanic groups, that is, whites, African Americans/blacks, Asians, and Native Americans (who could be American Indian, Alaskan, Hawaiian, or Pacific Island natives); ii) Hispanics of any race, but distinguishing whites from non-whites; and iii) others (non-Hispanics choosing other races or more than one race). Segregation will be measured separately for each gender due to the evidence of different occupational distributions of women and men of the same group membership. Regarding occupations, we have considered the detailed list of 469 occupations provided in the

[9] After a check to prevent the influence of outliers, four Hispanic male observations with $\Pr(W=0\,|\,z)$ of close to zero were discarded. Its inclusion would lead to disproportionally large counterfactual weights, according to expression (5).

[10] We considered 140 MSA with at least 4,000 sample observations. Workers were assigned to the MSA using the information of Public Use Microdata Area corresponding to the place of work (POWPUMA) available in publicly accessible ACS files, which, in some cases, required the assignment of a given POWPUMA to the MSA in which it has a larger population, according to the census. Workers with a job abroad were removed from the sample, and workers with a job but not currently working were assigned according to their area of residence. The remaining metropolitan and non-metropolitan areas were categorized, respectively, into the nine U.S. geographical regions, resulting in a total of 158 areas.

public PUMS files, which is based on the 2000 Standard Occupational Classification (SOC) System.

3.2 Unconditional segregation

It is well-known that there are high levels of occupational segregation of minorities with respect to whites in the U.S. The first four columns of Table 1 report for each demographic group (race or ethnicity by gender) the corresponding level of unconditional segregation as measured by the four indices described in the previous section. For all minorities, it is true that men are always more segregated than women with respect to (non-Hispanic) whites of the same gender.[11]

The ranking of groups according to their level of segregation is generally the same regardless of the index we use. In fact, most minorities can be ranked according to their segregation level using the segregation curves depicted in Figure 1 with no need to use indices. In the case of men, all distributions can be ranked despite some overlap between the curves of the most highly segregated minorities. Non-white Hispanic males show a higher level of segregation than any other group according to all indices, followed by Asian males and Hispanic white males. Workers of other races present the lowest level of segregation among men, followed by Native Americans and blacks. When no distinction is made among Hispanic males regarding their race, Asian males are slightly more highly segregated than they are, except for the dissimilarity index.

Regarding women, there are also clear dominance relationships for their segregation curves, but the curves are closer to one another, and the curve of white Hispanics crosses the corresponding curve for blacks and can hardly be distinguished from Asians'. According to all indices, non-white Hispanics and Asians also show the highest levels of segregation for women, but with white Hispanics and blacks having similar levels to those of Asians, especially for the dissimilarity index. Thus, Native American women and those of other races tend to show the lowest levels of segregation.

[11] In contrast, Alonso-Villar, Del Río, and Gradín (2010) found that Hispanic females are more segregated than Hispanic males when they are compared with the economy as a whole (instead of with whites).

Figure 1. Unconditional occupational segregation curves for whites/nonwhites in the US, 2005-07

