

# UC Irvine

## UC Irvine Previously Published Works

### Title

The Endogeneity of Race: Black Racial Identification and Men's Earnings in Mexico

### Permalink

<https://escholarship.org/uc/item/1km0v3sb>

### Journal

Social Forces, 98(4)

### ISSN

0037-7732

### Authors

Villarreal, Andrés

Bailey, Stanley R

### Publication Date

2020-06-12

### DOI

10.1093/sf/soz096

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Published in final edited form as:

*Soc Forces*. 2020 June ; 98(4): 1744–1772. doi:10.1093/sf/soz096.

## The Endogeneity of Race: Black Racial Identification and Men's Earnings in Mexico

Andrés Villarreal\*

University of Maryland-College Park

Stanley R. Bailey

University of California, Irvine

### Abstract

A growing body of sociological research has shown that racial identification is not only fluid, but crucially depends on other individual- and societal-level factors. When such factors are also associated with socioeconomic outcomes such as earnings, estimates of the disadvantage experienced by individuals because of how they identify racially obtained from standard regression models may be biased. We illustrate this potential bias using data from a large-scale survey conducted by the Mexican census bureau. This survey is the first by the government agency since the country's independence to include a question on black identification. We find evidence of a substantial bias in estimates of racial disadvantage. Results from our initial models treating racial self-identification as an exogenous predictor indicate that black men have higher earnings than non-black men. However, when we use an instrumental variables model that treats racial self-identification as endogenous, that is, as a function of the same unobserved characteristics as individuals' earnings, we find a significant *negative* effect of black identification on earnings. While previous studies have acknowledged the endogeneity of race, ours is the first to explicitly model racial self-identification as an endogenous predictor to obtain an unbiased estimate of its effect on individuals' socioeconomic conditions.

---

Governments throughout the world routinely classify individuals into racial categories using national censuses and specialized surveys. A prime motivation for doing so is to monitor the socioeconomic wellbeing of vulnerable populations (Loveman 2014; Morning 2008). Researchers rely on individuals' classification according to census racial categories to estimate differences in socioeconomic outcomes, such as educational attainment, occupational status and income, that are frequently taken as evidence of the disadvantage faced by minority populations. This analytical approach implicitly assumes that individuals' racial identification is fixed, and that it can be treated as an exogenous predictor of their life chances (see Stewart and Sewell 2011). Yet a growing body of work demonstrates that how a

---

\*Direct correspondence to Andrés Villarreal, Department of Sociology, 3131 Art-Sociology, University of Maryland, College Park, MD 20742, avilla4@umd.edu.

About the Authors

Andrés Villarreal is a professor of sociology at the University of Maryland-College Park and a faculty associate at the Maryland Population Research Center. His current research focuses on international migration, race and ethnicity, and social stratification. Stanley R. Bailey is a professor of sociology at the University of California, Irvine. His research focuses on ethnoracial attitudes, politics, demography, and stratification, principally in Latin America, as well as immigration and the sociology of religion.

person identifies racially can be fluid, and that it can crucially depend on other individual- and societal-level factors, some of which are also tied to socioeconomic outcomes (Eschbach, Supple and Snipp 1998; Liebler et al. 2017; Snipp 1986). Research in the United States, for example, reveals that individuals with lower earnings potential, such as those who have been previously incarcerated or unemployed, are more likely to identify themselves and be identified by others as black (Saperstein and Penner 2010, 2012). Selective identification of this type may increase estimates of the earnings disadvantage associated with black racial identification.

In this paper we examine the consequences of selective racial self-identification for the estimation of the racial disadvantage of the Afro-descent or black population in Mexico. For most of the twentieth century the contribution of individuals of African descent to Mexican history and to the racial composition of Mexican society were effectively hidden by a homogenizing racial ideology of *mestizaje* (Jones 2013; Lewis 2012, 2016; Sue 2013). This state-sponsored ideology held the *mestizo*, an individual of Spanish and indigenous descent, as the embodiment of the nation, and simply obviated African ancestry. However, under pressure from domestic activists and international organizations to more fully recognize the country's racial diversity and address existing racial inequality, the Mexican agency in charge of the national census included a question designed to identify individuals of African descent in 2015 (Loveman 2014; Velázquez and Iturralde 2016). This new racial self-identification question provides the first reliable estimates of the size and socioeconomic conditions of the Mexican black population, which has been historically disadvantaged (CONAPRED 2006; Velázquez and Iturralde 2012).<sup>1</sup>

However, because the black racial category is new to the Mexican census, it may be particularly susceptible to selection effects. Contrary to the experience in the U.S. noted above, our analysis suggests that the selective identification of individuals with higher earnings potential as black in Mexico leads to an *underestimation* rather than an overestimation of the disadvantage of this minority population. Because black identification is highly stigmatized, and to some extent foreign to the Mexican population (Sue 2013), individuals who possess higher levels of education and are perhaps more familiar with a societal trend toward multiculturalism may be more likely to identify as Afro-Mexican (see Eschbach et al. 1998; Telles and Paschel 2014; Villarreal 2014). Since these same individuals also have higher earnings potential, estimates of racial earnings disparities are likely to understate racial disadvantage in Mexico.

In statistical terms, the selective identification of individuals with higher or lower earnings potential as members of a racial minority means that their racial self-identification must be treated as an endogenous predictor in an earnings equation (see Angrist and Pischke 2009). Failing to account for endogeneity where there is evidence of selective identification based on individual characteristics that also affect socioeconomic outcomes will lead to biased estimates of the disadvantage experienced by individuals as a result of their minority status. Yet researchers for the most part continue to treat individuals' racial self-identification as an

---

<sup>1</sup>Following the wording of the survey question, we use the terms black (*negro/a*), Afro-Mexican, and Afro-descendant interchangeably, unless otherwise noted.

exogenous predictor. The few studies that acknowledge the endogeneity of racial identification limit their analysis to demonstrating that it is contingent on factors associated with individuals' life chances without specifically measuring the contribution of such selectivity to estimates of racial disadvantage (Bailey, Loveman, and Muniz 2013; Duncan and Trejo 2011; Eschbach et al. 1998; Liebler et al. 2017; Saperstein and Penner 2010, 2012; Telles and Paschel 2014).

Obtaining an unbiased estimate of the effect of racial self-identification on socioeconomic outcomes is difficult. A common approach to handling endogeneity bias is to use an experimental design in which the variable of interest, in this case an individual's racial identification, is randomly assigned (e.g., Bertrand and Mullainathan 2004; Pager 2003). When random assignment is not feasible, researchers may rely on quasi-experimental techniques such as instrumental variables modeling. This technique requires finding a variable that is associated with how an individual identifies racially, but is not directly associated with his or her socioeconomic status (Angrist and Pischke 2009; Wooldridge 2013). We use this technique to produce an unbiased estimate of the effect of racial self-identification on individuals' earnings in Mexico. We identify the model using as an instrumental variable the places where the Mexican government carried out a campaign specifically intended to increase Afro-Mexican self-identification in anticipation of the 2015 national survey.

## Contingent Racial Identification

Most sociologists today agree that race is socially constructed (American Sociological Association 2003). Although essentialist assumptions have lost some traction in recent years, many continue to be reflected in contemporary research (see Morning 2011). In particular, quantitative studies that model racial identification as an *explanans*, or causal mechanism of socioeconomic differences, implicitly assume that individuals' racial identification is stable over time and is independent of other processes of social stratification (on this point see Saperstein and Penner 2012, Sen and Wasow 2016, and Stewart and Sewell 2011).

However, an analytical approach to racial identification as an *explanandum*, or an outcome to be explained, is now core to contemporary constructivist perspectives (Telles and Paschel 2014; Wimmer 2009). The very exercise of modeling racial identification on a set of predictors overtly assumes contingency, not stability. Longitudinal perspectives frame the outcomes of racial contingency as identity change (Eschbach et al. 1998), boundary crossing (Loveman and Muniz 2007), identity switching (Eschbach and Gomez 1998), or response change (Liebler et al. 2017). Research using cross-sectional data focuses simply on correlates or predictors of self-reported race (Bailey and Telles 2006).

Analytical approaches to racial identification as *explanandum* entail the estimation of "propensities" to self-classify in one racial category compared to another (Saperstein and Penner 2012), and imply that some level of choice occurs at the micro-level (Eschbach et al. 1998). Earlier framings on identification contingency—symbolic, situational, optional, and subjective ethnicity (Gans 1979; Waters 1990; Alba 1990)—reflected relatively

unconstrained contexts, i.e., situations where boundaries display low salience, may be weak or dispersed, and/or mostly reflect individual interests. Although these studies focused on the ethnic identification of non-Hispanic whites, the constructivist principle of contingent identification applies to racial identification as well (Eschbach et al. 1998; Saperstein and Penner 2010; Telles and Paschel 2014; Liebler et al. 2017), i.e., where boundaries may demonstrate high salience and involve considerable costs resulting from processes of social closure (Wacquant 1997).

Snipp (1986) and Eschbach and colleagues (Eschbach et al. 1998; Eschbach and Gomez 1998) were early quantitative analysts of racial identification as an *explanandum*. They sought to account for large increases in American Indian identification in the 1980 and 1990 censuses compared to 1970; to do so, they delineated a population of “new” American Indians compared to those individuals who identified as American Indian in 1970. Their analytical approach revealed racial identification shifts associated with education and geographic dispersion, which they attribute in part to “increased variation in the meaning and function of Indian identity” after 1970. Importantly, the authors do not frame racial identification change as measurement error or inappropriate reporting of identity, but as an outcome or *explanandum* whose determinants the field is tasked with documenting and which policy need consider.

