



Published in final edited form as:

Socius. 2019 ; 5: . doi:10.1177/2378023119851016.

Placing Racial Classification in Context

Robert E. M. Pickett¹, Aliya Saperstein², Andrew M. Penner³

¹University of California, Berkeley, Berkeley, CA, USA

²Stanford University, Stanford, CA, USA

³University of California, Irvine, Irvine, CA, USA

Abstract

This article extends previous research on place-based patterns of racial categorization by linking it to sociological theory that posits subnational variation in cultural schemas and applying regression techniques that allow for spatial variation in model estimates. We use data from a U.S. restricted-use geocoded longitudinal survey to predict racial classification as a function of both individual and county characteristics. We first estimate national average associations, then turn to spatial-regime models and geographically weighted regression to explore how these relationships vary across the country. We find that individual characteristics matter most for classification as “Black,” while contextual characteristics are important predictors of classification as “White” or “Other,” but some predictors also vary across space, as expected. These results affirm the importance of place in defining racial boundaries and suggest that U.S. racial schemas operate at different spatial scales, with some being national in scope while others are more locally situated.

Keywords

racial classification; geographically weighted regression; place; culture; spatial statistics

In recent years the social sciences have experienced a kind of Cambrian explosion in data and methods that allow researchers to pursue questions that were previously unanswerable. Freed from reliance on thin sources of national data, scholars are now exploring spatial variation and subnational trends with greater interest (Alexander, Zagheni, and Barbieri 2017; Baumer et al. 2017; Chetty et al. 2016; Fording, Soss, and Schram 2011; Matthews and Parker 2013). We join this trend by using restricted-use geocoded data to explore racial classification, seeking to understand how this process may vary by place. Previous research

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages(<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Corresponding Author: Robert E. M. Pickett, University of California, Berkeley, Department of Demography, 2232 Piedmont Avenue, Berkeley, CA 94720–2120, USA, rpickett@berkeley.edu.

Authors' Note

An earlier draft of this article was presented at the 2015 annual meeting of the Population Association of America. This research was conducted with restricted access to BLS data. The views expressed here do not necessarily reflect the views of the BLS.

Supplemental Material

[Supplemental material](#) for this article is available online.

has demonstrated the importance of place in patterns of racial categorization, showing how both historical context and contemporary population composition play a role in how people racially identify or are racially classified (e.g., Bratter and O’Connell 2017; Liebler and Zacher 2016; Porter, Liebler, and Noon 2015). We take that observation one step further, relaxing the assumption that the characteristics of places will have consistent associations throughout the country. We draw on sociological theory suggesting that racial schemas are likely to differ qualitatively in different places and apply methods drawn from geography to test this in the United States.

Specifically, we use unique geocoded data from the 1979 National Longitudinal Survey of Youth (NLSY) that include racial classifications of respondents from each survey wave between 1979 and 1998. We use fixed-effects linear probability models to show that both individual and contextual factors influence how respondents were perceived racially by survey interviewers. We then dig deeper into these national associations by exploring whether characteristics of people and places predict racial classification differently in different places. To do so, we use models developed in the geography literature, specifically spatial-regime models and geographically weighted regression (GWR). Our results affirm the importance of places and their characteristics in defining racial boundaries and reveal patterns that would have been obscured by national average models alone.

Background

Our interest in studying the predictors of racial classification extends from the consensus position in the social sciences that racial categories are social constructions and not intrinsic or essential characteristics of people. A substantial body of work demonstrates that racial categorization varies across place and time. For instance, both Loveman (2014) and Nobles (2000) highlighted how national politics shape not only which racial categories are accepted in a particular country but also which individuals are seen as belonging to which categories. In this way, who counts as “White” in the Dominican Republic may be different from who counts as “White” in Puerto Rico or the mainland United States (Roth 2012). Haney Lopez (2006) provided another example, demonstrating that legal definitions of race changed substantially over time, as judges relied on varying combinations of commonsense understandings and scientific literature to determine who was allowed to be “White” for the purposes of naturalization in the United States. This evidence supports the widespread claim that race is “socially constructed” because everything from the names and number of categories to the criteria for category membership is contingent on understandings located in a particular country and particular period of time.

Research also shows that racial categorization can be highly contingent within the same country at the same point in time, such that people with similar characteristics may be categorized differently depending on where they live. For example, Liebler and Zacher (2016) showed that place-specific history, such as residing in American Indian territory or a former Confederate slave state, is related to racial identification. Similarly, Bratter and O’Connell (2017) found that living in a state with a history of antimiscegenation laws, living in areas with high Black relative to White poverty, or living in a highly diverse state increases the probability that parents will identify their biracial children only as “Black.”

Americans are also more likely to classify their neighbors in ways that “match” the predominant racial category in their local area (Porter et al. 2015), suggesting that a person may be perceived differently in different parts of the country.

Fluidity in racial categorization also occurs within individuals. Liebler et al. (2017) used linked data to follow the same people from one census to the next and found substantial variation in how Americans are recorded by both ethnicity and race. Renfrow (2004) argued that changes in racial identification do not necessarily have to reflect strongly held beliefs about personal identity. Instead, individuals may adapt or adopt the assumptions others make about their identities in order to “go with the flow” or more easily navigate potentially difficult social situations (see also Garcia 2014). Other research on the individual-level correlates of racial categorization reveals patterns that suggest racial categories draw significance from more than just differences in physical characteristics. For example, Vargas (2015) found that people who reported Hispanic origins were more likely to identify racially as “White” not only when they had lighter skin tone but also when they were older or were politically conservative. Saperstein and Penner (2012) also suggested that some within-person categorization fluidity is shaped by social status, consistent with widespread stereotypes about people with different social status characteristics (e.g., employment, poverty) belonging in different race categories. In all, prior research demonstrates not only the social construction of racial categories but also that the resulting categorization of individuals is more fluid and contingent than many would expect.

Culture, Cognition, and Schemas

Although not often explicitly discussed in research on racial classification, the broader sociological theory of cultural schemas is central to accounts of race as socially constructed (see, e.g., Brekhus et al. 2010; Brubaker, Loveman, and Stamatov 2004; Roth 2012). Schemas are the mental structures that both organize and modify perception (DiMaggio 1997). Thus, when we encounter an object in the world, we do not perceive it in its full complexity but instead quickly determine whether it is an instance of a familiar category and, if so, use our schematic knowledge of those *kinds of things* to fill in details. The same goes for when we encounter new people. For instance, Johnson, Lick, and Carpinella (2015) argued that classification of individuals relies on visual cues to categorize individuals, but which visual cues are important, and which are ignored, depends on stereotypical knowledge of what to expect. Racial classification, then, relies on knowledge about what the relevant racial categories are and the perception of individuals as falling into one category or another based on stereotypical expectations.

Schemas are cognitive, but they should not be thought of as purely psychological or individual. Instead, what the relevant categories are and the characteristics that are associated with them are culturally determined: schemas are intersubjective and constituted by the cultural environment in which they are embedded. They reflect a kind of typicality: schemas are defined and change according to the kinds of things they are used for. The contents of schemas are inherent in the materials that they are used to make sense of, and those materials themselves dictate which set of schemas are relevant (Sewell 1992, 2005). Thus, we come to see people as divided into racial groups (e.g., because they are so

classified by the census), and we are socialized to see certain people as falling into each category. As we use these categories in our day-to-day lives, however, the “webs of meaning” (Geertz 1973) of what kinds of people fall into which categories can shift, coming to reflect the people around to be classified.

However, culture, and its associated mental schemas, need not be monolithic. Instead, “anything we might designate as a ‘society’ or a ‘nation’ will contain, or fail to contain, a multitude of overlapping and interpenetrating cultural systems, most of them either subsocietal or trans-societal or both” (Sewell 2005:171). This suggests that, as researchers, we should not constrain our thinking to heavily stylized national cultures but instead allow for contestation and variation in how cultural schemas are deployed. Similarly, if schemas reflect “a mix of typicality and availability in a given location” (DiMaggio 1997:276), we should expect practices of classification to be shaped by local understandings (cf. DiMaggio 1987; Zerubavel 1991). We should not, then, simply assume that there is a single U.S. understanding of race. Some racial boundaries may take on nuances or significance from particularly local understandings, while others may be consistent not only within the United States but also have cross-national significance and application. Thus, the scale at which racial schemas operate should be treated as an empirical question, rather than assuming a single, uniform schema across all people and categories.