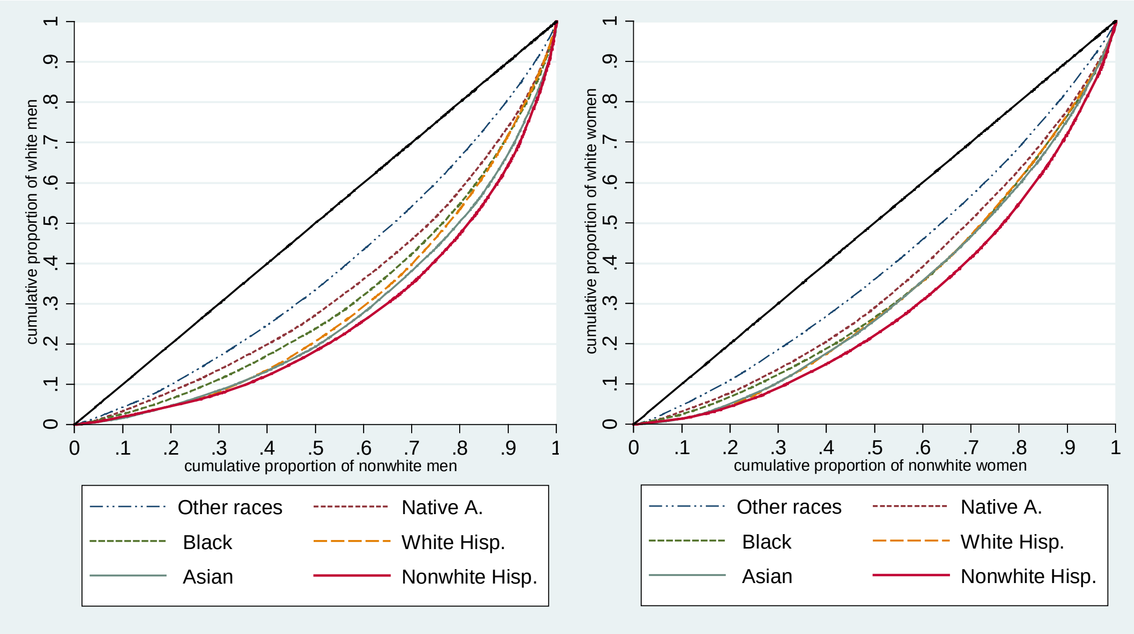

Table 1. Conditional and unconditional white/nonwhite occupational segregation in the US, 2005-07

	Unconditional				Conditional			
Group	**Duncan**	**Gini**	**Theil**	**Hutchens**	**Duncan**	**Gini**	**Theil**	**Hutchens**
Blacks								
males	0.280	0.383	0.242	0.060	0.255	0.347	0.202	0.050
females	0.245	0.333	0.192	0.046	0.220	0.306	0.163	0.039
Asians								
males	0.323	0.447	0.367	0.086	0.228	0.324	0.181	0.045
females	0.244	0.351	0.253	0.056	0.169	0.239	0.115	0.027
Native Americans								
males	0.241	0.334	0.185	0.047	0.170	0.252	0.120	0.030
females	0.211	0.294	0.156	0.038	0.171	0.249	0.121	0.030
Hispanics								
males	0.327	0.444	0.347	0.082	0.162	0.234	0.100	0.024
females	0.265	0.371	0.270	0.060	0.121	0.173	0.062	0.015
white males	0.308	0.420	0.318	0.074	0.167	0.239	0.105	0.025
white females	0.244	0.345	0.247	0.054	0.116	0.170	0.066	0.016
nonwhite males	0.351	0.476	0.389	0.095	0.201	0.291	0.166	0.039
nonwhite females	0.292	0.405	0.308	0.070	0.205	0.290	0.176	0.039
Other races								
males	0.167	0.236	0.091	0.023	0.126	0.180	0.055	0.014
females	0.142	0.200	0.068	0.017	0.118	0.167	0.053	0.013

3.3 Conditional segregation

It seems natural to ask to what extent the specific attributes of these groups along relevant dimensions such as education or geographic or demographic characteristics can explain their segregation levels. It is clear that groups such as Hispanics and Asians, which have in common a large share of recent immigration, tend to be far more segregated than the rest of groups with more native-born workers.[12] Indeed, groups with many immigrants tend to have different educational profiles, being either less educated (Hispanics) or more educated (Asians) than whites and any other group. To a lesser extent, other minorities, such as blacks and Native Americans, traditionally have faced lower educational outcomes compared with whites.[13] Further, immigrants face a series of barriers causing mismatches of educational attainment and occupation of employment (over- and under-education) more often than the native-born population: limited international transferability of skills, selectivity into migration, and labor-market discrimination (Chiswick and Miller, 2009). English proficiency tends to be more limited,[14] especially among Hispanics, considerably narrowing the range of available jobs and making promotion more difficult. Very often, jobs requiring low English skills are also those demanding low non-language skills (Maxwell, 2010). Hispanics are younger than any other group in age and, thus, have less experience, while Asians are similar in age to other minorities but are still younger than whites.[15] Minorities, additionally, tend to be located in specific areas of the country, with Asians and Hispanics being concentrated in geographical regions such as the Pacific, Middle Atlantic (Asians), and West South Central (Hispanics). Similarly, blacks are more overrepresented in the South Atlantic region and Native Americans in the West South Central area.

---

[12] While foreign-born workers comprised 82 percent of Asians of any sex, 65 percent of Hispanics males, and 53 percent of Hispanic females, they accounted for 5 percent of whites and less than 15 percent of blacks and Native Americans.

[13] For example, 39 percent of Hispanic men working have attained less than a secondary education (compared to 9 percent of whites and Asians, 13 percent of blacks, and 15 percent of Native Americans). Similarly, 54 percent of Asian male workers have achieved a bachelor's degree or higher compared to 32 percent of whites, 18 percent of blacks, and 11 percent of Latinos. Differences among women are similar.

[14] Among Hispanics, about 32 percent of men and 22 percent of women lack English skills (speak the language not well or not at all). This compares with 12 percent of Asian men (women: 14 percent), 4 percent of men of other races (women: 2 percent), and less than 1 percent in the rest of groups.