More recently, Liebler and her colleagues (Liebler et al. 2017) further examined American Indian identification and extended the analysis of “response change” across time to all other racial populations in the U.S. They leveraged unique linked, individual-level data from the 2000 and 2010 U.S. censuses. Among other things, their results 1) confirm earlier work on American Indian identity switching; 2) document the importance of age and region as principal determinants of response change; and 3) show that all racial and ethnic populations in the U.S. are affected by varying rates of response change. For example, whereas non-Hispanics’ white identification remained relatively stable (about a 3% response change rate between 2000 and 2010), Hispanics were as likely to change racial identification responses between 2000 and 2010 as not to do so (i.e., 50% change rate).

## The Endogenous Effect of Race

On the whole, this body of research leaves little doubt regarding the core constructivist assumption that individuals’ racial self-identification can be fluid and dependent on multiple individual and societal-level factors. Such contingency can present significant problems when racial self-identification is used as an *explanans*, that is, when causal claims are made about the disadvantage faced by individuals as a consequence of how they identify racially. For example, if individuals with higher educational attainment have become more likely to identify as American Indian (Eschbach et al. 1998), concomitant improvements in the income level of American Indians may result in the misleading impression that the historical disadvantage experienced by individuals as a consequence of identifying as American Indian has declined.

Saperstein and Penner’s (2010, 2012) research on racial fluidity and inequality illustrates a situation in which selective racial identification has the opposite effect on estimates of racial

disadvantage as Eschbach et al.'s (1998) study of American Indians. Examining changes in individuals' racial classification across waves of the National Longitudinal Survey of Youth (NLSY), they find that some respondents not only change their racial identification with surprising frequency, but they do so in response to factors associated with their social position. Specifically, individuals who had lost their jobs and those who had been incarcerated were more likely to identify as black in subsequent waves. Because both unemployment and prior incarceration reduce individuals' earnings (Western 2002; Wakefield and Uggen 2010), such changes in racial classification may lead to an *overestimation* of the negative effect of black identification on individuals' income.

Studies examining racial fluidity using a multigenerational approach also demonstrate how differences in the propensity to identify in a minority ethnoracial category may affect estimates of socioeconomic disadvantage. Duncan and Trejo (2011) document a process of ethnic attrition by which children of more educated parents of Mexican origin are less likely to identify as Mexican-American. The loss of Mexican-American identification can be largely explained by the higher intermarriage rates between more educated Mexican-Americans and non-Mexicans. The authors argue that such selective filtering of Mexican-American identification across generations may understate the socioeconomic advancement of this population because those individuals with higher likelihood of success are being systematically lost as members of this ethnoracial category. Similarly, in his analysis of the intergenerational transmission of indigenous ethnic identification in Mexico, Villarreal (2014) finds that children of indigenous women with higher educational attainment are more likely to be identified as indigenous. Because children of more educated mothers will in turn have better educational outcomes, selective ethnic identification in this case leads to an underestimation of the educational disadvantage experienced by indigenous children.

Perhaps nowhere is the need to account for the endogeneity of racial self-identification more pressing than in the Latin American context. An extensive body of research has documented a dynamic of racial "whitening" in countries throughout the region (e.g., Schwarzman 2007; Loveman and Muniz 2007; Telles and Paschel 2014). The movement of individuals from darker to lighter categories in local racial schemas over time and across generations is thought to be closely tied to socioeconomic factors. Schwartzman (2007), for example, finds that greater parental education leads to higher odds that children will be classified as white in Brazil. Notwithstanding the historical prevalence of whitening, research on contemporary racial identification in the region also suggests a relatively new dynamic: racial darkening, or movement of individuals from lighter to darker racial categories (Telles and Paschel 2014; Francis-Tan and Tanurri-Pianto 2015). Telles and Paschel (2014) find that educational attainment had a darkening effect on individuals' own racial identification in the Dominican Republic. This research clearly underscores the importance of accounting for selective racial self-identification in estimates of racial disparities.

## Endogeneity Bias and Instrumental Variables Models

As discussed in the preceding section, the selective identification of individuals with higher or lower earnings potential as racial minorities may result in an under- or over-estimation of the disadvantage associated with minority identification. In statistical terms, this means that

estimates of the causal effect of racial self-identification—the penalty paid by individuals because of how they identify racially—may be biased in regression models that ignore differences in the propensity to identify in a racial category based on characteristics that also affect socioeconomic outcomes. In this section we describe how an Instrumental Variables (IV) approach may be used to overcome this source of bias in models of racial earnings disadvantage. We begin by formally defining endogeneity bias as a form of omitted variable bias (Angrist and Krueger 2001: 72-73), and derive expressions for the direction and magnitude of the bias.

The simplest possible model of an individual's earnings as a function of his or her racial identification may be expressed as follows:

$$y_i = \alpha + \beta black_i + \varepsilon_i \quad (1)$$

where  $y_i$  is individual  $i$ 's logged earnings,  $black_i$  is a binary variable indicating his or her racial identification, and  $\beta$  is a coefficient that is meant to capture the effect of racial identification on earnings.<sup>2</sup> A key principle of regression analysis is that the estimate of the effect of a predictor will be biased if the predictor is correlated with the error term  $\varepsilon_i$  (Wooldridge 2013). This will occur if there are any variables omitted from our model that are correlated with both an individual's racial identification and his or her earnings. For example, as discussed in the previous section, Saperstein and Penner (2010, 2012) find that individuals who have been previously incarcerated are more likely to identify as black, and prior research has also shown that incarceration experience has a detrimental effect on earnings (Western 2002; Wakefield and Uggen 2010). Excluding incarceration from the regression model for earnings would therefore lead to a biased estimate of the earnings disadvantage associated with black identification in the United States.

If information regarding individuals' prior incarceration experience were available, we could obtain an estimate of the effect of black identification that is unbiased by prior incarceration experience by testing the following model:

$$y_i = \alpha + \rho black_i + \gamma incarceration_i + v_i \quad (2)$$

where  $incarceration_i$  is a binary variable indicating whether an individual  $i$  has been previously incarcerated, and  $\gamma$  is its corresponding coefficient. Standard path analysis indicates that the coefficient for black identification ( $\beta$ ) in equation (1) may be expressed as the sum of a direct and an indirect effect of black identification:

$$\beta = \rho + \gamma\eta \quad (3)$$

where  $\eta$  is a regression coefficient obtained by regressing incarceration on black identification, and  $\rho$  is the unbiased coefficient in equation (2).<sup>3</sup> The estimate of the earnings

<sup>2</sup>This model assumes a binary choice in racial identification. This assumption simplifies the presentation, and it reflects the conditions created by the introduction of the black identification question in Mexico as described below.

<sup>3</sup>For simplicity, this discussion assumes a linear model for binary outcomes.

disadvantage of black identification net of incarceration experience would therefore be biased by  $\gamma\eta$  if we omitted prior incarceration as a predictor in the earnings equation.

This calculation further allows us to predict the expected direction of the bias in our estimate of the effect of black identification when incarceration is omitted as a predictor of earnings. Because prior incarceration has a negative effect on earnings,  $\gamma$ , and incarceration has been shown to have a positive effect on black identification,  $\eta$ , we may expect a *downward* bias in the estimate of the effect of black identification on earnings based on the model expressed in equation (1). This situation is depicted in the path diagram in Figure 1a. In general, when either  $\gamma$  or  $\eta$  is negative and the other is positive we may expect a downward bias (the disadvantage of black identification is smaller than our estimate of  $\beta$ ). When  $\gamma$  and  $\eta$  are either both positive or both negative, we may expect an upward bias in our estimate (the disadvantage of black identification is larger than our estimate of  $\beta$ ). As shown in Figure 1b, a positive bias may occur if individuals' level of education is omitted from the model, and educational attainment is associated with both a greater likelihood of identifying as black and higher earnings.<sup>4</sup>

At root the problem of endogeneity bias in the context of racial stratification is therefore a form of omitted variable bias. When important variables affecting both racial identification and socioeconomic outcomes are omitted as predictors, we will obtain estimates that do not truly reflect the causal effect of black identification, that is, the disadvantage experienced by individuals as a result of their racial identification. While it may be possible to directly control for some factors that affect both individuals' racial identification and their socioeconomic outcomes, for example by including prior incarceration as a predictor if it is available, there are likely to be many other unobservable factors that are difficult or impossible to measure. Hence the need for a more general methodological strategy.

One way to overcome omitted variable bias is to use an experimental design where the variable of interest, namely individuals' racial identification, is assigned randomly. This approach has been used, for example, to estimate the disadvantage faced by black men and women in the U.S. labor market. Bertrand and Mullainathan (2004) conducted an audit study whereby nearly identical resumes were sent in response to job-wanted ads in which the name of the job applicant was randomly assigned a "white-sounding" name (e.g., Greg Baker) or an "African-American-sounding" name (e.g., Jamal Jones). The name of the applicant was the only signal of his or her racial identification. Bertrand and Mullainathan found that applications using African-American-sounding names received fewer callbacks. By randomly assigning a black racial identification to a set of fictitious applicants, such an experiment ensures that the estimate of the disadvantage of a black racial identification is not affected by any other individual characteristics that may otherwise be omitted from a statistical model of the difference in the likelihood of employment for individuals identified as black and white.<sup>5</sup>

<sup>4</sup>As shown below this is indeed the case in Mexico. Our analysis demonstrates that educational attainment increases the likelihood of identifying as black.

<sup>5</sup>See Pager (2003) and Pager, Western and Bonikowski (2009) for examples of similar experiments.