Quantitative studies of racial classification often prematurely foreclose the possibility of identifying subnational variation along these lines because standard approaches generally rely on regression analyses fit on national data. Standard regressions assume stability in their coefficients, constraining estimated factors to influence racial classification *in the same way* for all places. Exploring the potential for more local variation in racial schemas requires a different set of tools.

Analytic Approach

We begin by exploring how individual and contextual characteristics might predict racial classification by estimating a series of linear probability models. These models reveal which of our included factors are most closely associated with classification into or out of a given racial category across the full sample of respondents. From these results we infer the content of the implicit racial schemas being deployed by the survey interviewers (i.e., which factors matter for the interviewer’s perception of a given individual as a member of a particular racial group).

In our case, cultural theory suggests that we should expect spatial variation in these stereotypical associations, or “webs of meaning” underlying racial schemas, and thus provides an example in which standard regression assumptions of stable coefficients across space should be subjected to empirical test. To do so, we use two methods for studying subnational variation: spatial-regime models and GWR (Brunsdon, Fotheringham, and Charlton 1996; O’Loughlin, Flint, and Anselin 1994).

These methods are appealing in part for their intuitive approach. Both spatial-regime models and GWR allow for variation in regression coefficients by fitting a series of local regressions and examining how the estimated regression coefficients do or do not vary. Spatial-regime

models do so by splitting the data into mutually exclusive regions and generating estimates within each of these regions, akin to estimating a fully interacted regression model between the main predictors and a region identifier (Curtis, Voss, and Long 2012). GWR, on the other hand, fits local regressions on data within a given distance bandwidth surrounding a focal point. This splits the data into possibly overlapping regions whereby region size is determined inductively through maximizing model fit (Brunsdon, Fotheringham, and Charlton 1998; Fotheringham, Charlton, and Brunsdon 1998; Matthews and Yang 2012). Both methods fit regressions to local data, allowing for variation in regression parameters across space. Thus, they enable researchers to identify whether their estimates are highly variable (and likely local) or highly stable (and likely national).

Data and Methods

For this study, we use restricted-access geocoded data from the NLSY. This longitudinal survey began with a nationally representative sample of 12,686 men and women in the United States who were aged 14 to 22 years in 1979. Data have been collected about respondents each year (or every other year since 1994) and, importantly for our purposes, include the locations where interviews were conducted. The wealth of information in the NLSY allows us to delve into the process of racial classification of respondents by interviewers and explore the individual and contextual correlates of repeated classifications with its remarkable time series.

The dependent variables for our analyses are based on the racial classification recorded by interviewers each survey year between 1979 through 1998, which was the last year interviewers were asked to classify respondents.¹ We treat these racial classifications as proxies for how the respondents are typically perceived racially by others.² Respondents experienced notable variation in how their racial categories (“White,” “Black” or “Other”) were recorded; among survey participants, 20 percent experienced at least one change in their racial categorization at some point between 1979 and 1998, and, among those, 30 percent experienced five or more changes. This variation allows us to explore the factors that are associated with racial classification, by leveraging changes within the same respondent.

The geocoded nature of the data set allows us to extend previous research on racial classification in the NLSY by incorporating contextual data for the counties in which respondents live.³ We included the county unemployment rate and a measure of county-level poverty to provide contextual corollaries to the individual status characteristics included in previous research on racial classification in the NLSY (see Saperstein and Penner 2012).⁴

¹There is one exception during this time span: in 1987, NLSY interviews were conducted entirely over the phone, and interviewer-classified race was not recorded. Some interviews were conducted over the phone in other survey years as well, but the interviewer still recorded the racial classification of the respondent. We include these cases because previous research found that predictors of classification did not vary significantly by mode of interview (Saperstein and Penner 2012), but we do not include any data from 1987.

²The NLSY also includes respondent’s self-identified race, but only in 1979 and 2002. This makes it less feasible to study contextual characteristics as predictors of self-identification with these data. We do not include 2002 self-identified race as a predictor in our regression models because it comes four years after the last interviewer classification. We do not include 1979 self-identified race as a predictor, because the single data point per respondent would be nullified by the respondent-level fixed effects. For a discussion of similarities and differences between self-reports and interviewer classification, see Saperstein and Penner (2012).

³Because interviewers tend to be assigned to nearby respondents (within 50 miles or so), these contextual data provide information about the social environment the interviewer is embedded in as well.

We obtained estimates of the annual county-level unemployment rate from the Bureau of Labor Statistics (BLS) and estimates of average per capita income from the Regional Economic Accounts of the Bureau of Economic Analysis. County-level poverty was calculated by dividing the poverty rate for two adults and one child in each year by the county's average income. Thus, a score of 1 implies that the county's aggregate income is at poverty level, and scores greater than 1 indicate greater poverty. We also included county-level demographic characteristics such as population size and ethnoracial composition. Our measures of county population size, percentage Black residents, percentage Hispanic residents, and percentage foreign-born residents were all obtained from decennial census data. We created annual estimates of these quantities for each county and year through linear interpolation between censuses.

From data on ethnoracial composition, we also constructed estimates of the Simpson diversity index, whereby each county's diversity is measured as $\sum_i \frac{p_i(1-p_i)}{(1-\frac{1}{N})}$, where p_i is the

proportion of the population in racial-ethnic category i , and there are N categories (Reardon and Firebaugh 2002). In our case, we use four ethnoracial categories: Hispanic, non-Hispanic White, non-Hispanic Black, and non-Hispanic other. These categories mirror the racial classification scheme used to classify respondents but also account for Hispanic origins. We included the latter because previous research has found that racial fluidity is more common among Americans who report Hispanic origins (Liebler et al. 2017; Saperstein and Penner 2016), and thus counties with higher concentrations of residents reported as Hispanic in the census may have different patterns of racial classification than other places. The denominator of the diversity index standardizes for the number of groups, giving it a possible range of 0 to 1: 0 when everyone in a county is of the same ethnoracial group and 1 when everyone in a county is evenly distributed across all four groups. We multiplied by 100 to rescale this index to 0 through 100.

County summary statistics can be found in Table 1, including the minimum and maximum values for each characteristic and the standard deviation.⁵ On average, across counties and over time, respondents lived in counties with a population of about 330,000 (or a log population size of 12.7), an unemployment rate of 7.7 percent, and where average incomes were 61 percent higher than the poverty rate ($1/0.62 = 0.61$). According to the census, residents in these counties were, on average, 14.9 percent Black, 9.6 percent Hispanic, and 7.7 percent foreign born, and the counties had an average Simpson diversity index of 47.8. These figures are roughly in line with national averages for this period (the U.S. unemployment rate ranged from 4.5 percent in 1998 to 9.7 percent in 1982, the share of Black Americans ranged from 11.1 percent in 1970 to 12.3 percent in 2000, the share of Hispanic Americans ranged from 4.4 percent in 1970 to 12.5 percent in 2000, and the share of foreign-born Americans ranged from 4.7 percent in 1970 to 11.1 percent in 2000).

⁴We initially considered including contextual measures of crime as a corollary to individual experiences with incarceration but found that county-level data from the Federal Bureau of Investigation's Uniform Crime Reports were highly unreliable during this time period.

⁵Our analytic sample is 129,177 person-years because we drop cases with missing data. This is primarily a result of missing geographic information (91,305 observations) but also due to missing interviewer characteristics (22,518 observations) and missing individual characteristics (10,720).

Residential Mobility and Classification Fluidity

Historical evidence related to racial “passing”—the practice of hiding or downplaying one’s ancestry in order to be perceived as having another, typically more advantaged, background—suggests that permanent shifts from one racial category to another were facilitated by long-distance moves that allowed individuals to leave their families behind and invent their identities anew in a new place (see, e.g., Hobbs 2014). To ensure that within-person variation across space in racial classification is not driven primarily by such processes, we first examined whether racial classification fluidity in the NLSY is better attributed to people who move from place to place rather than place-based differences in racial schemas. Our results suggest that people who move from one county to another have relatively similar levels of year-to-year classification fluidity as those who stay in the same county from one year to the next (5.7 percent and 5.9 percent, respectively). We also find that average contextual characteristics are relatively similar for those who move and those who stay put and that the national-level predictors of classification are similar for both groups.⁶ Thus, we conclude that the fluidity observed in relatively contemporary NLSY data is not restricted to isolated experiences of long-term passing and can offer broader insight into how contextual characteristics shape racial classification. For additional details about this analysis, see the supplemental materials.