[15] The median Hispanic male worker is 35 years old (36 in the case of females), compared with 42 for whites, 39 for blacks and Asians, and 38 for Native Americans (the corresponding figures for women are roughly similar).

All of these factors altogether could differentially affect the opportunities of workers belonging to these minorities in the labor market, thus influencing segregation across occupations. For this reason, applying the methodology described in the previous section, we compute segregation conditional on a number of covariates, accounting for geographical location, education, English proficiency, immigration, and age, as described in the data description. Conditional segregation for each group is reported in the four last columns of Table 1, while Table 2 reports the change in the percentage of unconditional segregation after conditioning for all characteristics in the first block in the table.[16]

Segregation is, generally, reduced after accounting for covariates even if the extent varies significantly across groups and indices. These reductions are slightly higher in relative terms for women than for men in the cases of white Hispanics and Asians, of a similar magnitude in the case of blacks, and lower in the other cases (non-white Hispanics, Native Americans, and workers of other races). Taking the Duncan and Duncan dissimilarity index as a reference, it turns out that conditioning on covariates reduces more significantly the level of segregation for Hispanics than for any other group: about 47 percent for males and 54 percent for females. However, it is noteworthy that these reductions are similar for white and non-white Hispanic men but substantially lower for non-white females of the same ethnicity. The reductions for Asians (men and women) and Native American men are also high, around 30 percent. The lowest reductions are found in the case of blacks (9-10%). Native American females reduce segregation by 19% and people of other races by 17 percent (females) and 24 percent (males). The results from the Gini index are pretty similar to the those of the dissimilarity index, while those arising from using the Theil and Hutchens indices are qualitatively similar but differ in magnitude, with, generally, even larger reductions than those discussed above. For example, the reduction for white Hispanic males is about 65-67 percent and around 71-73 percent for females.

As a consequence, after controlling for covariates, according to conditional segregation, black males turn to be, now, the most segregated group of workers, followed by Asian males. Conditional segregation curves are presented in Figure 2. Clearly, in the case of men, the segregation curve for blacks is below the others, while the curves for Asians

[16] Estimates of the logit regressions for the probability of being white used in the computation of conditional segregation are reported in the Appendix.

and non-white Hispanics cross each other at the bottom of their employment distributions, and Native American and white Hispanics overlap for most of the range. In the case of women, segregation curves intersect in several cases, and there is a great range of overlap among them, indicating that the segregation levels are close. The most salient downward movement in the ranking after controlling for covariates according to segregation indices is observed in the case of white and non-white Hispanic males. The former has, now, a similar or lower level compared to those of Native Americans and Asian women. Similarly, white Hispanic women have become, along with workers of other races, the group with the lowest segregation levels. Still, female groups have lower levels of segregation than men's groups, except for Native Americans and non-white Hispanics, where they have similar levels.

Figure 2. Conditional occupational segregation curves for whites/nonwhites in the US, 2005-07