In many cases the random assignment of the variable of interest, and in particular individuals' racial identification, is not feasible or practical. In such cases, researchers may use an instrumental variables approach to obtain estimates that are not biased by omitted variables (Angrist and Krueger 2001; Angrist and Pischke 2009; Wooldridge 2013). The approach relies on finding another predictor, called an instrumental variable, that is correlated with individuals' racial identification, but otherwise unrelated to his or her earnings. The IV approach uses the variation induced in the endogenous predictor by the instrumental variable to obtain an unbiased estimate of its effect. As Angrist and Krueger (2001: 77) note, when the endogenous variable of interest is binary, "instrumental variables methods estimate causal effects for those whose behavior would be changed by the instrument if it were assigned in a randomized trial."<sup>6</sup> In the analysis below we will use as an instrument a coordinated campaign by Mexican government agencies specifically intended to increase Afro-Mexican self-identification in the lead-up up to the inter-decennial survey (see details below).

The IV model may be thought of as consisting of two stages: the first stage models individuals' racial identification as a function of a set of predictors that includes the instrumental variable, while the second stage models earnings as a function of racial identification and the same set of predictors excluding the instrumental variable. The fact that the instrumental variable is uncorrelated with earnings means that it can be excluded from the earnings equation (called the exclusion restriction), and that its entire effect on earnings is a result of its effect on individuals' odds of identifying as black. Notably, by using the IV model we obtain an unbiased estimate of the causal effect of identifying as black without having information regarding any omitted variable affecting both the probability of identifying as black and earnings. As Angrist and Krueger (2001: 73) note, "instrumental variables methods allow us to estimate the coefficient of interest consistently and free of asymptotic bias from omitted variables, *without actually having data on the omitted variables or even knowing what they are*" [emphasis added].

In the language of experimental studies, IV models allow us to isolate the causal effect of applying a "treatment" or intervention as if it were randomly assigned to individuals. The estimated coefficient for black identification in the second-stage equation captures the change in earnings we would observe in an imaginary experiment in which a random individual is "treated" with a black identification, that is, when his or her racial identification is changed from non-black to black, but everything else remains the same. We are, of course, not actually changing how individuals' identify racially, which would be inconsistent with a constructivist framework. Our attempt to isolate the effect of racial identification free of endogeneity bias due to omitted variables should not be confused with a search for individuals' "true" racial identification that is somehow different from that stated in their own responses to the survey question.<sup>7</sup>

---

<sup>6</sup>The instrumental variables model produces an estimate of what is called the Local Average Treatment Effect (LATE). To generalize the results to the entire population involves the assumption that everyone in the population has a similar response to the treatment. In the analysis below we are only able to estimate the effect of identifying as Afro-Mexican for those who may be persuaded to identify as such by the government campaign.

<sup>7</sup>Searching for a "true" racial identification would imply that individuals' declared racial identification contains measurement error. See our discussion below contrasting our use of an IV model to address endogeneity bias with the use of such models to correct for the effect of measurement error.

## The Afro-descent Population in Mexico

Mexico is a strategic context in which to investigate the endogenous effect of racial identification because both the newness of the Afro-Mexican category and the historical stigmatization of a black identity make strong selection effects more likely. Moreover, as we address below, the recent campaign by Mexican governmental agencies to promote black self-identification provides a plausible instrument with which to identify an endogenous model.

The African-descent population in Mexico is one of the oldest in the continent. In the mid-seventeenth century, the Kingdom of New Spain, which encompassed the territory of contemporary Mexico, was home to “the second-largest population of enslaved Africans and the greatest number of free blacks in the Americas” (Bennett 2003: 1). Slaves worked predominantly in agriculture and mining performing some of the most arduous tasks; however, they were also present in large urban centers, such as the nation’s capital (Carroll 2001; Martinez Montiel 1994). The practice of slavery declined following the disruption of the Portuguese slave trade to Spanish America after 1640, and with the recuperation of the indigenous population from the demographic disaster brought on by warfare and disease (Carroll 2001). Slavery was formally abolished in Mexico in 1824.

Despite the historical presence of a significant population of Afro-descendants, a racial ideology actively promoted by the Mexican state in the twentieth century effectively erased the contribution of individuals of African descent from national history and from the perceived racial composition of Mexican society (Lewis 2012; Sue 2013). This ideology defined the *mestizo*, an individual of Spanish and indigenous descent, as the embodiment of the nation, omitting any reference to African ancestry. As Vinson (2005: 3) notes, “blacks were literally written out of the national narrative.” The erasure of blacks was reflected in statistics compiled by the Mexican government. While Mexican censuses throughout the twentieth century attempted to identify the indigenous population, primarily through questions on language use, they did not include questions regarding black identification (Loveman 2014). The last available estimates of the size of the black population therefore date to the colonial period. Free blacks have been estimated to constitute approximately 10 percent of the population on the eve of Mexican independence in 1810 (Aguirre Beltrán 1972 [1946]: 223-230; Bennett 2003: 1; Vinson 2009: 101).

Mexico remained one of the few Latin American countries that did not identify individuals of African descent in the decennial censuses of 2010. However, pressure from international agencies and domestic civil society organizations led to the inclusion of a question meant to identify black respondents in the 2015 Intercensus Survey (*Encuesta Intercensal, EIC*). This large-scale survey was designed to replace the customary short-form version of the census conducted at the midpoint of each decade (INEGI 2015). Because the EIC was carried out by the Mexican National Institute of Statistics and Geography (INEGI), the same agency in charge of the census, it carries similar weight and is used for policy planning.

## Government Campaign to Increase Afro-Mexican Self-Identification

Anticipating low levels of Afro-Mexican identification in the Intercensus Survey due to the newness of the racial category and the stigmatization of a black identity in Mexico, the Mexican government launched a campaign specifically intended to promote black self-awareness and self-identification. The campaign, titled “I am Afro. I recognize myself and I count!” (*Soy Afro. ¡Me reconzco y cuento!*), was a joint effort between several government agencies including the National Council to Prevent Discrimination (*Consejo Nacional para Prevenir la Discriminación, CONAPRED*), the Ministry of the Interior (*Secretaría de Gobernación*), the National Institute of Social Development (*Instituto Nacional de Desarrollo Socioal, INDESOL*), and INEGI.<sup>8</sup> The campaign included billboards, posters, flyers and radio spots with standardized messages intended to instill pride in an Afro-Mexican identity and to highlight the importance of identifying as Afro-Mexican.<sup>9</sup> Campaign organizers also feared that the terms used for black racial identification varied regionally which could result in lower positive responses to the race question included in the Intercensus survey. The campaign material therefore used multiple terms to identify Afro-Mexicans and subsumed them under a general category of “Afro.”

The government’s campaign was limited to certain areas of the country thereby providing a natural experiment which we can use to identify our endogenous model. The target population was all Mexican adults in those areas, regardless of their socioeconomic level. Because the campaign may be expected to have increased individuals’ likelihood of identifying as black in targeted areas but not their earnings (i.e., campaign spending levels were not high enough to have a measurable effect on the local economies), a variable identifying the places where the campaign took place is used as an instrumental variable in the statistical analysis below. Importantly, campaign efforts were to some extent focused on areas with high expected concentrations of Afro-Mexicans. However, no other criteria such as the socioeconomic conditions or level of urbanization were used as factors for selecting communities where the campaign was carried out.<sup>10</sup> Thus, once the racial composition is controlled, the presence of the campaign should not be directly associated with individual’s income, thus satisfying the exclusion restriction (see details below).

## Data and Measurements

Our data are derived from the 2015 Mexican Intercensus Survey (*Encuesta Intercensal, EIC*). As noted above, the survey was intended to replace the customary mid-decennial census. The EIC sampled 6.1 million households, accounting for almost 20% of all households nationwide. The extremely large sample size is particularly useful for this study because the Afro-Mexican population is relatively small. Multivariate analysis of this population is therefore not possible using surveys with small sample sizes even if they include information regarding respondents’ racial identification. Another unique feature of the EIC is that it is

<sup>8</sup>Information about the campaign, including examples of the posters and radio spots used are available in the CONAPRED website: [http://www.conapred.org.mx/index.php?contenido=registro\\_encontrado&tipo=2&id=5364](http://www.conapred.org.mx/index.php?contenido=registro_encontrado&tipo=2&id=5364) [Accessed on October 1, 2018].

<sup>9</sup>The campaign also involved social media posts in Facebook and Twitter. Because these posts have a national reach, their effect on Afro-Mexican self-identification is not expected to vary geographically.

<sup>10</sup>Based on internal document and personal communication with CONAPRED staff.

representative at the municipal level for each of Mexico's 2,457 municipalities, and for all cities with more than 50,000 residents (INEGI 2015).<sup>11</sup> This allows us to compute estimates of the Afro-Mexican population in all municipalities nationwide as described below. Sample weights are provided by the INEGI and used throughout the analysis.

The dependent variable in all our regression models is the logged total monthly earnings from work as reported in the EIC. Earnings are recorded as the actual amount of pesos received rather than income intervals and therefore require no further transformation. Our key predictor is an individual's racial self-identification. The EIC questionnaire asks whether a household member considers him- or herself "black, that is, Afro-Mexican or Afro-descendant" according to his or her "culture, history and traditions." The survey's emphasis on cultural rather than phenotypic characteristics to identify the black population follows the practice used for identifying the indigenous population in past censuses. The questionnaire further allows household members to identify as "partly" black. For much of the analysis below we distinguish individuals who identify as black and partly black, and we test differences in socioeconomic outcomes between the two categories. However, because an instrumental variables model including two separate endogenous variables or a single categorical variable with multiple categories would be difficult to estimate, we dichotomize individuals' racial self-identification by grouping together all individuals who consider themselves black or partly black.<sup>12</sup>

An individual's indigenous identification is also included as a predictor in most models. The survey question for indigenous identification is similar to that for Afro-Mexicans. The EIC questionnaire asks whether each household resident considers him- or herself to be indigenous "according to his/her culture." The survey also allows for "partial" indigenous identification. A large research literature has shown the indigenous population in Mexico to be disadvantaged socioeconomically (e.g, INMUJERES 2006; Ramírez 2006). It is particularly important to control for indigeneity in our models because, as described below, a large percentage of survey respondents who identify as Afro-Mexican also identify as indigenous.