National Models

To explore the contextual factors that predict racial classification, we estimated a set of linear probability models with respondent-level fixed effects. Including respondent-level fixed effects allows our analyses to explore racial classification within respondents. Thus, we predict a person’s current racial classification as “White,” “Black,” or “Other,” net of his or her average racial classification.⁷ This estimation strategy also rules out common alternative hypotheses, most notably a hypothesis of classical measurement error (Saperstein and Penner 2016).⁸

Our inclusion of respondent-level fixed effects makes linear probability models preferable to the more common logit specification for binary outcomes. Logit fixed-effects models must be estimated conditional upon change within an individual, whereas linear probability models allow estimation on the full sample. We are primarily interested in how covariates predict classification into a given racial category, and linear probability models produce

⁶We do find a small positive relationship between moving between counties and racial classification as “Black,” but we do not find significant differences among the predictors when estimate split models for movers and stayers (see supplemental materials), and moving is not a significant predictor of classification as either “White” or “Other.”

⁷The dependent variables are coded as binary outcomes for “White” versus all else, “Black” versus all else, and so on. We prefer linear probability models of binary outcomes fit using ordinary least squares compared to multinomial models fit using maximum likelihood, in part because our later models are computationally intensive and we wish to preserve comparability between results. Previous research has found that specification of models as multinomial instead of as binary outcomes (using linear models or otherwise) does not substantially influence the results (Saperstein and Penner 2012).

⁸The hypothesis follows roughly: an individual has a stable characteristic that predicts his or her racial classification, but this classification is measured with some error. The error term of racial measurement will thus be associated with this fixed individual propensity. If researchers want to find factors that predict racial classification, they are faced with the possibility that observed associations with racial classification instead result from a correlation between those factors and the stable individual characteristic on one hand and the stable individual characteristic and the error term of racial classification on the other. If, however, researchers include individual-level fixed effects, these stable characteristics are effectively controlled, removing the possibility that observed relationships between various factors and changes in racial classification result from a person’s stable but unobserved propensity to have a given race.

reliable estimates of marginal effects, making it an appropriate modeling strategy even for our binary outcomes.⁹

Each model includes all of the county-level contextual variables described above and several key time-varying individual status characteristics also shown to be important predictors of racial classification in these data. In particular, we include measures of whether the respondent has ever been unemployed, ever experienced poverty, or ever been incarcerated. Additional time-varying controls include the respondent's age and the interviewer's age, race, education and gender (see Saperstein and Penner 2012 for more details on these measures).¹⁰ We also included year fixed effects to account for year-to-year changes in survey design and other temporal changes in how people came to be assigned to particular racial categories.¹¹

Spatial-Regime Models and GWR

We turn to spatial-regime modeling and GWR to explore spatial variation in regression estimates. Both methods allow us to compare coefficient estimates across place, either across predefined regions in spatial-regime modeling or across inductively generated regions in GWR. In both the spatial-regime models and GWRs, the underlying regression models are identical to the national models (i.e., linear probability models predicting classification as “White,” “Black,” or “Other,” with the inclusion of respondent and year fixed effects and controls listed above).¹²

For our spatial-regime models, we coded the respondent's state of residence into regions using a four-category division of “West,” “Midwest,” “Northeast,” and “South” and estimated models separately for each region. These results offer initial insight into whether our regression coefficients vary across space. If, for example, estimates are notably different in direction, magnitude, and/or statistical significance among regions, it would support the notion that the standard ordinary least squares assumption of stable coefficients in our national model is unfounded.

We then turn to GWR, which is analogous to a kernel regression across geographic space. We chose this approach because GWR allows data-driven regionalization that does not require the researcher to make a priori assumptions about the underlying spatial structure in the data, including assumptions about the size of relevant regions and the spatial structure of parameter error terms (discussed below).¹³ The results are otherwise analogous to spatial-

⁹Care should be taken if our estimates are used to produce predicted probabilities, however, as estimates may be outside the bounds of 0 to 1.

¹⁰We include both respondent and interviewer characteristics because our interest is in interviewers' perceptions of respondents. Thus, our unit of analysis is technically the interaction between a nationally representative sample of respondents and a nonrepresentative sample of interviewers. Interviewer controls are particularly important because NLSY interviewers, like survey interviewers generally, are predominantly “White,” female, and well educated.

¹¹It should be noted that modeling this large number of predictors after controlling for individual and year fixed effects is asking quite a lot of our data. We ran similar models with each variable added individually and tended to find similar substantive results but with larger magnitudes and more frequent statistical significance. Because we are interested in broader “webs of meaning,” however, we opted to focus on the full models to understand how a broad set of factors work together to produce patterns of racial classification.

¹²We also considered controlling for interviewer fixed effects, on the premise that perhaps our results are being driven by a small group of interviewers who differ in their method of classifying. We opted not to include these controls in our final models, because doing so may overcontrol for the patterns of interest, as most interviewers were assigned to specific geographic territories generally within 50 miles of their homes, making interviewer fixed-effects highly col-linear with local effects predicting classification. That said, models with interviewer fixed effects provide substantively similar results to those presented here.

regime models, in that both are estimating localized regressions, allowing easy and direct comparison between the two types of models.

The GWR proceeds in several steps. First, it is necessary to define the localized samples on which to run our models. We identified the optimal size of these localities by running a cross-validation exercise that relates the sum of squared errors of our models to the size of the locality that we have chosen. Once we found a suitable locality size (bandwidth), we ran our regressions for each place to identify the local regression coefficient for each variable in our model. We then tested whether the observed variation in regression coefficients constitutes significant variation by conducting a Monte Carlo simulation. Details of these steps are provided in the supplemental materials.

Methodological Limitations

We believe that our analysis provides strong evidence for spatial variation in the associations between individual and contextual correlates and racial classification, but we do not claim that it is comprehensive. We are limited in the number of parameters we can estimate simultaneously, so we relied on findings from previous research to select relevant individual and contextual predictors of racial classification. Restrictions in our data agreement with the BLS also require that we not report results below the state level. Thus, we cannot speak to potential variation in racial classification within states, although such variation seems plausible and is an important topic for future research. Furthermore, although the NLSY is a nationally representative survey, it is not necessarily representative at smaller levels of geography. Therefore, the results we present on subnational variation should be interpreted with caution, especially with regard to their generalization. Our results reflect the general relationships in our data, but we do not make strong inferential claims about where changes in racial classification are likely to take place. Rather, we present this analysis as evidence for the utility and importance of exploring racial classification as a spatial process.

Results

We first explore the states where changes in racial classification were most common. To do so, we identify all person-years in which a respondent's racial classification changed compared with the previous person-year and then aggregate these changes to the state level. Although our regression models predict classification into a given racial category, not changes in classification, racial fluidity provides the variation needed for our main analyses. Thus, the observed spatial variation in racial fluidity represents spatial variation in our ability to identify the factors that predict classification into a given category.

A descriptive map of where racial fluidity in the NLSY was more likely to occur is shown in Figure 1. The color ramp refers to the proportion of person-years in which changes in classification occurred within each state. Darker red states have a higher proportion of person-years in which respondents experienced changes in classification, while states with

¹³Researchers can explore nonstationarity of regression coefficients without a fixed regional analysis using other techniques as well, such as random slope models with a spatial correlation structure or multiscale GWR (Fotheringham, Yang, and Kang 2017). Some of the advantages of GWR are that it is easy to interpret and computationally much easier to estimate. We opted for this approach given the size of our data and the number of parameters we estimate.

less racial fluidity are colored with lighter reds. The map suggests that racial fluidity was more common in both the Southwest (California, Arizona, and New Mexico) and former “rust belt” areas of the northern Midwest (Michigan and Ohio). Interestingly, many of the states with high levels of racial fluidity in the NLSY in the 1980s and 1990s also have relatively high levels of multiple race reporting in 2010 census data (Jones and Bullock 2012). However, given that the NLSY is not nationally representative within states, the generalizability of these patterns should be interpreted with caution.

Predictors of Racial Classification: National Level

We next turn to national predictors of when a given racial classification is likely to occur (i.e., the classifications that are the result of the fluidity demonstrated above). Our results suggest that the characteristics of a respondent’s home county matter more for whether that respondent is classified as “White” or “Other” but that the individual’s own characteristics matter more for the probability of being classified as “Black.” Predictors of racial classification as “White” and “Other” are presented in Table 2, and predictors of racial classification as “Black” are presented in Table 3.

By comparing across the two national models in Table 2 (columns 1 and 6), it becomes clear that predictors of classification as “White” mirror those of classification as “Other.” Thus, a 1 percent increase in the county unemployment rate increases the probability of classification as “Other” by 0.2 percent, while the same increase in county unemployment also decreases the probability of classification as “White” by 0.2 percent. The only estimate in the national models for classification as “White” that is not mirrored with a statistically significant estimate in the opposite direction for classification as “Other” is percentage foreign born (though even here the coefficients have similar magnitudes in opposite directions). These results suggest that most factors are not related to either the probability of classification as “White” or “Other” alone but are instead jointly defining the boundary between the “White” and “Other” categories.

The national results also suggest the importance of contextual covariates for predicting “White” and “Other” classification, over and above individual status characteristics.¹⁴ Interviewers are less likely to classify respondents as “White” and more likely to classify them as “Other” in counties with higher unemployment rates, higher levels of aggregate poverty, greater shares of foreign born residents, and higher levels of ethnoracial diversity (as measured by the Simpson diversity index). At the same time, interviewers are more likely to classify respondents as “White” and less likely to classify them as “Other” in counties with greater shares of Hispanic and Black residents, all else being equal (including county ethnoracial diversity).¹⁵ The magnitude of these relationships is substantively quite small, so we do not want to overstate the role of contextual factors in the process of racial

¹⁴Although the individual status characteristics do not reach conventional levels of statistical significance in our national models predicting classification as “White” or “Other”—in part because of the simultaneous estimation of many variables—the magnitude and direction of the estimates are generally in line with previous research (see Saperstein and Penner 2012).

¹⁵The relationship with Hispanic population composition is consistent with a pre-1980 emphasis on classifying most people with Hispanic origins as “White” (Rodriguez 2000). We should note, however, that our models suggest that a higher proportion of Hispanic residents in a county is associated with higher probabilities that any given individual is classified as “White,” not only respondents of Hispanic origin.

classification. However, given the total number of controls included in our models (including respondent fixed effects), it is perhaps unsurprising that these estimates would be small. Furthermore, across the “White” and “Other” national models, 11 of the 14 contextual characteristic estimates are statistically significant; that contextual factors are significant predictors of racial classification at all is important evidence of how racial categories are socially constructed.

The national models predicting classification as “Black” (Table 3), on the other hand, tell a much different story than the models predicting classification as “White” or “Other.” First, county characteristics are not significant predictors of whether interviewers classify respondents as “Black” in general. Indeed, the only characteristic that approaches statistical significance is the individual covariate of whether the respondent had ever been incarcerated, which increases the probability of being seen as “Black” by 0.4 percent ($p = .14$). The lack of significant measured correlates could be due to the relative stability of the “Black” racial category, evidenced in part by the amount of variance explained by individual fixed effects. The higher proportion of variance explained in the national model predicting classification as “Black” (98 percent), compared with models for “White” (85 percent) and “Other” (48 percent), also suggests that unmeasured individual characteristics play a larger role in who is classified as “Black” in the United States.¹⁶ These could include relatively stable aspects of physical appearance, such as facial features or eye color, or racially stereotypical names (see, e.g., Bertrand and Mullainathan 2004; Cook, Logan, and Parman 2016).

Predictors of Racial Classification: Spatial Regimes

To assess whether these national results are stable across geographic space, we turn to spatial-regimes models, which fit a separate regression for each region. Our results from these models suggest that there is significant variation in how characteristics predict the boundary between “White” and “Other” but there is stability across regions in estimates predicting classification as “Black.”

The spatial-regime results for classification as “White” or “Other” can be found alongside the national estimates in Table 2. Although several predictors are consistent in direction across the country (e.g., the level of ethnracial diversity and the percentage of the county population that was recorded as Hispanic in the census), for other factors, the direction of the classification predicted varies by region. This variation is most noticeable for contextual characteristics that were not statistically significant in the national models. For instance, the national estimates for county population size indicate that interviewers are less likely to classify respondents as “White” and more likely to classify them as “Other” when they live in counties with larger populations, but the regional variation suggests that this relationship is statistically significant only in the Northeast, while in the West, there is a significant relationship of similar magnitude in the opposite direction. Similarly, the share of foreign-born residents in a county was not a statistically significant predictor of classification as “Other” in the national model, but this could be a reflection of regional patterns that run in

¹⁶It is worth noting that there are regional differences in the amount of variance that can be attributed to time-invariant individual characteristics. When predicting classification as “White” or “Other,” this variance is lowest in the West (i.e., the fraction μ is 63 percent and 34 percent, respectively), while for predicting classification as “Black,” the fraction μ_j is lowest in the Northeast.

opposite directions. We also see significant regional variation related to unemployment rates. On the national level, interviewers are significantly less likely to classify respondents as “White” and more likely to classify them as “Other” in counties with higher unemployment rates, but the regional models suggest this relationship holds only in the Northeast and in the South. In the West, the relationship is reversed: all else being equal, interviewers are more likely to classify respondents as “White” and less likely to classify them as “Other” in counties with higher unemployment rates. It is important to remember that these results are net of person fixed effects and thus indicate that the same individual is more likely to be classified one way or another depending on these county characteristics.

The results from spatial-regime models predicting classification as “Black” (Table 3) differ from those predicting classification as “White” and “Other.” Although some estimates in Table 3 vary in statistical significance across regions, compared with Table 2 we see fewer instances of regions with significantly positive or negative estimates that either average out to zero or run counter to predictions from the national models. There are predictors of classification as “Black” that are statistically significant only in specific regions. For instance, individual incarceration status is a significant positive predictor of classification as “Black” in the South, and the estimate is twice as large as the next largest positive estimate (in the West or Midwest). County aggregate poverty is also a significant positive predictor of classification as “Black” in the South, and all else being equal, interviewers are significantly more likely to classify respondents as “Black” in the Northeast when they live in counties with higher shares of the population recorded as Black in the census. However, these results are all substantively consistent with the national estimates. Just one regional estimate appears to contradict the national level results: only in the Northeast is the probability of “Black” classification significantly lower in counties with greater ethnorracial diversity. Given the number of models estimated, several of these estimates also could be statistically significant by chance.

Taken together, the variation in spatial-regime estimates raises the possibility that our national estimates may suppress regional patterns and violate the ordinary least squares stationarity assumption, especially in models predicting classification as “White” and “Other.” To more inductively identify the extent to which these coefficients vary across space, and whether the observed variation differs from what would be expected due to chance alone, we turn to GWR.

Predictors of Racial Classification: GW R

The GWR generally confirms and extends the spatial-regime findings: we find significant stability in the predictors of classification as “Black” and significant variation in the predictors of classification as “Other.” Results from the GWR can be found in Table 4. The table lists the rank of the observed geographic variation in regression coefficients for a given variable out of the distribution of simulated data in which geographic assignment is random without replacement.

The rankings clearly indicate that county characteristics have a higher degree of spatial variation in the models predicting classification as “Other” than we would expect at random.¹⁷ This suggests that some county characteristics may significantly predict classification as

“Other” in some places but not in others. Furthermore, it suggests that these characteristics may vary in *how* those characteristics are meaningful (i.e. a characteristic may positively predict classification as “Other” in some places but negatively predict classification elsewhere).

Figure 2, for instance, plots the observed local regression coefficients for county-level poverty in models predicting classification as “Other,” compared with similar plots drawn from simulations in which geographic position is randomly assigned. The larger map of observed variation indicates that county-level poverty is a particularly salient predictor both in the Northwest and Northeast, where it is positively associated with classification into the “Other” category, and in the Southwest and Midwest, where it is negatively associated with classification as “Other.” This contrast is obscured in regional regressions because the meaningful patterns do not overlap with the standard four-region boundaries. Furthermore, by comparing the observed distribution of regression coefficients in the larger map to the simulated plot with randomized geography in the smaller maps, we can see which subregional patterns drive the significant variation across space. For instance, the simulated plots suggest that we would not expect to see that people in the Southwest/Midwest cluster have lower probabilities of being classified as “Other” when they live in counties with higher levels of poverty, all else being equal, if our data were randomly distributed with respect to place. In all, the results suggest that classification as “Other” is determined by locally defined webs of meaning rather than a consistent national schema or even regional schemas.

In contrast, our results for classification as “Black” suggest that there is less clustering than we would expect on the basis of chance. That is, individual characteristics such as an experience of incarceration, poverty, or unemployment are significant in the GWR predicting classification as “Black” because the variation is significantly *lower* than we would expect. This suggests that classification as “Black,” especially its association with individual status characteristics, is much more a national phenomenon than a local one. Figure 3 plots the observed local regression coefficients for individual incarceration predicting classification as “Black,” compared with similar plots drawn from simulations in which geographic position is randomly assigned. If we were to randomly assign geographic location to individuals, we would expect to see lumpier distributions with high values in some places and low values in others, as shown in the smaller maps on the right. However, the observed pattern depicted in the largest map is more muted. Although there are regions of positive and negative associations, large spikes are fewer and farther between, and a moderate positive relationship between incarceration and classification as “Black” is present for most of the country. This regularity is suggestive of a consistent cultural schema that may be driving how individual status characteristics relate to classification as “Black.” What defines this “national culture” is difficult to determine but may include pervasive racial stereotypes that are influencing interviewer perception.

¹⁷Though coefficients for classification as “Other” tend to mirror those for classification as “White,” the Monte Carlo simulation suggests that the local variation in coefficients predicting classification as “White” is not significantly different from random chance. This suggests caution in making inferences about distinct patterns of classification as “White” from the regional models; we cannot rule out the possibility that those estimates are random perturbations around the true national results.

Discussion

Our study contributes several key findings to the literature on racial categorization. First, our results support prior research suggesting that racial classification and identification are best understood as reflecting cultural schemas about race, not inherent characteristics of people. Even after accounting for all factors that remain constant within individuals through respondent fixed effects, we find that social circumstances predict how an individual will be classified racially. This provides further evidence that race is not an obvious and apparent individual trait but a reflection of a cultural understanding of categories and what kinds of people fall within them.

Second, our findings suggest that the cultural schemas underlying racial categorization should not be thought of as monolithic. Instead some schemas may be particularly variable across places, whereas others may be remarkably stable. In particular, our results suggest that the racial schemas underlying classification as “Black” are remarkably stable, whereas schemas underlying classification as “Other” are particularly variable. In some places, for instance, unemployment rates may be strongly associated with notions of “Otherness,” whereas in other places high unemployment may suggest associations with “Whiteness” or “Blackness” or have no relationship to racial classification at all.

Third, this spatial heterogeneity in predictors of racial classification can drive down average results in data pooled across the entire country, making some factors wrongly appear “insignificant.” For instance, although we find that population size does not significantly predict classification as “White” at the national level, it is a significant predictor at the regional level: in the Northeast, people living in more populous places are less likely to be classified as “White,” all else being equal, whereas in the West we find the opposite relationship. Instead of a null relationship implied by the national estimate, then, population size may have a heterogeneous relationship with “Whiteness” that varies across the country. Beyond such broadly regional patterns, our data suggest significant subregional variation as well. In the full national sample, county-level poverty is a significant positive predictor of classification as “Other” (a one-unit increase in the ratio of the poverty line divided by average income increases the probability of being classified as “Other” by 2.9 percent), but local estimates vary significantly, ranging from 14.7 percent in Montana to –12.6 percent in California.¹⁸ County poverty appears to be a somewhat minor factor in predicting racial classification, on average, but looking at how this process unfolds in particular places may reveal more substantial associations. The more general point that national models can artificially attenuate associations by averaging together strong positive and strong negative effects has important implications beyond our particular case.

It is perhaps not terribly surprising that there is local variation in the factors driving classification as “Other.” First, it was offered to interviewees as an undefined residual category, and the ancestral origins of respondents (e.g., Asian vs. Hispanic) likely vary substantially by location as a result of historical patterns of migration and resettlement.

¹⁸We offer these as illustrative of the range of variation in estimates, but the specific magnitude reported should not be overinterpreted given that these estimates arise from smaller and nonrandomly sampled subsets of data.

Second, the years our data were collected, from 1979 to 1998, coincide with the rise of an official Hispanic category in the United States, which was forged out of uneasy alliances among Mexican-, Puerto Rican- and Cuban-origin Americans who not only varied in their regional concentrations but also in their advocacy for inclusion as “White” (Mora 2014). Nevertheless, given that our analyses account for both individual fixed effects and county-level census-reported ethnoracial composition, these results provide meaningful information about variation in the boundary between “Whites” and “Others” that may be related to, but also encompasses more than, variation in composition by ancestral origin. Indeed, our results are consistent with research suggesting that immigrants with the same ancestral origins were incorporated into the U.S. racial hierarchy differently in different places (see Fox and Bloemraad 2015).

In stark contrast, the schemas underlying the “Black” category appear to be remarkably stable throughout the country, with much less variation in the predictors of racial classification than we would expect to occur through random chance. Previous scholars have argued that the distinction between “White” and “Black” in the United States arose with the founding of the nation, as an attempt to resolve the contradictions between the nation’s promise of radical freedom for all and the institution of chattel slavery (Fields 1990; West 1982). Other literature suggests that a broad consensus about “who is Black” was forged both as a result of and in response to continued repression during the Jim Crow era (Davis 1991; Williamson 1995). It is not unlikely, then, that the consistency we find in predictors of racial classification as “Black” is a function of the particularly national character of the White-Black divide in the United States.

Conclusion

With newly developed methods and more detailed individual-level data, sociologists can go beyond the now commonplace demonstration that race is a socially constructed category because classifications can vary from one country to another. By drawing on spatial models, we can dive deeper into the implications of cultural theory and map how relationships between racial boundaries and contextual predictors vary across space within the same country. Rather than resting on researchers’ assumptions about meaningful regional boundaries (e.g., “the South”), the GWR identifies the regions that best fit our data through cross-validation. These tools demonstrate local variation in regression coefficients, and our use of Monte Carlo simulations allows us to measure the statistical significance of that variation.

Our results suggest that although there may be a broader U.S. cultural schema in defining who is “Black,” there is no such coherent national schema defining the border between “White” and “Other.” Instead these distinctions rely on local notions of what “Whiteness” or “Otherness” means. These subregional associations would have been concealed by standard assumptions of stability in regression parameters. They can be revealed, and subjected to further study, by connecting recent developments in the methodological literature with contemporary theoretical understandings of racial classification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Steve McClaskie for valuable background about the data; Jennifer Cassidy-Gilbert, Vijoy Abraham, and L. J. Laubenheimer for assistance with data access; and Monica Alexander, Jenifer Bratter, Paul Chung, David Harding, and Jennifer Johnson-Hanks for their thoughts and feedback.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a National Institute of Child Health and Human Development predoctoral training grant in demography (T32-HD007275), the Russell Sage Foundation, and the Stanford Institute for Research in the Social Sciences.

Author Biographies

Robert E. M. Pickett is a PhD candidate in the departments of sociology and demography at the University of California, Berkeley. His research explores ethnoracial categorization through the lens of cultural theory and quantitative methods.

Aliya Saperstein is an associate professor of sociology at Stanford University. Her research examines how categories of difference, such as race/ethnicity and sex/gender, are operationalized in survey research and the consequences of these methodological decisions for studies of stratification.

Andrew M. Penner is a professor of sociology at the University of California, Irvine. His research focuses on inequality, social categorization, and educational policy.

References

- Alexander Monica, Zagheni Emilio, and Barbieri Magali. 2017 "A Flexible Bayesian Model for Estimating Subnational Mortality." *Demography* 54(6):2025–41. [PubMed: 29019084]
- Baumer, Eric P, Ranson J. W. Andrew, Arnio Ashley N., Fulmer Ann, and De Zilwa Shane. 2017 "Illuminating a Dark Side of the American Dream: Assessing the Prevalence and Predictors of Mortgage Fraud across U.S. Counties." *American Journal of Sociology* 123(2):549–603.
- Bertrand Marianne, and Mullainathan Sendhil. 2004 "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4):991–1013.
- Bratter Jenifer L., and O'Connell Heather A.. 2017 "Multiracial Identities, Single Race History: Contemporary Consequences of Historical Race and Marriage Laws for Racial Classification." *Social Science Research* 68:102–16. [PubMed: 29108590]
- Brekhus Wayne H., Brunsmas David L., Platts Todd, and Dua Priya. 2010 "On the Contributions of Cognitive Sociology to the Sociological Study of Race." *Sociology Compass* 4(1):61–76.
- Brubaker Rogers, Loveman Mara, and Stamatov Peter. 2004 "Ethnicity as Cognition." *Theory and Society* 33(1):31–64.
- Brunsdon C, Fotheringham AS, and Charlton ME. 1996 "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28(4):281–98.
- Brunsdon C, Fotheringham AS, and Charlton ME. 1998 "Geographically Weighted Regression." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47:431–43.

- Chetty Raj, Stepner Michael, Abram Sarah, Lin Shelby, Scuderi Benjamin, Turner Nicholas, Bergeron Augustin, and Cutler David. 2016 “The Association between Income and Life Expectancy in the United States, 2001–2014.” *JAMA* 315(16):1750–66. [PubMed: 27063997]
- Cook Lisa D., Logan Trevon D., and Parman John M.. 2016 “The Mortality Consequences of Distinctively Black Names.” *Explorations in Economic History* 59:114–125.
- Curtis Katherine J., Voss Paul R., and David D 2012 “Spatial Variation in Poverty-Generating Processes: Child Poverty in the United States.” *Social Science Research* 41(1):146–59. [PubMed: 23017703]
- James Davis, F.. 1991 *Who Is Black? One Nation’s Definition*. Philadelphia: Pennsylvania State University Press.
- DiMaggio Paul. 1987 “Classification in Art.” *American Sociological Review* 52(4):440–55.
- DiMaggio Paul. 1997 “Culture and Cognition.” *Annual Review of Sociology* 23:263–87.
- Fields Barbara J. 1990 “Slavery, Race and Ideology in the United States of America.” *New Left Review* 181(May-June):95–118.
- Fording Richard C., Soss Joe, and Schram Sanford F.. 2011 “Race and the Local Politics of Punishment in the New World of Welfare.” *American Journal of Sociology* 116(5):1610–57.
- Fotheringham AS, Charlton ME, and Brunson C. 1998 “Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis.” *Environment and Planning A* 30(1):1905–27.
- Fotheringham A. Stewart, Wenbai Yang, and Wei Kang. 2017 “Multiscale Geographically Weighted Regression (MGWR).” *Annals of the American Association of Geographers* 107(6):1247–1265.
- Fox Cybelle, and Bloemraad Irene. 2015 “Beyond ‘White by Law’: Explaining the Gulf in Citizenship Acquisition between Mexican and European Immigrants, 1930.” *Social Forces* 94(1):181–207.
- Garcia Angela S. 2014 “Hidden in Plain Sight: How Unauthorized Migrants Strategically Assimilate in Restrictive Localities in California.” *Journal of Ethnic and Migration Studies* 40(12):1895–1914.
- Geertz Clifford. 1973 *The Interpretation of Cultures*. New York: Basic Books.
- Hobbs Allyson. 2014 *A Chosen Exile: A History of Racial Passing in American Life*. Cambridge, MA: Harvard University Press.
- Johnson Kerri L., Lick David J., and Carpinella Colleen M.. 2015 “Emergent Research in Social Vision: An Integrated Approach to the Determinants and Consequences of Social Categorization.” *Social and Personality Psychology Compass* 9(1):15–30.
- Jones Nicholas A., and Bullock Jungmiwha. 2012 “The Two or More Races Population: 2010.” 2010 Census Briefs. Retrieved December 20, 2018. <http://www.census.gov/prod/cen2010/briefs/c2010br-13.pdf>
- Liebler Carolyn A., Porter Sonya R., Fernandez Leticia E., Noon James M., and Ennis Sharon R.. 2017 “America’s Churning Races: Race and Ethnicity Response Changes between Census 2000 and the 2010 Census.” *Demography* 54(1):259–84. [PubMed: 28105578]
- Liebler Carolyn A., and Zacher Meghan. 2016 “History, Place, and Racial Self-Representation in 21st Century America.” *Social Science Research* 57:211–32. [PubMed: 26973041]
- Lopez Ian H. 2006 *White by Law: The Legal Construction of Race*. New York: New York University Press.
- Loveman Mara. 2014 *National Colors: Racial Classification and the State in Latin America*. Oxford, UK: Oxford University Press.
- Matthews Stephen A., and Parker Daniel M.. 2013 “Progress in Spatial Demography.” *Demographic Research* 28:271–312.
- Matthews Stephen A., and Yang Tse-Chuan. 2012 “Mapping the Results of Local Statistics: Using Geographically Weighted Regression.” *Demographic Research* 26:151–66. [PubMed: 25578024]
- Mora G. Cristina. 2014 *Making Hispanics: How Activists, Bureaucrats, and Media Constructed a New American*. Chicago: University of Chicago Press.
- Nobles Jenna. 2000 *Shades of Citizenship: Race and the Census in Modern Politics*. Stanford, CA: Stanford University Press.

- O'Loughlin John, Flint Colin, and Anselin Luc. 1994 "The Geography of the Nazi Vote: Context, Confession, and Class in the Reichstag Election of 1930." *Annals of the Association of American Geographers* 84(3):351–80.
- Porter Sonya R., Liebler Carolyn A., and Noon James M.. 2015 "An Outside View. What Observers Say about Others' Races and Hispanic Origins." *American Behavioral Scientist* 60(4):465–97.
- Reardon Sean F., and Firebaugh Glenn. 2002 "Measures of Multigroup Segregation." *Sociological Methodology* 32(1):33–67.
- Renfrow Daniel G. 2004 "A Cartography of Passing in Everyday Life." *Symbolic Interaction* 27(4): 485–506.
- Rodriguez Clara E. 2000 *Changing Race: Latinos, the Census and the History of Ethnicity*. New York: New York University Press.
- Roth Wendy D. 2012 *Race Migrations: Latinos and the Cultural Transformation of Race*. Stanford, CA: Stanford University Press.
- Saperstein Aliya, and Penner Andrew M.. 2012 "Racial Fluidity and Inequality in the United States." *American Journal of Sociology* 118(3):676–727.
- Saperstein Aliya, and Penner Andrew M.. 2016 "Still Searching for a True Race? Reply to Kramer et al. and Alba et al." *American Journal of Sociology* 122(1):263–85.
- Sewell William H. Jr, 1992 "A Theory of Structure: Duality, Agency, and Transformation." *American Journal of Sociology* 98(1):1–29.
- Sewell William H. Jr, 2005 *Logics of History. Social Theory and Social Transformation*. Chicago: University of Chicago Press.
- Vargas Nicholas. 2015 "Latina/o Whitening? Which Latina/os Self Classify as White and Report Being Perceived as White by Other Americans?" *Du Bois Review* 12(1):119–36.
- West Cornell. 1982 *Prophesy Deliverance! An Afro-American Revolutionary Christianity*. Lexington, KY: Westminster John Knox Press.
- Williamson Joel. 1995 *New People: Miscegenation and Mulattoes in the United States*. Baton Rouge: Louisiana State University Press.
- Zerubavel Eviatar. 1991 *The Fine Line: Making Distinctions in Everyday Life*. New York: Free Press.

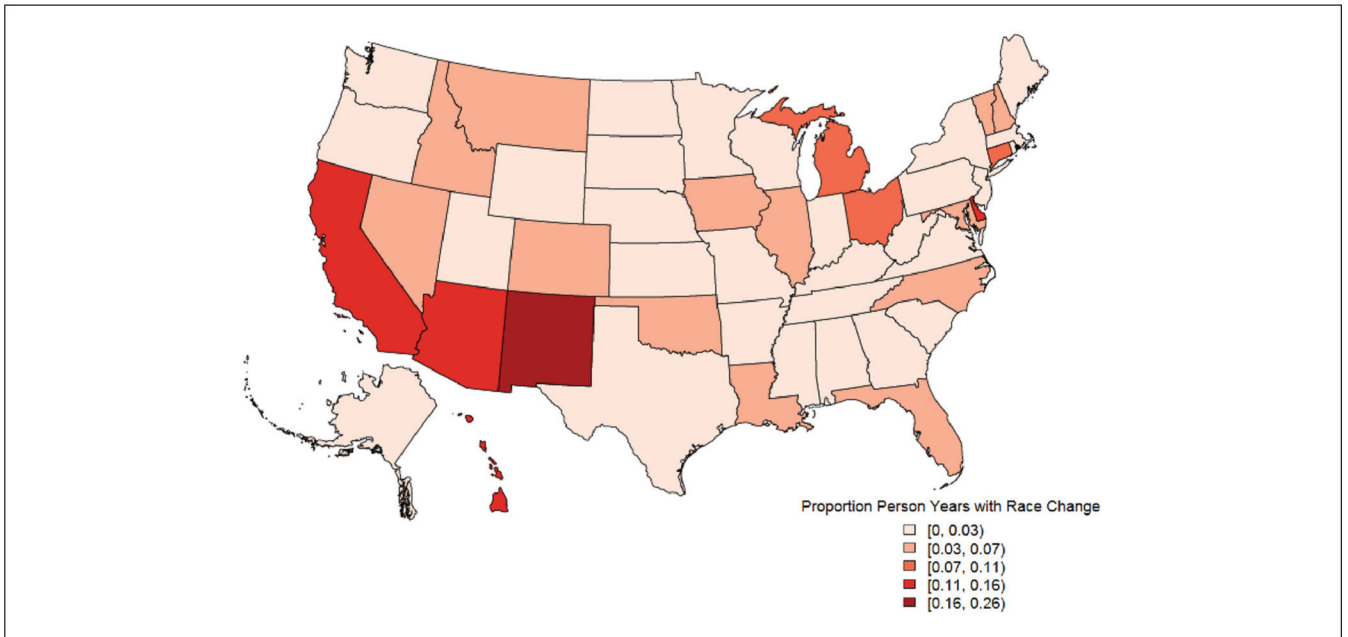


Figure 1. Frequency of racial classification fluidity by state in the 1979 National Longitudinal Survey of Youth.

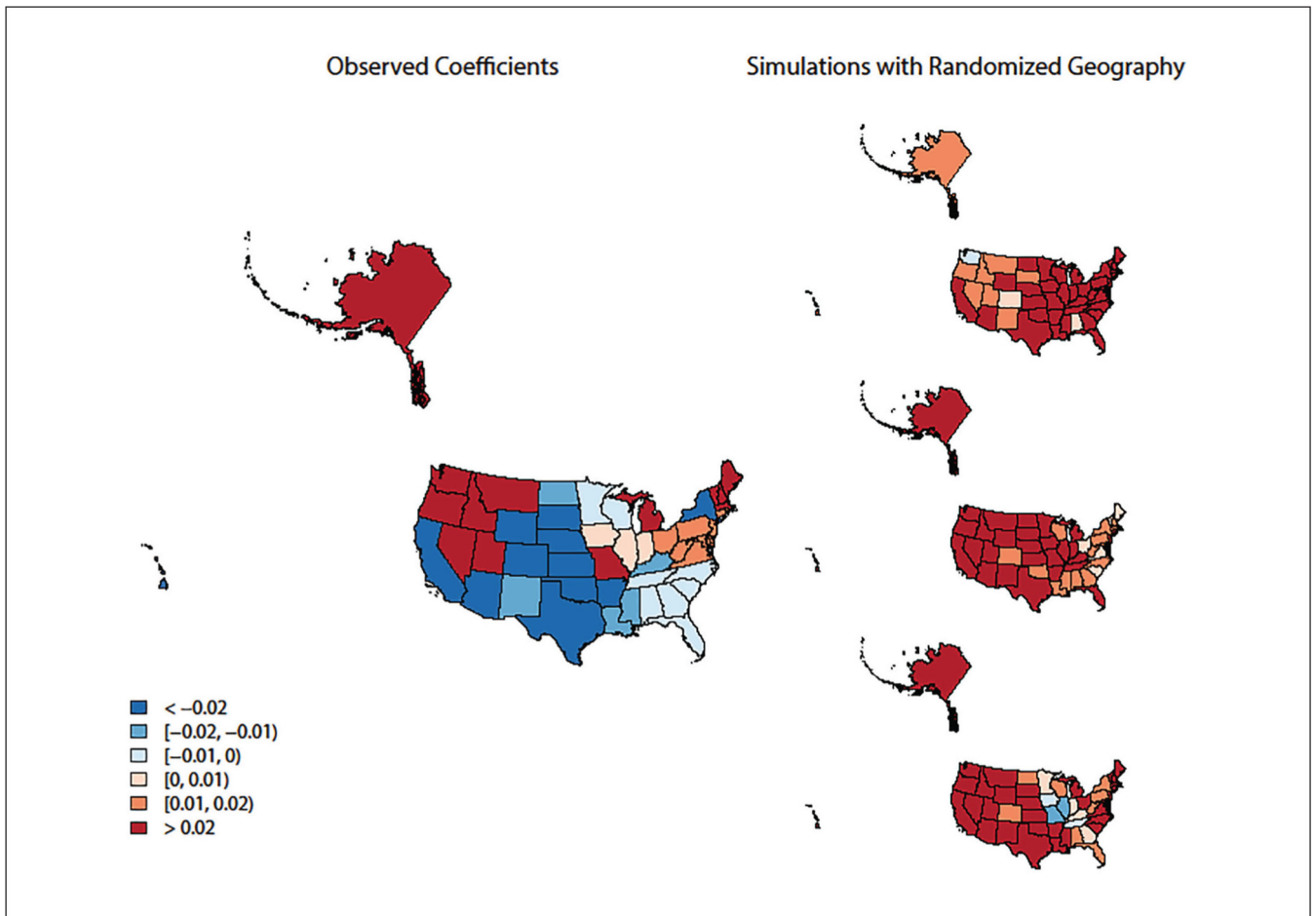


Figure 2. Maps comparing observed county poverty local regression coefficients with Monte Carlo simulations for models predicting classification as “Other.”

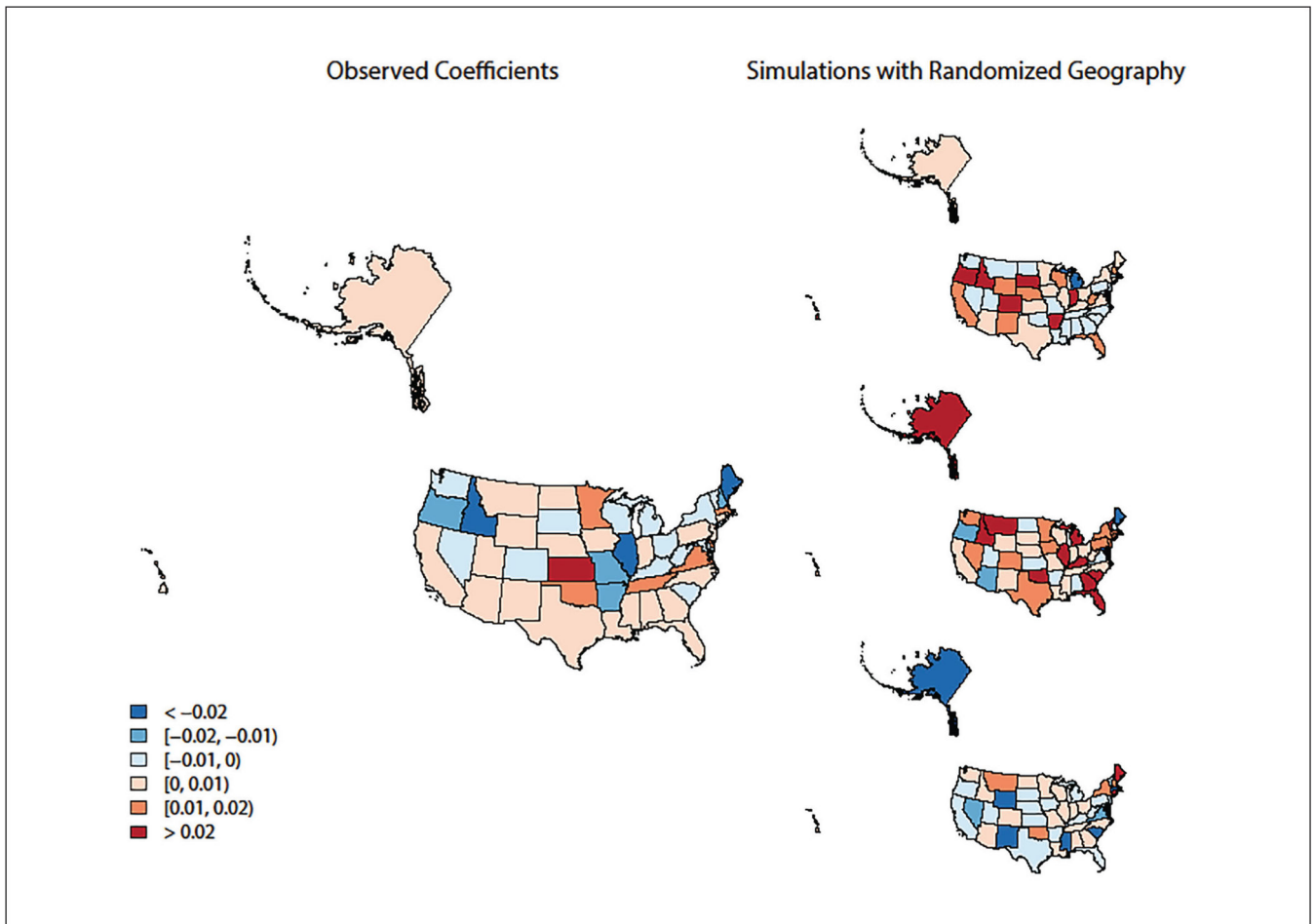


Figure 3. Maps comparing observed ever incarcerated local regression coefficients with Monte Carlo simulations for models predicting classification as “Black.”

Table 1.

County Summary Statistics.

| Contextual Characteristic | Minimum | Mean | Maximum | SD |
|-----------------------------------|---------|-------|---------|-------|
| Unemployment rate | .20 | 7.67 | 24.50 | 3.25 |
| Poverty level | .17 | .62 | 2.07 | .16 |
| Population size (log) | 8.03 | 12.70 | 16.05 | 1.56 |
| Simpson diversity index | .00 | 47.78 | 97.74 | 25.60 |
| Percentage Black residents | .00 | 14.91 | 85.61 | 14.31 |
| Percentage Hispanic residents | .00 | 9.61 | 96.86 | 14.45 |
| Percentage foreign-born residents | .00 | 7.67 | 44.08 | 8.16 |

Source: Restricted-use, geocoded data from the 1979 National Longitudinal Survey of Youth.

Note: $N = 129,177$. County unemployment comes from the Bureau of Labor Statistics; county poverty comes from the Regional Economic Accounts of the Bureau of Economic Analysis; and population size, ethnic and racial composition, percentage foreign born, and ethnoracial diversity come from interpolated decennial censuses.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Linear Probability Regression Models Predicting Classification as “White” and “Other.”

| | Classification as “White” | | | | | Classification as “Other” | | | | |
|-----------------------------------|---------------------------|-------------------|-------------------|--------------------|--------------------|---------------------------|--------------------|--------------------|-------------------|--------------------|
| | National | West | Midwest | Northeast | South | National | West | Midwest | Northeast | South |
| County unemployment rate | -.002*** (.000) | .005*** (.001) | -.000 (.000) | -.005*** (.001) | -.002** (.001) | .002*** (.000) | -.005*** (.001) | .000 (.000) | .005*** (.001) | .002*** (.001) |
| County poverty level | -.029* (.012) | -.067 (.049) | -.005 (.014) | -.025 (.041) | -.006 (.015) | .025* (.011) | .076 (.050) | .011 (.013) | .029 (.039) | -.011 (.014) |
| County population size (log) | -.003 (.002) | .014** (.005) | -.003 (.002) | -.013** (.005) | -.003 (.002) | .003 (.001) | -.013** (.005) | .003 (.002) | .011* (.005) | -.002 (.002) |
| Simpson diversity index | -.001*** (.000) | -.001 (.001) | -.000 (.000) | -.000 (.001) | -.001*** (.000) | .001*** (.000) | .001 (.007) | .000 (.000) | .001 (.001) | .001*** (.000) |
| Percentage Black in county | .000** (.000) | .001 (.002) | .000 (.000) | -.001 (.001) | .001*** (.000) | -.001** (.000) | -.001 (.002) | -.000 (.000) | .000 (.001) | -.001*** (.000) |
| Percentage Hispanic in county | .003*** (.000) | .002* (.001) | .002* (.001) | .000 (.001) | .004*** (.001) | -.003*** (.000) | -.002* (.001) | -.002* (.001) | .000 (.001) | -.004*** (.001) |
| Percentage foreign born in county | -.001*** (.000) | -.001 (.001) | .004*** (.001) | .002 (.001) | -.001 (.001) | .001 (.000) | .001 (.001) | -.003*** (.001) | -.003* (.001) | .001 (.001) |
| Respondent ever unemployed | -.003 (.003) | -.008 (.012) | -.005 (.003) | -.001 (.008) | -.002 (.003) | .002 (.006) | .007 (.012) | .004 (.003) | -.001 (.008) | .001 (.003) |
| Respondent ever impoverished | .001 (.003) | -.001 (.011) | .001 (.004) | .003 (.008) | .003 (.004) | -.001 (.003) | .002 (.011) | .001 (.004) | -.001 (.007) | -.004 (.003) |
| Respondent ever incarcerated | -.011 (.007) | .007 (.019) | -.011 (.008) | -.011 (.021) | -.010 (.006) | .006 (.007) | -.011 (.021) | .009* (.004) | .016 (.020) | .004 (.006) |
| Respondent fixed effects | × | × | × | × | × | × | × | × | × | × |
| Fraction due to μ_i | .85 | .63 | .92 | .84 | .91 | .48 | .34 | .55 | .50 | .51 |

Note: N = 129,177 (person-years) and 11,899 (respondents) for the national model, 24,613 (person-years) and 2,859 (respondents) for the West, 32,396 (person-years) and 3,284 (respondents) for the Midwest, 22,968 (person-years) and 2,587 (respondents) for the Northeast, and 49,200 (person-years) and 5,245 (respondents) for the South. Values in parentheses are standard errors. The statistically significant estimate for the share of “Black” residents in a county predicting classification as “White” is 0.0004, which rounds to 0 at three decimal places. All models also control for respondent age, interviewer characteristics, and year fixed effects (not shown).

* p < .05.

.100' < d

.10' < d
**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Linear Probability Regression Models Predicting Classification as “Black.”

| | National | West | Midwest | Northeast | South |
|-----------------------------------|-----------------|-----------------|-----------------|------------------|------------------|
| County unemployment rate | .000 (.000) | .000 (.000) | .000 (.000) | .000 (.001) | .000 (.000) |
| County poverty level | .004 (.005) | -.009 (.010) | -.006 (.008) | -.004 (.020) | .017** (.006) |
| County population size (log) | .000 (.001) | -.001 (.001) | .000 (.001) | .002 (.002) | .002 (.001) |
| Simpson diversity index | .000 (.000) | .000 (.000) | .000 (.000) | -.001* (.000) | .000 (.000) |
| Percentage Black in county | .000 (.000) | -.000 (.000) | -.000 (.000) | .001* (.000) | -.000 (.000) |
| Percentage Hispanic in county | -.000 (.000) | -.000 (.000) | -.000 (.000) | -.000 (.001) | -.000 (.000) |
| Percentage foreign born in county | .000 (.000) | .000 (.000) | -.001 (.001) | .001 (.000) | -.000 (.000) |
| Respondent ever unemployed | .001 (.001) | .001 (.002) | .001 (.001) | .002 (.004) | .001 (.002) |
| Respondent ever impoverished | .000 (.001) | -.001 (.002) | -.001 (.002) | -.002 (.003) | .001 (.002) |
| Respondent ever incarcerated | .004 (.003) | .003 (.009) | .003 (.007) | -.006 (.007) | .006* (.003) |
| Respondent fixed effects | × | × | × | × | × |
| Fraction due to μ_i | .98 | .97 | .98 | .95 | .98 |

Note: N = 129,177 (person-years) and 11,899 (respondents) for the national model, 24,613 (person-years) and 2,859 (respondents) for the West, 32,396 (person-years) and 3,284 (respondents) for the Midwest, 22,968 (person-years) and 2,587 (respondents) for the Northeast, and 49,200 (person-years) and 5,245 (respondents) for the South. Values in parentheses are standard errors. All models also control for respondent age, interviewer characteristics, and year fixed effects (not shown).

* p < .05.

** p < .01.

Table 4.

Evaluating Spatial Variation in Regression Coefficients with Monte Carlo Simulation.

| | Classification as "White" | | Classification as "Other" | | Classification as "Black" | |
|-----------------------------------|---------------------------|---------|---------------------------|---------|---------------------------|---------|
| | Rank | p Value | Rank | p Value | Rank | p Value |
| County unemployment rate | 1837 | .210 | 2,051 ^{***} | .001 | 442 | .431 |
| County poverty level | 1849 | .198 | 2051 ^{***} | .001 | 320 | .312 |
| County population size (log) | 1608 | .433 | 2051 ^{***} | .001 | 1,163 | .867 |
| Simpson diversity index | 1654 | .388 | 2,011 [*] | .040 | 215 | .210 |
| Percentage Black in county | 1957 | .093 | 2,040 [*] | .013 | 1,826 | .220 |
| Percentage Hispanic in county | 1211 | .820 | 2,051 ^{***} | .001 | 1,835 | .212 |
| Percentage foreign born in county | 1823 | .223 | 2,051 ^{***} | .001 | 1,811 | .210 |
| Respondent ever unemployed | 150 | .146 | 1,064 | .963 | 1 ^{***} | .001 |
| Respondent ever impoverished | 244 | .238 | 1,458 | .579 | 1 ^{***} | .001 |
| Respondent ever incarcerated | 732 | .714 | 1,341 | .693 | 11 [*] | .011 |

Note: Higher ranks suggest greater variation in a given predictor across space. The minimum *p* value with 2,051 simulations is .00098. It is therefore possible that the true empirical *p* value is smaller than what is listed here in some cases. Additional simulations would be needed to achieve higher precision, but this additional precision would not be substantively meaningful.

* *p* < .05.

*** *p* < .001.