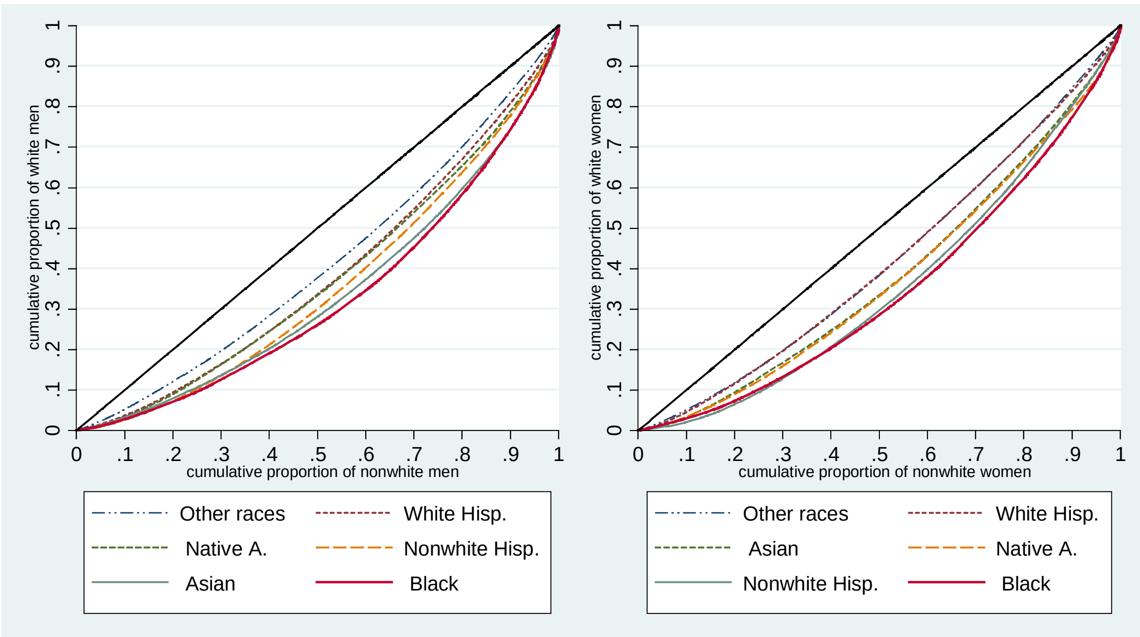

Table 2. Factors explaining white/nonwhite occupational segregation in the US, 2005-07
Percentage of change in unconditional segregation due to each set of characteristics (Shapley values)

	All characteristics				Geographic area				Education			
Group	**D**	**G**	**T**	**H**	**D**	**G**	**T**	**H**	**D**	**G**	**T**	**H**
Blacks												
males	-8.8	-9.4	-16.5	-17.3	4.8	4.6	9.6	8.8	-14.6	-14.0	-25.2	-25.3
females	-10.2	-7.9	-15.2	-14.8	8.8	7.3	16.1	13.9	-21.6	-16.5	-31.3	-29.6
Asians												
males	-29.5	-27.7	-50.8	-46.9	10.1	9.3	18.5	17.1	-10.1	-10.1	-14.4	-16.9
females	-30.8	-31.8	-54.6	-51.0	9.9	8.8	17.3	15.7	-6.5	-6.6	-11.9	-10.8
Native Americans												
males	-29.7	-24.5	-34.9	-35.4	-2.8	-1.8	-2.5	-1.6	-23.8	-20.0	-27.9	-30.1
females	-18.8	-15.2	-22.2	-19.8	0.6	2.0	4.7	5.6	-16.7	-15.8	-23.6	-23.9
Hispanics												
males	-50.5	-47.3	-71.1	-70.5	7.6	7.0	10.3	9.6	-23.3	-21.8	-31.1	-30.9
females	-54.2	-53.3	-77.0	-75.7	5.9	5.9	11.3	9.7	-27.9	-26.7	-37.1	-37.0
white males	-45.9	-43.1	-67.1	-65.5	7.6	7.0	9.3	9.0	-19.7	-18.5	-27.6	-26.9
white females	-52.5	-50.7	-73.1	-70.9	6.1	6.4	11.4	10.1	-24.5	-23.1	-32.7	-32.2
nonwhite males	-42.8	-38.7	-57.2	-58.7	8.6	8.4	14.1	12.7	-24.8	-23.1	-33.1	-34.0
nonwhite females	-29.9	-28.3	-43.0	-43.6	10.1	9.8	17.6	15.9	-31.1	-29.9	-54.0	-50.2
Other												
males	-24.5	-23.6	-39.6	-39.1	4.1	4.3	7.2	7.4	-8.4	-8.6	-14.5	-14.5
Females	-16.8	-16.5	-22.4	-23.2	10.6	8.8	16.6	15.2	-12.9	-12.9	-21.1	-20.5

	English proficiency				Immigration				Age			
Group	**D**	**G**	**T**	**H**	**D**	**G**	**T**	**H**	**D**	**G**	**T**	**H**
Blacks												
males	0.6	0.6	1.5	1.3	0.1	-0.6	-2.8	-1.9	0.3	0.0	0.5	-0.2
females	0.9	1.1	2.6	2.2	-0.3	-1.7	-5.9	-4.6	2.0	1.9	3.3	3.3
Asians												
males	-10.9	-9.7	-20.7	-17.3	-10.9	-11.1	-24.7	-20.2	-7.7	-6.1	-9.4	-9.7
females	-16.0	-15.2	-26.9	-25.1	-14.2	-15.2	-29.9	-27.0	-3.9	-3.6	-3.3	-3.8
Native Americans												
males	-2.6	-2.4	-4.4	-3.3	0.3	0.3	0.6	0.6	-0.8	-0.6	-0.7	-1.0
females	-6.6	-5.8	-11.8	-9.7	0.3	0.4	0.6	0.7	3.6	3.9	8.0	7.4
Hispanics												
males	-23.4	-22.0	-32.8	-31.7	-6.9	-6.6	-12.4	-11.5	-4.4	-3.9	-5.1	-6.0
females	-28.5	-28.5	-44.6	-41.9	-4.9	-5.0	-10.5	-9.2	1.4	0.9	3.9	2.7
white males	-22.1	-20.8	-30.8	-29.7	-6.6	-6.3	-11.8	-11.0	-5.1	-4.6	-6.3	-6.9
white females	-28.8	-28.3	-43.1	-40.4	-4.6	-4.6	-9.4	-8.3	-0.7	-1.0	0.7	-0.1
nonwhite males	-21.0	-19.3	-30.6	-29.0	-5.1	-4.9	-9.6	-8.8	-0.5	0.1	2.0	0.3
nonwhite females	-20.8	-19.4	-30.4	-29.4	-2.6	-2.3	-5.2	-4.6	14.4	13.5	28.9	24.7
Other races												
males	-2.0	-1.7	-3.0	-2.9	-1.4	-1.2	-1.7	-1.7	-16.8	-16.4	-27.6	-27.3
Females	-1.3	-1.4	-2.9	-2.6	0.3	0.0	-1.4	-0.7	-13.5	-11.1	-13.6	-14.6

