EQUALITAS
Economics of Inequality and Poverty Analysis

EQUALITAS Working Paper No. 18

Occupational Segregation by Race and Ethnicity in the U.S.: Differences across States

Carlos Gradín

Olga Cantó

Coral del Río

2013

Occupational Segregation by Race and Ethnicity in the U.S.: Differences across States

Carlos Gradín,
(Universidade de Vigo and EQUALITAS)
Coral del Río,
(Universidade de Vigo and EQUALITAS)
Olga Alonso-Villar
(Universidade de Vigo)

Abstract

Using the 2005–2007 American Community Survey, this paper analyzes the extent of geographical disparities in occupational segregation by race/ethnicity across the U.S. states. Our results show that there is a great geographical variation in segregation. A large part is driven by spatial disparities in workers' characteristics, mainly due to differences in the distribution of ethnic/racial minorities and their immigration/linguistic profiles. Taking these characteristics into account reduces this variation and re-shapes the segregation map, with the highest segregation moving from states in the Southwest to those in the East Central region, where minorities face more segregating labor markets.

Keywords: occupational segregation; race; ethnicity; states; United States

JEL Classification: J15; J71; D63

Departamento de Economía Aplicada, Facultade de CC. Económicas, Universidade de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain. E-mails: cgradin@uvigo.es, crio@uvigo.es and ovillar@uvigo.es.

Acknowledgements. The authors would like to thank the Editor and the referees for their helpful comments and suggestions. Financial support from the *Ministerio de Ciencia e Innovación* (grants ECO2010-21668-C03-03 and ECO2011-23460), *Xunta de Galicia* (grants 10SEC300023PR and CN2012/178), and FEDER is gratefully acknowledged.

1. Introduction

The United States has an outstandingly diverse population. The proportion of non-Hispanic whites decreased from 76% in 1990 to 65% in 2009.[1] In this multiracial society, residential and school segregation by race have been extensively documented. However, research on occupational segregation by race and ethnicity is scarcer and the analyses have been mainly undertaken at the national level. But the mechanisms driving segregation in the labor market may vary across the country due to differences in social contexts, characteristics of minorities, and the type of labor market they face, which suggests the convenience of analyzing occupational segregation at different subnational levels (Tomaskovic-Devey et al., 2006; Kaufman, 2010).

Several studies have tackled occupational segregation in the U.S. at the local level, exploring either variation across metropolitan areas or investigating selected cities (Catanzarite, 2000; Semyonov et al., 2000; Ovadia, 2003; Alonso-Villar et al., 2013). Certainly, urban areas are the natural scale on which to study the employment possibilities of individuals while taking commuting time into account. This is well understood by geographers, who emphasize the role played by residential and workplace maps in explaining urban labor market segmentation (Wyly, 1999; Ellis et al., 2004; Wang, 2010). Nevertheless, variation can also be found at broader geographical levels. In this regard, King (1992) and Kaufman (2010) provide evidence that occupational segregation by race in the South is higher than in the North.

Understanding segregation at the state level is also important given the role played by states in the labor outcomes of minorities (Beggs, 1995). On the one hand, many labor regulations are undertaken by states in crucial matters such as minimum wages, unemployment benefits, hours worked, and equal employment opportunities (Fitzpatrick and Perine, 2007). On the other hand, after the 1996 reform, states became the main actors in shaping the welfare system, some programs of which involve work requirements. This decentralization has created great inequality in welfare programs across states (Kail and Dixon, 2011), including immigrants' eligibility (Hero and Preuhs, 2007). In addition, state governments have ever-increasing legislative activity related to immigration policy (mainly directed to control and deter unauthorized immigrants), with remarkable effects on Hispanics (Raphael and Ronconi, 2009). One could expect that all these labor environment disparities across states would yield large differences in occupational segregation by race and ethnicity but, as far as we know, no study has analyzed it yet.

The aim of this paper is, first, to measure the extent of variation in occupational segregation by race and ethnicity across U.S. states and, second, to identify which states experience the lowest/highest levels of integration of racial/ethnic minorities in the labor market. For that purpose, several multigroup segregation indexes are used. Nevertheless, differences in segregation levels among states do not necessarily reflect disparities in the integration of minorities. When comparing two states, it could be plausible that the level of observed segregation of one was substantially higher than that of the other due to a compositional effect. This would occur if the state with the highest segregation level had a larger proportion of groups of workers who typically face stronger segregation. From previous research, minorities are known to be more unevenly distributed across occupations than non-Hispanic whites; highly and less-educated workers are more unevenly distributed than workers with intermediate grades; and recent immigrants, especially if they lack English proficiency, are excluded from many jobs (Hellerstein and Neumark, 2008; Alonso-Villar et al., 2012). States with larger proportions of highly segregated groups are more likely to show higher segregation even if the probability of a worker with certain attributes being over/underrepresented in some jobs is essentially the same. For this reason, this paper measures not only unconditional segregation but also conditional segregation in each state based on an estimated counterfactual distribution in which each state is given the relevant characteristics of a state of reference, as will be explained later.

This paper contributes to the segregation literature beyond its empirical findings. It shows that when it comes to comparing occupational segregation between two different areas (metropolitan areas, states, countries, etc.), the differential between them has two sources of a very different nature. The first refers to the spatial heterogeneity in workers' characteristics and industrial structures, which affects the set of occupations available for workers. The second involves disparities among areas in the labor opportunities that they bring to minorities, which makes some areas more segregative than others. Being both sources essential to quantifying the segregation level in a particular area, the second should be considered the most important when comparing segregation across areas. Failing to disentangle both sources will produce misleading comparisons if they go in opposite directions. Our methodological contribution to the literature is to adapt a regression-based decomposition technique that allows one to separate the two sources of spatial variability.

The paper is structured as follows. Section 2 establishes the theoretical framework in which the two sources of variability across states are identified. Section 3 introduces the methodology, presenting both the multigroup segregation indexes and the regression-based decomposition technique. Section 4 shows the spatial disparities for both the unconditional and conditional analysis once all states share a common distribution of characteristics. Section 5 summarizes the main conclusions.

2. Multigroup Occupational Segregation across States: A Theoretical Framework

Most occupational segregation studies have focused on disparities between the distributions of two population groups (mainly women versus men and blacks versus whites). In these cases, segregation by either sex or race arises when the corresponding distributions across occupations depart from each other (as measured, for example, with the index of dissimilarity proposed by Duncan and Duncan, 1955). Minimum segregation occurs when both groups share the same proportion of workers in each occupation while maximum segregation is achieved when the groups work in completely different occupations. In a multigroup context, such as our case with six race/ethnic groups, one could follow a similar approach by comparing each group with a reference one (e.g. whites) or, alternatively, making all possible pair-wise comparisons among all groups. For an analysis at the state level, the first alternative is problematic because whites are not always the majority of workers (e.g. Hawaii). On the other hand, pair-wise comparisons would be cumbersome (15 comparisons for each state).

Other approaches allow the analysis of multigroup segregation by simultaneously quantifying the disparities among all groups (Silber, 1992; Reardon and Firebaugh, 2002; Frankel and Volij, 2011). The advantage of these synthetic measures is that they offer a summarized picture of the segregation by race/ethnicity in each state. The price to pay is that this does not inform about the situation of each specific group. However, Alonso-Villar and Del Río (2010) provided segregation measures that allow conciliating both approaches, showing their complementarities. Multigroup segregation measures are just a weighted average of the segregation of each group, with weights equal to their share on the total workforce. Indeed, no segregation exists if every group is evenly distributed among occupations (i.e. in each occupation the population share of the group is the same for all groups). Consequently, segregation may increase through two distinctive channels. Ceteris paribus, segregation increases as the distribution of a group across occupations departs from that of the whole population. This component captures the segregation of each group. Likewise, segregation also increases with the population shares of those groups whose distributions lie further away from that of the whole population (these groups will be referred to as the most segregated).

By choosing a multigroup approach, we put the state and minorities as a whole (and not each particular group) at the center of the analysis. Thus, we try to answer the question of to what extent minorities in the US face a homogenous labor market or, on the contrary, the segregation they face substantially varies across states depending on several factors.

Processes generating segregation are complex. Kaufman (2010) surveys the major perspectives that explain the persistent employment differential among groups. Some approaches focus on the characteristics that workers bring to the labor market (supply-side factors). Among them, the *human capital/skills deficit* approach sees job differences among groups as the result of differences in workers' education, experience, and skills. In addition, according to the *worker preference* explanation, segregation could be the consequence of social and cultural differences among groups, although as Kaufman points out, this argument is often used to explain segregation by sex but not by race.

Other approaches emphasize the characteristics of the setting in which work occurs (demand-side factors). In this regard, several theories highlight the role played by discrimination against minorities in the labor market. Thus, segregation could be the result of *tastes for discrimination* exercised by employers, workers, and consumers. Employers could also discriminate based on racial stereotyping about workers' performance so that individuals are qualified according to the group to which they belong (*statistical discrimination*). This mechanism interacts with *queuing* processes that allocate "good" jobs to the advantaged group favoring segregation in the labor market. Other demand-side explanations call attention to the role played by market and organizational structures, personnel practices, and, in general, the social and economic context. Some perspectives combine demand- and supply-side factors, as is the case of the *spatial mismatch* approach, because labor disparities could arise from the mismatch between housing location of minorities and business location that results from residential segregation by race/ethnicity.

This paper classifies the disparities in occupational segregation between states, as coming from two main components. The first component summarizes differences in segregation that are due to a *compositional effect* of workers and industries. That is, it accounts for differences in the characteristics that minorities bring to the labor market (supply-side sources) as well as the kind of jobs available to all workers in the state (one demand-side explanation). The remaining differential in segregation between two states is labeled *intrinsic segregation effect*, and it captures segregation due to social, cultural, and political factors, and this is what we believe makes the labor market in some states more intrinsically segregative than others. Disentangling these two sources of differences in segregation will provide a better understanding of this phenomenon allowing identifying the states that make the integration of minorities easier or more difficult. As we will see in what follows, there are reasons to think that the U.S. states diverge in both the distribution of relevant workers' characteristics or industries, that produce the compositional effect, and the institutions and social attitudes, that shape how segregative the labor market is.

The compositional effect

The *compositional effect* is closely related to the supply-side sources and the industrial structure. As the American Community Survey shows (see Tables A1-A3 in the appendix for details),[2] there are wide differences among states in the attributes that workers bring to the labor market and the jobs available for them.

