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 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
1. Introduction

The United States has an outstanding and ever-increasing racially and ethnically diverse population. The proportion of non-Hispanic whites decreased from 76% in 1990 to 65% in 2009. In this multiracial society, residential and school segregation by race has been extensively documented. However, evidence about occupational segregation by race and ethnicity is scarcer and the analyses have been mainly undertaken at the national level although 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. This suggests that it is important to analyze occupational segregation at different subnational levels (Tomaskovic-Devey et al., 2006; Kaufman, 2010). Several studies have analyzed 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., 2012b). 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). However, 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. However, as far as we know, no study has analyzed occupational segregation in the U.S. at the state level.

Understanding segregation at the state level is very important given the role played by states in the labor outcomes of minorities. On the one hand, many labor regulations are undertaken by states, either by enacting specific laws or through common law, in crucial matters such as minimum wages, unemployment benefits, hours worked, equal employment opportunities, etc. (Fitzpatrick and Perine, 2007). These results in large spatial disparities across states in the employment opportunities they offer to minorities (Beggs, 1995). On the other hand, after 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.

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. In doing so, this paper analyzes whether occupational disparities among workers of different races/ethnicities share a common pattern across the country or instead vary state by state. 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 (Hellerstein and Neumark, 2008; Alonso-Villar et al., 2012a), 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. 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 makes a conceptual contribution to the segregation literature beyond its empirical findings. It shows that when it comes to comparing occupational segregation between two different areas (regions, states, countries, etc.), the differential between them has two sources of a very different nature. The first refers to 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.

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.


Thus far, 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.

There are many contexts, however, in which more than two groups are involved. This happens in this case because the focus is on segregation by race and ethnicity in the U.S. by considering 5 different groups: non-Hispanic whites, blacks, Asians and Native Americans, and Hispanics of any race. From now on, these populations will be simply referred to as groups. In this multigroup context, segregation analyses based on one-to-one comparisons become cumbersome and a more synthetic approach is desirable. Fortunately, in recent years, new approaches have been proposed that allow the analysis of multigroup segregation by simultaneously quantifying the disparities among all groups. As in the two-group case, there is no segregation 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). There are two distinctive channels through which segregation may increase. Ceteris paribus, segregation
increases as the distribution of a group across occupations departs from that of the whole population (Alonso-Villar and Del Rio, 2010). 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.

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. Other 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.

These segregation theories identify several explicative factors related with either workers’ characteristics or social and labor environments. These factors are expected to vary greatly across the country, which supports the need for subnational analyses. As already noted, states bear an active role in labor market regulation, welfare, and immigration control. The great variability across states in citizen and government ideology is also well established among 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 coolest attitudes of the country toward African Americans, Hispanics, and Asians, while the direct opposite is found in the Pacific region. All of this is likely to lead to different integration levels of minorities in the states.

Additionally, workers’ characteristics vary across the country following geographic patterns, so analysis at the state level seems a reasonable way of capturing these patterns. Racial and ethnic groups are not evenly distributed across states: Hispanics are more concentrated in California, Texas, and Florida; African Americans in southeastern states; 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. 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). Regarding immigration, states also show different 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).

To analyze disparities in occupational segregation between states, this paper distinguishes two components. The compositional effect component is the segregation differential that comes from states' differences in workers’ races/ethnicities and human capital, together with states' differences in sectoral structures. This component measures the segregation gap that results from disparities in both the endowments that workers bring to the labor market and the industrial structure they face since it affects the set of occupations available to workers. To take into account disparities in human capital, one needs to consider not only 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., 2012a). The intrinsic segregation effect component is the result of one state labor market being more segregative than another. This effect is directly associated with discrepancies in the opportunities that the labor environment of states bring to minorities, and
it may arise from cross-state disparities in issues such as citizens’ attitudes, unionization, government policies, or social capital.

Disentangling these two effects will help to better understand the segregation phenomenon since it allows to identify the states in which minorities are better/worst integrated regardless of the industrial structures and workers’ characteristics of 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 is more segregative 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 segregation each of them face. This is why it is important to separate the two effects.