Note: **D**=Duncan and Duncan Dissimilarity, **G**=Gini, **T**=Theil, **H**=Hutchens square root.

3.4 Main explicative factors

Ethnic and racial groups differ not only in the magnitude of the reduction due to controlling for characteristics but also in the nature of the underlying factors. The percentage of reduction in unconditional segregation that is induced by each set of variables (using the Shapley decomposition) is shown in the remaining blocks in Table 2. Here, we discuss the case of the dissimilarity index.[17]

The most relevant factors underlying segregation are the ability to speak English and educational attainment. Each of these factors alone accounts for a reduction in the dissimilarity index of about 20-30 percent in the segregation of Latino workers. While education seems to be more relevant in explaining the segregation of non-white Hispanics, English proficiency appears to be more relevant for whites of the same origin. This compares to values ranging from about 3-7 percent of the reduction for these groups due to their immigration status. Thus, the time of arrival in the U.S. seems to be not very important in explaining the segregation of Latinos, once the gap in observed skills (language and education) has been taken into account.

English proficiency, education, and immigration status explain about 10 percent of the segregation of Asian men and, respectively, 16, 14, and 6.5 percent for women of the same race. Thus, education and language skills are much less relevant in explaining the segregation of Asians than they are for Hispanics, while immigration status is more relevant for others, especially for Asian women. Age seems also to play a role in segregation for Asians (8 percent of reduction for men and 4 percent for women).

For Native Americans and blacks, most explained segregation can be attributed to their educational gaps: 15 and 22 percent of reduction for, respectively, black males and females and 24 and 17 percent in the case of Native Americans. Among the other factors, only English proficiency appears to be relevant for Native American females (7 percent).

Unlike the other attributes, controlling for geographical location generally increases rather than decreases the segregation of Hispanics, Asians, and blacks. That is, segregation would be larger (by 6-10 percent) had Latinos had the same geographical

[17] Note that, again, the Gini index provides similar results in sign and magnitude to those of the dissimilarity index, while the Theil and Hutchens indices provide similar qualitative results but with generally higher values.

distribution of whites. Similar percentages apply to the other groups. Similarly, controlling for age increases the segregation of non-white Hispanic women (14 percent), while it decreases segregation for white Hispanic men, having no effect on others of the same ethnicity.

Conclusions

In this paper, we have adapted a propensity-score technique, initially proposed in the literature with respect to wage differentials, for the analysis of conditional segregation. By measuring the segregation of a counterfactual occupational distribution in which non-whites are given the characteristics of whites, we quantify the segregation that can and cannot be explained by individual characteristics. This counterfactual is simply constructed by reweighting the original distribution of non-whites using predictions from a logit model of the probability of being white. Further, we are able to identify the individual contribution of each factor to overall segregation by following a Shapley approach.

Our technique is used to measure the conditional segregation of various minorities in the U.S. Our results show that the segregation of Hispanics and Asians show the largest levels of unconditional segregation. However, this high segregation can be, to a large extent, attributed to their specific characteristics, especially their lack of English proficiency, education attainment, and high percentage of recent immigration to the U.S. These three factors explain at least 50-60 percent of observed segregation for Hispanics and at least 30-35 percent for Asians. Among these factors, education and English proficiency are more important to Hispanics, while immigration status is more relevant for Asians. By contrast, blacks, with a larger share of native-born workers, show relatively low segregation levels compared to the other groups before conditioning, but a smaller share of that can be explained by their attributes, ending up with higher conditional segregation than any other group. In this case, education is the only salient factor. In all cases, except for Native American male workers, if minorities had the same geographical distribution across the country as whites, their segregation would be even larger; that is, the uneven geographical distribution mitigates the segregation that would be observed otherwise.