Following a standard practice in most censuses worldwide, a single household informant may answer the survey questions for all household members in the EIC, including the questions related to their ethnic and racial self-identification. The EIC therefore generates proxy measures of individuals' ethnic and racial self-identification except in those cases where the household informant is the individual in question. To avoid a potential discrepancy between the informant's perception of a particular household member's ethnic and racial self-identification and his or her own self-identification, we restrict our sample to household informants.<sup>13</sup>

---

<sup>11</sup>Municipalities are political and administrative units similar to counties in the U.S., although they are somewhat smaller in size. In 2015 the average municipality had a population of 48,649 residents.

<sup>12</sup>In separate analyses not presented here we tested a model in which individuals who identified as partly black were grouped together with those who do not identify as black. The results were consistent with those reported below.

<sup>13</sup>Household informants are not identified in the publicly-released version of the EIC data. We obtained this information by special request to INEGI.

Because our regression models examine individuals' earnings, we further limit our sample to working-age men (ages 18 to 55). We exclude women from our analysis because their inclusion would require us to account for differences in their selective participation in the labor market, which is difficult to accomplish.<sup>14</sup> We also exclude from our sample individuals with very low earnings (less than one fourth the minimum wage) because they are likely to be only weakly attached to the labor market (Kopczuk, Saez and Song 2010). Finally, we restrict the sample to native-born Mexicans to avoid including black foreign expatriates from countries such as the U.S. who are likely to have higher earnings.

Following standard human capital theory, we include men's years of education and age as predictors of their earnings. In addition, because there are significant differences in the cost of living between regions of Mexico and across communities of different levels of urbanization, we also control for the region of residence and the local population size. Mexican states are grouped into five regions according to the Mexican National Institute of Statistics and Geography (INEGI 2009). The southern region, which generally has the lowest income levels and includes a higher number of black and indigenous residents, is used as the baseline category.

### Endogenous Models

One of the key contributions of our study is to explicitly model individuals' racial self-identification as an endogenous predictor of earnings using a quasi-experimental design. As discussed in a previous section, individuals' probability of identifying as black may be a function of the same unobserved characteristics that affect their earnings. If this is the case, then our estimates of the disadvantage of identifying as black in Mexico will be biased. To address this possibility, we test models in which black identification is treated as an endogenous predictor. Because our endogenous predictor is binary, we use a non-linear specification for the first stage equation. A linear equation is used in the second stage predicting individuals' logged earnings.

We identify our primary model using as an instrument a variable indicating whether an individual lives in a state where the Mexican government carried out its campaign to increase the self-identification of Afro-Mexicans in anticipation of the survey.<sup>15</sup> If the campaign was successful, individuals living in those states will be more likely to identify themselves as black, and yet the campaign should have no measurable effect on individuals' earnings as discussed earlier.

Because the government agencies in charge of the campaign targeted areas of the country thought to have a larger presence of Afro-Mexicans, we control for the racial composition of the municipal population. Controlling for the relative size of the black population is also warranted because previous studies in other national contexts have demonstrated that the

<sup>14</sup>We nevertheless replicated our models for Mexican women instead of men. The results were fully consistent with those presented below.

<sup>15</sup>The instrumental variable is defined as binary because the total amount of campaign spending by state is not available. However, given that the campaign involved spending on items that are likely to vary in their efficacy (e.g., billboards, radio spots and flyers), the total amount spent by state may not adequately capture the impact of the campaign on individuals' self-identification even if it were available.

local racial composition may affect both how individuals identify racially (Lee and Bean 2004; Xie and Goyette 1997), and their economic opportunities (Beggs, Villemez and Arnold 1997; Cohen 1998).

To insure that the results of our primary IV model are not sensitive to the instrumental variable used, we also replicate our results using an alternative instrumental variable obtained from a pioneering study of the genetic composition of the Mexican population. Moreno-Estrada, et al. (2014) examined the subcontinental (African, European, and Native American) genomic ancestry of the Mexican population using DNA samples extracted from a wide range of individuals located in 10 Mexican states. They found important differences in the African ancestry of the *mestizo* population across these states, from a low of 1.9% for the state of Campeche to a high of 6.9% in the state of Guerrero.<sup>16</sup> Although not synonymous with racial self-identification, ancestry may nevertheless be expected to affect how individuals identify racially. Specifically, we expect individuals living in states with a higher African admixture to be more likely to identify as Afro-Mexican. Yet African admixture in a state should not directly affect earnings. Because information regarding African admixture is only available for 10 states, the analysis using this alternative instrumental variable is limited to those states.<sup>17</sup> However, one important advantage of this alternative instrument used as a robustness check is that it does not depend on the concentration of black Mexicans, therefore allowing us to exclude the municipal racial composition as a control variable.

## Descriptive Results

Table 1 shows the percentage of the Mexican population that identifies as black or partly black. In order to give a full picture of the racial composition of the country, the sample used to calculate these percentages is not restricted by age, gender, national origin, or household informant status. Nationally 1.16% of all Mexicans identified as black and 0.50% identified as partly black according to their “culture, history and traditions”. While seemingly small, these percentages together represent nearly 2 million Mexican residents. Because this is the first time that the INEGI, the official agency in charge of the national census, asks respondents about their racial identification, it is not possible to validate these findings using previous censuses. However, the percentage of the population identifying as black is consistent with estimates from the AmericasBarometer survey conducted by the Latin American Public Opinion Project (LAPOP). This survey is among the very few to include a question about racial self-identification in Mexico. The question asked by the AmericasBarometer survey includes black and mulatto among the five mutually-exclusive racial categories from which respondents can choose. According to the 2012 survey 0.56% of respondents (8 out of 1,431) identified as black, while 1.19% (17 out of 1,431) identified

<sup>16</sup>The data used in our analysis are extracted from Table S5 in the Supplementary Materials for Moreno-Estrada, et al. (2014) available online (<http://science.sciencemag.org/content/sci/suppl/2014/06/11/344.6189.1280.DC1/Moreno-Estrada.SM.revision.1.pdf>). The variable used is based on their *mestizo* sample.

<sup>17</sup>Both instrumental variables cannot be included in the same model without encountering estimation problems since information on African admixture is only available for 10 states and both variables are defined at the state level, as are the regional dummy variables described above. We lack a sufficient number of states to include both state-level predictors as instruments.

as mulatto. Together these two categories account for 1.75% of all respondents which is comparable to estimates from the EIC.<sup>18</sup>

The geographic distribution of the Afro-Mexican population also helps to validate estimates derived from the EIC. The map in Figure 2 shows the percentage of the municipal population that identifies as either black or partly black. Consistent with historical and ethnographic accounts (e.g., Aguirre Beltrán 1972 [1946]; Jones 2013; Lewis 2000, 2012, 2016; Sue 2013; Vaughn 2004), coastal areas in the southern states of Oaxaca and Guerrero and in the gulf state of Veracruz have high concentrations of Afro-Mexicans (for example, 20.5% of residents of the municipalities belonging to the coastal region known as the *Costa Chica* spanning the states of Oaxaca and Guerrero identify as black or partly black). Perhaps more surprisingly, the Federal District, which forms the central part of the Mexico City metropolitan area, and the surrounding state of Mexico, have the next largest concentrations of black residents, respectively. As discussed in a previous section, Mexico City contained a large slave population during the colonial period. The nation's capital has also received large internal migration flows from various states including those with historical black populations.

Table 2 shows a breakdown of the percentage of black, partly black, and non-black individuals who also identify as indigenous or partly indigenous. A much higher percentage of individuals who identify as black also identify as indigenous (67.5% of blacks, compared to 22% of non-blacks). This finding is consistent with previous studies that suggest historical admixture between Afro-Mexicans and indigenous peoples, many of whom live in close proximity. Moreover, as noted by Lewis (2000, 2016), Afro-Mexicans may also identify as indigenous as a way to preserve their national identity given the erasure of blacks from the national narrative.

Table 3 compares the educational attainment, occupational status and earnings of black and non-black men included in our regression models. The results indicate few statistically significant differences between black, partly black, and non-black men. First, black men have similar average years of education as non-black men. However, black men are significantly less likely to have only a primary education or less, and more likely to have completed high school than their non-black counterparts. Second, a significantly lower percentage of black and partly black men work in agricultural and low-skilled occupations, respectively. Finally, black men receive similar earnings as non-black men. Although not shown in Table 3, the differences between black and non-black men are larger when they are disaggregated according to their indigenous identification. For example, once their indigenous identification is taken into account, black men have significantly higher levels of education, and indigenous black men have significantly higher earnings than their non-black counterparts.

Overall, these descriptive results indicate that Afro-Mexicans have similar levels of educational attainment, occupational status and earnings compared to the rest of the

---

<sup>18</sup>The comparatively small sample of Afro-Mexicans in that survey (a combined total of 25) does not allow for a detailed statistical analysis of their socioeconomic conditions. Data for the AmericasBarometer can be found online at <http://www.vanderbilt.edu/lapop/>.