With respect to workers' characteristics, the differences are remarkable. Racial and ethnic groups are not evenly distributed across states: Hispanics are more concentrated in California, Texas, and Florida; African Americans in some states of the East South Central and South Atlantic areas; Asians in California, Hawaii, and New York; and Native Americans (including American Indian, Alaskan, Hawaiian, and Pacific Islander natives) in Alaska, Arizona, New Mexico, Oklahoma, and Hawaii. These differences could give rise to geographical disparities in segregation not because of a better integration of minorities in some states but to lower numbers of those groups usually facing higher segregation. This is why in our empirical analysis we take into account the racial and ethnic composition of states.

Cross-state disparities in education are notable as well (e.g., the share of workers holding a bachelor's degree in Massachusetts doubles that of Nevada or Washington). States also show different immigration patterns (e.g., the percentage of workers with less than 5 years of residence in Florida is more than 5 times that of Ohio), which should be taken into account given that language and cultural differences might affect the range of jobs that immigrants are offered (Maxwell, 2010), especially if the number of years of residence in the U.S. is low. Moreover, the job opportunities of newly arrived immigrants are likely to depend on migrant networks (Hellerstein et al., 2010), which may reinforce the concentration of immigrants of a race/ethnic group in jobs with a high presence of that group (Patel and Vella, 2007). For these reasons, this paper not only controls for attained education but also immigration profile and English proficiency since recent immigrants lacking English skills are expected to be more segregated than native-born workers (Alonso-Villar et al., 2012).

Regarding industrial structure, as an example, the District of Columbia stands out for having the largest public administration (26% of the work force compared with around 5% in New York and in other states), while Indiana, Michigan and Wisconsin stands out in manufacturing (around 20% of their employment).

The intrinsic segregation effect

The *intrinsic segregation effect* is directly associated with the opportunities that the labor environment of states bring to minorities. It may arise from cross-state disparities in issues such as citizens' attitudes, unionization, government policies, or social capital and, therefore, it accounts for demand-side explanations, other than the industrial structure.

As already noted, states bear an active role in labor market regulation, welfare, and immigration control, which gives rise to notable spatial disparities. Thus, for example, while 23 states have minimum wages equal to the federal minimum wage, 18 states, including California, Connecticut, Illinois, Massachusetts, Oregon and Washington, and the District of Columbia have minimum wages above it.[3] Unemployment insurance benefits also vary widely across states: Massachusetts, New Jersey, Pennsylvania, and Washington offer (maximum) weekly unemployment benefits that more than double those of Mississippi, Arizona, Louisiana, or Alabama.[4] One would expect that states implementing labor and welfare policies more supportive of vulnerable groups (as the low-paid and the least employable), were those with more egalitarian institutional environments. This is relevant since, as shown by Beggs (1995), states with more egalitarian institutional environments tend to exhibit less inequality by race in terms of both earnings and access to good jobs.

Another example arises from the E-verify program, a voluntary employment verification program created by the U.S. government to reduce the hiring of unauthorized workers. By the end of 2012, 20 states required it for at least some employers (Arizona requires the use of E–

Verify for all employers while Florida, Oklahoma, and Pennsylvania only require it for state agencies, public employers, and/or public contractors).[5] As Amuedo-Dorantes and Bansak (2012, p. 547) point out, "E-Verify mandates, particularly statewide mandates, significantly curtail the employment likelihood of likely unauthorized male and female workers […] and appear to redistribute likely unauthorized labor toward industries often benefiting from specific exclusions, such as agriculture or food services." Consequently, this regulation may bring important consequences in terms of segregation. In addition, immigration laws recently passed in some states (like Arizona, Georgia, Alabama, and Utah) designed to expand the power of local police to check the immigrant status of population may also affect the labor opportunities of undocumented workers (mainly Hispanics).

Consistently with the above differences, there is also great variability across states in citizen ideology, as has been shown by political scientists (Berry et al, 1998). Population attitudes toward minorities follow a clear geographical pattern as well. According to the 2002 General Social Survey, the East South Central region has the most prejudiced attitudes of the country toward African Americans, Hispanics, and Asians, while the opposite is found in the Pacific region. All of this is likely to lead to different integration levels of minorities in the states.

According to our framework, if two states have the same industrial composition and the same distribution of characteristics among workers, all the difference in segregation between these states would result from the *intrinsic segregation effect*. Therefore, our interpretation is that the labor market in one state brings fewer opportunities for integration than the other (i.e., subgroups with the same characteristics in both states work in a more restricted set of occupations in one of them). On the contrary, if each subgroup of workers sharing similar characteristics has the same occupational opportunities in both states, all the segregation differential would arise from cross-state differences in the relative size of these subgroups and, therefore, from the *compositional effect*. In practice, states may diverge in both aspects- the composition of groups and the level of integration they bring to minorities. This is why it is important to quantify the two effects separately.

3. Data and Methods

3.1 Data

The data used in this study come from the 2005-2007 Public Use Microdata Sample (PUMS) files of the American Community Survey (ACS) conducted by the U.S. Census Bureau. This survey provides a variety of information on demographic and labor-related characteristics reflecting the labor market performance right before the 2008 economic recession.

Regarding race and ethnicity, people are asked to choose the race or races with which they most closely identify and answer whether they are of Spanish/Hispanic/Latino origin. This produces six mutually exclusive groups of workers composed of the four major single race groups that do not have a Hispanic origin, plus Hispanics of any race, and others: whites; African Americans or blacks; Asians; American Indians, Alaskan, Hawaiian, and Pacific Islander natives (hereinafter referred to as Native Americans); Hispanics; and other races (those non-Hispanics reporting some other race or more than one race). This study will omit the "non-Hispanic" origin of the groups from this point on, for simplicity.

The classification of occupations is consistent with the Current Population Survey, based on a detailed occupation recode of the Standard Occupational Classification System (SOC). The list includes 52 occupations.[6] Multigroup segregation measurement requires focusing the analysis on 32 states out of 50, together with the District of Columbia, with a significant sample for most demographic groups. Dropped states, those having two or more minorities with less than 520 observations, represent 9% of workers in the survey and are mainly those with smaller and less demographically diverse populations, mostly in the central and northwest areas of the country.[7] Working with this restricted set of states prevents the small-unit bias problem that leads to overestimation of the segregation level of groups with small samples while dropping states do not affect either the conditional or unconditional segregation measurement of the remaining states. The final sample used in our analysis includes 3,747,905 employed workers (from a minimum of 18,692 observations in Hawaii to a maximum of 467,119 in California).

3.2 Multigroup segregation indexes

To compute segregation in each state, three multigroup segregation indexes are used: *M*, *IP*, and *G* (the subscript referring to state is dropped for simplicity). The mutual information index, *M*, is the multigroup generalization of the index proposed by Theil and Finizza (1971) that has been recently characterized by Frankel and Volij (2011) in terms of basic axioms. It measures the reduction in the uncertainty of the distribution of employment among occupations due to knowledge of the distribution of population among racial/ethnic groups. We can write this as follows:

$$M = \sum_g \frac{C^g}{T} \log\left(\frac{T}{C^g}\right) - \sum_j \frac{t_j}{T}\left[\sum_g \frac{c_j^g}{t_j} \log\left(\frac{t_j}{c_j^g}\right)\right],$$

where C^g is the size of racial/ethnic group *g*; *T* represents total population; c_j^g is the number of individuals of group *g* in occupation *j*; and t_j is the size of occupation *j*.

The *IP* index proposed by Silber (1992),

$$IP = \frac{1}{2}\sum_g \sum_j \left|\frac{c_j^g}{T} - \frac{C^g}{T}\frac{t_j}{T}\right|,$$

is the generalization of the popular index of dissimilarity to the multigroup case, according to the proposal by Karmel and MacLachlan (1988) in the dichotomous case. Finally, the Gini index, *G*, corresponds to the unbounded version of the measure proposed by Reardon and Firebaugh (2002):

$$G = \frac{1}{2}\sum_g \sum_{i,j} \frac{t_i}{T}\frac{t_j}{T}\left|\frac{c_i^g}{t_i} - \frac{c_j^g}{t_j}\right|.$$

As shown by Alonso-Villar and Del Río (2010), multigroup segregation indexes *M*, *IP*, and *G* can be written as the sum of the segregation level of each group into which the economy is partitioned, weighted by the group's share on the total population. For example, we can write the mutual information index as $M = \sum_g \frac{C^g}{T} M^g$, where $M^g = \sum_j \frac{c_j^g}{C^g} \ln\left(\frac{c_j^g / C^g}{t_j / T}\right)$ represents the segregation of group *g* (according to the Theil index that results from comparing the distribution of group *g* with the distribution of total jobs across occupations). Consequently, the contribution of each race/ethnicity to the overall segregation of the state depends on both its population share and the disparities between the employment distribution of that group and the occupational structure of the state.

The three indexes measure the extent to which the distributions of racial and ethnic groups across occupations depart from the employment structure of the economy, but each index gives a different weight to these discrepancies. The *IP* and *G* indexes pay more attention to discrepancies that occur in those occupations in which groups have an intermediate presence, while index *M* is more affected by discrepancies that take place in occupations in which the groups have a low representation.[8] For the sake of simplicity, the results are presented based on the *M* index, but the other two have also been calculated -and discussed whenever they yield different results -to provide greater robustness to the analysis.

3.3 Measuring conditional segregation

This section presents the procedure that allows the computation of conditional segregation in each state when controlling for both workers' characteristics (race/ethnicity, education, immigration profile, and English proficiency) and industrial structure. It is a propensity score method initially proposed by Di Nardo et al. (1996) for the decomposition of wage differentials and later adapted by Gradín (2012) to measure conditional occupational segregation of nonwhites versus whites in the U.S. at the national level.

To obtain the counterfactual distribution of each state, we first partion workers in each state into several mutually exclusive subgroups or "cells," each a specific combination of attributes (e.g., Hispanic immigrants who have lived up to 5 years in the U.S., have a university degree, and work in the manufacturing sector). Let $z \equiv (z_1,...,z_k)$ denote a vector of k covariates describing these attributes. The counterfactual distribution is the density function the state would have were it given the same distribution of attributes of the state of reference while keeping unchanged the distribution of every subgroup across occupations in that state. This density function involves a weighting scheme according to which each subgroup in the state is given the same relative size as the corresponding subgroup in the state of reference. These weights, Ψ_z, can be easily estimated from the data:

$$\Psi_z = \frac{\dfrac{\Pr(D = \text{NewYork}|z)}{\Pr(D = \text{NewYork})}}{\dfrac{\Pr(D = s|z)}{\Pr(D = s)}} = \frac{\Pr(D = s)}{\Pr(D = \text{NewYork})} \frac{\Pr(D = \text{NewYork}|z)}{\Pr(D = s|z)},$$

where D is the categorical variable representing state membership. The first component can be directly approximated by the ratio between the population samples in both states. The second component can be obtained by estimating the probability of an individual with attributes z belonging to New York (rather than to its own state s) using a binary probability model. Therefore, we estimate the following logit model:

$$\Pr(D = \text{New York}|z) = \frac{\exp(z\hat{\beta})}{1+\exp(z\hat{\beta})},$$

over the pooled sample with observations from both states, where $\hat{\beta}$ is the associated vector of estimated coefficients.