The new technique proposed here allows us to say that ethnic and racial segregation in the U.S. is, to a large extent, explained by individual attributes of non-whites, especially those related to recent immigration that are expected to eventually vanish with the progressive assimilation of foreign-born workers. However, a substantial level of segregation is explained by some minorities' having low educational profiles compared to whites. This remarkably affects also those minorities with larger shares of native-born workers. Further, a notable share of observed segregation still remains unexplained and could be caused by any form of racial/ethnic discrimination faced by these groups in the labor market.

References

Albelda, Randy (1986), "Occupational Segregation by Race and Gender, 1958-1981", *Industrial and Labor Relations Review*, vol. 39(3): 404-411.

Alonso-Villar, Olga and Del Río, Coral (2010), "Local versus Overall Segregation Measures", *Mathematical Social Sciences*, 60(1): 30-38.

Alonso-Villar, Olga, Del Río, Coral and Gradín, Carlos (2010), "The Extent of Occupational Segregation in the US: Differences by Race, Ethnicity, and Gender", ECINEQ Working Papers Series, 180.

Aslund, Olof and Skans, Oskar N. (2009), "How to measure segregation conditional on the distribution of covariates", *Journal of Population Economics*, 22: 971-981.

Carrington, William J., and Kenneth R. Troske (1998), "Interfirm Segregation and the Black/White Wage Gap", *Journal of Labor Economics*, 16(2): 231-260.

Chantreuil, F. and Trannoy, Alain (1999), "Inequality Decomposition Values: the Trade-Off Between Marginality and Consistency", THEMA Working Papers 99-24, THEMA, Université de Cergy-Pontoise.

Chiswick, Barry R., and Paul W. Miller (2009), "Educational Mismatch: Are High-Skilled Immigrants Really Working at High-Skilled Jobs and the Price They Pay If They Aren't?" in Barry R. Chiswick (ed.), *High Skilled Immigration in a Globalized Labor Market*, Washington D.C.: American Enterprise Institute (forthcoming).

DiNardo, John, Fortin, Nicole .M., and Lemieux, Thomas (1996), "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach", *Econometrica* 64:1001–1044.

Duncan, Otis D. and Duncan, Bervely (1955), "A Methodological Analysis of Segregation Indexes", *American Sociological Review*, 20(2): 210-217.

Gradín, Carlos (2010), "Race and income distribution: Evidence from the US, Brazil and South Africa", ECINQ Working Paper Series, No. 179.

Hellerstein, Judith, K., and Neumark, David (2008), "Workplace Segregation in the United States: Race, Ethnicity, and Skill", *Review of Economics and Statistics*, 90(3): 459-477.

Hutchens, Robert M. (2004), "One Measure of Segregation", *International Economic Review* 45(2): 555-578.

Hutchens, Robert M. (1991), "Segregation Curves, Lorenz Curves, and Inequality in the Distribution of People across Occupations", *Mathematical Social Sciences*, 21: 31- 51.

Kalter, Frank (2000), "Measuring Segregation and Controlling for Independent Variables", MZES Working Paper, 19, Mannheim.

King, Mary, C. (1992), "Occupational Segregation by Race and Sex, 1940-88", *Monthly Labor Review*, 115: 30-37.

Lemieux, Thomas (2002), "Decomposing Wage Distributions: a Unified Approach", *Canadian Journal of Economics*, 35, pp. 646-88.

Massey, Douglas S. (1979), "Effects of Socioeconomic Factors on the Residential Segregation of Blacks and Spanish Americans in U.S. Urbanized Areas", *American Sociological Review*, vol. 44(6), pp. 1015-1022.

Maxwell Nan L. (2010), "English Language and Low-Skilled Jobs: The Structure of Employment", *Industrial Relations*, vol. 49(3), pp. 457-465.

Mora, Ricardo and Ruíz-Castillo, Javier (2009), "The Statistical Properties of the Mutual Information Index of Multigroup Segregation", Working Paper 09-84, Economic Series (48), Universidad Carlos III de Madrid.

Queneau, Hervé (2009), "Trends in Occupational Segregation by Race and Ethnicity in the USA: Evidence from Detailed Data", *Applied Economics Letters*, 16: 1347-1350.

Reardon, Sean F. and Firebaugh, Glenn (2002), "Measures of multigroup segregation", *Sociological Methodology*, 32: 33-76.