Mexican population even before controlling for other individual characteristics. This apparent socioeconomic parity of Afro-Mexicans is surprising given the historical stigmatization of black identification in Mexican society (Lewis 2012; Sue 2013), and the disadvantage faced by individuals of darker skin color in Mexico (Bailey, Saperstein, and Penner 2014; Villarreal 2010). In the multivariate analysis below we examine the earnings differences between Mexican black and non-black men in greater detail. We specifically test whether any racial differences in earnings may be explained by other confounding factors, as well as by the selective self-identification as Afro-Mexican of individuals with higher earnings potential.

## Multivariate Results

Table 4 shows the results of the regression models predicting men's logged earnings. Model 1, which only controls for individuals' age and years of education, indicates no statistically significant racial gap in earnings. However, in Model 2 where we control for individuals' indigenous identification, we find that black men have significantly *higher* earnings than non-black men. As noted earlier, because a large percentage of Mexican black men are also indigenous and indigenous men have significantly lower earnings, failing to control for indigenous identification results in an overestimation of the disadvantage of identifying as black. Similarly, when we control for the level of urbanization and the region of the country in which men reside in Model 3, we also find that black Mexican men have significantly higher earnings. The results of Model 4, which controls for both indigenous identification, and the level of urbanization and region of residence indicates that black men's earnings are 7 percent higher than those who are not black. This is comparable to having one more year of education.

The higher earnings of black men remain after controlling for the percentage of the municipal population that identifies as either black or partly black in Model 5. Interestingly, the apparent advantage of identifying as black is not shared by men who identify as only partly black. The earnings of partly black men are similar to those of non-black men in all models shown in Table 4. Model 6 includes an interaction term between men's racial and ethnic identification. This model allows us to test whether the apparent earnings advantage of black men varies depending on whether they also identify as indigenous. The non-significant interaction term indicates that the earnings gap between black and non-black men is the same regardless of their ethnicity, and that men's race and ethnicity may be treated as separate predictors. The final model in Table 4 includes fixed effects for 3-digit occupations (156 occupational categories). The positive and significant coefficient for the black racial category indicates that black men earn more than non-black men even within the same occupation. It also means that the apparently higher earnings of black men cannot be explained by their sorting into occupations that pay more.

## Endogenous Model

Table 5 presents the results of our primary IV model in which black identification is explicitly modeled as an endogenous predictor using the government's campaign as an instrumental variable. The results indicate that identifying as black in Mexico has a



statistically significant *negative* effect on individuals' earnings. The startling reversal of the effect of black identification suggests that the earnings advantage of Afro-Mexicans in earlier models may indeed be attributed to a greater tendency for Mexicans with greater earnings potential to embrace a black identity. This is further confirmed by the positive and significant correlation between the error terms for the equations corresponding to black identification and men's earnings ( $\rho$ ). The positive correlation between the errors means that characteristics omitted from our model increase both individuals' probability of identifying as black and their earnings. The Wald test reported at the bottom of Table 5 indicates that we can reject the null hypothesis that the two equations in the model are independent (see the online Appendix for a full discussion of diagnostic tests).

We cannot, of course, identify the specific omitted variables leading to the biased estimate of the effect of black racial self-identification in our earlier models. However, the direction of the bias and the positive correlation between the error terms in Table 5 allow us to rule out some possibilities. For example, although our models omit individuals' skin color as a predictor because it is not available in the survey, the omission of skin color cannot be the primary factor leading to the observed bias. Because a darker skin color is expected to be positively associated with a black racial self-identification and negatively associated with earnings, omitting skin color should lead to a *negative* bias in our estimate of racial disadvantage. Yet we find a positive bias as well as a positive correlation between the error terms for the first- and second-stage equations in Table 5. A more likely explanation for the observed bias is that individuals with more cultural and human capital are more likely to both embrace a black racial identity and have higher earnings. As Villarreal (2014) has argued with regards to indigenous identification in Mexico, individuals with more human capital are likely to be more aware of the new multicultural message that destigmatizes an ethnoracial minority status. Individuals with more cultural and human capital may also be more familiar with the specific terms used to identify Afro-Mexicans in the Intercensus survey questionnaire.

Our instrumental variable, which captures the places where Mexican government agencies carried out their campaign to promote black identification prior to the survey, is a significant and positive predictor of black identification suggesting that the campaign was effective, and that the variable may in fact serve as a valid instrument.<sup>19</sup> The results of the first-stage equation also indicate that individuals living in municipalities with a high concentration of Afro-Mexicans, those who identify as indigenous, are younger, more educated and live in the Southern and Central regions of the country are significantly more likely to identify as black.

Finally, the results of the IV model using the percent African admixture as an alternative instrumental variable are shown in Table 6. Importantly, the results confirm our finding that a black racial identification has a significant negative effect on individuals' earnings. The percentage of African admixture in the state, which is used as an instrumental variable, has a positive and statistically significant effect. An F-test for the first-stage equation is 103.24,

---

<sup>19</sup>The F-statistics for the first-stage equation is 64.08, well above the standard minimum threshold of 10 for a strong instrument (Staiger and Stock 1997). See the Appendix for further diagnostic tests of the instrumental variables model.

well above the accepted minimum threshold of 10 for a strong instrument (Staiger and Stock 1997). A Wald test also indicates that we can reject the null hypothesis that the two equations are independent.

## Conclusions

A growing body of sociological research has shown that racial identification is not only fluid, but crucially depends on other individual- and societal-level factors. When such factors are also associated with socioeconomic outcomes, estimates of the disadvantage experienced by individuals because of how they identify racially obtained from standard regression models may be biased. The results of our statistical analysis vividly illustrate this potential risk. Our initial models, which ignored differences in the propensity to identify in a racial category based on characteristics that also affect earnings, indicated that black Mexican men were not disadvantaged and in fact earned more than non-black men. However, once racial self-identification was considered a function of the same omitted variables as individuals' earnings, our model showed a significant negative effect of black identification on earnings.

Results from our instrumental variables model provide a better estimate of the causal effect of racial self-identification, by which we mean the disadvantage (or advantage) that individuals experience as a result of identifying in a racial category. For example, an individual's own racial identification in everyday life might lead others to deny him or her a job or a promotion, or prevent him or her from living in a particular neighborhood or attending a good school. Estimates of the socioeconomic gap obtained from ordinary regression models do not accurately reflect the disadvantage experienced by those who identify in a minority racial category because they conflate socioeconomic differences in who chooses to identify as a racial minority, and how much individuals who identify in that racial category suffer socioeconomically as a result.

By analogy, when we seek to measure the payoffs of a college education on earnings we are not merely interested in estimating the earnings differences between individuals with and without a college degree. Instead, we are interested in estimating how much more an individual earns because of, or as a direct result of, his or her college education (i.e., the value added by a college degree). For such an estimation, we need to take into account the fact that individuals who have higher earnings potential—because of their class background or abilities, for example—are more likely to attend and graduate from college. Similarly, our estimation of the effect of the disadvantage experienced by individuals as a result of their racial identification must take into account the higher or lower propensity of individuals with higher earnings potential to identify in that category.

The objective of our statistical analysis should not be confused with the search for the effect of an underlying “true” racial identification. By adjusting for the selective identification of individuals as black in our estimates of earnings disadvantage we are not implying that individuals' “true” racial identification is different from that stated in the census form any more than adjusting for the selectivity into college when estimating the payoff of a college education implies that individuals who state that they hold a college degree actually do not

have one. Consistent with a constructivist framework, we do not assume that individuals' racial identification contains measurement error that needs to be corrected to get at an underlying essential race. On the contrary, our approach acknowledges that individuals' racial identification will, by definition, depend on various individual and societal factors. We seek to statistically adjust for the selective identification based on these factors only for the purpose of obtaining an unbiased estimate of the causal effect of individuals' racial identification net of selectivity. To put it differently, it is not the measure of racial identification that we think is biased, but rather it is the *estimate of the effect* of racial identification that is biased in ordinary regression models.

Although IV models have also been used in other contexts to correct for the effect of measurement error in a predictor, that is expressly not our intention. Notably, the fact that our IV model results in a reversal in the estimated direction of the effect of black racial identification on earnings rather than merely an attenuation of its effect supports our interpretation that the IV model is adjusting for endogeneity bias due to omitted variables, rather than bias due to measurement error. Under customary assumptions measurement error in a predictor only leads to the attenuation of its estimated effect on the dependent variable, that is, a bias in the absolute value of the regression coefficient towards zero, and not a reversal in its estimated direction (e.g., Wooldridge 2013: 310-312).

Although our analysis allows us to conclude that individuals who identify as black are disadvantaged economically, we are not able to determine the reasons why. Ethnographic studies have already begun to explore the negative views associated with a black identity in contemporary Mexico (Jones 2013; Lewis 2000, 2012, 2016; Sue 2013). Yet more research is still required to specifically understand the multiple ways in which a black racial identification results in a disadvantage in earnings in the Mexican context. Second, our analysis has focused on the effect that individuals' own racial identification has on their earnings. Like most other surveys worldwide, the EIC lets respondents choose their own racial identification. However, the disadvantage that individuals face in the labor market and in other contexts is more likely to be affected by the way others identify them racially than by how they identify themselves. Unfortunately, we are unable to examine the extent to which racial self-identification coincides with racial identification by others in Mexico using data from the EIC. Yet research in other national contexts suggests that estimates of the effect of racial self-identification on income may in fact underestimate the effect of racial identification by others (Bailey et al. 2013; Telles and Lim 1998).