Let S_i and S_0 be the (unconditional) levels of segregation in state i and in the reference state, respectively, and S_i^* the conditional segregation level obtained in the counterfactual distribution in which state *i* is given the characteristics of the reference. The differential in segregation between both states, $S_0 - S_i$, can be decomposed into two terms. The *compositional effect* is the change in segregation in state *i* after being given the characteristics of the state of reference, $CE = S_i^* - S_i$. It quantifies the part of the differential that can be explained by the set of covariates *z*. The *intrinsic segregation effect* is the difference in segregation between both states using the same distribution of characteristics, $ISE = S_0 - S_i^*$. It captures the unexplained segregation, i.e., the differences due to disparities in the labor opportunities that states bring to minorities. Both effects sum up the total differential[9]:

$$S_0 - S_i = \left(S_i^* - S_i\right) + \left(S_0 - S_i^*\right) = CE + ISE .$$

After completing the same exercise for every state, while keeping the reference state unchanged, it is possible to explore segregation disparities across states under a similar distribution of workers' characteristics and industrial structures by comparing their conditional segregation levels. Note that by doing so, the *intrinsic segregation effects* of any two states *i* and *j* are compared because $S_i^* - S_j^* = \left(S_0 - S_j^*\right) - \left(S_0 - S_i^*\right)$.

On the other hand, the *compositional effect* can be disaggregated into the detailed contribution of each factor (a subset of covariates) to identify which are more explicative (Gradín, 2012). We obtain these contributions using the Shapley decomposition (Shorrocks, 2013; Chantreuil and Trannoy, 2013). The main advantage of this decomposition, widely used in income distribution analyses, is that the contributions of covariates are path independent and sum up the overall explained segregation.

4. Segregation at state level

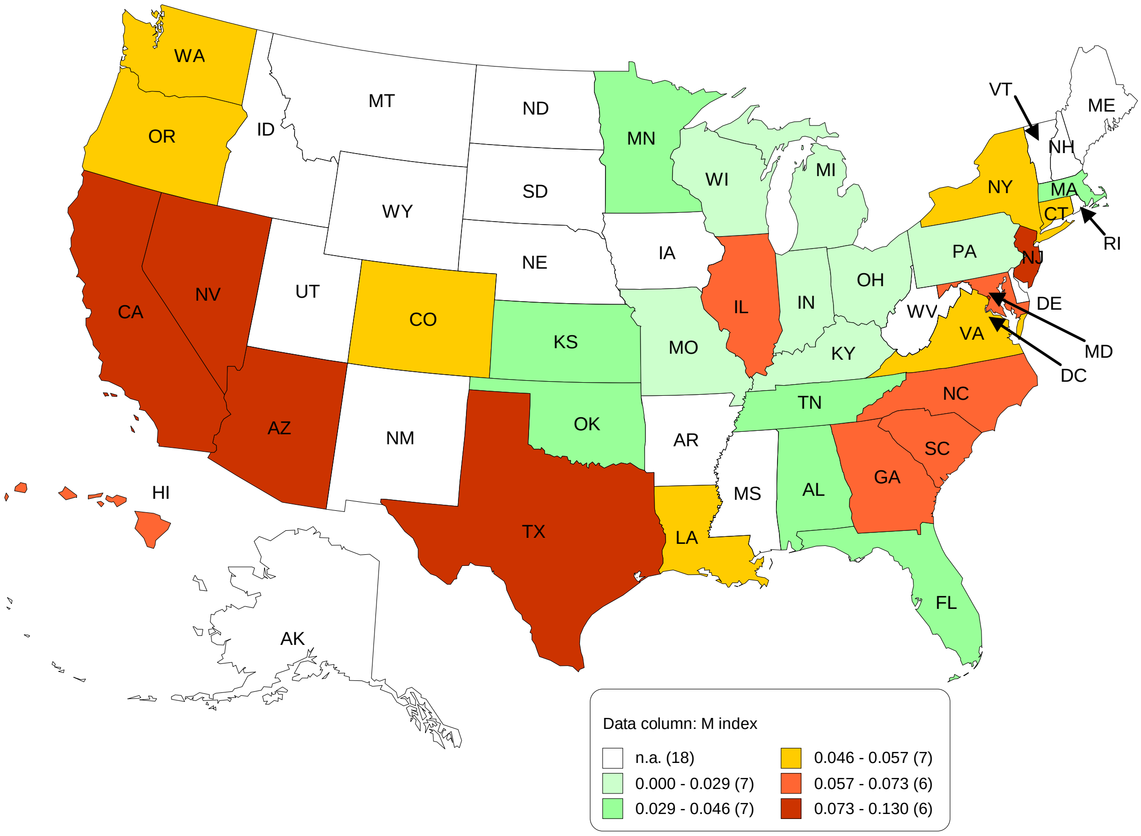
In this section, we first compute occupational segregation by race and ethnicity in each state to explore variation across the U.S. Then, we disentangle the two sources of variation in segregation across states: the *intrinsic segregation* and *compositional* effects,.

4.1 Unconditional segregation at state level

Using index *M*, Map 1 shows the unconditional segregation levels of states classified into five groups, each including six or seven states (the corresponding values for indexes *M*, *IP*, and *G* are given in Table A4 in the appendix). The map shows a great geographical variation in segregation, the coefficient of variation of segregation being equal to 0.482 (see Table 1). The highest level of segregation is found in the District of Columbia, which is more than double the average segregation (0.052), distantly followed by several western states (such as California, Nevada, and Arizona) and Texas. In the east, only New Jersey joins the District of Columbia in this highly segregated group. The lowest segregation levels can be found in Midwestern states such as Ohio, Wisconsin, Missouri, Indiana, and Michigan, together with Pennsylvania and Kentucky (which barely reach half of the average segregation). The *G* and *IP* indexes produce a similar ranking of states, except that they also include Hawaii in the former group and Minnesota in the latter (the coefficients of variation using these indices are 0.407 and 0.408, respectively).

There seems to be a clear link between the level of segregation of a state and its racial and ethnic composition. Highly segregated states share a relatively low presence of whites; some states have a large proportion of Hispanics, while others show remarkable racial diversity. By contrast, low-segregated states are more likely to have higher proportions of whites. This pattern suggests that the greater the degree of racial/ethnic diversity, the greater the segregation in a state. This could be driven by the high representation of the most segregated groups.

This is not the first time that the race/ethnic composition has shown itself crucial to explaining variation in segregation in the U.S. At the national level, Queneau (2009) found that the reduction in segregation for blacks and the increase for Hispanics between 1983 and 2002 were mainly due to a change in racial and ethnic composition rather than to changes in occupational structure. In the case of residential segregation, Iceland (2004) found that metropolitan areas with greater growth in Hispanic and Asian and Pacific Islander populations experienced greater growth in segregation for these groups.



Map 1. Unconditional occupational segregation by race/ethnicity in selected states (*M* index).
Note: White states have not been assigned a value due to lack of data in the survey.

But racial and ethnic diversity alone does not explain the whole geographical variation in segregation, because states with groups of a similar size have different segregation levels. Indeed, on the one hand, Tennessee, Alabama, Louisiana, Georgia, North and South Carolina, Virginia, and Maryland do not experience a similar segregation level (segregation is remarkably lower in Tennessee, Alabama, and Virginia) despite their large proportions of African-Americans and low proportions of other minorities (see Table A1 in the appendix). On the other hand, Florida has much lower segregation than California, Arizona, Texas, and Nevada, notwithstanding the similar large presence of Hispanics. The next subsection will discuss in more detail whether the great cross-state variability in segregation is just the result of differences in workers' characteristics and industrial structures (*compositional effect*), or is associated with different levels of *intrinsic segregation*.

4.2 Conditional segregation at state level

To disentangle the sources of variation of segregation across states, this paper initially measures conditional segregation using New York as the state of reference, whose characteristics are close to the characteristics of the whole country and could be interpreted as the "average" state. Later, it will be shown that the main results remain unaltered when an alternative benchmark is considered (California, which has a rather different demographic composition).

After pooling the sample of each target state and that of New York, the probability of a worker belonging to New York was estimated using a logit regression. The explicative variable is a dummy that has a value of 1 if the worker belongs to the New York sample and 0 if the worker belongs to the target state. Explaining variables are an array of 34 dummies accounting for four factors: race and ethnicity (six groups); attained education (less than high school, high school diploma, some college, and bachelor's degree or higher); immigration profile (born in the U.S., immigrant with up to 5 years of residence, between 6 and 10, between 11 and 15, or more than 15); English proficiency (speaking only English, speaking English very well, well, not well, not at all); and industry (the 14 groups of the North American Industry Classification System at one digit). For simplicity, gender and age were not used as control variables, because although relevant for segregation, the cross-state variation in these variables is small and so is their contribution to explain cross-state differences in segregation.

Using the probabilities predicted by the logit model as explained in Section 3, we obtain the counterfactual density of each state as if it had the same distribution of characteristics as New York. This density function is used to measure conditional segregation. We first compare unconditional and conditional segregation to measure the magnitude of the *compositional effect* in each state and identify which observable factors contribute to this variability in segregation across states. Second, we compare conditional segregation across states to identify those with the most and least segregative labor markets once the compositional effect is removed.

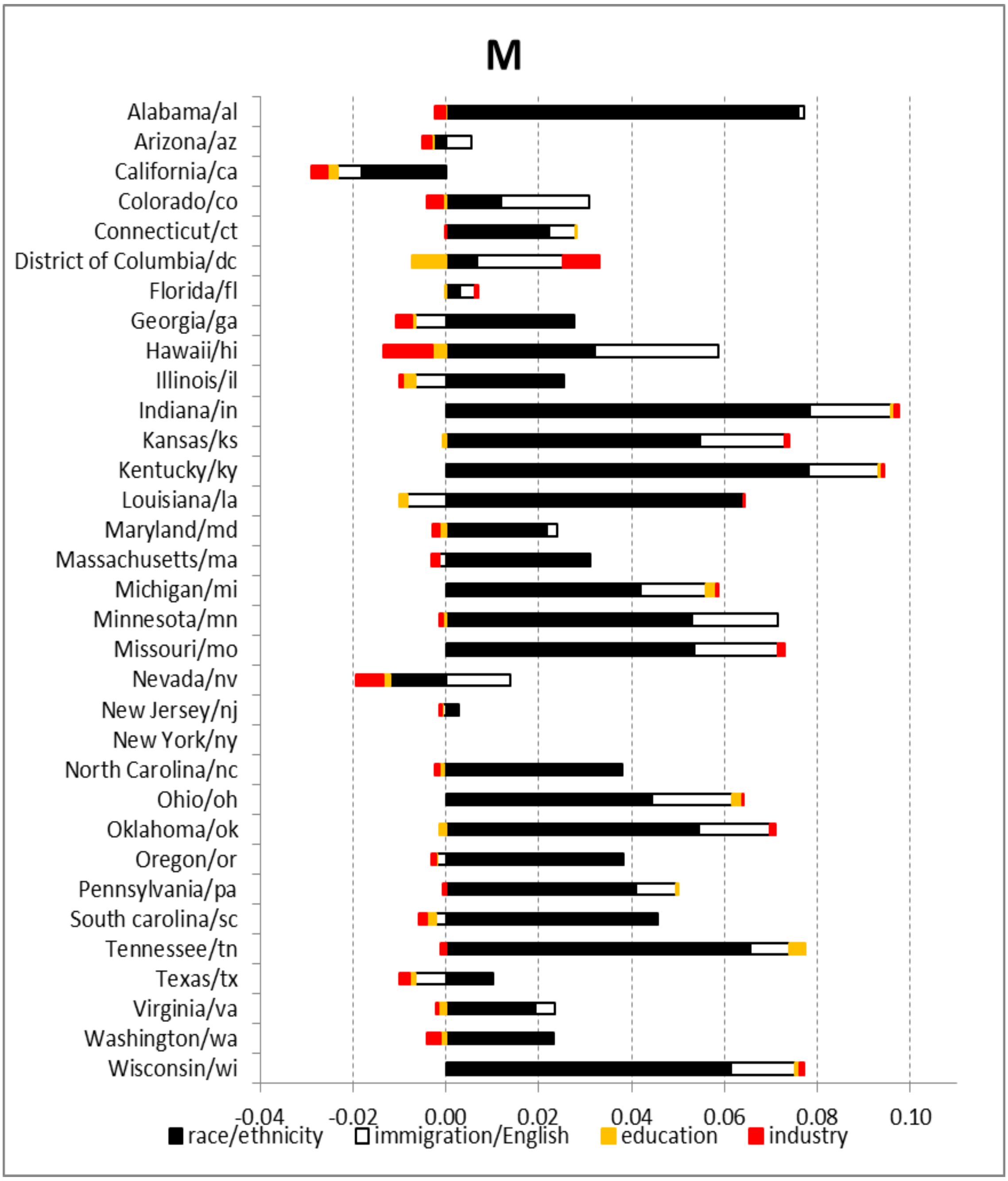


Figure 1. The *compositional effect*: Conditional-unconditional segregation gap in selected states (M index). Factors' contributions measured using the Shapley decomposition.

Figure 1 displays the change in segregation experienced by each state after conditioning on characteristics. This figure also shows the contribution to the overall change (estimated using Shapley decomposition) of each set of explanatory factors: race/ethnicity composition, educational level, immigration profile/English proficiency, and industry structure. Positive (negative) values indicate that segregation increases (decreases) after conditioning for that factor; the distribution of that characteristic in the target state leads to less (more) segregation than in New York. Obviously, segregation in the state of reference, New York, does not change at all by construction. The sum of contributions by the four factors (either positive or negative) represents the net overall change in segregation, the *compositional effect* of the differential in segregation between each state and the reference.[10]

Most states experience a net increase in segregation after conditioning on characteristics, the largest being Alabama (from 0.042 to 0.117, with *M* index), Indiana (from 0.025 to 0.123), and Kentucky (from 0.023 to 0.117), indicating that their distributions of characteristics, compared with that of New York, partially offset the underlying level of segregation faced by their minorities in their labor markets. The main exceptions are western states such as California and Nevada, where the net effect is negative; their distributions of characteristics produce more segregation than that of New York. Other states such as Arizona, Texas, and New Jersey experience virtually no net change, because positive and negative effects cancel each other. Florida also shows a very small variation.

	Unconditional Segregation	Compositional Effect (ref. New York)									
M		Total	Δ%	Race/ ethnicity	Δ%	Immigration	Δ%	Education	Δ%	Industry	Δ%
Mean	0.052	0.091	73.4	0.085	63.4	0.059	13.1	0.052	-1.2	0.051	-1.9
Standard deviation	0.025	0.022	-12.4	0.021	-14.8	0.026	4.8	0.024	-3.8	0.026	1.4
Coef. Of Variation (St. dev. / mean)	0.482	0.243	-49.5	0.264	-45.3	0.463	-3.9	0.471	-2.3	0.492	2.1
IP											
Mean	0.096	0.142	47.3	0.137	41.9	0.105	8.8	0.095	-1.7	0.095	-1.7
Standard deviation	0.039	0.021	-46.4	0.023	-42.7	0.038	-2.1	0.039	-2.1	0.040	0.5
Coef. Of Variation (St. dev. / mean)	0.408	0.149	-63.6	0.174	-57.5	0.382	-6.5	0.405	-0.7	0.413	1.2
Gini											
Mean	0.132	0.193	45.7	0.185	39.5	0.144	8.8	0.131	-1.1	0.130	-1.5
Standard deviation	0.054	0.026	-51.4	0.028	-48.5	0.053	-1.9	0.053	-1.7	0.054	0.7
Coef. Of Variation (St. dev. / mean)	0.407	0.136	-66.7	0.159	-60.9	0.383	-6.1	0.404	-0.9	0.412	1.2

Table 1. Summary of statistics for segregation indexes across states.

Table 1 reveals the impact conditioning has on the mean and dispersion of segregation across states. On average, segregation increases by 73% with the *M* index (around 46-47% with the other two measures). However, the geographical dispersion of conditional segregation, measured by the coefficient of variation, is much lower than that in the unconditional case: it is reduced by 50% (*M*) or more (64% and 67% for *IP* and *G*, respectively). This means that at least half of the relative variability observed among the unconditional segregation levels of states can be explained by the *compositional effect*, with the remaining being the *intrinsic segregation effect*.

As supply-side theories of segregation predict, characteristics workers bring to the labor market matter in explaining a substantial part of segregation and, consequently, segregation variability across states. The states' racial/ethnic structures, fueled by their immigration and linguistic profiles, turns out to be clearly the most important reason behind the strong *compositional effect*. This is a consequence of recent immigrants with poor English fluency who belong to a racial/ethnic minority (mainly Hispanics and Asians) being generally more segregated than others (Alonso-Villar et al., 2012, Gradín, 2012). Similarly, Hellerstein and Neumark (2008) found workplace segregation (i.e., at the establishment level) of Hispanics to be more closely associated with poor language proficiency than with lower education. Thus, the larger these subgroups are in the state's population, the higher the level of segregation.

The racial and ethnic differences explain an increase of 63% in the average segregation after conditioning and a reduction of about 45% in its geographical variation (coefficient of variation), while the immigration profile accounts for an additional 13% increase in the average level and 4% reduction in the coefficient of variation. This is the case of Louisiana and the most of the states in the East South Central area, such as Alabama, Kentucky, and Tennessee. After controlling for race/ethnicity, these states have more than double the initial level of segregation, while immigration/English has a more diverse effect. A similar pattern is found in Oklahoma and in most of the Midwestern states (Indiana, Kansas, Minnesota, Missouri, Ohio, Wisconsin and, to a lesser extent, Michigan). In both regions, the listed states differ from the rest of the country by having less immigration and smaller populations of Hispanics and Asians. Given that these groups are in general highly segregated, increasing their weight to make states share the same structure raises segregation.

The opposite occurs in those states with an overall negative effect on segregation after conditioning on characteristics, which is the case of California and Nevada. California is characterized by strong and recent immigration flows (35% of its workers are born outside the U.S. and 12% speak English not well or not at all; in New York, these figures are 29% and 6%, respectively; see Table A2 in the appendix). Moreover, this state has larger shares of Hispanic (33%) and Asian (13%) populations than the state of reference (15% and 7%, respectively). In the case of Nevada, where Hispanics also represent a large minority (22%), the share of immigrants is lower than in New York, which explains the positive impact of this factor.

Among the other factors accounted for in the *compositional effect*, neither the other supply-side determinants nor the industrial composition (which is the only demand-side explanation considered in the compositional effect) seems to be crucial to explaining the cross-state variation in segregation (their impacts on the mean and dispersion of segregation are small). The weak association between racial/ethnic segregation in the U.S. and education could seem counterintuitive at first because of strong evidence of workers' sorting on skills across occupations. However, recent research conducted for occupational and workplace segregation at the national level (Alonso et al., 2012a and Gradín, 2012; Hellerstein and Neumark, 2008) has shown that while the immigration and linguistic profile of Hispanics explained most of their segregation, the educational gap did not help explain much of segregation for these groups or for blacks. According to Hellerstein and Neumark (2008), the segregation of blacks could be more associated with non-skill-based explanations such as discrimination, residential segregation, or labor market networks, which in our case are captured by the *intrinsic segregation* term.

Beyond this general pattern, a few more facts are noteworthy. Education and industry play a significant role in explaining segregation in the District of Columbia, even though they are of opposite sign, thus canceling each other. The effect of education on segregation could be the result of a higher level of education in this district, which explains why segregation decreases when controlling for this factor. Washington, D.C. has the largest relative concentration of workers with a university degree in the country (59% of the labor force versus 37% in New York, which is also one of the largest shares), and individuals with either high or low education tend to be more unevenly distributed across occupations at the national level than people with intermediate grades (Alonso-Villar et al., 2012). The large size of public administration in the District of Columbia as compared to New York may also explain why segregation increases when its weight is reduced.[11]

With respect to the rest of the country, education is also of some relevance in states such as California and Tennessee. In California, controlling for education has the same effect as in the District of Columbia, but for a different reason: the population with only primary education is around 50% higher than in New York. In Tennessee, there is an opposite impact of education, because Tennessee has a higher population of intermediately educated workers than New York. Industry is also an important factor in Nevada and Hawaii, where high segregation appears to be partially connected to industrial structures. The former state places much weight on construction, nearly twice that of New York; the latter emphasizes active duty military

(6% of the work force). Both share important entertainment-related activities: 24% and 14% of employment, respectively (compared with just 8% in New York).

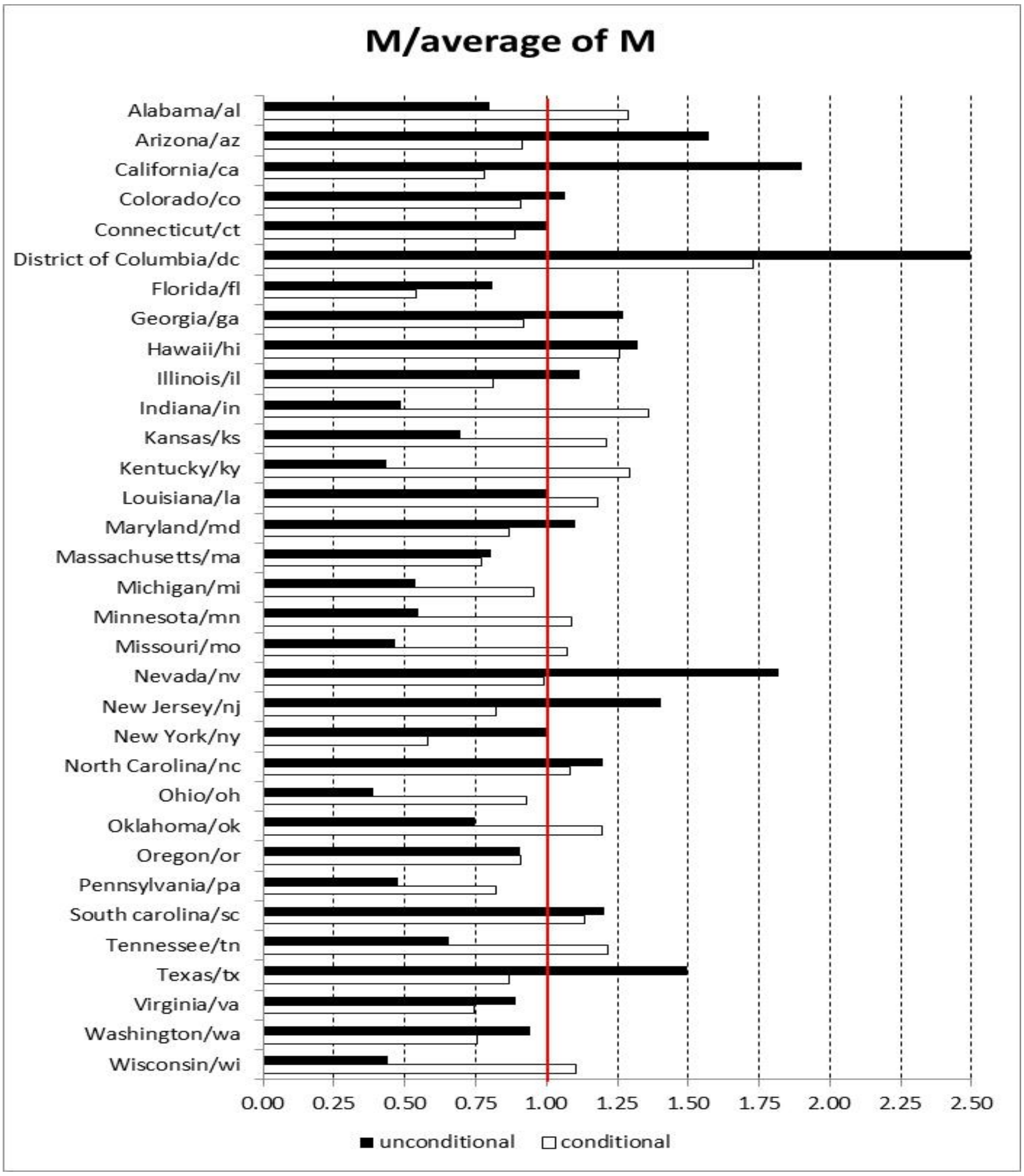
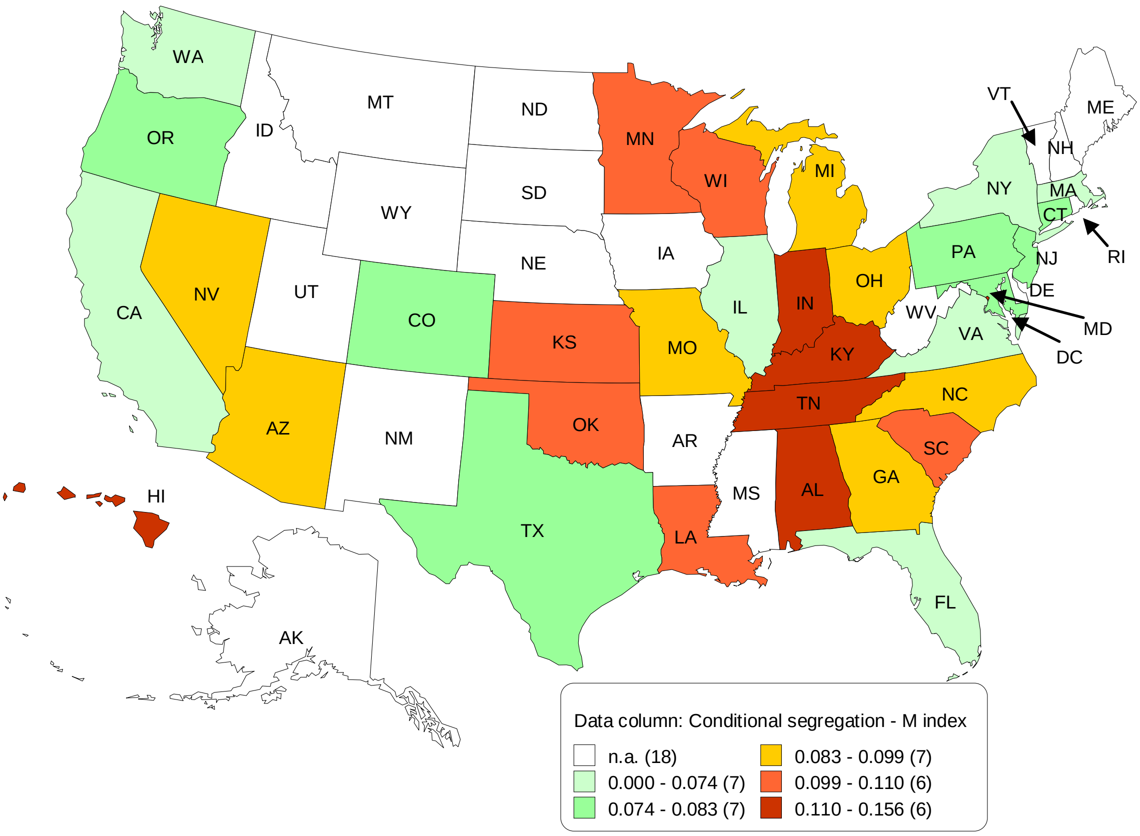


Figure 2. Conditional and unconditional occupational segregation in selected states (*M* index expressed relative to the average of states).

Conclusions about the impact of the *compositional effect* in explaining the variation in segregation across states can be derived by looking at changes in the segregation ranking after conditioning on workers' characteristics and industrial composition. Figure 2 reports the segregation of states, relative to the average segregation, both before and after conditioning. Because most states experience increments after conditioning, a state is expected to raise its level of segregation when it increases more than the average. It is unsurprising that this is the case of most states in the east central region that have strong race/ethnicity and/or immigration effects. Similarly, the relative level of a state decreases when segregation either decreases or increases less than the average. The most significant reductions in relative segregation occur in some states on the east coast (New York, New Jersey, Georgia, Florida, Maryland, and the District of Columbia) and in most southwestern states. Illinois and Texas also present a remarkable reduction.



Map 2. Conditional occupational segregation by race/ethnicity in selected states (*M* index).

Note: White states have not been assigned a value due to the small sample size for some demographic groups in the survey.

Map 2 shows the resulting geographical distribution of conditional segregation, thus reflecting cross-state variation only in *intrinsic segregation*, that is, the segregation persisting after the compositional effect has been removed. It identifies the area with the highest *intrinsic segregation* around the vertical line in the east central region running from Indiana down to Alabama, passing through Kentucky and Tennessee. This is in addition to the particular cases of Hawaii and the District of Columbia. States with intermediate-high levels of segregation can be found both in the north central area (Minnesota, Wisconsin and Kansas) and the south central area (Oklahoma and Louisiana). A similar segregation level is found in South Carolina and North Carolina. The group with the lowest conditional segregation is comprised of states on the east coast (Florida, Virginia, New York, and Massachusetts), the west coast (especially Washington and California), and Illinois.

To summarize, our analysis shows how misleading comparisons based on the unconditional levels of segregation can be. On the one hand, states with relatively low segregation display very different patterns. While in some states (Indiana, Kentucky, Tennessee, and Alabama) this was just the consequence of their low racial diversity and immigration/English profile (a large *compositional effect*, combined with high *intrinsic segregation*), in others (Florida, Pennsylvania, Massachusetts, and Washington), the *compositional effect* was rather small, and low segregation was mainly driven by their more integrative labor markets (low *intrinsic segregation*). Similarly, states with high unconditional segregation also displayed different patterns. While a considerable part of the high segregation found in California, Texas, New Jersey, and Illinois turned out to be the result of a higher presence of minorities, this is not the case of the District of Columbia and Hawaii (and to a lesser extent, South and North Carolina), which also show high levels of *intrinsic segregation*.[12]

We checked the robustness of these findings by using California as the state of reference, which has a different distribution of characteristics compared to the national average. The qualitative results remained unchanged (except in the case of Hawaii). The Spearman rank correlation coefficient and the Pearson correlation coefficient between segregation levels across states using the New York and California benchmarks are 0.88 and 0.92, respectively, when using the M index. Discrepancies between both benchmarks are mainly due to the race/ethnicity factor. In the case of Hawaii, when using California as the state of reference, the performance of this state improves substantially (with a change of 17 positions in the ranking with respect to New York). The remarkably low segregation of Hispanics in Hawaii (who make up 7% of workers) makes conditional segregation decrease notably when using California as the state of reference because in California, Hispanics represent 33% of the work force but only 15% in New York.