Rawlston, Valerie, and William E. Spriggs (2002), "A Logit Decomposition Analysis of Occupational Segregation: An Update for the 1990s of Spriggs and Williams", *Review of Black Political Economy*, Spring, 91-96.

Sethi, Rajiv, and Somanatham, Robini (2009), "Racial Inequality and Segregation Measures: Some Evidence from the 2000 Census", *Review of Black Political Economy*, (36): 79-91.

Shorrocks, Anthony (1999), "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value", Mimeo, University of Essex.

Springgs, William E., and Williams, Rhonda M. (1996), "A Logit Decomposition Analysis of Occupational Segregation: Results for the 1970s and 1980s", *Review of Economics and Statistics*, 78; 348-355.

Tomaskovic-Devey, Donald; Zimmer, Catherine; Stainback, Kevin; Robinson, Corre; Taylor, Tiffany; and McTague, Tricia (2006), "Documenting Desegregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966-2003", *American Sociological Review*, 71: 565-588.

Appendix

Table A1. Logit regressions of the probability of being white: pool samples of whites and each minority

	Hispanic		Black		Asian		Native American		Other races	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
No school	-1.275	-1.103	-0.615	-0.508	-1.122	-1.477	-0.391	-0.550	-0.717	-0.833
	(0.053)	(0.073)	(0.070)	(0.071)	(0.085)	(0.084)	(0.172)	(0.199)	(0.123)	(0.171)
Nursery to Grade 4	-1.771	-1.383	-0.866	-0.816	-0.059	-0.331	-0.287	-0.789	-1.051	-0.642
	(0.070)	(0.100)	(0.084)	(0.111)	(0.124)	(0.143)	(0.239)	(0.322)	(0.146)	(0.175)
Grades 5-6	-2.111	-1.963	-0.676	-0.852	-0.385	-0.813	-0.581	-0.903	-1.052	-0.835
	(0.048)	(0.063)	(0.058)	(0.077)	(0.083)	(0.090)	(0.196)	(0.201)	(0.132)	(0.145)
Grades 7-8	-0.573	-0.599	0.049	-0.324	0.042	-0.275	-0.202	-0.363	-0.488	-0.312
	(0.029)	(0.041)	(0.035)	(0.041)	(0.058)	(0.066)	(0.089)	(0.111)	(0.074)	(0.094)
Grade 9	-0.852	-0.796	0.075	-0.135	-0.136	-0.296	-0.338	-0.421	-0.313	-0.473
	(0.024)	(0.034)	(0.031)	(0.035)	(0.070)	(0.073)	(0.085)	(0.096)	(0.069)	(0.081)
Grade 10	-0.307	-0.289	0.009	-0.232	-0.118	-0.092	-0.252	-0.238	-0.252	-0.171
	(0.024)	(0.030)	(0.022)	(0.023)	(0.052)	(0.053)	(0.065)	(0.076)	(0.055)	(0.060)
Grade 11	-0.341	-0.248	-0.310	-0.508	0.096	0.015	-0.232	-0.381	-0.229	-0.339
	(0.023)	(0.027)	(0.020)	(0.019)	(0.057)	(0.056)	(0.058)	(0.064)	(0.048)	(0.051)
Grade 12	-0.280	-0.319	-0.247	-0.403	-0.231	-0.467	-0.156	-0.208	-0.161	-0.289
(no diploma)	(0.026)	(0.034)	(0.023)	(0.025)	(0.046)	(0.050)	(0.069)	(0.080)	(0.056)	(0.066)
Some college	0.239	0.209	0.251	0.191	-0.068	0.107	0.225	0.234	-0.040	-0.058
(<1 year)	(0.018)	(0.018)	(0.015)	(0.013)	(0.035)	(0.037)	(0.044)	(0.042)	(0.035)	(0.034)
1+ year in college	0.376	0.291	0.159	0.028	-0.370	-0.136	0.339	0.236	-0.092	-0.109
(no degree)	(0.014)	(0.014)	(0.011)	(0.010)	(0.024)	(0.025)	(0.033)	(0.032)	(0.026)	(0.026)
Associate's degree	0.483	0.527	0.295	0.286	-0.412	-0.137	0.384	0.307	0.008	0.065
	(0.019)	(0.018)	(0.014)	(0.012)	(0.029)	(0.029)	(0.043)	(0.039)	(0.034)	(0.032)
Bachelor's degree	1.082	1.038	0.882	0.756	-0.889	-0.665	1.023	1.004	0.337	0.336
	(0.015)	(0.015)	(0.011)	(0.010)	(0.020)	(0.021)	(0.039)	(0.039)	(0.026)	(0.026)
Master's degree	1.532	1.318	1.036	0.797	-0.929	-0.519	1.141	1.093	0.393	0.336
	(0.025)	(0.022)	(0.017)	(0.013)	(0.024)	(0.025)	(0.058)	(0.056)	(0.038)	(0.035)
Professional degree	1.255	1.256	1.501	1.207	-1.156	-0.900	1.405	1.167	0.450	0.229
	(0.034)	(0.041)	(0.031)	(0.030)	(0.033)	(0.040)	(0.095)	(0.109)	(0.056)	(0.060)
Doctorate degree	2.004	1.718	1.344	1.147	-0.955	-0.546	1.374	1.168	0.644	0.296
	(0.052)	(0.063)	(0.039)	(0.041)	(0.034)	(0.049)	(0.123)	(0.161)	(0.074)	(0.089)
English: very well	-3.188	-3.299	0.169	0.463	-1.968	-1.874	-2.009	-2.022	-0.436	-0.419
	(0.012)	(0.012)	(0.019)	(0.019)	(0.018)	(0.020)	(0.034)	(0.033)	(0.036)	(0.037)
English: well	-3.419	-3.411	0.399	0.584	-2.413	-2.313	-1.871	-1.871	-0.397	-0.411
	(0.019)	(0.020)	(0.031)	(0.033)	(0.023)	(0.025)	(0.065)	(0.071)	(0.054)	(0.065)
English: not well	-3.991	-3.928	0.724	0.627	-2.788	-2.732	-1.084	-1.035	-0.730	-0.712
	(0.023)	(0.025)	(0.046)	(0.046)	(0.030)	(0.032)	(0.113)	(0.127)	(0.069)	(0.075)
English: not at all	-4.758	-4.972	1.665	1.269	-2.911	-2.886	-0.548	-0.698	-1.149	-0.994
	(0.053)	(0.061)	(0.138)	(0.131)	(0.071)	(0.076)	(0.438)	(0.474)	(0.118)	(0.158)
Immigrant	-0.815	-0.260	-1.513	-1.336	-2.595	-2.759	0.008	0.030	-1.335	-1.099
(0-5 years)	(0.030)	(0.036)	(0.036)	(0.040)	(0.031)	(0.035)	(0.111)	(0.121)	(0.065)	(0.075)
Immigrant	-0.810	-0.335	-1.405	-1.366	-2.590	-2.722	0.440	0.420	-1.307	-1.103
(6-10 years)	(0.024)	(0.027)	(0.029)	(0.030)	(0.026)	(0.029)	(0.099)	(0.111)	(0.055)	(0.065)
Immigrant	-0.835	-0.392	-1.275	-1.253	-2.829	-2.965	0.337	0.513	-1.224	-0.973
(11-15 years)	(0.025)	(0.028)	(0.031)	(0.030)	(0.028)	(0.030)	(0.118)	(0.113)	(0.061)	(0.069)
Immigrant	-1.185	-0.911	-0.973	-0.911	-3.098	-3.275	0.013	0.143	-1.352	-1.327
(>15 years)	(0.015)	(0.017)	(0.016)	(0.016)	(0.018)	(0.019)	(0.056)	(0.060)	(0.034)	(0.035)
Age	-0.016	-0.044	-0.026	-0.056	0.059	0.041	-0.045	-0.066	0.031	0.026
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)
Age^2 (x100)	0.063	0.097	0.045	0.082	-0.030	-0.011	0.066	0.091	-0.004	0.004
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.004)	(0.004)
Intercept	2.620	2.856	1.629	1.783	3.342	3.608	6.687	6.822	3.023	2.872
	(0.040)	(0.045)	(0.033)	(0.031)	(0.063)	(0.068)	(0.134)	(0.138)	(0.076)	(0.076)
Pseudo R2	61.3	56.0	13.0	14.6	55.4	56.7	15.3	16.1	9.5	9.2
Wald chi2(182)	233126	190,637	60,501	71,571	153,119	139,165	14,249	14,770	18,121	17,490
Probability > chi2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
No. of observations	1,893,055	1,644,408	1,787,360	1,641,872	1,727,314	1,538,436	1,653,474	1,470,185	1,663,251	1,480,047

Notes. Omitted categories: high school graduate, speaking only English, born in the US. Dummies for geographical areas also included in all regressions but omitted here. Separate regressions were run for white and nonwhite Hispanics but are also omitted here for the sake of presentation.