Over the past decade numerous countries throughout Latin American and other regions of the world have begun to identify and enumerate new ethnoracial categories in their national censuses (Loveman 2014). These efforts constitute an important step towards addressing ethnoracial inequality and incorporating historically marginalized populations into national political life. However, estimates of ethnoracial disparities obtained from tabulations and standard regression models using these new ethnoracial categories should be interpreted with caution. Such estimates may not always reflect the disadvantage experienced by individuals as a result of their ethnoracial identification. In particular, because individuals with greater human and cultural capital may be more familiar with the new terms used by the census, and with the new multicultural message that destigmatizes an ethnoracial minority status,

estimates from tabulations and standard regression models may underestimate the disadvantage experienced by ethnoracial minorities.

Previous work attempting to isolate the causal effect of racial identification on social outcomes has relied on experimental designs in which an individual's race is randomly assigned (Bertrand and Mullainathan 2004; Pager 2003). However, random assignment through experimental designs is often not feasible or too costly. We have proposed a strategy whereby an instrumental variables approach may be used instead. However, the implementation of this approach requires finding an exogenous factor affecting individuals' racial identification through a naturally occurring experiment. The Mexican government's intentional campaign to promote black self-identification provided such an experiment. Other government policies, such as affirmative action programs have been shown to increase black identification in countries such as Brazil and may also serve this purpose (Bailey 2008; Francis-Tan and Tannuri-Pianto 2015). More generally, as governments' in Latin America and around the world seek to identify new ethnoracial categories in their national censuses, they should consider building in experimental designs into their strategies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

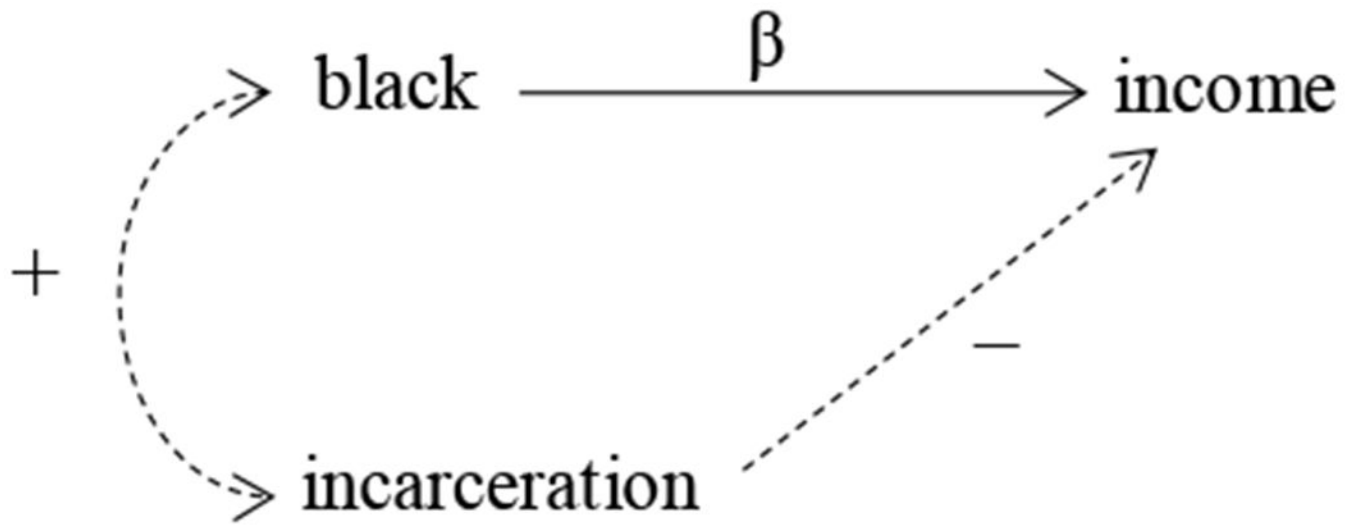
## References

- Aguirre Beltrán Gonzalo. 1972 [1946]. *La población negra de México: Estudio Etnohistórico*. Mexico: Fondo de Cultura Económica.
- Alba Richard D. 1990 *Ethnic Identity: The Transformation of White America*. New Haven, CT: Yale University Press.
- American Sociological Association. 2003 *The Importance of Collecting Data and Doing Social Scientific Research on Race*. Washington, DC: American Sociological Association.
- Angrist Joshua D., and Krueger Alan B.. 2001 "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15: 69–85.
- Angrist Joshua D., and Pischke Jörn-Steffen. 2009 *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton NJ: Princeton University Press.
- Bailey Stanley R. 2008 "Unmixing for Race Making in Brazil." *American Journal of Sociology* 114: 577–614.
- Bailey Stanley R., Loveman Mara, and Muniz Jeronimo O.. 2013 "Measures of 'Race' and the Analysis of Racial Inequality in Brazil." *Social Science Research* 42: 106–119. [PubMed: 23146601]
- Bailey Stanley R., Saperstein Aliya, and Penner Andrew M.. 2014 "Race, color, and income inequality across the Americas." *Demographic Research* 31: 735–756.
- Bailey Stanley R., and Telles Edward E.. 2006 "Multiracial versus Collective Black Categories: Examining Census Classification Debates in Brazil." *Ethnicities* 6: 74–101.
- Beggs John J., Villemez Wayne J., and Arnold Ruth. 1997 "Black Population Concentration and Black-White Inequality: Expanding the Consideration of Place and Space Effects." *Social Forces* 76:65–91.
- Bennett Herman L. 2003 *Africans in Colonial Mexico: Absolutism, Christianity, and Afro-Creole Consciousness, 1570–1640*. Bloomington, IN: Indiana University Press.

- 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: 991–1013.
- Carroll Patrick. 2001 *Blacks in Colonial Veracruz: Race, Ethnicity, and Regional Development*. Austin, TX: University of Texas Press.
- Cohen Philip N. 1998 “Black Concentration Effects on Black-White Gender Inequality: Multilevel Analysis for U.S. Metropolitan Areas” *Social Forces* 77: 207–229.
- Consejo Nacional para Prevenir la Discriminación (CONAPRED). 2006 *Afrodescendientes en México; reconocimiento y propuestas antidiscriminación*. Mexico: CONAPRED.
- Duncan Brian, and Trejo Stephen J.. 2011 “Who Remains Mexican? Selective Ethnic Attrition and the Intergenerational Progress of Mexican Americans” Pp. 285–320 in *Latinos and the Economy: Integration and Impact in Schools, Labor Markets, and Beyond*, edited by Leal David L. and Trejo Stephen J.. New York: Springer.
- Eschbach Karl and Gómez Christina. 1998 “Choosing Hispanic Identity: Ethnic Identity Switching among Respondents to High School and Beyond.” *Social Science Quarterly* 79: 74–90.
- Eschbach Karl, Supple Khalil, and Matthew Snipp C. 1998 “Changes in Racial Identification and the Educational Attainment of American Indians, 1970-1990.” *Demography* 35: 35–43. [PubMed: 9512908]
- Francis-Tan Andrew M. and Tannuri-Pianto Maria. 2015 “Inside the Black Box: Affirmative Action and the Social Construction of Race in Brazil.” *Ethnic and Racial Studies* 38: 2771–2790.
- Gans Herbert J. 1979 “Symbolic Ethnicity: The Future of Ethnic Groups and Cultures in America.” *Ethnic and Racial Studies* 2: 1–20.
- Instituto Nacional de Estadística, Geografía e Informática (INEGI). 2009 *Compendios Estadísticos Regionales 2009*. Mexico: INEGI.
- Instituto Nacional de Estadística y Geografía (INEGI). 2015 *Encuesta Intercensal 2015: Síntesis metodológica y conceptual*. Mexico: INEGI.
- Instituto Nacional de las Mujeres (INMUJERES). 2006 *Las mujeres indígenas de México: Su contexto socioeconómico, demográfico y de salud*. México: INMUJERES.
- Jones Jennifer Anne Meri. 2013 “‘Mexicans will take the jobs that even blacks won’t do’: An analysis of blackness, regionalism and invisibility in contemporary Mexico.” *Ethnic and Racial Studies* 36: 1564–1581.
- Kopczuk Wojciech, Saez Emmanuel, and Song Jae. 2010 “Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937.” *Quarterly Journal of Economics* 125: 91–128.
- Lee Jennifer, and Bean Frank D.. 2004 “America’s Changing Color Lines: Immigration, Race/Ethnicity, and Multiracial Identification.” *Annual Review of Sociology* 30: 221–242.
- Lewis Laura A. 2000 “Blacks, black Indians, Afromexicans: the dynamics of race, nation, and identity in a Mexican moreno community (Guerrero).” *American Ethnologist* 27: 898–926.
- Lewis Laura A. 2012 *Chocolate and Corn Flour: History, Race, and Place in the Making of ‘Black’ Mexico* Durham, NC: Duke University Press.
- Lewis Laura A. 2016 “Indian allies and white antagonists: toward an alternative mestizaje on Mexico’s Costa Chica.” *Latin American And Caribbean Ethnic Studies* 3: 222–241.
- 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 200 and the 2010 Census.” *Demography* 54: 259–284. [PubMed: 28105578]
- Loveman Mara. 2014 *National Colors: Racial Classification and the State in Latin America*. New York: Oxford University Press.
- Loveman Mara, and Muniz Jeronimo O.. 2007 “How Puerto Rico Became White: Boundary Dynamics and Intercensus Racial Reclassification.” *American Sociological Review* 72: 915–939.
- Montiel Martínez, María Luz, ed. 1994 *Presencia Africana en México*. Mexico: Consejo Nacional para la Cultura y las Artes.
- Moreno-Estrada Andrés, et al. 2014 “The genetics of Mexico recapitulates Native American substructure and affects biomedical traits.” *Science* 344, 6189: 1280–1285. [PubMed: 24926019]