To disentangle the intrinsic segregation and compositional effects among states, this paper takes one state as the reference and constructs for each of the other states a counterfactual occupational distribution with the same workers’ characteristics (and industrial structures) as the benchmark. This is done adapting to our context a methodology initially proposed by Di Nardo et al. (1996) to analyze wage disparities and later adapted by Gradín (2012) for analysis of conditional segregation of each non-white group with respect to non-Hispanic whites at the national level. According to this propensity score procedure, the original observations of workers in the target state are reweighted by their probability, predicted by a logit model, of belonging to the state of reference based on their own characteristics. This procedure allows one to compare segregation across states by considering a common distribution of relevant characteristics, those of the state of reference, shared by all of them.
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). Occupations are classified consistently with the Current Population Survey, based on a detailed occupation recode of the Standard Occupational Classification System (SOC). The list includes 52 occupations. 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 are mainly those with smaller and less demographically diverse populations, mostly in the central and northwest areas of the country. 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. 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. It can be written as
\[ M = \sum_{g} \frac{C^g}{T} \log \left( \frac{T}{C^g} \right) - \sum_{j} t_j \left[ \sum_{g} \frac{c_j^g}{t_j} \log \left( \frac{t_j}{c_j^g} \right) \right], \]

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),

\[ IP = \frac{1}{2} \sum_{g} \sum_{j} \left| \frac{c_j^g}{T} t_j - \frac{C^g}{T} T \right|, \]

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):

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

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, the mutual information index can be written as

\[ M = \sum_{g} \frac{C^g}{T} M^g, \]

where \( M^g = \sum_{j} \frac{c_j^g}{C^g} \ln \left( \frac{c_j^g}{C^g} \right) \)

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.\(^{vi}\) 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 one to compute 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 Gradin (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, workers in each state are first partitioned into several mutually exclusive subgroups or “cells,” with each being 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 \( z = (z_1, \ldots, z_k) \) denote a vector of \( k \) covariates describing these attributes. Next, it is obtained the density function across occupations that 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, \( \Psi_z \), can be easily estimated from the data:

\[
\Psi_z = \frac{\Pr(D = \text{New York} | z)}{\Pr(D = \text{New York})} = \frac{\Pr(D = s) \Pr(D = \text{New York} | z)}{\Pr(D = s) \Pr(D = \text{New York})},
\]

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. Thus, the following logit model is estimated,

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

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:\[ S_0 - S_i = (S_i^* - S_i) + (S_0 - S_i^*) = CE + ISE. \]

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 since $S_i^* - S_j^* = (S_0 - S_i^*) - (S_0 - S_j^*)$.

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). These contributions are obtained by using the Shapley decomposition (Shorrocks, 2012; Chantreuil and Trannoy, 2012). 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, first occupational segregation by race and ethnicity in each state is computed to explore variation across the U.S. Second, the two sources of variation in segregation across states, the *intrinsic segregation* and *compositional* effects, are disentangled.

#### 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 the appendix). The map shows a great geographical variation in segregation, the coefficient of variation of segregation being equal to 0.482. 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 southwestern 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 northeastern states such as Ohio, Wisconsin, Missouri, Kentucky, Indiana, Michigan, and Pennsylvania (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. On the contrary, 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 was found to be 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.
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 proportion of African-Americans and low proportion of other minorities. On the other hand, Florida has much lower segregation than California, Arizona, Texas, and Nevada, notwithstanding the similar large presence of Hispanics. In the next subsection, it will be discussed 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. 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 in the latter was estimated using a logit regression. By reweighting the original distribution using the predicted probabilities, it is obtained 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. Unconditional and conditional segregation are first compared to measure the magnitude of the compositional effect in each state and to identify which observable factors are contributing to this variability in segregation across states. Second, conditional segregation across states are compared in order to identify those with the most/least segregative labor markets once the compositional effect was removed.

The change in segregation experienced by each state after conditioning on characteristics is reported in Figure 1. This figure 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, meaning that 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, that is, the compositional effect of the differential in segregation between each state and the reference.
Figure 1. The *compositional effect*: Conditional-unconditional segregation gap in selected states (M index). Factors’ contributions using the Shapley decomposition.

Most states experience a net increase in segregation after conditioning on characteristics, the largest being in 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 the labor market. 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.

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Segregation</th>
<th>Race/ethnicity</th>
<th>Immigration</th>
<th>Education</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>M Mean</td>
<td>0.052</td>
<td>0.091</td>
<td>73.4</td>
<td>0.085</td>
<td>63.4</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.025</td>
<td>0.022</td>
<td>-12.4</td>
<td>-14.8</td>
</tr>
<tr>
<td></td>
<td>Coef. Of Variation (St. dev. / mean)</td>
<td>0.482</td>
<td>0.243</td>
<td>-49.5</td>
<td>-45.3</td>
</tr>
<tr>
<td>IP</td>
<td>0.096</td>
<td>0.142</td>
<td>47.3</td>
<td>0.137</td>
<td>41.9</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.039</td>
<td>0.021</td>
<td>-46.4</td>
<td>-42.7</td>
</tr>
<tr>
<td></td>
<td>Coef. Of Variation (St. dev. / mean)</td>
<td>0.408</td>
<td>0.149</td>
<td>-83.6</td>
<td>-57.5</td>
</tr>
<tr>
<td>Gini</td>
<td>0.132</td>
<td>0.193</td>
<td>45.7</td>
<td>0.185</td>
<td>39.5</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.054</td>
<td>0.026</td>
<td>-51.4</td>
<td>-48.5</td>
</tr>
<tr>
<td></td>
<td>Coef. Of Variation (St. dev. / mean)</td>
<td>0.407</td>
<td>0.136</td>
<td>-66.7</td>
<td>-60.9</td>
</tr>
</tbody>
</table>

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

Table 1 reveals the impact that 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 that workers bring to the labor market matter to explain a substantial part of segregation and, consequently, segregation variability across states. The states’ racial/ethnic structure fueled by their immigration and linguistic profile 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., 2012a, Gradin, 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 indeed 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 most of the states in the southeast such as Alabama, Kentucky, Louisiana, 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). 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 rest of 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 analysis) 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 counter-intuitive 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 Gradin, 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 do 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., 2012a). The District of Columbia also stands out for having the largest public administration (26% of the work force compared with around 5% in New York and in other states); it is expected that this industry has less segregation by race and ethnicity than does the private sector. Therefore, segregation increases as its weight is reduced.

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).
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. Moving to the center of the country, states with intermediate-high levels of segregation can be found both in the north (Minnesota and Wisconsin) and south (Kansas, 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, very different patterns are identified in states with relatively low segregation. 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, different patterns were found among states with high unconditional segregation. 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.

The robustness of these findings was checked 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). In fact, 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 workforce 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, and southwestern states. 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 an industrial structure 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, the highest conditional segregation was found in the east central region, 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 more (less) segregative labor markets, this could be undertaken in future research.

References


Appendix

<table>
<thead>
<tr>
<th>States</th>
<th>M</th>
<th>IP</th>
<th>Gini</th>
<th>M conditional*</th>
<th>IP conditional*</th>
<th>Gini conditional*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>0.042</td>
<td>0.091</td>
<td>0.126</td>
<td>0.117</td>
<td>0.169</td>
<td>0.228</td>
</tr>
<tr>
<td>Arizona</td>
<td>0.082</td>
<td>0.146</td>
<td>0.200</td>
<td>0.083</td>
<td>0.129</td>
<td>0.179</td>
</tr>
<tr>
<td>California</td>
<td>0.099</td>
<td>0.166</td>
<td>0.236</td>
<td>0.071</td>
<td>0.127</td>
<td>0.176</td>
</tr>
<tr>
<td>Colorado</td>
<td>0.056</td>
<td>0.106</td>
<td>0.143</td>
<td>0.083</td>
<td>0.137</td>
<td>0.187</td>
</tr>
<tr>
<td>Connecticut</td>
<td>0.053</td>
<td>0.093</td>
<td>0.132</td>
<td>0.081</td>
<td>0.134</td>
<td>0.188</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>0.130</td>
<td>0.194</td>
<td>0.260</td>
<td>0.156</td>
<td>0.211</td>
<td>0.282</td>
</tr>
<tr>
<td>Florida</td>
<td>0.042</td>
<td>0.098</td>
<td>0.139</td>
<td>0.049</td>
<td>0.101</td>
<td>0.144</td>
</tr>
<tr>
<td>Georgia</td>
<td>0.066</td>
<td>0.122</td>
<td>0.164</td>
<td>0.083</td>
<td>0.139</td>
<td>0.183</td>
</tr>
<tr>
<td>Hawaii</td>
<td>0.069</td>
<td>0.133</td>
<td>0.184</td>
<td>0.114</td>
<td>0.151</td>
<td>0.203</td>
</tr>
<tr>
<td>Illinois</td>
<td>0.058</td>
<td>0.107</td>
<td>0.148</td>
<td>0.074</td>
<td>0.122</td>
<td>0.169</td>
</tr>
<tr>
<td>Indiana</td>
<td>0.025</td>
<td>0.048</td>
<td>0.064</td>
<td>0.123</td>
<td>0.167</td>
<td>0.218</td>
</tr>
<tr>
<td>Kansas</td>
<td>0.036</td>
<td>0.070</td>
<td>0.094</td>
<td>0.110</td>
<td>0.182</td>
<td>0.215</td>
</tr>
<tr>
<td>Kentucky</td>
<td>0.023</td>
<td>0.040</td>
<td>0.055</td>
<td>0.117</td>
<td>0.159</td>
<td>0.217</td>
</tr>
<tr>
<td>Louisiana</td>
<td>0.053</td>
<td>0.107</td>
<td>0.151</td>
<td>0.107</td>
<td>0.161</td>
<td>0.218</td>
</tr>
<tr>
<td>Maryland</td>
<td>0.057</td>
<td>0.105</td>
<td>0.148</td>
<td>0.079</td>
<td>0.127</td>
<td>0.176</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>0.042</td>
<td>0.077</td>
<td>0.104</td>
<td>0.070</td>
<td>0.126</td>
<td>0.175</td>
</tr>
<tr>
<td>Michigan</td>
<td>0.028</td>
<td>0.058</td>
<td>0.079</td>
<td>0.087</td>
<td>0.134</td>
<td>0.179</td>
</tr>
<tr>
<td>Minnesota</td>
<td>0.029</td>
<td>0.051</td>
<td>0.069</td>
<td>0.099</td>
<td>0.156</td>
<td>0.212</td>
</tr>
<tr>
<td>Missouri</td>
<td>0.024</td>
<td>0.050</td>
<td>0.070</td>
<td>0.097</td>
<td>0.147</td>
<td>0.198</td>
</tr>
<tr>
<td>Nevada</td>
<td>0.095</td>
<td>0.157</td>
<td>0.210</td>
<td>0.090</td>
<td>0.136</td>
<td>0.188</td>
</tr>
<tr>
<td>New Jersey</td>
<td>0.073</td>
<td>0.127</td>
<td>0.176</td>
<td>0.075</td>
<td>0.128</td>
<td>0.177</td>
</tr>
<tr>
<td>New York</td>
<td>0.053</td>
<td>0.110</td>
<td>0.156</td>
<td>0.053</td>
<td>0.110</td>
<td>0.156</td>
</tr>
<tr>
<td>North Carolina</td>
<td>0.063</td>
<td>0.113</td>
<td>0.149</td>
<td>0.098</td>
<td>0.151</td>
<td>0.199</td>
</tr>
<tr>
<td>Ohio</td>
<td>0.020</td>
<td>0.044</td>
<td>0.062</td>
<td>0.084</td>
<td>0.135</td>
<td>0.183</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>0.039</td>
<td>0.080</td>
<td>0.109</td>
<td>0.109</td>
<td>0.163</td>
<td>0.220</td>
</tr>
<tr>
<td>Oregon</td>
<td>0.047</td>
<td>0.076</td>
<td>0.102</td>
<td>0.083</td>
<td>0.128</td>
<td>0.177</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>0.025</td>
<td>0.050</td>
<td>0.071</td>
<td>0.074</td>
<td>0.131</td>
<td>0.176</td>
</tr>
<tr>
<td>South Carolina</td>
<td>0.063</td>
<td>0.123</td>
<td>0.166</td>
<td>0.103</td>
<td>0.155</td>
<td>0.208</td>
</tr>
<tr>
<td>Tennessee</td>
<td>0.034</td>
<td>0.068</td>
<td>0.093</td>
<td>0.110</td>
<td>0.161</td>
<td>0.218</td>
</tr>
<tr>
<td>Texas</td>
<td>0.078</td>
<td>0.150</td>
<td>0.208</td>
<td>0.078</td>
<td>0.130</td>
<td>0.179</td>
</tr>
<tr>
<td>Virginia</td>
<td>0.046</td>
<td>0.095</td>
<td>0.132</td>
<td>0.068</td>
<td>0.123</td>
<td>0.168</td>
</tr>
<tr>
<td>Washington</td>
<td>0.049</td>
<td>0.080</td>
<td>0.110</td>
<td>0.068</td>
<td>0.124</td>
<td>0.169</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>0.023</td>
<td>0.044</td>
<td>0.059</td>
<td>0.100</td>
<td>0.150</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Table A1. 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/).
3 In what follows, the “non-Hispanic” origin of these groups will be omitted for the sake of simplicity.
4 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.
5 The 18 states dropped are those having two or more minorities with less than 520 observations: 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. These states represent 9% of workers in our survey. The fact that these states are not included in our analysis does not affect either the conditional or unconditional segregation level of the remaining states.
6 For a more detailed discussion of the differences/similarities among these indexes, see Alonso-Villar and Del Río (2011), in which they are used to quantify the spatial concentration of employment.
7 This is in line with the conventional wage gap decomposition in the explained and unexplained effects (characteristics and coefficients, respectively).
The explicative variable is a dummy that has a value of 1 if the worker belongs to the New York sample and 0 if she/he belongs to the target state. Explaining variables are an array of dummies accounting for four factors: race and ethnicity (six groups omitting whites); attained education (less than high school (omitted), high school diploma, some college, and bachelor's degree or higher); immigration (born in the U.S. (omitted), immigrant with up to 5 years of residence, between 6 and 10, between 11 and 15, or more than 15) and English proficiency (speaking only English (omitted), speaking English very well, well, not well, not at all); and industry (NAICS at one digit, 14 groups (omitting group 10)).

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).

Main results do not change using the JP and G indexes.