Conclusions

This paper has analyzed the extent of geographical disparities in occupational segregation by race and ethnicity across U.S. states. The unconditional analysis resulted in great spatial discrepancies, with segregation being highly concentrated in the District of Columbia, New Jersey, Hawaii, Texas, and several western states (such as California, Nevada, and Arizona).

Because these disparities may arise from an uneven distribution of workers' characteristics and industrial structures across states, this paper has also estimated conditional segregation by using a distribution of the relevant attributes of individuals (race/ethnicity, attained education, immigration profile, and English proficiency) and industrial structures that are similar across states. The study has revealed that the geographical dispersion of segregation is significantly reduced after conditioning for these factors, of which the racial/ethnic composition appears to be the most relevant. Moreover, the segregation map dramatically changes when conditional segregation is considered, with higher segregation moving toward the east. Thus, apart from the District of Columbia, which retained its high segregation level, and Hawaii, the east central region displayed the highest conditional segregation, mostly in Alabama, Kentucky, Tennessee, and Indiana. Our analysis suggests that the low levels of unconditional segregation in this region arise from its low racial diversity (*compositional effect*) rather than from a wider integration of minorities into its labor markets (*intrinsic segregation*). On the other hand, Washington, Pennsylvania, Massachusetts, and Florida have low segregation levels even when controlling for the mentioned attributes, indicating that this is not the result of a compositional effect.

This approach has substantial implications for understanding segregation, especially regarding its geographical variation in a largely heterogeneous and highly decentralized country like the U.S. A large *compositional effect* turned out to be responsible for about half or more of the segregation disparities across the states. Removing this effect not only reduces the variation in segregation, but also, and even more importantly, dramatically changes the segregation map, altering the areas that show the largest/lowest levels. This provides wide support for the relevance of supply-side factors as determinants of observed levels of segregation. However, a large part of segregation remains unexplained after removing the *compositional effect*. This is, by construction, the *intrinsic segregation effect*, which could arise from cross-state disparities in citizens' attitudes, government policies, or social capital, among other (mostly demand-side) factors highlighted in the literature. The role of each of these factors in determining variation in *intrinsic segregation* remains unclear and cannot be addressed in our setting. However, after having identified which states show the most or least segregative labor markets, this could be undertaken in future research.

References

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

Alonso-Villar, O. and Del Río, C. (2013): "Concentration of economic activity: An analytical framework," *Regional Studies* 47(5), 756-772.

Alonso-Villar, O., Del Río, C., and Gradín, C. (2012): "The Extent of Occupational Segregation in the US: Differences by Race, Ethnicity, and Gender," *Industrial Relations* 51(2), 179-212.

Alonso-Villar, O., Gradín, C. and Del Río, C. (2013): "Occupational segregation of Hispanics in U.S. metropolitan areas," *Applied Economics* 45(30), 4298-4307.

Amuedo-Dorantes, C. and Bansak, C. (2012): "The Labor Market Impact of Mandated Employment Verification Systems," *American Economic Review: Papers & Proceedings* 102(3), 543-548.

Beggs, J. (1995): "The Institutional Environment: Implications for Race and Gender Inequality in the U.S. Labor Market," *American Sociological Review* 60(4), 612-633.

Berry, W., Ringquist, E., Fording, R., and Hanson, R. (1998): "Measuring Citizen and Government Ideology in the American States, 1960-93," *American Journal of Political Science* 42(1), 327-348.

Catanzarite, L. (2000): "Brown-Collar Jobs: Occupational Segregation and Earnings of Recent-Immigrant Latinos," *Sociological Perspectives* 43(1), 45-75.

Chantreuil, F. and Trannoy, A. (2013): "Inequality Decomposition Values: The Trade-Off between Marginality and Consistency," *Journal of Economic Inequality* 11(1), 83-98.

DiNardo, J., Fortin, N., and Lemieux, T. (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica* 64(5), 1001-1044.

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

Ellis, M., Wright, R., and Parks, V. (2004): "Work Together, Live Apart? Geographies of Racial and Ethnic Segregation at Home and at Work," *Annals of the Association of American Geographers* 94(3), 620-637.

Fitzpatrick , J. and Perine, J. (2007): "State labor legislation enacted in 2007," *Monthly Labor Review*, January, 3-31.

Frankel, D. M. and Volij, O. (2011): "Measuring School Segregation," *Journal of Economic Theory* 146(1), 1-38.

Gradín, C. (2012): "Conditional Occupational Segregation of Minorities in the US," *Journal of Economic Inequality*, DOI 10.1007/s10888-012-9229-0.

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

Hellerstein, J., McInerney, M., and Neumark, D. (2010): "Spatial Mismatch, Immigrant Networks, and Hispanic Employment in the United States," *Annales d'Economie et de Statistique* 99-100, 141-167.

Hero, R. and Preuhs, R. (2007): "Immigration and the Evolving American Welfare State: Examining Policies in the U.S. States," *American Journal of Political Science* 51(3), 498-517.

Iceland, J. (2004): "Beyond Black and White. Metropolitan Residential Segregation in Multi-Ethnic America," *Social Science Research* 33, 248-271.

Kail, B. and Dixon, M. (2001): "The Uneven Patterning of Welfare Benefits at the Twilight of AFDC: Assessing the Influence of Institutions, Race, and Citizen Preferences," *The Sociological Quarterly* 52, 376-399.

Kaufman, R. L. (2010): *Race, Gender, and the Labor Market: Inequalities at Work*. Lynne Rienner Publishers, Boulder (CO), USA.

Karmel, T. and MacLachlan, M. (1988): "Occupational Sex Segregation—Increasing or Decreasing?," *The Economic Record* 64, 187-195.

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

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

Ovadia, S. (2003): "The Dimensions of Racial Inequality: Occupational and Residential Segregation Across Metropolitan Areas in the United States," *City & Community* 2(4), 313-333.

Patel, K. and Vella, F. (2007): "Immigrant Networks and Their Implications for Occupational Choice and Wages," IZA Discussion Paper No. 3217.

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

Raphael, S. and Ronconi, L. (2009): "The Labor Market Impact of State-Level Immigration Legislation Targeted at Unauthorized Immigrants." UC Berkeley Labor Center Working Paper.

Reardon, S. and Firebaugh, G. (2002): "Measures of Multigroup Segregation," *Sociological Methodology* 32, 33-76.

Semyonov, M., Haberfeld, Y., Cohen, Y., and Lewin-Epstein, N. (2000): "Racial Composition and Occupational Segregation and Inequality across American Cities," *Social Science Research* 29, 175-187.

Shorrocks, A. (2013): "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value," *Journal of Economic Inequality* 11(1), 99-126.

Silber, J. (1992): “Occupational Segregation Indices in the Multidimensional Case: A Note,” *The Economic Record* 68, 276-277.

Theil, H. and Finizza, A.J. (1971): “A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts,” *Journal of Mathematical Sociology* 1, 187-194.

Tomaskovic-Devey, D., Zimmer, C., Stainback, K., Robinson, C., Taylor, T., and McTague, T. (2006): “Documenting Desegregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966-2003,” *American Sociological Review* 71, 565-588.

Wang, Q. (2010): “How Does Geography Matter in the Ethnic Labor Market Segmentation Process? A Case Study of Chinese Immigrants in the San Francisco CMSA,” *Annals of the Association of American Geographers* 100(1), 182-201.

Wyly, E. (1999): “Local Labor Markets and Occupational Sex Segregation in an American Metropolis,” *Journal of Urban Affairs* 21(1), 1-33.

Appendix

States	Whites	African Americans	Asians	Native Americans	Hispanics	Other Races	Less than High School	High School	Some College	Bachelor
	Race/ethnicity						Education			
Alabama/al	71.9	23.2	1.1	0.5	2.5	0.9	12.9	31.1	31.7	24.4
Alaska/ak	73.2	3.5	5.0	9.0	5.3	4.1	8.9	28.7	36.5	25.9
Arizona/az	63.4	3.0	2.5	3.0	27.0	1.1	14.5	25.3	34.3	25.9
Arkansas/ar	79.5	12.7	1.1	0.8	4.8	1.1	13.3	34.9	30.8	21.0
California/ca	46.8	5.4	12.6	0.7	32.7	1.8	15.9	22.8	30.7	30.7
Colorado/co	75.5	3.2	2.5	0.7	16.7	1.3	10.4	23.4	31.3	34.9
Connecticut/ct	76.7	8.2	3.5	0.2	10.0	1.4	8.7	27.9	27.6	35.9
Delaware/de	74.4	16.9	2.9	0.3	4.6	0.9	10.8	30.5	28.6	30.2
District of Columbia/dc	49.2	33.6	6.9	0.2	8.5	1.6	6.1	15.1	20.2	58.7
Florida/fl	61.5	13.8	2.3	0.3	20.9	1.2	12.2	29.8	31.5	26.5
Georgia/ga	62.0	26.6	2.8	0.2	7.3	1.0	12.8	29.5	28.8	28.9
Hawaii/hi	27.5	2.6	39.6	7.6	7.2	15.4	6.9	29.5	33.7	29.9
Idaho/id	87.8	0.4	1.0	1.2	8.4	1.2	11.6	28.3	35.9	24.2
Illinois/il	69.7	11.2	4.4	0.2	13.6	0.9	10.6	25.9	31.3	32.3
Indiana/in	86.5	7.0	1.3	0.2	4.2	0.9	11.2	34.3	30.8	23.7
Iowa/ia	92.7	1.9	1.5	0.2	3.2	0.5	9.2	31.0	34.4	25.4
Kansas/ks	83.5	5.2	2.0	0.7	7.3	1.3	9.7	27.6	34.2	28.5
Kentucky/ky	88.9	7.0	1.1	0.2	2.1	0.7	11.4	34.1	31.0	23.4
Louisiana/la	66.8	26.9	1.5	0.5	3.5	0.9	13.6	34.2	29.5	22.7
Maine/me	95.7	0.8	1.0	0.5	1.0	1.0	7.9	33.9	31.0	27.3
Maryland/md	64.3	23.5	4.9	0.3	5.8	1.2	9.9	26.9	27.9	35.2
Massachusetts/ma	82.1	4.9	4.6	0.2	6.6	1.7	8.5	25.4	25.9	40.2
Michigan/mi	81.9	10.6	2.5	0.4	3.5	1.1	8.5	28.3	35.3	28.0
Minnesota/mn	88.8	3.2	3.0	0.7	3.3	1.0	7.8	25.3	35.4	31.5
Mississippi/ms	63.5	32.3	1.0	0.4	2.3	0.5	14.3	31.5	32.6	21.6
Missouri/mo	85.2	9.1	1.5	0.4	2.8	1.0	10.3	30.4	31.7	27.6
Montana/mt	91.0	0.5	0.9	3.7	2.4	1.6	8.2	31.2	33.5	27.1
Nebraska/ne	88.0	3.2	1.5	0.5	5.9	0.9	9.2	27.7	36.0	27.1
Nevada/nv	62.3	6.2	6.4	1.3	22.1	1.8	15.2	30.8	32.7	21.3
New Hampshire/nh	93.8	0.9	1.9	0.3	2.3	0.7	9.2	29.8	30.4	30.6
New Jersey/nj	64.6	11.8	6.8	0.2	15.5	1.1	10.2	28.8	26.1	34.9
New Mexico/nm	46.9	1.9	1.5	6.9	41.7	1.1	13.7	27.6	32.8	26.0
New York/ny	63.6	12.9	7.2	0.3	14.7	1.4	10.4	26.1	27.0	36.6
North Carolina/nc	71.2	18.6	1.8	1.0	6.5	0.9	12.6	28.8	31.2	27.3
North Dakota/nd	92.3	0.7	0.7	3.9	1.8	0.7	7.7	26.8	38.6	26.9
Ohio/oh	85.9	9.3	1.6	0.2	2.1	0.9	9.1	33.7	30.9	26.3
Oklahoma/ok	75.2	6.4	1.9	5.8	6.6	4.1	12.1	31.1	32.6	24.2
Oregon/or	82.8	1.4	3.7	1.1	9.2	1.9	10.9	25.5	35.2	28.4
Pennsylvania/pa	85.4	7.9	2.5	0.1	3.4	0.7	8.9	34.9	26.9	29.4
Rhode Island/ri	83.8	3.7	2.6	0.3	8.1	1.6	11.2	26.9	29.9	32.0
South Carolina/sc	68.4	25.3	1.3	0.3	3.9	0.8	12.4	32.1	30.4	25.1
South Dakota/sd	91.3	0.8	0.7	4.4	2.0	0.9	9.8	31.6	33.2	25.4
Tennessee/tn	79.8	14.7	1.4	0.3	3.1	0.8	11.9	33.6	29.6	24.9
Texas/tx	52.7	10.3	3.5	0.4	32.2	1.0	16.9	26.5	30.4	26.3
Utah/ut	84.5	1.0	2.0	1.6	10.1	0.8	10.8	25.7	38.0	25.5
Vermont/vt	96.3	0.6	1.0	0.3	1.1	0.8	7.4	30.8	28.9	32.8
Virginia/va	68.7	19.2	4.5	0.3	6.2	1.2	10.7	27.6	28.8	32.9
Washington/wa	78.5	3.0	6.8	1.4	8.1	2.2	9.5	24.1	35.1	31.2
West Virginia/wv	94.7	2.9	0.8	0.1	0.9	0.6	9.8	39.3	29.1	21.8
Wisconsin/wi	89.1	3.9	1.6	0.7	3.9	0.7	9.1	31.2	33.0	26.6
Wyoming/wy	88.8	0.7	0.7	1.5	6.9	1.4	9.1	30.7	37.2	22.9

Table A1. Demographic structure and educational level (all rows in percentage).

		Years of residence				English proficiency				
States	Born in the US	0-5 years	6-10 years	11-15 years	16+ years	only English	very well	well	not well	not at all
AL	95.26	0.94	0.98	0.45	2.37	95.17	2.53	0.81	0.98	0.5
AK	89.47	1.33	1.34	1.6	6.27	86.01	8.22	3.65	1.93	0.18
AZ	79.82	3.61	3.96	2.93	9.68	72.51	14.5	4.59	4.98	3.42
AR	93.83	1.1	1.24	0.8	3.03	93.32	3.01	1.36	1.63	0.68
CA	64.52	3.52	5.14	4.59	22.23	58.32	21.25	8.9	7.69	3.83
CO	86.75	1.95	2.9	2	6.4	83.99	8.27	3.26	3.16	1.33
CT	81.5	2.41	3.55	2.77	9.77	80.51	11.38	4.51	2.82	0.78
DE	89.84	1.42	2.45	1.28	5.02	88.79	6.78	2.02	1.61	0.79
DC	77.44	3.53	4	2.73	12.3	78.55	13.92	4.22	2.57	0.74
FL	73.41	3.72	5.13	3.81	13.93	72.71	14.58	5.73	4.64	2.34
GA	86.61	2.54	3.21	1.98	5.66	87.11	6.1	2.76	2.77	1.27
HI	77.34	2.48	2.81	3	14.38	75.85	13.83	6.53	3.42	0.36
ID	92.22	0.99	1.26	1.12	4.4	89.72	5.76	1.94	1.91	0.67
IL	81.64	2.13	3.58	2.95	9.7	77.76	11.59	5.15	4.06	1.43
IN	94.43	0.94	1.32	0.8	2.51	92.7	4.17	1.4	1.33	0.39
IA	95.06	0.82	1.17	0.71	2.24	93.84	3.2	1.41	1.19	0.36
KS	92.31	1.25	1.65	0.93	3.86	90.71	4.94	1.81	1.75	0.79
KY	95.56	0.91	1.04	0.6	1.88	95.22	2.64	1.08	0.77	0.29
LA	94.94	0.88	0.81	0.5	2.86	91.44	5.69	1.51	1.01	0.35
ME	95.55	0.47	0.61	0.4	2.98	93.3	5.08	0.93	0.57	0.1
MD	84.39	2.59	3.08	2.07	7.87	85.24	8.6	3.27	2.18	0.72
MA	81.22	2.71	3.68	2.6	9.78	80.48	11.05	4.57	2.87	1.03
MI	92.28	1.07	1.66	1.17	3.83	91.16	5.51	1.93	1.09	0.3
MN	92.04	1.37	1.81	1.27	3.5	91.25	5.06	1.94	1.38	0.37
MS	96.48	0.85	0.62	0.33	1.72	95.73	2.39	0.65	0.81	0.41
MO	95.07	0.81	0.98	0.76	2.38	94.13	3.63	1.14	0.83	0.27
MT	97.23	0.3	0.28	0.3	1.89	96.04	3.16	0.49	0.25	0.06
NE	92.91	1.22	1.36	1.16	3.35	91.81	3.72	1.75	1.78	0.94
NV	74.89	3.66	4.29	3.72	13.45	73.01	12.79	6.4	5.51	2.28
NH	92.78	0.96	1.28	0.88	4.11	91.72	5.69	1.35	1.17	0.07
NJ	73.18	3.4	4.83	4.12	14.46	71.83	15.59	6.21	4.8	1.57
NM	87.28	1.91	2.02	1.53	7.25	65.07	25.14	4.46	3.45	1.88
NY	71.3	2.98	4.61	4.55	16.57	71.95	15.54	6.37	4.64	1.49
NC	90.28	1.89	2.63	1.46	3.73	89.95	4.76	1.89	2.34	1.07
ND	96.7	0.44	0.61	0.3	1.95	95.16	3.62	0.84	0.3	0.08
OH	95.11	0.74	1.02	0.59	2.54	93.93	4.02	1.19	0.71	0.14
OK	92.35	1.3	1.66	1.07	3.63	91.2	4.26	2.03	1.87	0.64
OR	87.19	1.73	2.45	1.99	6.64	85.98	6.9	2.91	2.75	1.45
PA	92.61	1.09	1.32	1.01	3.97	91.28	5.36	1.84	1.17	0.36
RI	84.4	1.76	2.58	1.93	9.33	81.8	10.3	3.71	2.95	1.24
SC	93.14	1.51	1.54	0.81	3.02	93.13	3.39	1.32	1.51	0.64
SD	96.9	0.89	0.45	0.42	1.34	94.62	3.25	1.12	0.78	0.24
TN	94.02	1.21	1.29	0.86	2.62	93.88	3.19	1.26	1.23	0.44
TX	78.72	2.77	4.08	3.22	11.21	66.74	18.67	5.76	5.53	3.29
UT	88.78	1.75	2.59	1.71	5.17	85.35	8.31	2.78	2.61	0.95
VT	95.66	0.57	0.58	0.53	2.66	95.29	3.44	0.76	0.43	0.08
VA	86.07	1.98	2.89	1.81	7.25	86.59	7.5	3	2.2	0.71
WA	83.99	2.14	3.11	2.35	8.42	83.73	8.47	3.92	2.73	1.14
WV	97.69	0.38	0.28	0.28	1.37	97.37	1.71	0.45	0.4	0.07
WI	94.79	0.73	1.16	0.7	2.63	92.8	4.33	1.4	1.18	0.29
WY	95.89	0.71	0.66	0.23	2.51	93.75	4.3	0.94	0.78	0.23

Table A2. Immigration status: years of residence in the US and English proficiency (all rows in percentage.)

States	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
AL	1.8	8.1	15.0	3.6	12.3	5.4	1.9	5.9	8.0	20.1	7.2	5.1	5.1	0.6
AK	4.9	8.0	4.0	1.9	11.0	8.1	2.1	4.4	8.2	20.3	7.9	5.3	10.0	3.9
AZ	1.3	10.7	7.9	3.1	12.0	4.8	1.8	8.4	11.0	18.4	9.8	4.7	5.3	0.7
AR	3.4	7.6	15.9	3.2	13.2	5.8	1.9	5.0	6.3	21.4	7.1	4.7	4.2	0.5
CA	1.9	7.9	10.6	3.7	11.1	4.6	3.0	7.4	11.8	18.7	8.9	5.2	4.4	0.8
CO	2.1	9.7	7.3	3.2	11.4	4.6	3.5	8.0	12.3	17.7	10.0	5.0	4.4	1.0
CT	0.4	6.7	12.5	3.0	11.4	3.7	2.6	9.6	10.3	23.1	7.9	4.5	3.7	0.5
DE	1.0	8.2	10.2	2.9	12.2	4.1	1.7	11.8	9.9	20.5	8.0	4.3	4.7	0.5
DC	0.2	3.7	1.4	0.6	2.7	3.3	4.5	5.9	20.3	15.4	7.0	7.4	25.8	1.9
FL	1.1	10.3	5.9	3.5	12.8	5.0	2.4	8.5	11.4	18.4	10.3	5.1	4.7	0.7
GA	1.2	8.7	11.5	3.6	11.8	6.1	2.8	7.0	10.1	18.4	7.8	4.8	5.1	1.2
HI	1.4	7.7	3.3	2.5	11.3	5.4	1.9	6.3	9.4	18.4	13.9	4.4	8.1	6.0
ID	5.4	10.1	10.4	2.9	12.2	4.4	2.2	5.9	8.9	19.6	8.1	4.2	5.1	0.7
IL	1.0	6.6	13.5	3.9	10.8	6.0	2.4	8.0	10.3	20.6	8.2	4.7	3.8	0.3
IN	1.3	6.8	20.7	3.2	11.3	5.2	1.9	5.5	7.0	20.8	8.2	4.7	3.4	0.1
IA	4.0	6.7	15.6	3.5	11.9	4.9	2.2	7.3	6.2	23.2	7.3	4.3	3.0	0.1
KS	3.7	6.3	14.0	3.4	11.3	5.2	3.1	6.1	7.7	22.3	7.3	4.3	4.6	0.8
KY	3.1	7.1	14.8	3.3	11.6	6.0	1.7	5.7	7.0	21.6	7.6	4.6	4.7	1.1
LA	4.1	9.2	8.4	3.2	12.0	5.3	1.7	5.7	8.3	21.5	9.1	5.2	5.5	0.9
ME	2.5	8.2	10.3	2.9	13.8	3.7	2.0	6.1	7.2	25.6	8.4	4.5	4.4	0.6
MD	0.6	8.1	6.1	2.7	11.1	4.5	2.5	7.2	13.3	21.9	7.4	5.1	8.7	0.9
MA	0.4	6.7	10.6	3.2	10.8	3.8	2.9	8.1	12.3	25.1	7.9	4.4	3.9	0.2
MI	1.1	5.9	19.1	3.1	11.4	4.1	1.9	5.9	8.6	22.0	8.8	4.6	3.5	0.1
MN	2.3	6.8	14.5	3.5	11.3	4.6	2.3	7.6	9.1	22.5	7.7	4.5	3.2	0.1
MS	2.7	7.6	15.4	3.1	11.7	4.8	1.7	5.1	6.0	22.2	8.8	4.7	5.3	1.1
MO	1.8	7.2	12.1	3.2	11.8	5.5	2.4	7.4	8.5	21.5	8.6	4.9	4.5	0.6
MT	7.4	9.4	5.0	2.8	12.7	4.9	2.0	5.9	7.2	21.5	10.1	4.9	5.7	0.6
NE	5.0	6.7	11.0	3.5	11.4	6.6	2.0	7.7	7.6	21.6	7.8	4.4	4.0	0.7
NV	1.5	11.3	4.6	2.8	10.5	4.9	1.7	7.1	10.0	13.2	24.0	3.9	4.2	0.6
NH	0.8	7.8	12.7	3.5	14.9	3.6	2.3	6.7	8.8	22.5	8.3	4.5	3.5	0.1
NJ	0.4	6.4	10.4	4.1	11.5	6.1	3.0	8.3	11.2	21.7	7.6	4.5	4.7	0.2
NM	3.9	9.0	5.4	2.5	11.7	4.4	1.9	5.3	10.7	22.4	10.3	4.4	7.3	0.9
NY	0.6	6.1	7.6	3.1	10.5	5.4	3.5	9.2	10.9	25.2	8.2	4.8	4.7	0.3
NC	1.5	8.8	14.0	3.2	11.4	4.4	2.1	6.6	8.7	21.2	8.0	4.4	4.1	1.7
ND	8.0	6.8	8.4	3.4	12.0	5.1	2.1	5.6	6.1	24.3	8.4	3.9	4.6	1.4
OH	1.1	6.0	16.9	3.4	11.5	5.0	2.0	6.8	8.5	22.0	8.4	4.4	3.8	0.2
OK	4.4	7.1	10.2	3.5	11.4	5.2	2.4	6.1	7.7	21.2	8.7	5.1	5.8	1.1
OR	3.4	7.6	12.6	3.6	12.5	4.5	2.1	6.5	9.8	19.5	8.8	4.5	4.4	0.1
PA	1.3	6.4	13.2	3.3	11.8	5.2	2.1	6.6	9.4	24.1	7.8	4.7	4.0	0.1
RI	0.4	7.0	11.7	2.6	11.2	3.6	2.1	8.0	8.7	25.3	9.5	4.8	4.5	0.6
SC	1.0	8.8	14.9	3.1	11.8	4.8	1.8	5.9	8.3	19.5	9.0	4.8	4.8	1.7
SD	7.1	6.4	10.2	3.1	11.3	4.2	2.0	8.8	6.4	21.9	8.6	4.8	4.9	0.6
TN	1.1	7.7	15.4	3.6	12.0	6.7	2.1	6.2	8.3	19.8	8.1	5.0	3.9	0.3
TX	2.8	9.1	10.0	3.6	11.6	5.6	2.3	6.9	10.1	19.8	8.1	5.3	4.2	0.8
UT	1.8	8.8	11.0	3.2	12.5	5.0	2.6	7.0	10.2	19.8	8.2	4.3	5.4	0.3
VT	2.3	8.2	12.3	2.8	11.6	3.2	2.3	5.0	7.2	26.0	9.2	4.6	5.1	0.2
VA	1.2	8.2	8.7	2.4	11.1	4.3	2.7	6.7	13.1	18.9	7.5	4.9	7.5	2.8
WA	2.5	7.7	11.0	3.4	11.3	4.8	2.9	6.3	10.7	19.7	8.5	4.5	5.1	1.5
WV	4.7	7.6	9.2	2.6	12.5	5.6	1.7	4.8	7.4	24.4	8.9	4.5	6.1	0.1
WI	2.6	6.4	19.1	3.3	11.7	4.5	2.0	6.3	7.3	21.4	8.2	4.0	3.2	0.1
WY	11.7	9.0	4.6	2.5	11.5	6.0	1.5	4.3	6.4	21.6	9.6	4.4	6.0	1.0

Table A3. Industrial structure (all rows in percentage).

Note: NAICS industry codes are: 1) Agriculture, Forestry, Fishing, and Mining; 2) Construction; 3) Manufacturing; 4) Wholesale Trade; 5) Retail Trade; 6) Transportation; 7) Information; 8) Finance, Insurance, and Real Estate; 9) Professional, Scientific, and Management; 10) Education and Health, 11) Arts, Entertainment, Accommodation, and Food Services; 12) Other Services; 13) Public Administration; and 14) Military.

States	M	IP	Gini	M conditional*	IP conditional*	Gini conditional*
Alabama	0.042	0.091	0.126	0.117	0.169	0.228
Arizona	0.082	0.146	0.200	0.083	0.129	0.179
California	0.099	0.166	0.236	0.071	0.127	0.176
Colorado	0.056	0.106	0.143	0.083	0.137	0.187
Connecticut	0.053	0.093	0.132	0.081	0.134	0.188
District of Columbia	0.130	0.194	0.260	0.156	0.211	0.282
Florida	0.042	0.098	0.139	0.049	0.101	0.144
Georgia	0.066	0.122	0.164	0.083	0.139	0.183
Hawaii	0.069	0.133	0.184	0.114	0.151	0.203
Illinois	0.058	0.107	0.148	0.074	0.122	0.169
Indiana	0.025	0.048	0.064	0.123	0.167	0.218
Kansas	0.036	0.070	0.094	0.110	0.162	0.215
Kentucky	0.023	0.040	0.055	0.117	0.159	0.217
Louisiana	0.053	0.107	0.151	0.107	0.161	0.218
Maryland	0.057	0.105	0.148	0.079	0.127	0.176
Massachusetts	0.042	0.077	0.104	0.070	0.126	0.175
Michigan	0.028	0.058	0.079	0.087	0.134	0.179
Minnesota	0.029	0.051	0.069	0.099	0.156	0.212
Missouri	0.024	0.050	0.070	0.097	0.147	0.198
Nevada	0.095	0.157	0.210	0.090	0.136	0.188
New Jersey	0.073	0.127	0.176	0.075	0.128	0.177
New York	0.053	0.110	0.156	0.053	0.110	0.156
North Carolina	0.063	0.113	0.149	0.098	0.151	0.199
Ohio	0.020	0.044	0.062	0.084	0.135	0.183
Oklahoma	0.039	0.080	0.109	0.109	0.163	0.220
Oregon	0.047	0.076	0.102	0.083	0.128	0.177
Pennsylvania	0.025	0.050	0.071	0.074	0.131	0.176
South Carolina	0.063	0.123	0.166	0.103	0.155	0.208
Tennessee	0.034	0.068	0.093	0.110	0.161	0.218
Texas	0.078	0.150	0.208	0.078	0.130	0.179
Virginia	0.046	0.095	0.132	0.068	0.123	0.168
Washington	0.049	0.080	0.110	0.068	0.124	0.169
Wisconsin	0.023	0.044	0.059	0.100	0.150	0.200

Table A4. Segregation indexes.

* Conditional analysis reweighting observations in each state using New York's distribution by race/ethnicity, education, immigration profile, English proficiency, and industry.

[1] See the U.S. Census Bureau population estimates by race and ethnicity (http://www.census.gov/).
[2] Source: 2005-07 American Community Survey, Census Bureau.
[3] Source: National Conference of State Legislatures (http://www.ncsl.org). January, 2013.
[4] Source: U.S. Department of Labor (http://workforcesecurity.doleta.gov/unemploy/comparison2013.asp), January, 2013.
[5] Source: National Conference of State Legislatures (http://www.ncsl.org).
[6] A higher level of detailed (3-digit SOC with 469 categories) was not used because it would be problematic in most states due to the relatively small number of observations for various demographic groups.
[7] The 18 states dropped are: Alaska, Arkansas, Delaware, Idaho, Iowa, Maine, Mississippi, Montana, Nebraska, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, Utah, Vermont, West Virginia, and Wyoming.
[8] For a more detailed discussion of the differences/similarities among these indexes, see Alonso-Villar and Del Río (2013), in which they are used to quantify the spatial concentration of employment.
[9] This is in line with the conventional wage gap decomposition in the explained and unexplained effects (characteristics and coefficients, respectively).
[10] The values of the conditional segregation can be found in Table A4 in the appendix.

[11] Indeed, according to our calculations, occupational segregation by race and ethnicity in the public administration at the national level is half the level in the remaining sectors (0.018 compared to 0.043, *M* index).
[12] Main results do not change using the *IP* and *G* indexes.