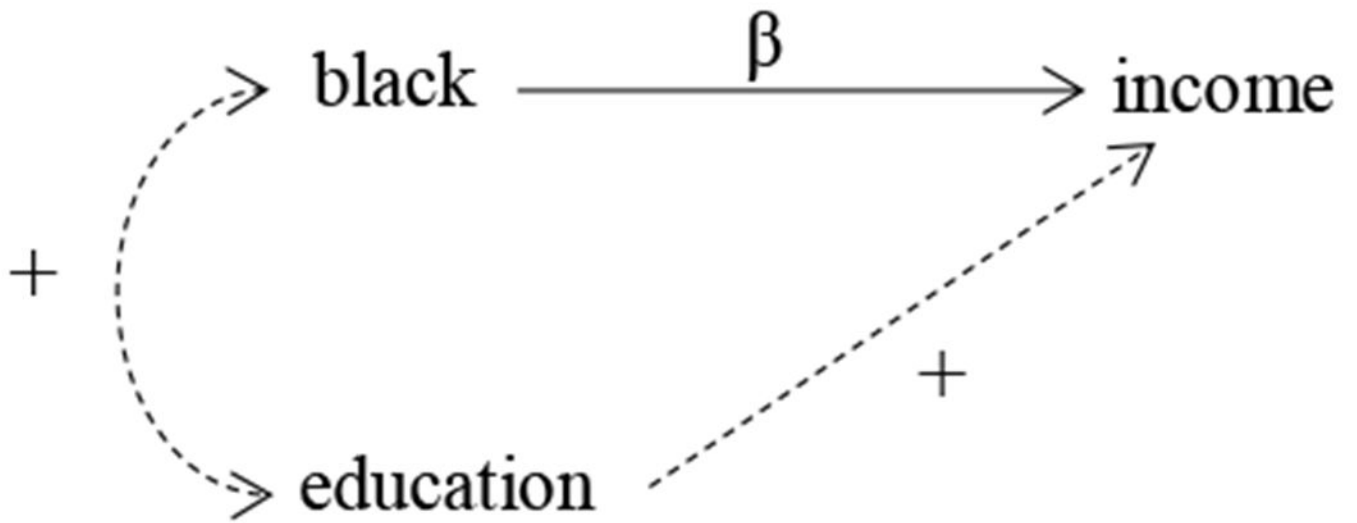
- Morning Ann. 2008 "Ethnic Classification in Global Perspective: A Cross-National Survey of the 2000 Census Round." *Population Research and Policy Review* 27: 239–272.
- Morning Ann. 2011 *The Nature of Race: How Scientists Think and Teach about Human Difference*. Berkeley, CA: University of California Press.
- Pager Devah. 2003 "The Mark of a Criminal Record." *American Journal of Sociology* 108: 937–975.
- Pager Devah, Bonikowski Bart, and Western Bruce. 2009 "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74: 777–799. [PubMed: 20689685]
- Ramírez Alejandro. 2006 "Mexico" Pp. 150–198 in *Indigenous Peoples, Poverty and Human Development in Latin America*, edited by Hall Gillette, and Patrinos Harry A.. New York: Palgrave Macmillan.
- Saperstein Aliya, and Penner Andrew M.. 2010 "The Race of a Criminal Record: How Incarceration Colors Racial Perceptions" *Social Problems* 57: 92–113.
- Saperstein Aliya, and Penner Andrew M.. 2012 "Racial Fluidity and Inequality in the United States." *American Journal of Sociology* 118: 676–727.
- Sen Maya, and Wasow Omar. 2016 "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics." *Annual Review of Political Science* 19: 499–522.
- Snipp Matthew C. 1986 "Who are American Indians? Some observations about the perils and pitfalls of data for race and ethnicity." *Population Research and Policy Review* 5: 237–252.
- Schwartzman Luisa Farah. 2007 "Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil." *American Sociological Review* 72: 940–963.
- Staiger Douglas, and Stock James H.. 1997 "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65: 557–586.
- Stewart Quincy Thomas and Sewell Abigail A.. 2011 "Quantifying Race: On Methods for Analyzing Social Inequality" Pp. 209–234 in *Rethinking Race and Ethnicity in Research Methods*, edited by Stanfield John H. II. Walnut Creek, CA: Left Coast Press.
- Sue Christina A. 2013 *Land of the Cosmic Race: Race Mixture, Racism, and Blackness in Mexico*. New York: Oxford University Press.
- Telles Edward E., and Lim Nelson. 1998 "Does it Matter Who Answers the Race Question? Racial Classification and Income Inequality in Brazil." *Demography* 35: 465–474. [PubMed: 9850470]
- Telles Edward, and Paschel Tianna. 2014 "Who is Black, White, or Mixed Race? How Skin Color, Status, and Nation Shape Racial Classification in Latin America." *American Journal of Sociology* 120: 864–907.
- Vaughn Bobby. 2004 "Los negros, los indígenas y la diáspora. Una perspectiva etnográfica de la Costa Chica" Pp. 75–96 in *Afroméxico: El pulso de la población negra en México: Una historia recordada, olvidada y vuelta a recordar*, edited by Vinson Ben III and Vaughn Bobby. Mexico: Fondo de Cultural Económica.
- Velázquez María Elisa, and Iturralde Gabriela. 2012 *Afrodescendientes en México: Una historia de silencio y discriminación*. Mexico: Consejo Nacional para Prevenir la Discriminación.
- Velázquez María Elisa, and Iturralde Gabriela. 2016 "Afromexicanos: reflexiones sobre las dinámicas del reconocimiento." *Anales de Antropología* 50: 232–246.
- Villarreal Andrés. 2010 "Stratification by Skin Color in Contemporary Mexico." *American Sociological Review* 75: 652–678.
- Villarreal Andrés. 2014 "Ethnic Identification and its Consequences for Measuring Inequality in Mexico." *American Sociological Review* 79: 775–806. [PubMed: 25999600]
- Vinson Ben III. 2005 "Afro-Mexican History: Trends and Directions in Scholarship." *History Compass* 3: 1–14.
- Vinson Ben III. 2009 "From Dawn 'til Dusk: Black Labor in Late Colonial Mexico" Pp. 96–135 in *Black Mexico: Race and Society from Colonial to Modern Times*, edited by Vinson Ben III and Restall Matthew. Albuquerque, NM: University of New Mexico Press.
- Wacquant Loic. 1997 "For an Analytic of Racial Domination." *Political Power and Social Theory* 11: 221–234.
- Wakefield Sara, and Uggen Christopher. 2010 "Incarceration and Stratification." *Annual Review of Sociology* 36: 387–406.

- Waters Mary C. 1990 *Ethnic Options: Choosing Identities in America*. Berkeley, CA: University of California Press.
- Western Bruce. 2002 "The Impact of Incarceration on Wage Mobility and Inequality." *American Sociological Review* 67: 526–546.
- Wimmer Andres. 2009 "Herder's Heritage and the Boundary-Making Approach: Studying Ethnicity in Immigrant Societies." *Sociological Theory* 27: 244–270.
- Wooldridge Jeffrey M. 2013 *Introductory Econometrics: A Modern Approach*. Fifth Edition Mason OH: South-Western.
- Xie Yu, and Goyette Kimberly. 1997 "The Racial Identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census." *Social Forces* 76: 547–570.

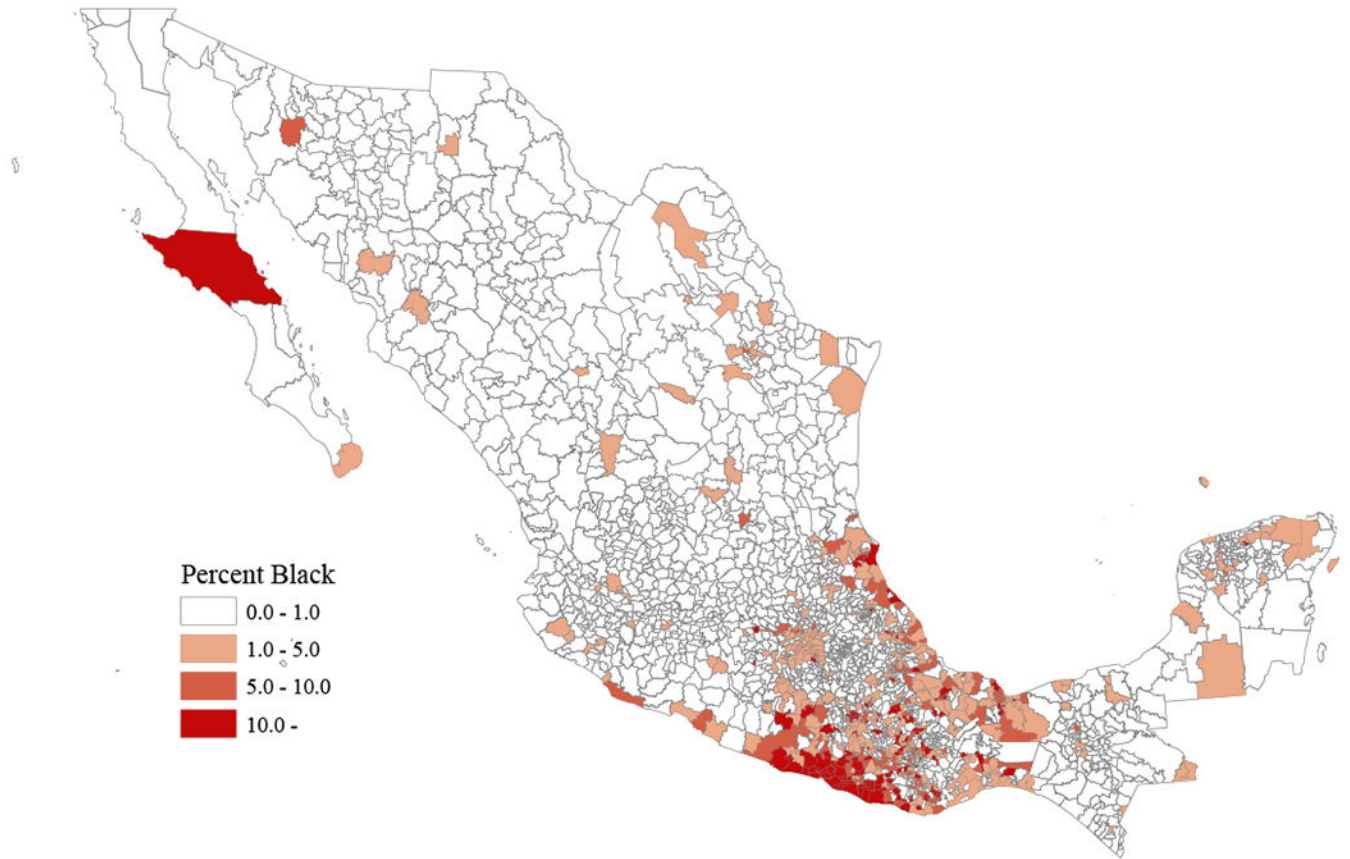


**Figure 1a:**  
Omitted variable (incarceration) leads to a downward bias in the estimate of the effect of black identification on earnings





**Figure 1b:**  
Omitted variable (education) leads to an upward bias in the estimate of the effect of black identification on earnings



**Figure 2:**  
Percent of municipal population identified as black or partly black, 2015

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1:**

Percent of Mexican population identifying as black, 2015

	<b>Population</b>	<b>Percent</b>
Black	1,381,853	1.16
Partly black	591,702	0.50
Not black	114,783,562	96.03
Don't know	1,681,817	1.41
Not specified	1,091,819	0.91
Total	119,530,753	100.00

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2:**

Percent of black and non-black population identifying as indigenous, 2015

<b>Percent Indigenous</b>	
Black	67.5
Partly black	62.0
Not black	22.5
Total	23.2

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3:**

Education, occupation and earnings of Mexican men by race

	<b>Black</b>	<b>Partly black</b>	<b>Not black</b>
Education			
Avg. years of education	10.8	10.8	10.7
Primary or less	28.8 <sup>*</sup>	28.9	30.3
Middle school	37.8	39.3	38.0
High school	36.7 <sup>**</sup>	33.1	33.1
College or more	22.5	23.9	23.5
Occupation			
Agricultural workers	7.7 <sup>**</sup>	8.0	8.4
Low-skill workers	38.7	35.5 <sup>**</sup>	39.6
Professional and Directors	27.2	28.5	27.0
Earnings			
Avg. monthly earnings	8,199	8,566	8,438

Notes: All numbers are percentages unless otherwise specified. See text for description of analytical sample. The black category includes individuals who are partly black. Significance levels are relative to non-blacks. Weighted sample.

\*  
p<.05

\*\*  
p<.01 (two-tailed tests)

**Table 4:**  
Linear regression models predicting men’s log monthly earnings

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Black identification							
Black	-0.002 (0.011)	0.051** (0.011)	0.038** (0.011)	0.063** (0.011)	0.043** (0.011)	0.037* (0.018)	0.037** (0.011)
Partly black	-0.010 (0.016)	0.020 (0.016)	0.010 (0.016)	0.025 (0.016)	0.019 (0.016)	0.014 (0.020)	0.018 (0.016)
Proportion blacks in municipality					0.553** (0.038)	0.553** (0.038)	0.517** (0.036)
Indigenous identification							
Indigenous		-0.132** (0.003)		-0.071** (0.003)	-0.070** (0.003)	-0.070** (0.003)	-0.058** (0.003)
Partly indigenous		-0.021** (0.008)		-0.009 (0.008)	-0.009 (0.008)	-0.009 (0.008)	-0.005 (0.008)
Interaction							
Black and indigenous						0.009 (0.019)	
Age	0.008** (0.000)	0.008** (0.000)	0.008** (0.000)	0.008** (0.000)	0.008** (0.000)	0.008** (0.000)	0.006** (0.000)
Years of education	0.088** (0.000)	0.086** (0.000)	0.081** (0.000)	0.080** (0.000)	0.080** (0.000)	0.080** (0.000)	0.052** (0.000)
Urbanization							
2,500 to 14,999 residents			0.098** (0.004)	0.097** (0.004)	0.096** (0.004)	0.096** (0.004)	0.049** (0.003)
15,000 a 49,999 residents			0.148** (0.004)	0.143** (0.004)	0.143** (0.004)	0.143** (0.004)	0.073** (0.004)
50,000 a 99,999 residents			0.188** (0.004)	0.183** (0.004)	0.182** (0.004)	0.182** (0.004)	0.100** (0.004)
100,000 and more residents			0.253** (0.003)	0.246** (0.003)	0.244** (0.003)	0.244** (0.003)	0.152** (0.003)
Regions							
Northwest			0.230** (0.005)	0.217** (0.005)	0.232** (0.005)	0.232** (0.005)	0.235** (0.005)
Northeast			0.185** (0.004)	0.171** (0.004)	0.183** (0.005)	0.183** (0.005)	0.173** (0.004)
Center			0.108** (0.004)	0.099** (0.004)	0.107** (0.004)	0.107** (0.004)	0.096** (0.004)
Center-west			0.178** (0.004)	0.166** (0.004)	0.180** (0.004)	0.180** (0.004)	0.171** (0.004)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Occupations							
Occupation fixed-effects							yes
Constant	7.454 ** (0.006)	7.502 ** (0.007)	7.234 ** (0.007)	7.270 ** (0.007)	7.254 ** (0.007)	7.254 ** (0.007)	7.693 ** (0.008)
R-squared	0.2692	0.2747	0.2988	0.3007	0.3010	0.3010	0.3757
N	709,317	700,580	709,317	700,580	700,580	700,580	700,580

\* p<.05

\*\* p<.01 (two-tailed tests)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5:**

Instrumental variables model treating black identification as endogenous using states' exposure to Afro-identity campaign as an instrument

	<b>black</b>	<b>Log income</b>
Black identification		
Black or partly black		-0.188** (0.020)
Proportion blacks in municipality	7.057** (0.099)	0.811** (0.047)
Indigenous identification		
Indigenous	0.731** (0.015)	-0.059** (0.003)
Partly indigenous	0.903** (0.026)	0.005 (0.008)
Age	-0.003** (0.001)	0.008** (0.000)
Years of education	0.011** (0.001)	0.080** (0.000)
Urbanization		
2,500 to 14,999 residents	0.073** (0.019)	0.097** (0.004)
15,000 a 49,999 residents	0.086** (0.024)	0.144** (0.004)
50,000 a 99,999 residents	0.161** (0.024)	0.183** (0.004)
100,000 and more residents	0.307** (0.017)	0.246** (0.003)
Regions		
Northwest	-0.210** (0.036)	0.233** (0.005)
Northeast	-0.086** (0.024)	0.185** (0.005)
Center	0.074** (0.016)	0.109** (0.004)
Center-west	-0.200** (0.021)	0.181** (0.004)
Campaign		
State with promotion campaign	0.139** (0.016)	



	<b>black</b>	<b>Log income</b>
Constant	-2.745 ** (0.035)	7.250 ** (0.007)
$\rho$		0.163 **
N		700,580

Wald test of indep. eqns. ( $\rho = 0$ ):  $\chi^2(1) = 156.59$ , Prob >  $\chi^2 = 0.0000$

\*  
p<.05

\*\*  
p<.01 (two-tailed tests)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 6:**

Instrumental variables model treating black identification as endogenous using African admixture in state as an instrument

	<b>black</b>	<b>Log income</b>
Black identification		
Black or partly black		-0.232 ** (0.025)
Indigenous identification		
Indigenous	0.697 ** (0.022)	-0.044 ** (0.005)
Partly indigenous	0.938 ** (0.042)	0.014 (0.013)
Age	-0.004 ** (0.001)	0.008 ** (0.000)
Years of education	0.014 ** (0.002)	0.072 ** (0.000)
Urbanization		
2,500 to 14,999 residents	0.083 ** (0.021)	0.136 ** (0.005)
15,000 a 49,999 residents	-0.045 * (0.021)	0.182 ** (0.005)
50,000 a 99,999 residents	-0.027 (0.030)	0.198 ** (0.006)
100,000 and more residents	0.128 ** (0.023)	0.285 ** (0.005)
Regions		
Northwest	-0.734 ** (0.082)	0.150 ** (0.007)
Northeast	-0.604 ** (0.047)	0.020 * (0.008)
Center-west	-0.489 ** (0.024)	0.210 ** (0.005)
Admixture		
Percent African admixture in state	0.159 ** (0.006)	
Constant	-2.680 ** (0.052)	7.322 ** (0.011)
P		0.190 **
N		281,233

Note: Sample restricted to 10 states for which African admixture information is available.

\*  
p<.05

\*\*  
p<.01 (two-tailed tests). Wald test of indep. eqns. ( $\rho = 0$ ):  $\chi^2(1) = 103.24$ , Prob >  $\chi^2 = 0.0000$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript