Race Perceptions in the Labor Market

Pedro C. Sant'Anna* Sulin Sardoschau† Aiko Schmeißer‡

11th December 2025
preliminary - do not circulate

Abstract

This paper applies a constructivist approach to race in the labor market, whereby racial classification is shaped by social perception rather than a fixed attribute. We combine information from two large-scale administrative data sets to obtain three distinct racial identity measures for 330,000 workers in Brazil between 2003 and 2015: employer classification, self-identification, and an algorithmic measure of skin tone extracted from photographs. First, we develop a simple conceptual framework that clarifies when self- and employer-perceptions matter, why racial perception is important for wage-setting, and how it can affect standard estimates of racial disparity. Then, we take this to the data and show that there is a large and systematic divergence between self-identified and employer-ascribed race and that the same worker can be classified differently across employers. We estimate a "race function" that characterizes how employers map phenotypic cues, self-identification, education, and employment histories into racial categories, revealing that productivity-relevant factors shape race perceptions. Finally, we compare measures of racial wage disparities under essentialist and constructivist approaches.

Keywords: Race, identity, fluidity, wage gap, Brazil
JEL Codes: J15; J50; J71; Z10

*MIT Department of Economics and Núcleo de Estudos Raciais (INSPER). E-mail: p_stanna@mit.edu
†Humboldt Universität zu Berlin, RFBerlin and IZA. E-mail: sulin.sardoschau@hu-berlin.de
‡Columbia University. E-mail: as8091@columbia.edu

1 Introduction

A large empirical literature documents racial disparities in labor markets, the judicial or schooling system, and many other domains. Most analyses implicitly take an essentialist stance, treating race as a fixed individual attribute that can be straightforwardly measured and held constant. A constructivist perspective starts from a different premise. It views racialization as a process in which individuals are seen and evaluated through shared cultural meanings, expectations, and stereotypes that have developed historically and are reinforced by institutions (Bohren et al., 2025; McGee, 2025).

In this view, the cues that shape racial perception can extend beyond physical appearance or ancestry and include signals about social or economic status. For example, two individuals with similar phenotypic features may be read differently depending on their level of schooling. A person with limited formal education may be more readily perceived as now-White than an otherwise identical person with a university degree. This creates a measurement problem for empirical work. If education is both relevant for the outcome of interest and an input into how race is perceived by the decision-maker, then conventional estimates of racial disparities conflate disparate treatment with differences in the signals that influence racial classification (Rose, 2023; Hu and Kohler-Hausmann, 2025).[1]

This paper studies the construction of race and its implications to measuring disparities in the context of Brazil's labor market between 2003 and 2015. We combine information on 330,000 workers from two large-scale administrative data sets to observe three distinct racial identity measures for the same individual: employer classification, self-identification, and an algorithmic measure of skin tone extracted from photographs. Our analysis proceeds in four steps. First, we develop a simple conceptual framework that clarifies when self- and employer-perceptions matter, why racial perception is important for wage-setting, and how it can affect standard estimates of racial disparity. Second, we document systematic divergence between self-identified and employer-ascribed race and show that the same worker is frequently classified differently across employers. Third, we estimate a race function that characterizes how employers' race perceptions are shaped by workers' skin tone, self-identification, and productivity-relevant factors such as education and employment histories, as well as local prevailing norms. Finally, we compare measures of racial wage disparities under the three racial identity measures. The results highlight the importance of addressing the endogeneity of racial categories and quantifying how racial cues translate into racial perceptions for interpreting their relevance in wage setting.

Our data comes from the *Relação Anual de Informações Sociais* (RAIS), which provides matched employer-employee records for the universe of formal jobs in Brazil with detailed information on workers' employment histories, firm characteristics and, crucially, includes the worker's race as reported by the employer. We merge RAIS to data on the universe of political candidates from the *Tribunal Superior Eleitoral* (TSE), for whom we observe self-declared race and a measure of skin tone extracted from standardized electoral photographs using the algorithmic pipeline developed by Adukia et al. (2023).[2]

[1] By conventional estimates we mean comparing differences in outcomes for a given recorded race (typically self-reported or enumerator assessed), either unconditional or in benchmarking regressions that hold productivity-relevant factors, such as education, constant.

[2] Relative to the full RAIS workforce, this linked sample is slightly older and more male but comparable

Brazil's fluid system of racial classification, shaped by its colonial history and socioeconomic structures (Telles, 2014), provides an ideal setting for examining how racial identity is constructed and interpreted across different institutional contexts, particularly in the labor market. At the same time, Brazil's experience reflects broader global patterns in the malleability of racial boundaries (Davenport, 2020; Rose, 2023). Shifts in self-identification in response to economic conditions, policy incentives, and evolving social norms appear in settings such as the United States (Adukia et al., 2025; Baron et al., 2024; Antman and Duncan, 2023; Dahis et al., 2019b; Antman and Duncan, 2015), India (Atkin et al., 2021; Cassan, 2015), China (Jia and Persson, 2021), Indonesia (Rademakers and van Hoorn, 2021), and Romania (Mitrut et al., 2025). A prominent example is ethnic attrition among Hispanics in the United States, where both self-identification and enumerator-ascribed race vary systematically with economic status (Duncan and Trejo, 2011, 2025a,b).

We develop a simple framework that makes the process of racial perception explicit and clarifies how it shapes measured disparities, building on an emerging literature on constructivism in economics (Rose, 2023; Hu and Kohler-Hausmann, 2025; Adukia et al., 2025). The framework distinguishes self-identification from employer-ascribed race and models each as the outcome of a classification process that maps phenotypic and status cues into racial categories. In some instances, these measures coincide, and in others, they diverge in systematic ways that make one measure more relevant than the other.[3] The framework highlights that divergence may not simply reflect noise in reporting but captures meaningful differences in how workers and employers internalize or interpret the same cues.

It also clarifies why standard estimates of racial wage gaps can be difficult to interpret. Many cues that shape employers' perceptions are simultaneously productivity relevant. For instance, if education affects both wages and the employer's racial perception of the worker, then omitting education induces bias in the estimated impact of race on earnings. Conditioning on such variables (as practiced in typical benchmarking regressions) removes omitted-variable bias but also holds fixed the very inputs through which racialization operates. Reduced-form estimates with respect to a single cue, such as skin tone or distinctly Black names, face a related challenge: they capture an intent-to-treat effect that masks heterogeneity in how that cue is interpreted across workers of different productivity. Decision makers may, for example, place less weight on skin tone as a racial signal for more educated workers. Consequently, smaller reduced-form estimates for educated workers may reflect both less discrimination experienced by non-White workers in that group and a lower probability of being perceived as non-White in the first place.

The main goal of the paper is to take this constructivist view of race to the data, beginning with a documentation of mismatched racial identities in the labor market. We show that

in education and average earnings. Labor market attachment remains strong because most candidates run for municipal council positions that require no more than 10 hours per week, and roughly 85% of candidates do not win office.

[3]Racial identities may coincide when individuals internalize repeated feedback about how others categorize them, causing self-identification to converge toward external perceptions. Even when they do not coincide, the relevant measure depends on the question at hand: self-identification may be more appropriate for studying social and political dynamics across race groups, whereas in settings where the decision-maker's perception matters, such as the employer's racial perceptions of workers in wage setting decisions, this may be the more relevant construct.

self-identified and employer-ascribed race are related but not interchangeable race concepts. Only about three quarters of workers are classified in the same broad category (White or non-White), and mismatches are strongly asymmetric: employers are much more likely to record self-identified non-White workers as White than to classify self-identified Whites as non-White. These discrepancies concentrate among individuals with intermediate skin tones and are closely tied to job mobility, with a sizable share of workers changing racial classification when they move between firms.

To quantify the role of firms in racial ascription, we extend the analysis to the full RAIS sample of about 45 million workers and exploit workers' moves across establishments in an AKM-style framework (Abowd et al., 1999). We model employer-ascribed race with worker and firm fixed effects, where worker effects capture portable racial cues that are perceived similarly across employers, and establishment effects capture employer-specific classification practices. Establishment effects account for about one-third of the variation in employer-ascribed race. If race in RAIS reflected an innate worker attribute, firm effects would play no role; instead, their contribution is large, even higher than the role of firms in explaining wage dispersion in Brazil (Gerard et al., 2021).[4]

In order to examine how employers ascribe race, we estimate a race function that depends on workers' skin tone, self-identification, and a broad set of demographic and labor market characteristics, including age, gender, education, past wages and tenure, along with an extensive set of region, industry, occupation, and firm fixed effects. The probability of being classified as non-White increases monotonically with skin tone. A one standard deviation increase in skin tone raises the likelihood of being ascribed non-White by 11 percentage points. Nevertheless, productivity-relevant factors reshape this race function. First, for a given skin tone, more educated workers are substantially more likely to be classified as White. Second, the elasticity of racial ascription with respect to skin tone is smaller among workers with a university degree than among workers with less than a high-school degree. This means that the same phenotypic cue is weighted differently across workers of different productivity. For a subsample of workers, we also observe careers at their previous employer. Focusing on past wages and tenure as direct markers of productivity, we find that these factors also contribute to the whitening of workers. Our most demanding specification includes firm-by-occupation fixed effects, comparing racial ascription across workers in the same three-digit occupation within the same firm. Even within this narrow comparison group, status cues continue to shift employer perceptions in similar ways. This shows that race perceptions arise at the job level – from the interaction of workers and employers – and are not driven solely by sorting of workers across employers that uniformly whiten or darken their workforce.

Finally, we turn to the relationship between self-perception and employer ascription, augmenting the employer race function with self-identified race.[5] As expected, self-identification

[4] A battery of robustness checks, including comparisons across subsamples with higher reporting quality, placebo tests using gender, and split-sample instrumental variable strategies suggest that these patterns are unlikely to be driven by random reporting error.

[5] We also estimate an analogous race function for self-identification alone. We find that skin tone, education, and regional differences explain self-perception in a similar direction as employer ascription, though to varying degrees. While characteristics of workers – their skin tone and education – are more predictive for self-identification, local social norms appear to matter more for how employers perceive the racial identities of workers.

is a strong predictor of employer perceptions conditional on skin tone and other socioeconomic cues, either because workers send racial signals that employers partially incorporate into their assessments or because self-identification captures unobserved traits that shape both self-perception and employer-ascription. Conditioning on self-identification also reveals which additional cues employers use when they depart from workers' stated racial identities. Even within the same firm and three-digit occupation cell and holding self-identification fixed, employers rely on skin tone and education as key inputs into racial ascription. Together, these patterns suggest that, consistent with the constructivist view, race perception is shaped by socioeconomic signals that correlate with productivity and that self-identification and employer perception capture related but distinct aspects of racial identity.

These findings motivate an approach that incorporates the role of racial perception into the measurement of disparities. We introduce a perception-normalized disparity measure (PND) that scales observed wage differences across a racial cue by the extent to which the cue shifts the probability of being perceived as non-White (Rose, 2023). In terms of an instrumental variable framework, the PND rescales the reduced-form effect of a racial cue on the outcome of interest by the first-stage effect of that racial cue on racial perception. Conceptually, it expresses disparities in units of racialization rather than in units of physical appearance. This allows for meaningful comparisons across workers and settings for which racial cues carry different weights. The PND requires a racial instrument that moves the decision maker's perception of the worker's race. When one can argue that the cue affects wages only through its effect on perceptions, the PND isolates the causal effect of being perceived as non-White. Otherwise, it is best seen as a scaling metric.

To assess whether these conceptual differences translate into economically meaningful differences in measured disparities, we compare estimates under conventional and constructivist approaches. We begin by estimating conventional wage gaps in RAIS using employer-ascribed race, successively introducing the same controls and fixed effects as in the race-function analysis. Unconditional wage disparities lie at 14.8%. These disparities persist even within narrowly defined firm-by-occupation cells and amount to 2.1% for workers that are recorded as non-White by the employer. Finally, we augment the wage disparity specification with self-declared race and find that conditional on self-declared race, employer-ascribed race carries an additional wage penalty of 1.7%, emphasizing that employer perception is a important margin.

We then turn to estimates typically used in the colorism literature that use skin tone as a racial cue and study wage disparities along the skin tone distribution. Using our algorithmic measure of skin tone from standardized photographs, we document a monotonic wage gradient that remains within firms and occupations. We interpret these estimates as intent-to-treat effects of a racial cue whose impact depends on how strongly it translates into employer race perceptions. The unconditional colorism disparity measure suggests that a one standard deviation increase in skin tone is associated with a 2.6% wage penalty. Analogous to our previous results, we find that conditional on self-identification, darker-skinned workers suffer from an added wage penalty in the same firm-by-occupation cell.

Building on these insights, we leverage skin tone as an instrument for employer-ascribed race to recover a perception-based disparity that accounts for heterogeneity in the mapping from racial cues to racial classification. The unconditional PND reveals that non-White workers earn 23.5% less on average, compared to 14.8% when using employer-ascription dir-

ectly. Similarly, in our most stringent specification that includes firm-by-occupation fixed effects and self-declared race, we detect a wage penalty of 8.4% – which is four times larger than the conventional estimate. Importantly, the PND attenuates less than conventional measures when adding controls, in line with the idea that it restores perception-driven variation among workers of otherwise similar socio-economic status. Finally, we show that noise in skin tone measurement and in employer-recorded race leads to substantial attenuation in estimates, implying that both conventional race-based gaps and cue-based disparities understate the magnitude of perception-driven wage penalties. We also discuss the caveats and limitations of the PND measure, focusing on threats to excludability and resulting challenges for measuring racial disparities with observational data.

Overall, the central contribution of our analysis is to show that race perceptions are systematically shaped by productivity-relevant factors and that conventional disparity estimates, even when they reflect the perceptions of decision makers, yield patterns that differ meaningfully from perception-normalized measures. Despite some of the limitations in our study, which uses just one, incomplete racial cue and cannot account for the joint perceptions of all potential decision makers (including customers or co-workers), we hope to clarify and show empirically that the choice of racial measure and estimand is not innocuous but depends on the underlying model of racial classification that researchers need to make explicit.

This paper contributes to several strands of the literature. First, it contributes to a literature on racial identity and its implications for empirical analyses of race (Rose, 2023; Kim and Loury, 2019; Sen and Wasow, 2016; Holland, 2008; Baron et al., 2024). The constructivist turn is consistent with research on the formation of social identity and its economic consequences. Since the seminal work of Akerlof and Kranton (2000), economists have modeled identity as part of agents' utility functions, shaping cooperation, discrimination, and prosocial behavior (Ben-Ner et al., 2009; Shayo, 2020). Identity influences consumption patterns (Atkin, 2016) and labor supply (Oh, 2023), while experimental evidence demonstrates how shared or conflicting identities structure collaboration (Lowe, 2021; Del Carpio and Guadalupe, 2022). In labor markets, gendered and racial perceptions influence both labor-force participation and occupational sorting (Bursztyn et al., 2020; Glover et al., 2017).

In parallel, sociological and political science work documents the fluidity of racial identification and the malleability of social boundaries (Schwartzman, 2007; Saperstein and Penner, 2012; Davenport, 2020). Studies across diverse contexts show that identity can shift with incentives and social norms, for instance, in the United States (Antman and Duncan, 2023; Dahis et al., 2019b; Antman and Duncan, 2015; Cassan, 2015; Atkin et al., 2021; Jia and Persson, 2021). These findings echo the argument that classification systems are historically contingent and responsive to policy incentives. This paper contributes by investigating race as ascribed by others, focusing particularly on the role of employers. Studies of hetero-identification of race mostly do so in laboratory settings (e.g., Rodeheffer et al., 2012; Freeman et al., 2011). We complement this literature by studying the determinants of racial perceptions in a real-world setting, with administrative data used to inform policies and statistics on racial wage gaps.

Another strand focuses on racial disparities in the labor market. The canonical literature links differences in wages and employment to discrimination, human capital, and firm heterogeneity (Bayer et al., 2025; Abowd et al., 1999; Card et al., 2018; Bonhomme et al., 2019; Kline et al., 2020). Field experiments document how small changes in identity

cues (names, photos, speech) alter employer responses (Bertrand and Mullainathan, 2004; Agan and Starr, 2018; Hersch, 2006; Doleac and Stein, 2013; Grogger, 2019; Ajzenman et al., 2025). A related literature emphasizes that race itself is multidimensional and that phenotypic traits such as skin tone generate substantial within-race disparities in wages, employment, and marriage outcomes, consistent with colorism and differential treatment based on perceived racial status (Goldsmith et al., 2006, 2007; Kreisman and Rangel, 2015; Hersch, 2008). Complementing this work, economic historians document that racial identity is not fixed: individuals have actively passed for white when incentives were strong, and these endogenous shifts in racial classification carry large consequences for income, longevity, and intergenerational mobility (Nix and Qian, 2015; Dahis et al., 2019a; Noghanibehambari and Fletcher, 2025). Large-scale correspondence and audit studies show substantial heterogeneity across employers and industries, suggesting that discrimination may be concentrated in certain firms or decision environments (Kline and Walters, 2021; Kline et al., 2022, 2024). At the same time, structural decompositions reveal that firm sorting and institutional factors explain an increasing share of wage inequality (Card et al., 2008; Hsieh et al., 2019; Bayer and Charles, 2018). Within this broader literature, the perception of race by others, rather than self-identification, has received comparatively little empirical attention outside of controlled laboratory settings (e.g., Adukia et al., 2025).[6] In real-world markets, decision-makers act on their own ascriptions, and understanding these processes requires observing how cues translate into perceptions. Our analysis contributes by studying these mechanisms in an administrative context with economic consequences, where perceptions shape both reported race and wage setting.

2 Conceptual Framework

In what follows, we build on previous insights in Rose (2023) and Hu and Kohler-Hausmann (2025) to develop a framework that makes the process of racial perception explicit. We distinguish self-identification from employer-ascribed race, consider when these measures coincide or diverge, and examine how observable cues such as skin tone or education feed into each. This foundation allows us to formalize how disparities arise when outcomes are determined on the basis of racial perceptions, and to propose a perception-normalized measure of disparity.

Race Concept under Constructivism In much of empirical economics, race enters analysis as if it were a stable trait that individuals possess; an attribute like education or age

[6] Adukia et al. (2025) use detailed "complexion" descriptions recorded by Freedman's Bank clerks alongside Census racial categories to study how socioeconomic status shapes racialization during Reconstruction. In their setting, both appearance and race are assessed by human perceivers operating in different institutional contexts, so discrepancies between complexion and Census race reflect cross-observer disagreement rather than perceptions within a single economic relationship. Complementing this approach, we examine employer perceptions in a contemporary labor market. We construct a continuous, algorithmic measure of skin tone from photographs and link it to employer-ascribed race reported within firms by staff directly involved in wage setting and promotion. Unique personal identifiers (CPF) allow exact matches across photographs, self-declared race, and administrative employer-employee records with detailed information on wages, occupations, contracts, firm characteristics, and complete employment histories.

that can be observed, coded, and held constant. This may be owed to data constraints or an implicit essentialist view of race. A constructivist perspective starts from a different premise. It views race as a social and relational status that comes into being through perception and interaction. To be "racialized" is not simply to have certain physical features or ancestry, but to be seen and treated through a set of shared cultural meanings, expectations, and stereotypes that have accumulated historically and are sustained by institutions (Bohren et al., 2025; McGee, 2025).

When an employer, landlord, or police officer "perceives race," what is perceived is rarely limited to skin tone or ancestry. Perception activates a richer interpretive frame: race perceptions are determined by appearance, language, dress, and behavior to socially learned ideas about competence, trustworthiness, or respectability. In this sense, perceiving race and perceiving the broader situation are often intertwined rather than separable. A decision-maker who sees a worker as Black or White does not just register a color category; she apprehends the candidate through a lens that gives additional meaning to other cues such as education, speech, or demeanor. Empirically, this means that what counts as a "racial cue" and what counts as a "non-racial control" are not self-evident, but depend on a theory of what race perception entails (Hu and Kohler-Hausmann, 2025; Rose, 2023).

Constructivist accounts thus differ from essentialist views in two ways. First, they emphasize that racial classification is context-dependent: the same person may be read differently across settings or by different observers. Second, they recognize that the line between "race" and "productivity-relevant" characteristics is porous, since the social meaning of education, speech, or dress can itself vary with racialization. For empirical researchers, this means that any estimate of a "racial effect" necessarily rests on decisions (often implicit) about which differences are to be treated as part of race perception and which as legitimate grounds for differential treatment. In what follows, we conceptualize this issue in the context of the labor market. In particular, we are interested in how status cues and other productivity-relevant factors affect an employer's perception of a given worker's race. While this is an important application case, it extends to many other settings, including, for instance, sentencing decisions in the judiciary or educational outcomes. It is also not necessarily restricted to race but could extend to cues that lead to ascriptions of other social identities, such as gender, sexual orientation, religion, or caste.

Mismatched Racial Identities. Racial disparities are typically measured using self-reported race, either from interviews, census data, or administrative data. Consider the case where recorded race is coded as White or non-White $R^g \in \{r^0, r^1\}$, and is either self-reported or employer ascribed with $g \in \{S, E\}$. If self- and employer-perceptions of race are determined by the same underlying latent race function $A(Z)$, where Z collects the observable and contextual characteristics that shape racial classification, then racial disparity in outcome Y based on R^S or R^E is equivalent and any differences would reflect noise rather than distinct notions of race.

This equivalence may hold in some settings: for instance, individuals may eventually internalize the feedback they repeatedly receive about how others categorize them, leading their self-identification to converge to dominant external perceptions; similarly, some racial cues such as very dark or very light skin tone, particular hair texture, or certain linguistic

or phenotypical traits may be so unambiguous in the local social context that all observers, including the individual, interpret them in the same way; in such cases, the distinction between self- and employer-ascribed race effectively vanishes and $Y(R^S) = Y(R^E)$.

In other settings, however, this equivalence may not hold. Different observers may interpret the same racial cues through distinct social and institutional lenses, producing related but not identical outputs. These distinct race functions $A_S(Z)$ and $A_E(Z)$ may share inputs and even functional form, but differ in the weights observers place on specific cues or in the thresholds that define category boundaries. For instance, employers may form judgments from observable cues rather than private information about ancestry, family history, or social affiliations that inform an individual's self-understanding. In this case, their divergence is informative, reflecting heterogeneity in how social contexts translate observable traits into racial categories rather than measurement error in an underlying "true" race.

When studying racial disparities in the labor market, two perspectives are conceivable. If the researcher is interested in understanding labor market disparities more broadly, and believes that self-identification better captures unobserved factors that influence wages such as access to social networks, internalized motivation, or behavioral responses to anticipated or past discrimination, then R^S may represent the more relevant dimension of identity. Conversely, if the researcher is more interested in perception-based disparities or (direct) discrimination, then employer-ascribed identity R^E may be the more relevant margin. In either case, understanding the forces that shape self-identities and perceptions is necessary to choose the appropriate measure.

Race Function and Implications for Racial Disparity Measures. Let us examine the case of employers' perceptions, since we are interested in disparate treatment on the basis of a protected characteristic of the worker as it is *perceived* by the decision-maker. In particular, consider an employer who sets a wage $Y \in P$ that maximizes their utility $U(Y, R^E, X)$ given the employer-perceived race $R^E \in \{r^0, r^1\}$ and a set of other productivity-relevant characteristics X. In a potential outcomes framework, the researcher studying racial earnings disparities in the labor market is interested in examining whether $Y(r^0, X) \neq Y(r^1, X)$.

If race perceptions are indeed a function of social and economic status signals, then Z could contain factors that are productivity-relevant and co-determine wages (i.e., $Z \cap X \neq \varnothing$). In that case, a fundamental identification problem arises as one cannot separate directs of Z from discrimination based on R^E. Any estimates of racial wage disparities are biased with the sign governed by the covariances of Z with R and Y. If the researcher decides to control for Z to examine whether $Y(r^0, X|Z) \neq Y(r^1, X|Z)$, then the estimate will be purged from OVB but the estimand holds fixed the very cues that generate the perception. If R^E is (nearly) a deterministic function of Z, then the residual identifying variation in R^E is largely classification noise, rendering estimates fragile and typically attenuated toward zero. In short, unbiasedness need not imply substantive informativeness: conditioning on Z can benchmark away the identity contrast one aims to measure.

To illustrate this issue, let us take the example of education again. Education increases worker productivity, but low schooling also makes employers substantially more likely to classify a worker as non-White. If researchers lack data on education, the estimated racial gap conflates the effect of perceived race with the effect of an omitted productivity-relevant

variable that also shapes racial classification. However, if researchers do observe education and condition on it to estimate $Y(w, X|Z) \neq Y(b, X|Z)$, they remove this source of omitted-variable bias, but they do so by holding fixed a key input into the perception process itself. In settings where education strongly predicts employers' racial ascriptions, conditioning on education strips away much of the variation that generates perceived racial identity. Take the most extreme case, where race perception depends entirely on education. Conditioning on education means that racial disparities based on R^E will vanish entirely; not because decision-makers do not consider race when taking action but because education absorbs the social meaning of race.

Another approach in empirical studies is to focus on a single salient dimension of appearance or identity to examine whether $Y(z^0, X) \neq Y(z^1, X)$, where $z^{(0,1)} \in Z$ could include, for instance, skin tone, accent, hairstyle, or name. This approach, common in the colorism and audit-study literatures, measures the total outcome gradient with respect to that cue. However, the elasticity of the race function $A_g(\cdot)$ with respect to z may vary systematically across productivity-relevant characteristics X. For example, a darker skin tone may increase the likelihood of being perceived as non-White more sharply among low-educated workers than among highly educated ones. In that case, the reduced-form coefficient of z is a weighted average of heterogeneous perception effects across groups with different X, conflating variation in identity salience with variation in productivity-relevant traits.

Consequently, when the "racial elasticity" of a cue depends on X, the wage gradient in z reflects both genuine differences in how appearance shapes perceptions and the compositional effects of who is most affected by those perceptions. In other words, smaller estimated skin tone disparities among, for instance, highly educated workers may reflect two aspects: genuinely lower discrimination or racial disparities within this group or a lower probability of being perceived as non-White in the first place. If the researcher is interested in measuring disparity based on a single racial cue z, that is an appropriate approach. Instead, if the researcher is interested in measuring disparities in context where that racial cue actually translates into being perceived as a minority group, then one may have to take seriously the process of racialization with respect to that racial cue. Thus, to isolate disparities that stem from the process of racialization by the decision maker, it is therefore necessary to understand the inputs to the race function and incorporate this classification process into the construction of the race measure itself.

In sum, conventional estimates of racial disparities that use self-reported or ascribed race may fail or at least not provide the relevant estimand if Z contains factors that are productivity-relevant and/or if the interpretation of racial cues differs across contexts and worker characteristics. This motivates a scaling approach, which we label the *Perception-Normalized Disparity* (PND), that accounts for how strongly a given cue shifts racial perceptions. This approach normalizes the reduced form effect of the racial cue $Y(z|X)$ by the elasticity of that racial cue in the race function $A_g(z|X)$. Intuitively, the PND measures the change in wages per unit change in the probability of being perceived as non-White. This is equivalent to a local average treatment effect (LATE) for worker-employer pairs whose perceived racial identity is marginal with respect to the racial cue z. The PND should be viewed as a scaling exercise, rather than the causal effect of perceptions, unless one can argue that conditional on X the racial cue z affects outcomes solely through its impact on employer-perceived race. Under this exclusion restriction, PND identifies the causal effect of being

perceived as non-White on wages. In general, however, it is best viewed as a standardized measure of disparity that expresses observed wage differences in units of social perception rather than in units of physical appearance.

3 Background and Data

3.1 Race in Brazil

Race in Brazil is shaped by the country's colonial history and its institutional legacy. Brazil's racial system originated during the colonial era, when Portuguese colonizers enslaved millions of Africans and forced Indigenous peoples into labor. The country was the last in the Americas to abolish slavery in 1888, but formal segregation policies after abolition never matched those of the United States. Yet, the absence of formal segregation did not eliminate racial hierarchy. Influential scholars such as Freyre (1933) advanced the idea of "racial democracy," suggesting that widespread miscegenation fostered more harmonious race relations than in the United States. Critics like Schwarcz (2013), Fernandes (1965) and Holanda (1936) highlighted persistent inequalities and argue that Brazil's so-called racial democracy obscured socioeconomic gaps separating White and non-White populations.

Unlike binary distinctions common in the United States, Brazil's racial classification operates along a continuum of phenotypical traits, creating ambiguity and context-dependence in how individuals are classified or classify themselves. In official statistics, the Brazilian Institute of Geography and Statistics (IBGE) recognizes five racial categories: *branco* (White), *preto* (Black), *pardo* (Brown), *amarelo* (Asian), and *indígena* (Indigenous). The *pardo* category covers a broad spectrum of individuals with mixed ancestry, reflecting the country's history of racial mixing. Scholars such as Telles (2004) argue that this category can conceal the scale of racial inequality, since many individuals with darker skin tones may identify as *pardo*, reducing the size of the population identifying as *preto* and complicating efforts to measure disparities. *Preto* and *pardo* workers are shown to exhibit relatively similar levels of education, employment, and wages, and are therefore often pooled into a combined non-White category (e.g., Miller and Schmutte, 2023; Gerard et al., 2021; Silva, 2000). Compared to White individuals, non-White individuals have substantially different socio-economic outcomes: they earn 35% less, are more likely to be unemployed, are underrepresented in leadership positions, and have fewer years of schooling (França and Portella, 2023; Gerard et al., 2021; IBGE, 2022; Firpo et al., 2021).

Moreover, race in Brazil is not static. Individuals and observers may classify or perceive the same person differently depending on economic conditions or institutional settings. In the last two decades, many Brazilians changed their self-identification towards *preto* and *pardo* (Miranda, 2015), possibly in response to increased opportunities and lower stigma (De Micheli, 2021) or policy incentives (Francis and Tannuri-Pianto, 2013). At the same time, improvements in socio-economic status lead self- and intergenerational-identification to move towards White (Cornwell et al., 2017; Schwartzman, 2007; Maggie and Fry, 2004).

3.2 Data

We combine information from two large-scale administrative data sources that allow us to observe three racial identity measures for the same individual. First, we draw on the *Relação Anual de Informações Sociais* (RAIS) that provides matched employer-employee data on all formal jobs in Brazil and, importantly, records workers' race *as reported by the employer*. Second, we merge RAIS with data on the universe of political candidates in Brazil, for whom we observe their *self-declared race* as well as an algorithmic measure of *skin tone* derived from standardized electoral photographs. This section describes the two data sources, the three racial identity measures, and the steps used to construct our estimation sample.

Employer-Ascribed Race. We use RAIS records covering all formal sector employment relations in Brazil between 2003 and 2015. The data is collected annually by Brazil's Ministry of Labor and Employment for the administration of various social security programs. All employers are mandated to submit information on their employees via an online platform provided by the Ministry (Appendix Figure C.1 shows a screenshot of the interface). Compliance with reporting requirements is high because incomplete submissions can trigger substantial penalties (MTE, 2024).[7] Reporting is done at the establishment level, by a dedicated HR or administrative employee, or by the owner or manager themselves in the case of smaller establishments (Silveira, 2022; Osório, 2003). Each record in the data corresponds to a worker-establishment contract in a given year and contains detailed information on average monthly earnings, contracted hours, occupation, contract type, the start and end date of the contract, along with workers' sociodemographic characteristics (gender, age, education) and the firm's industry and location. Workers are identified by their unique tax identifier *Código de Pessoa Física* (CPF) and their names, and can be tracked across employers, which are identified by their *Cadastro Nacional da Pessoa Jurídica* (CNPJ).

Crucially, RAIS collects data on employees' race since 1999. The race question asked in RAIS and the available racial categories are the standard ones in Brazilian statistics, as discussed in the previous section. There is no specific guidance on how to report a worker's race, either in the official RAIS Manual or in the reporting interface, and the reported race information is not subject to any systematic auditing. (MTE, 2024). Most employers do not seem to consult workers and instead classify race based on their own perceptions (Silveira, 2022; Osório, 2003). See Appendix C.1 for more details on how employers collect and report worker information. Importantly, there are also no direct economic incentives for reporting a more or less racially diverse workforce (e.g., no tax incentives or affirmative action laws in the private labor market). Hence, we do not expect strategic reporting in the sense of changes in employees' ascribed race to achieve a desired racial composition. At the same time, the presence of a race question in RAIS is not unusual for respondents, as Brazilians have long been accustomed to being asked about their race and that of others in official

[7] The Ministry of Labor conducts inspections as an enforcement tool. While the Brazilian labor market is characterized by numerous regulations that include minimum wage, maximum working hours, or disability quotas, there are no incentives for firms to report a certain racial composition (Szerman, 2022; Ponczek and Ulyssea, 2022). However, because RAIS reporting is high-stakes overall, firms pay close attention when completing these forms.

statistics.[8] Taken together, these features suggest that employers take reporting of race in RAIS seriously and that employer-reported race is a good proxy for how employers perceive the racial identity of their workers. Nevertheless, we will subject our main results to a series of robustness checks designed to explicitly account for potential measurement error in employer-ascribed race.

Self-Declared Race. To enrich RAIS with independent measures of racial identity, we link it to administrative records from Brazil's *Tribunal Superior Eleitoral* (TSE), which contain information on the universe of individuals who run for political office. Beginning in 2014, the TSE registry includes each candidate's self-declared race using the five standard categories. Importantly, only about 15% of all candidates are elected, and about 90% run for municipal council positions – typically part-time roles during which councilors maintain their private-sector jobs.[9] The strong labor market attachment makes the candidate sample well-suited for comparing self-declared and employer-ascribed racial identities. Candidates in the TSE records are identified by their tax identifier (CPF), enabling a deterministic linkage to RAIS and avoiding any potential mismatch in racial identities that might arise under probabilistic record-linkage procedures.

Skin Tone. The candidate registry also provides standardized photographs used in electronic voting machines for nearly all candidates between 2004 and 2024, yielding close to three million photos. We use these head-shots to construct an algorithmic measure of skin tone that is derived directly from the color value of each pixel in a photo. In contrast to existing measures in the literature, which typically rely on enumerator-reported skin tone (Woo-Mora, 2025; Bueno and Dunning, 2017; Telles, 2014), this approach delivers an objective and reproducible measurement of skin tone. This circumvents the issue that characteristics and experiences of the enumerator or respondents' other racial cues affect how racial cues are reported in the first place.

The algorithmic pipeline is illustrated in Figure 1. We first isolate the skin region in each photo using Google MediaPipe's Image Segmentation tool, which segments the skin part from other elements of the image (background, hair, clothing, accessories). We then summarize the color of the skin region by closely following the algorithm of Adukia et al. (2023). We use a k-means clustering approach to group pixels into $k = 5$ clusters and exclude the smallest two clusters that often refer to shadows or other non-skin facial features. For each remaining cluster, we compute its characteristic color and express these values in the perceptually linear $L^*a^*b^*$ color space.[10] A weighted average of these cluster colors produces a single representative skin color for each candidate. Appendix C.2 provides more information on the characteristics of electoral photos, details on the measurement procedure, as well

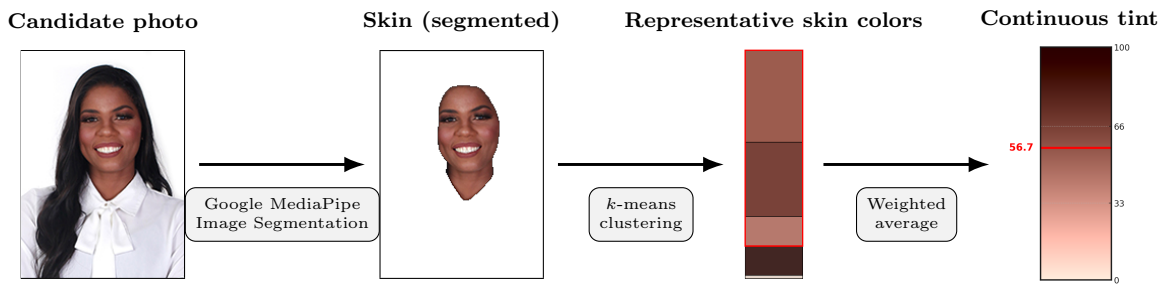
[8] The Brazilian Census Bureau asks respondents' race with the same alternatives as today since at least the 1960 census.

[9] On average, municipal councilors are required to attend the council only four days per month (Ferraz and Finan, 2011).

[10] The $L^*a^*b^*$ (or CIELAB) color space is a device-independent system defining colors with three values: $L*$ (lightness), $a*$ (red-green axis), and $b*$ (yellow-blue axis). Created by the International Commission on Illumination, it's perceptually uniform, meaning equal distances between colors represent similar perceived differences.

as a validation exercise that compares our algorithmic skin tone measure to assessments of candidates' skin tone by human raters.[11]

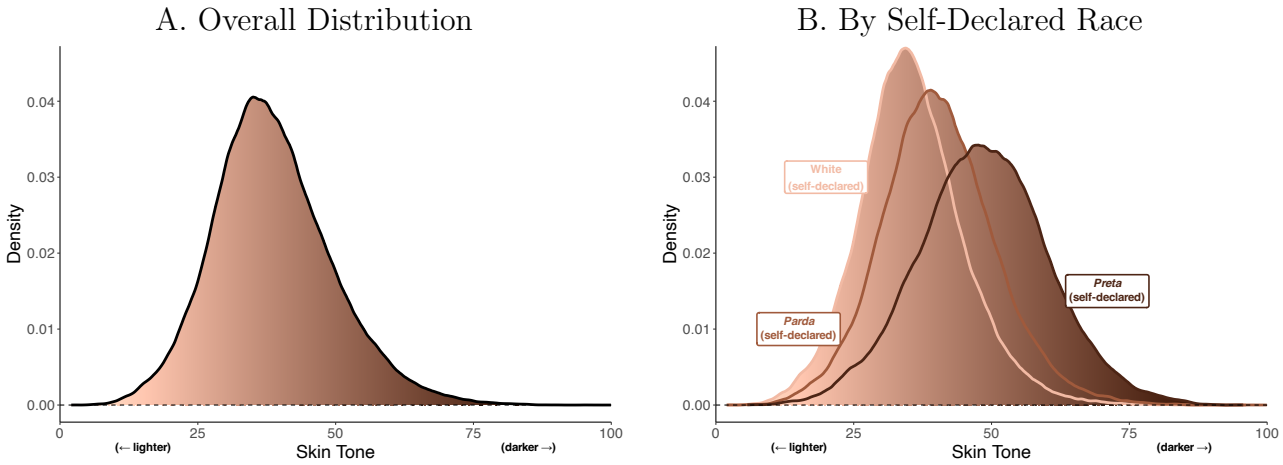
Figure 1: Skin Tone Extraction Pipeline



Notes: The figure illustrates how we extract skin tone from electoral photos. The example shown corresponds roughly to the 90th percentile of the skin tone distribution. Appendix Figure C.2 provides additional examples across the distribution.

Figure 2 depicts the distribution of skin tone in our data, separating self-declared White, *Parda* (Brown), and *Preta* (Black) candidates (Appendix Figure C.2 also shows example photos across the skin tone distribution). Skin tone has a mean of 38.6 and standard deviation of 10.8, with self-declared White and Brown (Brown and Black) candidates differing on average by 0.45 (0.75) standard deviations. Throughout the analysis, we standardize skin tone and report estimates in standard deviation units.

Figure 2: Skin Tone Distribution



Notes: The figure shows the distribution of skin tone as measured from the electoral photos of candidates. As will be done in all figures in paper, the skin tone distribution is visualized by varying the L^* value of a skin tone with $a^* = 17.2$ and $b^* = 19.1$, which correspond to the sample averages of a^* and b^* across all our colored photos. Panel A depicts the skin tone distribution among all candidates. Panel B splits candidates by their self-declared race, distinguishing White, *Pardo* (Brown), and *Preto* (Black) candidates.

[11] Data on the race assessment for a sub-sample of candidates among human raters are collected by Bueno and Dunning (2017). In addition, about half of all candidate photos in our estimation sample are in grayscale. In Appendix C.2, we assess the measurement quality across grayscale and colored photos, providing evidence that motivates the pooling of both types of photos in our main analysis.

Estimation Sample. Our estimation sample consists of approximately 330,000 unique individuals who appear as political candidates at least once between 2014 and 2024 and who can be linked to at least one formal work contract in RAIS. For these matched individuals, we construct a yearly panel capturing all jobs of an individual between 2003 and 2015.[12] Our analysis compares employer-ascribed race in these jobs to the candidate's self-declared race and skin tone measured at the time of candidacy.[13] If an individual has run for a political office multiple times, we assign their modal self-declared race (across 2014-2024) and their mean skin tone (across 2004-2024). This approach implicitly treats self-identity and skin tone as time-invariant individual characteristics, which we compare to potentially divergent racial perceptions of different employers.[14] In robustness checks, we relax this assumption and restrict comparisons to periods in which the different racial identity measures overlap in time.

Appendix Table B.1 presents summary statistics for our estimation sample of political candidates and compares them with the full RAIS workforce. Relative to other workers, candidates are more likely to be male (by 13 percentage points) and slightly older (by 0.84 years), but they are similar in education and average earnings. Moreover, as shown in Appendix Table B.2, the wage gap between White and non-White workers (based on employer-ascribed race) is 10 percentage points smaller among candidates than among other workers, suggesting that our estimated disparities likely represent lower bounds for population-level gaps. Although political candidates are not representative of the Brazilian workforce, this sample is uniquely suited for our purposes because it allows us to compare three distinct measures of racial identity: race as perceived by the employer, self-declared racial identity, and skin tone as a phenotypical marker.

4 Mismatched Racial Identities

In this section, we document how employer-ascribed race differs from self-declared race and how the same worker is classified across employers. Throughout, we focus on the binary distinction between White and non-White. Appendix Table B.3 also reports results when disaggregating the non-White category into *pretos* and *pardos*.

[12] We focus on private-sector, full-time ($\geq$ 30 hours/week), open-ended contracts held by individuals aged 20 to 54. For each year, we retain the primary job, defined as the contract with the highest annual earnings.

[13] We exclude observations where either the employer-ascribed or self-declared race is Asian or Indigenous, following standard practice in the literature (e.g., Gerard et al., 2021). Asian and Indigenous individuals make up 1.05% of the population of formal workers and are geographically concentrated. Moreover, the formation of racial identities and perceptions for these categories may be different from that of Black identity.

[14] We acknowledge that individuals' self-identification can change over time. Among individuals who run for a political office multiple times between 2014 and 2024, we observe a change in self-declared race (between White and non-White) across subsequent elections in 16.9% of cases. This is substantially lower than the mismatch rates between self-declared and employer-ascribed race and between ascribed race across subsequent employers that we will document in Section 4. While a large literature has documented the fluidity of self-identity (Saperstein and Penner, 2012; Liebler et al., 2017; Miranda, 2015; Muniz et al., 2024), in our paper we focus on the construction of racial perceptions by employers.

Self-Declared versus Employer-Ascribed Race. Panel A of Table 1 cross tabulates self-declared and employer-ascribed race. Roughly 73% of workers fall on the diagonal, that is, they self-identify and are classified by their employer as belonging to the same racial group (White or non-White). The remaining 27% exhibit a mismatch between self-identification and employer-ascription. The mismatch is asymmetric. Among self-declared Whites, about 15% are recorded as non-White by their employer. Among self-declared non-Whites, about 41% are reported as White. In other words, conditional on how workers describe themselves, employers are much more likely to "whiten" non-White workers than to "darken" self-identified Whites.

Skin tone helps to interpret these discrepancies. Within each cell of Panel A, we report the standardized skin tone measure. Workers who self-identify as non-White and are also classified as non-White have, on average, the darkest skin tones. Self-identified Whites who are classified as White have the lightest skin tones. The two off-diagonal cells, where self-identification and employer-ascription differ, are characterized by intermediate skin tone levels. Panel (a) of Appendix Figure A.3 shows the full skin tone distribution, confirming that mismatches concentrate among workers with mid-range skin tones rather than at the extremes of the distribution. Notably, the average tendency of employers to "whiten" rather than "darken" their workers does not appear to be driven by skin tone differences: self-identified non-Whites classified as White by their employer have skin tones that are 0.28 SD darker than those of self-identified Whites classified as non-White.

Taken together, these facts show that self- and employer-perceptions are tightly linked but not interchangeable. For many workers, especially those with ambiguous skin tone, the racial category encoded in administrative data is not a mechanical reflection of their self-understanding.

Mismatch across Employers. We next examine how the same individual is classified across employers. Panel B of Table 1 restricts the sample to workers who change jobs and compares employer-ascribed race in the previous and current establishment across consecutive years. About 77% of job switchers are assigned the same racial category by both employers. The remaining 23% change racial classification when they move. As in Panel A, the off-diagonal cells are sizable on both sides. Roughly 30% of workers who were previously classified as non-White are recorded as White by their new employer, while close to 20% of previously White workers are coded as non-White in their subsequent job. Panel (b) of Appendix Figure A.3 shows that mismatches across employers are again concentrated among individuals with intermediate skin tones.[15]

These patterns indicate that employer-ascribed race is not a fixed attribute of the worker but the outcome of an interaction between the worker's appearance and the classification practices of each firm. To assess the role of firms' racial classification practices in more detail, we exploit workers' moves across firms in the full RAIS sample. Specifically, we run an AKM model (Abowd et al., 1999) that regresses an indicator for being ascribed non-White by the current employer on a full set of worker and firm fixed effects. This allows

[15]While Table 1 focuses on the matched sample of political candidates, Appendix Table B.4 shows that the mismatch rates across employers are very similar within the full sample of workers in RAIS, suggesting that the candidate sample is not atypical in terms of racial ambiguity.

Table 1: Mismatch in Ascribed Race

[A] Mismatch between Self-Declared and Employer-Ascribed Race

	Employer-Ascribed Race			
	White		Non-White	
Self-Declared Race	%	Avg. Skin Tone	%	Avg. Skin Tone
White	46.48%	-0.31	8.43%	-0.10
Non-White	18.63%	0.18	26.46%	0.45

[B] Mismatch across Subsequent Employers

	Employer-Ascribed Race, Current			
	White		Non-White	
Employer-Ascribed Race, Previous	%	Avg. Skin Tone	%	Avg. Skin Tone
White	49.82%	-0.22	12.18%	0.15
Non-White	10.88%	0.15	27.12%	0.40

Notes: The table displays the (mis)match shares for different race measures in our sample, as well as the average skin tone for each cell. Skin tone is standardized to have a mean of zero and a standard deviation of one across the whole sample (higher values represent darker shades, while lower values represent lighter shades). Panel A presents the distribution of workers by self-declared race (rows) and employer-ascribed race (columns). It considers all worker-year observations ($N = 1,510,833$). Panel B presents the distribution of ascribed race for the same worker across different employers in consecutive years. It considers all workers who changed jobs ($N = 153,366$) and presents the distribution of workers by their ascribed race in the previous employer (rows) and current employer (columns).

us to separate variation in firm-specific practices from variation in workers' portable racial cues. Appendix D provides details on the empirical specification, tests of the identification assumptions, and estimation results. The variance decomposition shows that only about half (46.9-51.5%) of the variation in employer-ascribed race is explained by worker effects – i.e., characteristics workers carry with them across jobs – while firm effects account for another third (30.3-34.0%) of the variation (see Appendix Table D.1).[16] The magnitude of firms' contribution is large, even higher than their role in explaining the variance of wages in the Brazilian labor market (Gerard et al., 2021). This underscores that firms play a quantitatively important role in the racial ascription of workers: some firms systematically "whiten" their workforce, whereas others tend to "darken" them.

Interpreting Mismatch and Measurement Concerns. A natural concern is that the observed mismatches may simply reflect noisy reporting rather than differences in latent race concepts – for instance, if employers pay little attention to reporting in RAIS. In the following, we discuss several pieces of evidence indicating that measurement error alone is unlikely to account for the observed mismatch.

[16]The remaining variation is largely captured by the residual, given a very small covariance between worker and firm effects. When correcting for limited mobility bias in a leave-one-out connected set following the approach of Kline et al. (2020), the share of variation explained by firm effects decreases slightly (to 25.5-27.3%), while the share explained by worker effects is reduced even more (to 32.1-33.3%,) and the positive covariance between worker and firm effects becomes more pronounced (8.4-9.7%) (see Appendix Table D.2).

First, as noted above, the mismatch is concentrated among intermediate skin tones, suggesting that self- and employer-ascribed racial categories diverge when phenotypical features allow for ambiguity. This is inconsistent with a scenario whereby employers record race with noise uniformly across the skin tone distribution. Instead, it suggests systematic divergence in ascribed race in cases where visual cues are ambiguous.

Second, changes in workers' racial ascription are tightly linked to job mobility. As shown in Appendix Figure A.4, about 80% of race changes in the full RAIS occur following a job change, in particular in more recent years. In contrast, there is little variation in reported race within ongoing worker-firm matches. The AKM decomposition has shown that firms differ in their propensity to "whiten" or "darken" workers relative to how they are typically seen in the labor market. Together, these patterns are consistent with systematic heterogeneity in race perceptions across employers, rather than idiosyncratic reporting error within firms.

Third, we show that the mismatch patterns are similar across sub-samples that should exhibit higher reporting quality. Appendix Table B.4 repeats the comparison between self-declared and employer-ascribed race under a series of restrictions. We begin by excluding firms that uniformly assign the same race to their entire workforce.[17] Because these firms might plausibly be less attentive in their reporting, we re-estimate mismatch rates in the subset of establishments where we observe workers assigned to different racial groups. We then restrict the sample to firms with dedicated human-resources staff to address concerns that variation in classification firms' capacity to collect information on workers. We also limit the sample to worker-year observations in which both workers and employers report race in the same period, ruling out that discrepancies arise from changes in self-declared race over time. Finally, we focus on candidates for local councilor positions, among whom potential strategic incentives to misreport self-identified race are arguably weaker than for higher political offices. Across all these robustness checks, both the magnitude and the asymmetry of mismatches remain virtually unchanged. As an additional benchmark, the same table reports mismatches in gender, showing that disagreement between self-declared and employer-reported gender and changes in reported gender across employers are negligible. If employers filled out workers' demographic information in RAIS with large inattention, we would expect to observe similar mismatches in gender reporting, which we do not.

Fourth, we exploit the presence of multiple employer observations for many workers and aggregate employer-ascribed race across jobs. Appendix Table B.5 compares several measures of individual "market race" – including the modal employer classification and the share of employers that code the worker as non-White – to self-declared race.[18] To address measurement error in these aggregated measures, we estimate split-sample IV regressions in which the measure constructed from one random half of a worker's yearly observations is instrumented with the measure constructed from the other half. Even after averaging across employers and applying measurement error corrections, we find that the association between employer-ascribed and self-declared race remains far from one-to-one.

Overall, the evidence points to a world in which racial ascription has both a person

[17] Most establishments do not assign the same racial category to all their workers: Appendix Figure A.1 shows that less than 8% of all workers are employed in fully-segregated establishments that classify their entire workforce as exclusively White.

[18] Using the modal classification is a standard approach in studies with RAIS data (e.g., Gerard et al., 2021).

component and an employer component. Workers carry with them a bundle of traits that makes them more or less likely to be seen as non-White, but firms differ in how they map those traits into racial categories. Self-identification adds a third layer, which often aligns with employer perceptions but diverges precisely where classification is most ambiguous. In the next section, we examine the determinants of racial ascription by employers.

5 Race Function

In the previous section we documented that self-declared and employer-ascribed race often diverge, and that the same worker can change racial category when moving across firms. These patterns suggest that race can be ambiguous and the outcome of a classification rule that employers apply to observable worker characteristics. In this section we characterize this employer "race function." We first relate the probability that a worker is coded as non-White to continuous skin tone, both in the full sample and within self-declared racial groups. We then show how this mapping shifts with other worker characteristics, making explicit how employers construct racial categories from workers' skin tone, self-identity, and socio-economic status, as well as the social context that employers face.

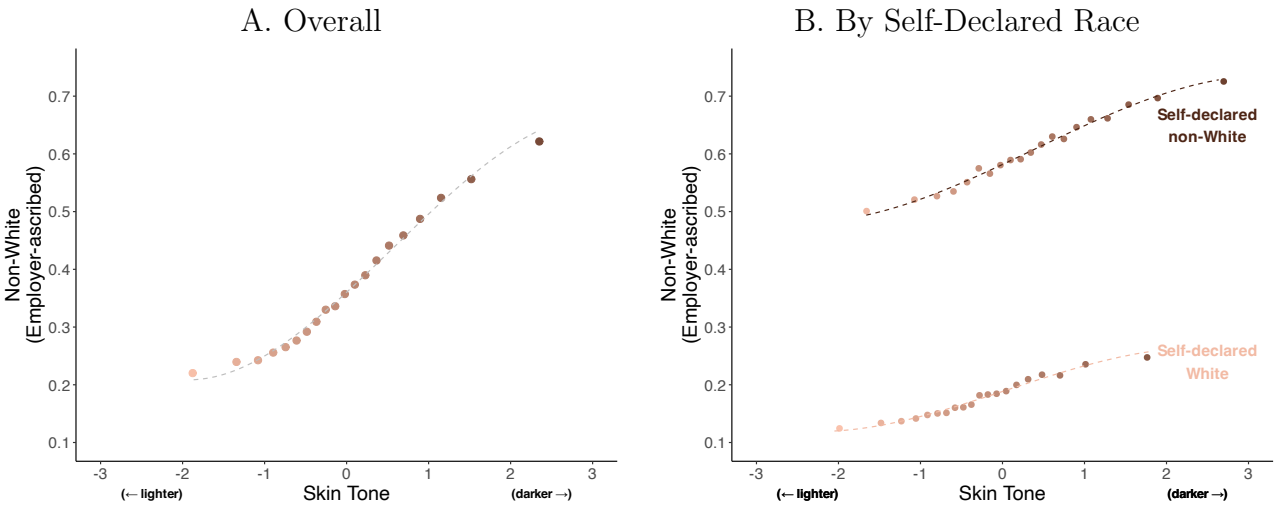
Skin Tone and Self-Identification Figure 3 shows that skin tone is a central input into employers' racial ascription. Panel (a) shows a binned scatterplot of the probability that a worker is coded as non-White against skin tone. The relationship is steep and monotone: darker-skinned workers are much more likely to be classified as non-White, while very light-skinned workers are usually classified as White. Moving from the 10th to the 90th percentile of skin tone increases the likelihood of being ascribed non-White by 32 percentage points. Panel (b) splits the sample by self-declared race and shows that this pattern holds even within self-identification. Among workers who identify as White, the likelihood of being ascribed non-White rises with darker skin tone. Among workers who self-identify as non-White, lighter-skinned individuals are less likely to be coded as non-White. Employers therefore do not simply reproduce how workers describe themselves but they incorporate their own visual impressions into racial classification.

At the same time, skin tone is far from a complete input of the race function. Both panels reveal substantial heterogeneity in the tails of the distribution. About 20% of the lightest-skinned workers are still classified as non-White, while over 30% of the darkest-skinned workers are coded as White. The patterns suggest that other worker characteristics also shape racial ascription even at a given level of skin tone.[19]

Education and Context. Figure 4 illustrates that worker characteristics and local social context systematically shift how employers map skin tone into racial categories along two central dimensions: education and region. We focus on how these factors reshape the race-skin tone relation, considering both its level and its slope. The level captures who is

[19] Alternatively, measurement error from our skin tone algorithm could explain some of the residual variation. We discuss measurement error in skin tone in more detail in Section 6.

Figure 3: Relationship between Employer-Ascribed Race and Skin Tone



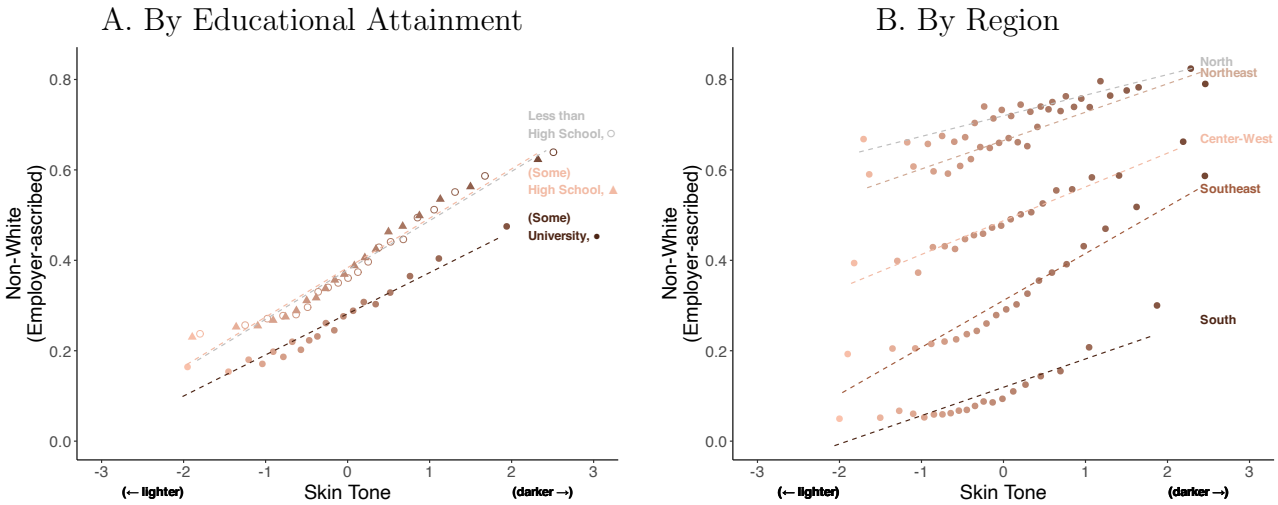
Notes: The figures show binned scatterplots of employer-ascribed race (an indicator equal to one if the worker is ascribed non-White) and workers' skin tone (standardized), both unconditionally (Panel A) and splitting the sample by self-declared race (Panel B). The plots are constructed following the procedure in Cattaneo et al. (2024). For each sample, we divide the skin tone distribution in twenty equal-sized bins and, for each bin, plot the average of employer-ascribed race. We fit third-degree polynomials based on the underlying data of each sample.

considered White or non-White at a given tone, while the slope captures how strongly employers translate incremental skin tone differences into racial categories. Together, these two components reveal not only who is coded as what, but how sensitive racial ascription is to variations in phenotype.

Panel (a) shows that education changes the entire race function. Among workers with low schooling, the probability of being classified as non-White rises sharply with darker skin tone. For university-educated workers, the gradient is flatter and uniformly lower, especially at the darker end of the skin tone distribution. More educated workers are therefore less likely to be coded as non-White than workers of the same skin tone who have less schooling. This pattern indicates two mechanisms: education raises the likelihood of being coded as White at any given skin tone (a lower threshold for Whiteness) and it attenuates the weight employers place on skin tone (a flatter gradient).

Panel (b) shows that regional context also alters the mapping from skin tone to race. For a given skin tone, employers in the North classify workers as non-White at substantially higher rates than employers in the South, with the Northeast, Center-West and Southeast in between. This suggests that "Whiteness" includes individuals with much darker skin tones in the South compared to the North. This classification practice reflects the local racial context: in the North, where the Afro-descendant population is largest, the racial norm is non-White and only very light-skinned workers stand out as White. In the South, shaped by European immigration, the racial norm is White, and only very dark-skinned workers are classified as non-White. The curves also differ in slopes: the Southeast exhibits a steeper gradient, suggesting that small tone differences carry more weight in racial ascription. Again, this is in line with the local racial context in the Southeast, characterized by a highly ethnically diverse population and thus sensitivity to racial categories. Together, both panels

Figure 4: Relationship between Employer-Ascribed Race and Skin Tone, by Education and Region of Birth



Notes: The figures show binned scatterplots of employer-ascribed race (an indicator equal to one if the worker is ascribed non-White) and workers' skin tone (standardized), splitting the sample by workers' educational attainment (Panel A) and region (Panel B). The plots are constructed following the procedure in Cattaneo et al. (2024). For each sample, we divide the skin tone distribution in twenty equal-sized bins and, for each bin, plot the average of employer-ascribed race. We fit third-degree polynomials based on the underlying data of each sample.

underscore that racial ascription in Brazil's labor market is not a mere reflection of phenotype but is mediated by worker characteristics and local racial context, which jointly shape how employers interpret the same underlying racial cue.

Multivariate race function. To quantify how employers translate multiple worker attributes into racial classifications, we estimate a multivariate "race function" using a linear probability model. Table 2 reports coefficients relating employer-ascribed race to a broad set of worker and labor market characteristics. All specifications include year fixed effects to absorb changes in racial coding across all worker employer pairs over time. Our goal is twofold: to isolate the direct effects of these characteristics and to examine how the coefficient on skin tone evolves as richer controls and fixed effects are added.

We begin with the direct effects. Column 1 reports the unconditional association between skin tone and employer-ascribed race: a one-standard-deviation increase in skin tone raises the probability of being classified as non-White by 11 percentage points. Column 2 introduces demographic controls, that is, education, gender, and a squared age term. Conditional on skin tone, university graduates are 10 pp less likely to be coded as non-White than those with less than high school education, which confirms that status signals shape race perceptions. Column 3 adds regional fixed effects. Relative to the North and the Northeast, where the shares of Afro-descendant and Indigenous populations are highest, employers in all other regions whiten workers more readily at a given skin tone.

Columns 4 and 5 introduce progressively richer labor-market fixed effects, which allow us to compare workers who face similar task requirements, pay scales, and employer environments but differ in demographic and phenotypic attributes. Column 4 includes industry fixed

effects; column 5 adds 3-digit occupation by industry fixed effects.[20] Across all these specifications, the qualitative pattern remains: education decreases the likelihood of being classified as non-White conditional on skin tone, even within narrowly defined labor market segments. These results support a central claim of the constructivist approach: productivity-relevant characteristics shape race perceptions. To substantiate this claim further, we leverage information from previous employers for a sub-sample of workers who hold multiple jobs over our observation period. Appendix Table B.6 replicates Table 2 but adds labor market seniority (tenure conditional on age) and the employee's final wage at their previous job to the race function. Seniority reduces the likelihood of being classified as non-White, and higher wages at the previous firm are also associated with whitening, providing direct evidence of the productivity-racialization path.

Notably, column 6 includes establishment-by-3-digit-occupation fixed effects, effectively holding the employer, i.e., the interpreter of race, constant and exploiting variation in phenotypical and status cues among workers that hold the same job. Darker-skinned individuals remain more likely to be classified as non-White, but the magnitude of the skin tone coefficient decreases by approximately 40% in column 6 compared to column 1, and by 13% compared to column 5. This attenuation indicates that some of the unconditional correlation could reflect sorting across more or less whitening employers. At the same time, the large remaining effect size within establishments shows that racial categorization is specific to worker-employer pairs, thus operating at a fine-grained level.

Finally, column 7 introduces self-declared race. Although self-identification may itself be influenced by the same factors that shape others' perceptions, and may therefore be a "bad control", our goal here is to hold constant workers' own identity and examine the remaining role of phenotype and demographics in generating divergence between employer and worker perceptions.[21] As expected, self-identifying as non-White increases the likelihood of being coded as such by employers, either because workers and employers interpret the same underlying cues similarly or because workers signal their identity in ways that employers (partially) incorporate. The coefficient on skin tone falls by more than half once self-identification is included, consistent with shared underlying determinants of racial identity, as well as with the idea that employers interpret phenotypical cues independently of workers' stated identities. Yet, even when holding self-identification constant, employer-ascribed race is affected by skin tone and by other social cues, such as education. This reinforces the conclusion that self-identification and perception by employers are not the same object; indeed, employer-ascribed race still responds to phenotypical and social cues even conditional on self-perception.

[20] We consider 9 industry categories following the Brazilian national classification of economic activities: agriculture, extractive industries, manufacturing, construction, retail, transport & communication, finance & real state, education & health, and other services. Three-digit occupations correspond to the 192 subcategories in the Brazilian Occupation Classification (2002 CBO).

[21] We estimate a separate race function for self-identification in Appendix Tables B.7 and B.8, showing that self-identification places different weights on observable characteristics and thus emphasizing that self-identification and employer perception are related but distinct outputs of different race functions.

Table 2: Correlates of Employer-Ascribed Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Skin Tone	0.110***	0.107***	0.082***	0.081***	0.078***	0.068***	0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Self-Declared							0.235***
							(0.003)
Demographics							
Male		0.002	-0.004**	-0.011***	-0.013***	-0.014***	-0.012***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Age		-0.025***	-0.014***	-0.016***	-0.013***	-0.012***	-0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2		-0.007***	-0.003***	-0.003***	-0.002***	-0.002**	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Some) High School		-0.009***	-0.034***	-0.027***	-0.012***	-0.007**	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
(Some) University		-0.096***	-0.115***	-0.110***	-0.058***	-0.032***	-0.014**
		(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)
Region							
Northeast			-0.052***	-0.053***	-0.055***		
			(0.004)	(0.004)	(0.004)		
Southeast			-0.400***	-0.398***	-0.399***		
			(0.003)	(0.003)	(0.003)		
South			-0.584***	-0.578***	-0.578***		
			(0.003)	(0.003)	(0.003)		
Center-West			-0.226***	-0.224***	-0.223***		
			(0.004)	(0.004)	(0.004)		
Constant	0.370***	0.392***	0.752***	0.750***	0.737***	0.390***	0.277***
	(0.001)	(0.002)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No	No
Industry × Occupation FE	No	No	No	No	Yes	No	No
Firm × Occupation FE	No	No	No	No	No	Yes	Yes
R-squared	0.057	0.065	0.209	0.213	0.223	0.779	0.793
Adjusted R-squared	0.057	0.065	0.209	0.213	0.222	0.719	0.736

Notes: The table reports regression estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and skin tone. All columns include year fixed effects. The first column displays the unconditional relationship (apart from year fixed effects), while the remaining columns progressively add controls. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Racial Wage Disparities

Our results so far are consistent with a constructivist view of race: racial identities diverge between workers and employers and across employers. Moreover, racial ascriptions are a function of individuals' skin tone, self-identification, socio-economic status signals, and local social norms. In the following, we shed light on the implications of this constructivist view of race for measures of racial disparities in the labor market.

6.1 Conventional Estimates

Disparities across Racial Groups. We start by quantifying racial wage gaps that are commonly estimated in the literature. Panel A of Table 3 shows differences in log hourly wages between workers who are ascribed White vs. non-White by their employer, following existing studies on racial wage gaps with the RAIS data. The unconditional wage gap in our sample, reported in column 1, is 14.8%.[22] In Appendix Table B.10, we compare wage gaps in terms of employer-ascribed race to wage gaps in terms of workers' self-declared race within the same sample of workers. In fact, almost all racial wage gap estimates in the literature rely on self-reported information. The unconditional wage gap for self-declared race lies at 16.3% and is thus modestly larger than for employer-ascribed race.

While unconditional wage gaps are thought to capture the disparate impact of race that entails all forms of discrimination – both direct and systemic ones – researchers often run "benchmarking regressions" in order to account for potential productivity differences across workers and quantify disparate treatment by employers based on race (Blinder 1973; Corcoran and Duncan 1979; Oaxaca and Ransom 1994; Darity, Guilkey, and Winfrey 1996). To examine this aspect, we estimate conditional wage gaps, successively adding the same covariates and fixed effects discussed in Section 5 (and shown in Table 2). We find that the wage gap based on employer-ascribed race shrinks by almost two thirds when controlling for sociodemographic characteristics and region fixed effects and further declines when accounting for sorting across industries and occupations. Finally, column 6 shows that employees working in the same firm and 3-digit occupation receive a 2.1% lower wage when they are ascribed non-White by their employer.[23]

In a last step, we run a horse-race regression, jointly including employer-ascribed and self-declared race in the most stringent model with firm-by-occupation fixed effects. Column 7 reveals that both race concepts matter independently for wages: self-identifying as non-White is associated with 1.5% lower wages, while being perceived as non-White by the employer comes with an additional wage penalty of 1.7%. These results show that the racial mismatch documented in Section 4 translates into wage disparities: two workers that both self-identify as White and both work in the same firm and occupation, receive different wages if one of the workers is perceived as non-White by the employer.

The results from Section 5 suggest that these conventional disparity measures should be interpreted with caution. Employers' race perceptions and individuals' self identification are determined by factors that are productivity-relevant, such as their education or previous wages. Thus, the same set of covariates that enter the wage setting equation also shape the very race measures typically used to quantify disparities. The endogeneity of race creates a fundamental identification problem for measuring disparities across identity groups. While benchmarking regressions account for some of the status cues that shape race perceptions, there may be other unobserved factors that correlate with both productivity and employer-

[22] We always condition on year fixed effects. Our wage gap estimate is smaller than typical estimates in the literature (e.g., Gerard et al., 2021). Appendix Table B.2 shows that our sample of candidates exhibits smaller racial wage gaps (in terms of employer-ascribed race) than the full RAIS sample, suggesting that all our result should be viewed as lower bounds for wage disparities in the broader population.

[23] After adding the fine-grained set of firm-by-occupation fixed effects for self-identified race, we estimate a slightly smaller wage gap of 1.9%.

recorded race. More fundamentally, if one is interested in estimating the disparate impact of race perceptions that does not control away some of the differences arising from systemic discrimination experiences – such as in the education system – one would have to address the endogeneity issue even in the unconditional estimate.Another concern arises from potential measurement error: If race perceptions are measured with noise – in case employers are inattentive in reporting the racial identity of their workers – even the conditional estimates will be attenuated.

Disparities by Skin Tone. In order to solve the endogeneity of race perceptions, a common approach is to use specific racial cues that plausibly affect the likelihood of being seen as non-White. Correspondence experiments, for example, have used name-based markers of racial and ethnic background in fake resumes to study their impact on hiring decisions (e.g., Bertrand and Mullainathan, 2004; Kline and Walters, 2021; Kline et al., 2022, 2024). Another prominent racial cue in the colorism literature is skin tone (e.g., Woo-Mora, 2025; Hersch, 2024; Abramitzky et al., 2023). Most existing studies use interviewer assessed skin tone Adukia et al. (2025); Nix and Qian (2015); Noghanibehambari and Fletcher (2025), which may be influenced by productivity-relevant factors and status cues and thus subject to the same endogeneity concerns. In contrast, the standardized photographs allow us to circumvent this issue and leverage an algorithmic measure of skin tone instead.

Figure 5 presents a binned scatterplot for the relationship between wages and skin tone. We find that wages decrease almost monotonically across the skin tone distribution.[24] Panel B of Table 3 shows results from linear regressions of wages on skin tone. On average, we find that a one standard deviation darker skin tone is associated with receiving 2.6% lower wages. This relation changes very little when controlling for demographic characteristics and region and industry fixed effects, but shrinks when adding more fine-grained fixed effects. Comparing workers in the same firm and occupation and with the same demographics, we find that a one standard deviation darker skin tone is still significantly related with a lower wage of about 0.5%. These results are consistent with the findings of the colorism literature that has documented differences across labor market outcomes for individuals with varying skin tones. Our findings add to this literature by documenting the relevance of skin tone for wages across workers in the same firms and jobs. Analogous to Panel A, column 7 of Panel B suggests that a worker with a one standard deviation darker skin tone will receive a 0.3% lower wages compared to a worker in the same firm-by-occupation cell and same racial self-identification.

Interpretation of Conventional Measures. The interpretation of these wage differences by workers' skin tone is complicated by the fact that they do take into account how skin tone translates into the perceptions of employers about the racial identity of their workers when making decisions about their wages. More generally, the impact of any racial cue should be interpreted as an intent-to-treat "effect" that may be higher or lower depending on how much of a racial contrast it generates in the perception of the decision maker. The race function results in Section 5 have revealed one important challenge when it comes to

[24]For the lightest skin tones we find that the relation is less clear. Measurement error in skin tone could explain this behavior at the tails of the distribution (see our discussion of measurement error below).

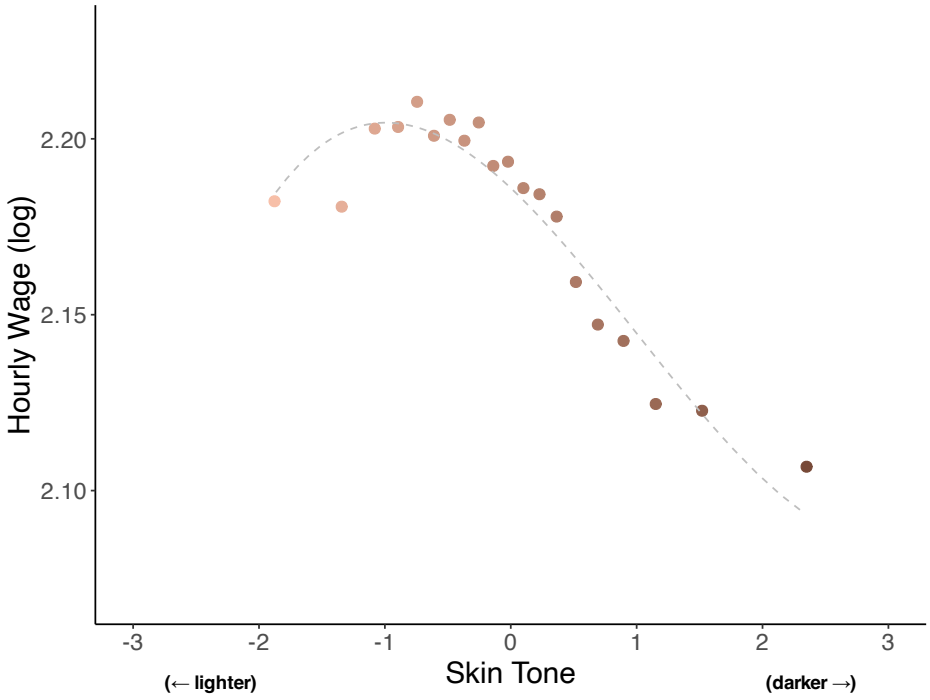
Table 3: Wage Disparities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
[A] OLS: Conventional Disparity							
Non-White Employer-Ascribed	-0.148***	-0.103***	-0.054***	-0.064***	-0.038***	-0.021***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Non-White Self-Declared							-0.015***
							(0.002)
[B] Reduced Form: Colorism Disparity							
Skin Tone	-0.026***	-0.032***	-0.026***	-0.028***	-0.015***	-0.005***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Self-Declared							-0.017***
							(0.002)
[C] First Stage: Race Function							
Skin Tone	0.110***	0.107***	0.082***	0.081***	0.078***	0.068***	0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Self-Declared							0.235***
							(0.003)
[D] 2SLS: Perception-Normalized Disparity							
Non-White Employer-Ascribed	-0.235***	-0.301***	-0.314***	-0.348***	-0.193***	-0.080***	-0.084***
	(0.011)	(0.010)	(0.014)	(0.014)	(0.012)	(0.013)	(0.022)
Non-White Self-Declared							0.003
							(0.006)
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No	No
Industry × Occupation FE	No	No	No	No	Yes	No	No
Firm × Occupation FE	No	No	No	No	No	Yes	Yes

Notes: The table reports different estimates of racial wage gaps based on employer-ascribed race. All regressions include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. The final column includes all fixed-effects, as well as controls for self-declared race. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced form regressions of log hourly wage on skin tone. Panel C presents the first stage coefficient from regressing employer-ascribed race on skin tone. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by skin tone. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

interpreting typical benchmarking regressions. As shown in Table 2 (and replicated in Panel C of Table 3), the relation between skin tone and employer-ascribed race shrinks by almost half when going from an unconditional contrast to one controlling for demographics and detailed firm × occupation fixed effects. This attenuation in the first-stage impact of skin tone highlights that the previous skin tone wage differentials have to interpreted carefully. The decline in wage disparities by skin tone after adding controls could not only be explained

Figure 5: Relationship between Wages and Skin Tone



Notes: The figure shows a binned scatterplot of log hourly wage on worker's skin tone (standardized). The plots are constructed following the procedure in Cattaneo et al. (2024). We divide the skin tone distribution in twenty equal-sized bins and, for each bin, plot the average log hourly wage. We fit a third-degree polynomial based on the underlying data.

by making workers more similar in terms of productivity-relevant characteristics, but also by the declining relevance of skin tone for how employers distinguish workers into racial groups. Consequently, benchmarking regressions do not only run into the risk of overcontrol when it comes to identifying the disparate impact of skin tone – due to the introduction of bad control concerns (Ayres, 2020) – but also when researchers seek to identify the disparate treatment of skin tone – due to changes in how skin tone differences map into race perceptions under conditional contrasts Rose (2023).

Another concern arises from the possibility the our skin tone measure could be noisy. While the algorithmic photo-based measurement is an objective approach to identify workers' skin tone, it can be confounded, for example, by variation in lighting across photos or non-skin facial features (eyes, teeth, etc.). Note that classical measurement error may not only be a concern for our algorithmic skin tone measure but also for interviewer assessment of workers' skin tone. As we discuss in Appendix C.2, there is strong variation in how humans rate the skin color of political candidates. To examine the potential attenuation arising from noisy measures of skin tone in our data, we implement a split-sample IV approach. For that, we focus on a sample of individuals who have run for political office multiple times, i.e. for whom we have several electoral photos. We use the average skin tone from one randomly selected half of the individuals' photos and instrument it by the average skin tone measured from the other half of photos. Results are reported in Table B.9. We find a substantial role of measurement error in attenuating skin tone differences. The split-sample IV results

show that a one standard deviation darker skin tone is associated with a 34.6 percentage point increase in the likelihood to be reported non-White by the employer and with a decline in wages by 11.4%. The IV estimates are more than twice as high compared to the OLS results in the same sample of repeated candidates, highlighting that measurement error can a serious concern when interpreting the effects of racial cues, such as skin tone.

6.2 Perception-Normalized Disparities

PND Estimates. The findings motivate an approach that takes into account how racial cues map into race perceptions when measuring disparities. In Section 2, we have introduced the concept of perception-normalized disparities (PND). PND expresses outcome differences by skin tone in units of the implied racial contrast arising from skin tone differences. In other words, we scale the reduced-form relation of skin tone with wages by the first-stage impact of skin tone on employers' race perceptions using a 2SLS estimation. In Panel D of Table 3, we show the resulting estimates. We find an unconditional PND of -0.235, indicating that workers who are perceived non-White by their employer as a consequence of their skin tone receive 23.5% lower wages than workers whose skin tone makes them be perceived White by the employer. Thus, measuring the impact of skin tone on wages in terms of the implied race perception leads to a disparity that is almost 60% larger compared to the direct contrast by perceived race from the OLS estimate in Panel A.

The usefulness of the PND approach becomes particularly clear when when turning towards conditional contrasts. In the most stringent specification that compares workers with the same demographics who work in the same job and self-identify with the same racial category, we estimate a PND of 8.4% (see column (7)) which is a reduction of 64% relative to the unconditional contrast. In comparison, the reduced-form impact of skin tone on wages changes from -0.026 (see column (1)) to -0.003 (see column (7)), i.e. shrinks by about 88%. The difference in these changes is explained by the attenuation in the first-stage relationship between skin tone and employers' race perception which the PND estimates effectively accounts for. Notably, columns (2) to (4) also show that conditional PND estimates can be substantially larger than the unconditional estimate, while the corresponding conditional skin tone differentials remain similar to the unconditional one, highlighting how the same reduced-form impact of a racial cue can have very different interpretations when taking into account their implications for race perceptions.

Furthermore, the PND estimates also effectively take into account any measurement error in skin tone. To see this, we can go back to the split-sample IV estimates from Table B.9) of how race perceptions and wages are impacted by skin tone after purging measurement error. The estimates would imply a PND of $-0.114/0.346 = -0.329$. It is the same as the PND that would result when using the uncorrected OLS estimates of the impact of skin tone on race perceptions (15.2 percentage points) and wages (-5.1%).[25] Measurement error attenuates the reduced-form and the first-stage impacts of skin tone to the same extent, and the PND estimate adjusts for this attenuation as it scales the reduced-form by the first-stage.

In Appendix Table B.10, we also perform an alternative scaling exercise in which we

[25]The PND differs from our main results because it is based on the restricted sample of individuals with multiple electoral photos.

express wage differences by skin tone in units of implied changes in self-declared rather than employer-ascribed race. Given that skin tone is more predictive for individuals' self-identification than for employers' racial perceptions, the resulting PND estimates are smaller (as the reduced form impact of skin tone on wages remains the same). For example, we find an unconditional PND for self-declared race of 16.8%, about one third smaller than the unconditional PND for employer-ascribed race. These results highlight again that self-declared and employer-ascribed racial identities are distinct concepts that can matter differently for wages, and that PND estimates crucially depend on the racial identity units in which wage differentials are expressed.

Interpretation of PND. As mentioned above, the PND expresses disparities in units of racialization rather than in units of physical appearance. Because the PND takes into account how the racial cue translates into perceptions, it allows for meaningful comparisons across workers and settings for which racial cues carry different weights. Therefore, we view the PND as a useful scaling metric. When one can argue that the cue affects wages only through its effect on perceptions, the PND isolates the causal effect of being perceived as non-White. It can then be interpreted as the local average treatment effect for worker-employers pairs for which skin tone determines race perceptions.[26]

In our data, the key perceiver is the employer, since race in RAIS is recorded by the firm. We view this measure as one salient institutional perception of the worker's race and, more broadly, as a proxy for how the worker is seen in the labor market, including by co-workers and customers. This broader interpretation requires an additional assumption, namely that employer perceptions are correlated with the perceptions of other relevant actors. Under this view, the PND captures a perception-normalized disparity that is anchored in employer perceptions and reflects the component of wage differences that is mediated by this broader perception of race. If other actors' perceptions also affect wages and respond to skin tone, the PND need not identify the causal effect of employer perception alone. In that case, skin tone influences wages both through employers' racialization and through the behavior of co-workers or customers, and the PND recovers a weighted average effect of these perception channels. For these reasons, we interpret the PND as a perception-based disparity that is tightly linked to employer-recorded race and, under stronger exclusion restrictions, as the causal effect of being perceived as non-White by the employer.

Let us also emphasize that skin tone is only one and incomplete input into the race function. Our advantage relative to many existing studies is that we do not rely on enumerator or interviewer assessments of skin tone, which are themselves socially constructed, but instead use an algorithmic measure. Nonetheless, important caveats remain when using skin tone as a racial instrument. First, although the algorithmic measure is less socially constructed and likely more precise than interviewer ratings, it is still measured with error. Second, as with many instrumental variables, the exclusion restriction may not be satisfied. Even when we interpret employer-recorded race as a broader proxy for a worker's "market race" and conditional on a stringent set of controls and fixed effects, there is still the possibility that workers differ in their productivity.

[26] In contrast, colorism disparity measures (i.e., the reduced form) can be interpreted as an intent-to-treat effect that does not take into account in how far skin tone affects the perceptions of the decision-maker.

In addition, it is not obvious what should be held constant and what the relevant estimand ought to be. One possibility is that researchers are interested in disparate treatment based on a protected characteristic as it is perceived by the decision maker. Another is that they seek to quantify disparities arising from a "neutral" phenotypic trait that is plausibly unrelated to productivity. A further possibility is that they wish to capture perceptions held by a broader set of actors, not only employers. Each of these conceptual targets implies a different design and interpretation of both the reduced form and the PND.

Despite these caveats and limitations, the central contribution of our analysis is to show that race perceptions are systematically shaped by productivity-relevant factors and that conventional disparity estimates, even when they reflect the perceptions of decision makers, yield patterns that differ meaningfully from perception-normalized measures. This underscores that the choice of racial measure and estimand is not innocuous but depends on the underlying model of racial classification that researchers need to make explicit.

7 Conclusion

This paper adopts a constructivist perspective on race in the labor market and shows that racial classification is an outcome of perception and socio-economic status cues, and context rather than a fixed worker attribute. Using linked Brazilian administrative data, we document substantial divergence between self-declared and employer-ascribed race, as well as systematic reclassification of the same worker across firms. We also estimate a race function that shows how commonly used racial measures are endogenous to productivity-related signals and to firm-specific classification practices, complicating the interpretation of standard wage gap estimates.

Several limitations are important for interpretation. First, employer-ascribed race in RAIS reflects perceptions formed within the firm and may not coincide with how workers are perceived by other market actors such as customers or coworkers. Second, while algorithmic skin tone avoids enumerator subjectivity, it remains an imperfect proxy for racial cues and captures only one dimension of racialization. Third, perception-normalized disparities rely on strong assumptions for causal interpretation. Absent a credible exclusion restriction, they should be understood as a scaling device that expresses wage differentials in units of racial perception rather than as a clean estimate of discrimination. Finally, the linked sample of political candidates exhibits smaller wage gaps than the full workforce, so the magnitudes we report are best viewed as lower bounds.

These caveats point to several directions for future research. A first priority is to measure racial perception directly across multiple actors and settings, allowing researchers to distinguish employer-specific, customer-facing, and peer-based channels. A second avenue is to exploit institutional changes, reporting rules, or policy interventions that shift how race is classified or recorded, thereby generating plausibly exogenous variation in racial perception. A third direction is to extend the framework to dynamic and general equilibrium settings, where classification, sorting, and wage setting coevolve. More broadly, our results suggest that progress in the empirical study of racial inequality requires treating race as a socially produced variable and making explicit the perception processes embedded in the data.

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333.

Abramitzky, Ran, Jacob Conway, Roy Mill, and Luke Stein, "The Gendered Impacts of Perceived Skin Tone: Evidence from African-American Siblings in 1870–1940," NBER Working Paper 31016, National Bureau of Economic Research 2023.

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz, "What We Teach About Race and Gender: Representation in Images and Text of Children's Books," *The Quarterly Journal of Economics*, November 2023, *138* (4), 2225–2285.

_ , Richard Hornbeck, Daniel Keniston, and Benjamin Lualdi, "The Social Construction of Race during Reconstruction," Working Paper 33502, National Bureau of Economic Research February 2025.

Agan, Amanda and Sonja Starr, "Ban the box, criminal records, and racial discrimination: A field experiment," *The Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

Ajzenman, Nicolás, Bruno Ferman, and Pedro C Sant'Anna, "Discrimination in the formation of academic networks: A field experiment on #EconTwitter," *American Economic Review: Insights*, 2025, *7* (3), 357–375.

Akerlof, George A and Rachel E Kranton, "Economics and Identity," *The Quarterly Journal of Economics*, 2000, *115*, 715–753.

Antman, Francisca and Brian Duncan, "Incentives to identify: Racial identity in the age of affirmative action," *Review of Economics and Statistics*, 2015, *97* (3), 710–713.

_ and _ , "American Indian Casinos and Native American Self-Identification," *Journal of the European Economic Association*, 2023, *21* (6), 2547–2585.

Atkin, David, "The caloric costs of culture: Evidence from Indian migrants," *American Economic Review*, 2016, *106* (4), 1144–1181.

_ , Eve Colson-Sihra, and Moses Shayo, "How do we choose our identity? a revealed preference approach using food consumption," *Journal of Political Economy*, 2021, *129* (4), 1193–1251.

Ayres, Ian, "Testing for Discrimination and the Problem of "Included Variable Bias"," *Unpublished manuscript*, 2020.

Baron, E Jason, Jr Doyle Joseph J, Natalia Emanuel, Peter Hull, and Joseph Ryan, "Discrimination in Multi-Phase Systems: Evidence from Child Protection*," *The Quarterly Journal of Economics*, 02 2024, p. qjae007.

Bayer, Patrick and Kerwin Kofi Charles, "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *Quarterly Journal of Economics*, 2018, *133* (3), 1459–1501.

_ , _ , and Ellora Derenoncourt, "Racial Inequality in the Labor Market," Technical Report, National Bureau of Economic Research 2025.

Ben-Ner, Avner, Brian P McCall, Massoud Stephane, and Hua Wang, "Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence," *Journal of Economic Behavior & Organization*, 2009, *72* (1), 153–170.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bohren, J Aislinn, Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement*," *The Quarterly Journal of Economics*, 05 2025, *140* (3), 1743–1799.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87* (3), 699–739.

Bueno, Natálie S and Thad Dunning, "Race, Resources, and Representation: Evidence from Brazilian Politicians," *World Politics*, 2017, *69* (2), 327–365.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review*, 2020, *110* (10), 2997–3029.

Cantoni, Enrico and Vincent Pons, "Does context outweigh individual characteristics in driving voting behavior? Evidence from relocations within the United States," *American Economic Review*, 2022, *112* (4), 1226–1272.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 2008, *123* (1), 177–218.

_ , Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

_ , Jörg Heining, and Patrick Kline, "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 2013, *128* (3), 967–1015.

Cassan, Guilhem, "Identity-based policies and identity manipulation: Evidence from colonial Punjab," *American Economic Journal: Economic Policy*, 2015, *7* (4), 103–131.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng, "On binscatter," *American Economic Review*, 2024, *114* (5), 1488–1514.

Cornwell, Christopher, Jason Rivera, and Ian M. Schmutte, "Wage Discrimination When Identity Is Subjective: Evidence from Changes in Employer-Reported Race," *Journal of Human Resources*, 2017, *52* (3), 719–755.

Dahis, Ricardo, Emily Nix, and Nancy Qian, "Choosing Racial Identity in the United States, 1880–1940," NBER Working Paper 26465, National Bureau of Economic Research November 2019. Revised June 2020.

_ , _ , and _ , "Choosing racial identity in the United States, 1880-1940," *Working Paper*, 2019.

Davenport, Lauren, "The Fluidity of Racial Classifications," *Annual Review of Political Science*, 2020, *23*, 221–240.

de Holanda, Sérgio Buarque, *Raízes do Brasil*, Rio de Janeiro: Livraria José Olympio Editora, 1936.

Del Carpio, Lucia and Maria Guadalupe, "More women in tech? Evidence from a field experiment addressing social identity," *Management Science*, 2022, *68* (5), 3196–3218.

do Valle Silva, Nelson, "A research note on the cost of not being white in Brazil," *Studies in Comparative International Development*, 2000, *35*, 18–27.

Doleac, Jennifer L and Luke CD Stein, "The visible hand: Race and online market outcomes," *The Economic Journal*, 2013, *123* (572), F469–F492.

Duncan, Brian and Stephen J. Trejo, "Tracking Intergenerational Progress for Immigrant Groups: The Problem of Ethnic Attrition," *American Economic Review*, May 2011, *101* (3), 603–08.

_ and Stephen J Trejo, "Immigrant Age at Arrival and the Intergenerational Transmission of Ethnic Identification Among Mexican Americans," in "AEA Papers and Proceedings," Vol. 115 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2025, pp. 451–456.

_ and _ , "Which Mexicans Are White? Enumerator-Assigned Race in the 1930 Census and the Socioeconomic Integration of Mexican Americans," *ILR Review*, 2025, *78* (1), 62–85.

Fernandes, Florestan, *A Integração do Negro na Sociedade de Classes*, São Paulo: Dominus Editora, 1965.

Ferraz, Claudio and Frederico Finan, "Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance," *Unpublished Manuscript*, 2011.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of geographic variation in health care: Evidence from patient migration," *The quarterly journal of economics*, 2016, *131* (4), 1681–1726.

Firpo, Sergio, Michael França, and Alysson Portella, "Racial Inequality in the Brazilian Labor Market and the Role of Education," *Available at SSRN 3967828*, 2021.

França, Michael and Alysson Portella, *Números da discriminação racial: Desenvolvimento humano, equidade e políticas públicas*, Editora Jandaíra, 2023.

Francis, Andrew M and Maria Tannuri-Pianto, "Endogenous Race in Brazil : Affirmative Action and the Construction of Racial Identity among Young Adults," *Economic Development and Cultural Change*, 2013, *61*, 731–753.

Freeman, Jonathan B, Andrew M Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady, "Looking the part: Social status cues shape race perception," *PloS one*, 2011, *6* (9), e25107.

Freyre, Gilberto, *Casa-Grande & Senzala*, São Paulo: Companhia das Letras, 1933.

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil," *Econometrica*, 2015, *83* (2), 423–464.

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card, "Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil," *American Economic Review*, 2021, *111* (10), 3418–3457.

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores," *The Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.

Goldsmith, Arthur H., Darrick Hamilton, and Jr. Darity William A., "Shades of Discrimination: Skin Tone and Wages," *American Economic Review*, May 2006, *96* (2), 242–245. Papers and Proceedings issue (May 2006).

_ , _ , and _ , "From Dark to Light: Skin Color and Wages Among African-Americans," *Journal of Human Resources*, 2007, *42* (4), 701–738.

Grogger, Jeffrey, "Speech and wages," *Journal of Human Resources*, 2019, *54* (4), 926–952.

Hersch, Joni, "Skin-Tone Effects Among African Americans: Perceptions and Reality," *American Economic Review*, May 2006, *96*, 251–255.

_ , "Profiling the New Immigrant Worker: The Effects of Skin Color and Height," *Journal of Labor Economics*, April 2008, *26* (2), 345–386.

_ , "Colorism and Immigrant Earnings in the United States, 2015–2024," IZA Discussion Paper 17397, Institute of Labor Economics (IZA) 2024.

Holland, Paul W, "Causation and race," *White logic, white methods: Racism and methodology*, 2008, pp. 93–109.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and US Economic Growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hu, Lily and Issa Kohler-Hausmann, "What is Perceived When Race is Perceived and Why It Matters for Causal Inference and Discrimination Studies," *Law & Society Review*, 2025, *59* (2), 239–264.

IBGE, "Desigualdades Sociais por Cor ou Raça no Brasil," *Estudos e Pesquisas -Informação Demográfica e Socioeconômica*, 2022, *48* (2nd edition), 1–16.

Jia, Ruixue and Torsten Persson, "Choosing ethnicity: The interplay between individual and social motives," *Journal of the European Economic Association*, 2021, *19* (2), 1203–1248.

Kim, Young-Chul and Glenn C Loury, "To be, or not to be: Stereotypes, identity choice and group inequality," *Journal of Public Economics*, 2019, *174*, 36–52.

Kline, Patrick and Christopher R Walters, "Reasonable Doubt: Experimental Detection of Job-Level Employment Discrimination," *Econometrica*, 2021, *89* (2), 765–792.

_ , Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

_ , Evan K. Rose, and Christopher R. Walters, "A Discrimination Report Card," *American Economic Review*, August 2024, *114* (8), 2472–2525.

Kline, Patrick M, "Firm Wage Effects," Technical Report, National Bureau of Economic Research 2024.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten, "Leave-out estimation of variance components," *Econometrica*, 2020, *88* (5), 1859–1898.

Kreisman, Daniel and Marcos A. Rangel, "On the Blurring of the Color Line: Wages and Employment for Black Males of Different Skin Tones," *The Review of Economics and Statistics*, March 2015, *97* (1), 1–13.

Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis, "America's Churning Races: Race and Ethnic Response Changes between Census 2000 and the 2010 Census," *Demography*, 2017, *54* (1), 259–284.

Lowe, Matt, "Types of contact: A field experiment on collaborative and adversarial caste integration," *American Economic Review*, 2021, *111* (6), 1807–1844.

Maggie, Yvonne and Peter Fry, "Qual Cor, Qual Raça?," *Estud. Av.*, 2004, *18* (50), 57–67.

McGee, Dan, "Exploitation Through Racialization*," *The Quarterly Journal of Economics*, 01 2025, *140* (2), 1581–1631.

Micheli, David De, "Racial reclassification and political identity formation," *World Politics*, 2021, *73* (1), 1–51.

Miller, Conrad and IM Schmutte, "The Dynamic Effects of Co-Racial Hiring," *Unpublished Manuscript*, 2023.

Miranda, Vítor, "A resurgence of black identity in Brazil? Evidence from an analysis of recent censuses," *Demographic Research*, 2015, *32*, 1603–1630.

Mitrut, Andreea, Gabriel Kreindler, Margareta Matache, Andrei Munteanu, and Cristian Pop-Eleches, "Education and Selection into Ethnic Identification: Evidence from Roma People in Romania," Working Paper 34383, National Bureau of Economic Research October 2025.

MTE, *Manual de Orientação da RAIS*, Brasília: Ministério do Trabalho e Emprego, 2024.

Muniz, Jerônimo, Aliya Saperstein, and Bernardo L. Queiroz, "Racial classification as a multistate process," *Demographic Research*, 2024, *50* (17), 457–472.

Nix, Emily and Nancy Qian, "The Fluidity of Race: "Passing" in the United States, 1880–1940," NBER Working Paper 20828, National Bureau of Economic Research January 2015.

Noghanibehambari, Hamid and Jason Fletcher, "Passing as White: Racial Identity and Old-Age Longevity," NBER Working Paper 33394, National Bureau of Economic Research January 2025.

Oh, Suanna, "Does identity affect labor supply?," *American Economic Review*, 2023, *113* (8), 2055–2083.

Osório, Rafael Guerreiro, "O sistema classificatório de cor ou raça do IBGE," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2003.

Ponczek, Vladimir and Gabriel Ulyssea, "Enforcement of labour regulation and the labour market effects of trade: Evidence from Brazil," *The Economic Journal*, 2022, *132* (641), 361–390.

Rademakers, Robbert and André van Hoorn, "Ethnic switching: Longitudinal evidence on prevalence, correlates, and implications for measuring ethnic segregation," *Journal of Development Economics*, 2021, *152.* 102694.

Rodeheffer, Christopher D, Sarah E Hill, and Charles G Lord, "Does this recession make me look black? The effect of resource scarcity on the categorization of biracial faces," *Psychological Science*, 2012, *23* (12), 1476–1478.

Rose, Evan K, "A constructivist perspective on empirical discrimination research," *Journal of Economic Literature*, 2023, *61* (3), 906–923.

Saperstein, Aliya and Andrew M Penner, "Racial fluidity and inequality in the United States," *American journal of sociology*, 2012, *118* (3), 676–727.

Schwarcz, Lilia Moritz, *Nem preto nem branco, muito pelo contrário: cor e raça na sociabilidade brasileira*, Editora Companhia das Letras, 2013.

Schwartzman, Luisa Farah, "Does money whiten? Intergenerational changes in racial classification in Brazil," *American Sociological Review*, 2007, *72* (6), 940–963.

Sen, Maya and Omar Wasow, "Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics," *Annual Review of Political Science*, 2016, *19*, 499–522.

Shayo, Moses, "Social identity and economic policy," *Annual Review of Economics*, 2020, *12*, 355–389.

Silveira, Leonardo, "Imputação da informação de raça/cor na Rais para o setor público brasileiro," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2022.

Szerman, Christiane, "The labor market effects of disability hiring quotas," *Available at SSRN 4267622*, 2022.

Telles, Edward E., *Race in Another America: The Significance of Skin Color in Brazil*, Princeton: Princeton University Press, 2004.

_ , *Pigmentocracies: Ethnicity, race, and color in Latin America*, UNC Press Books, 2014.

Woo-Mora, L. Guillermo, "Unveiling the Cosmic Race: Skin tone and intergenerational economic disparities in Latin America and the Caribbean," *Journal of Development Economics*, 2025.

Online Appendix to
"Race Perceptions in the Labor Market"

Pedro C. Sant'Anna Sulin Sardoschau Aiko Schmeißer

11th December 2025

A Additional Figures **2**

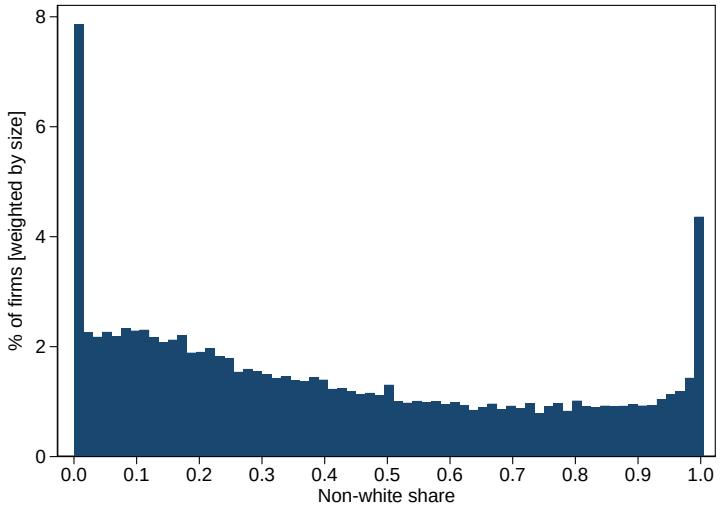
B Additional Tables **5**

C Details on Data **15**
C.1 Employer-Ascribed Race 15
C.2 Skin Tone 16

D AKM Model **21**
D.1 Empirical Strategy 21
D.2 Results 23

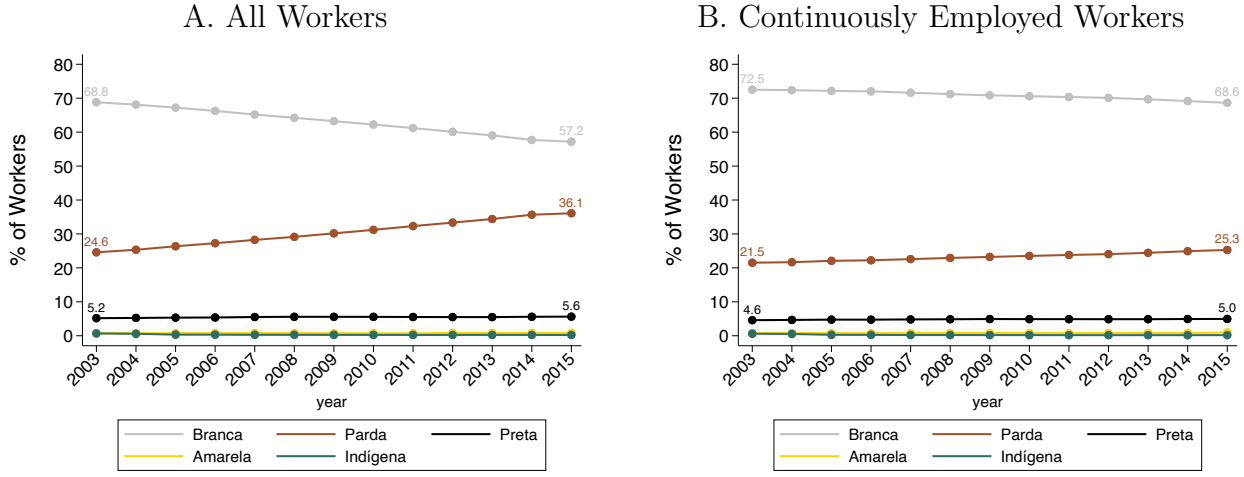
A Additional Figures

Figure A.1: Histogram of Non-White Share Across Establishments



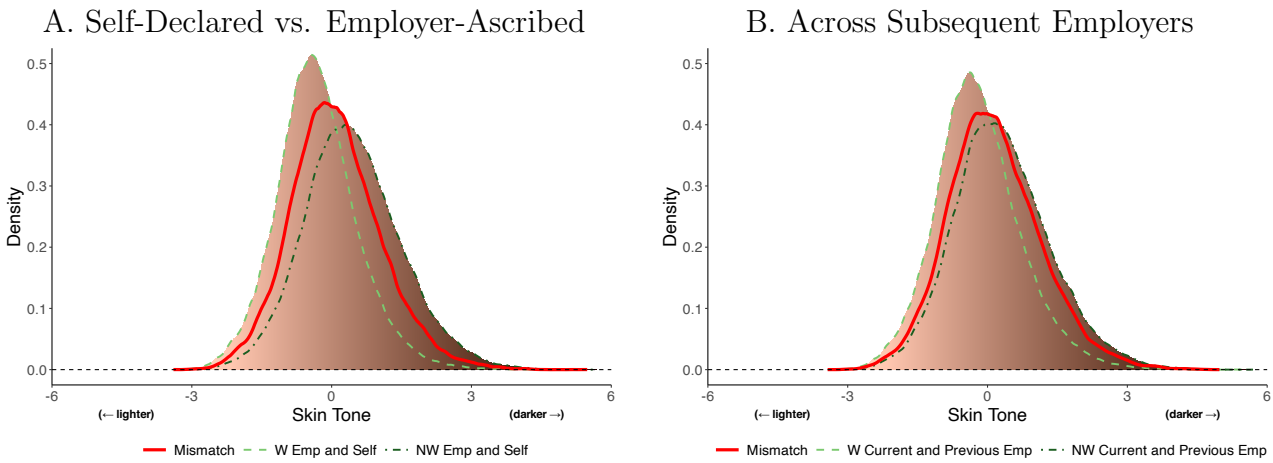
Notes: The figure displays a histogram of the non-White (Black or Brown) share of workers across establishments in the full RAIS sample. Establishments are weighted by their number of workers.

Figure A.2: Distribution of Employer-Ascribed Race in RAIS, 2003 to 2015



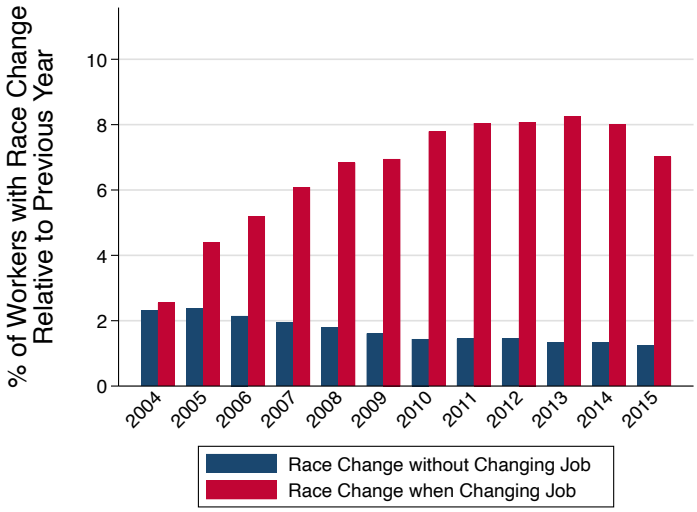
Notes: The figures show time series of the race distribution among workers in RAIS. Panel A considers all workers in the full RAIS sample, while Panel B is restricted to workers employed in all years between 2003 and 2015 (i.e., a balanced panel).

Figure A.3: Skin Tone Distribution by Mismatch Category



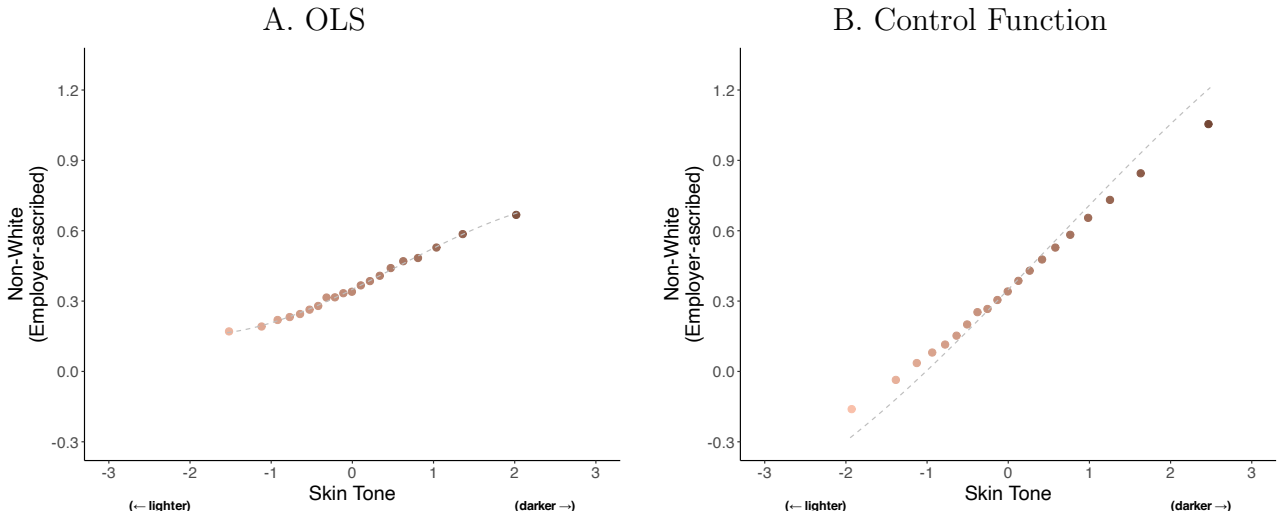
Notes: The figures show histograms of skin tone by categories of racial mismatch. The samples are the same as in Table 1. Panel A shows the skin tone distribution by cells of self-declared and employer-ascribed race, while Panel B shows the skin tone distribution by cells of ascribed race by the current and previous employer. The skin tone measure is standardized to have a mean of zero and a standard deviation of one within our sample.

Figure A.4: Race Changes in RAIS, 2004 to 2015



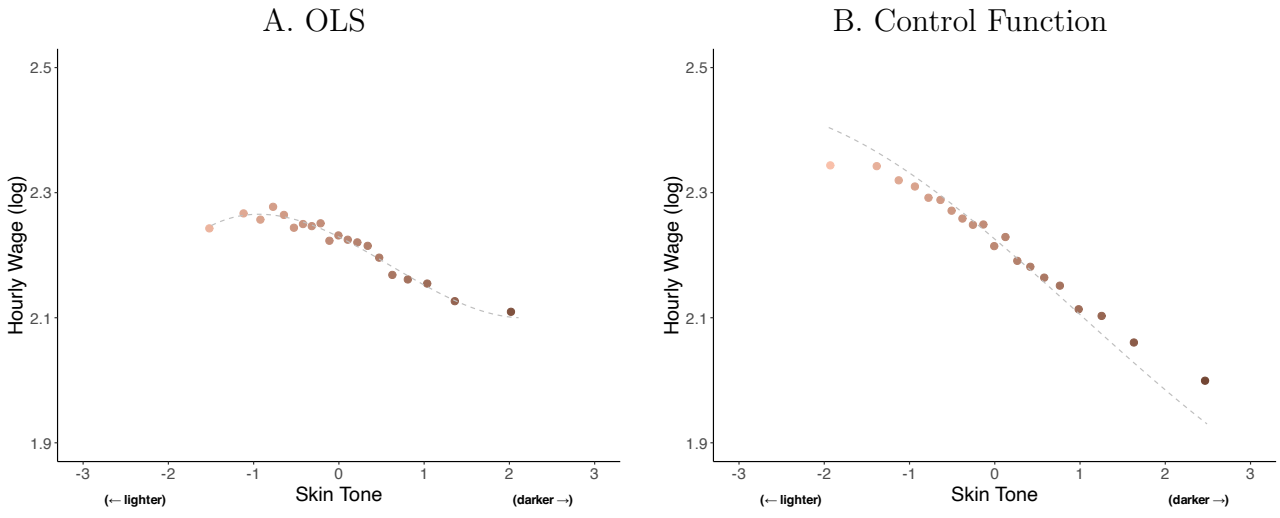
Notes: The figures show a time series of the probability that a worker's ascribed race changed between two consecutive years, separately by workers who stayed in the same job and workers who transitioned to a new job. Numbers are based on the full RAIS sample, 2003 to 2015.

Figure A.5: Relationship between Employer-Ascribed Race and Skin Tone, Accounting for Measurement Error in Skin Tone



Notes: The figures show binned scatterplots of ...

Figure A.6: Relationship between Wages and Employer-Ascribed Race, Accounting for Measurement Error in Skin Tone



Notes: The figures show binned scatterplots of ...

B Additional Tables

Table B.1: Summary Statistics of Candidate Sample and Full RAIS Sample

	(1) Candidates Mean	(2) Non-candidates Mean	(3) Difference (1) - (2)
Worker Characteristics			
Non-White (Employer-Ascribed)	0.37	0.37	-0.00**
Male	0.75	0.62	0.13***
Age (years)	34.48	33.65	0.84***
Log Hourly Wage (2018 BRL)	2.18	2.20	-0.02***
Highest Degree Attained			
Less than Elementary	0.03	0.04	-0.01***
Elementary School	0.14	0.15	-0.01***
Middle School	0.23	0.23	0.00***
High School	0.52	0.49	0.03***
Undergraduate or more	0.09	0.10	-0.01***
Region of Birth			
North	0.05	0.04	0.01***
Northeast	0.17	0.16	0.01***
Center-West	0.51	0.55	-0.04***
Southeast	0.19	0.18	0.02***
South	0.08	0.07	0.01***
Establishment Characteristics			
Firm Size	536.54	580.17	-43.62***
Average Log Hourly Wage	2.15	2.21	-0.06***
Sector			
Agriculture	0.03	0.02	0.01***
Extractive industry	0.01	0.01	0.00***
Manufacturing	0.24	0.25	-0.01***
Construction	0.06	0.07	-0.00***
Retail	0.25	0.25	-0.01***
Transport & communication	0.13	0.13	-0.00***
Finance & real state	0.15	0.17	-0.02***
Education & health	0.09	0.07	0.01***
Other services	0.05	0.03	0.02***
No. of worker-year obs.	1,510,833	310,051,787	
No. of workers	331,531	61,075,388	
No. of establishments	358,899	5,616,123	

Notes: The table shows mean characteristics of workers who have run for a political office between 2014 and 2024 (our estimation sample, column (1)) and all other workers in RAIS (column (2)). Column (3) shows the difference in means between both samples. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Racial Gaps in Candidate Sample vs Full RAIS Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Candidates			Non-candidates		
	White Mean	Non-White Mean	Racial gap (1)-(2)	White Mean	Non-White Mean	Racial gap (4)-(5)
Worker Characteristics						
Male	0.73	0.77	-0.04***	0.60	0.66	-0.06***
Age (years)	34.72	34.09	0.63***	33.87	33.25	0.62***
Log Hourly Wage (2018 BRL)	2.22	2.10	0.13***	2.28	2.06	0.23***
Highest Degree Attained						
Less than Elementary	0.02	0.04	-0.02***	0.03	0.06	-0.03***
Elementary School	0.13	0.15	-0.02***	0.14	0.16	-0.03***
Middle School	0.23	0.23	0.00	0.23	0.24	-0.01***
High School	0.51	0.52	-0.01***	0.49	0.49	-0.01***
Undergraduate or more	0.11	0.06	0.05***	0.12	0.05	0.07***
Region of Birth						
North	0.02	0.11	-0.08***	0.02	0.09	-0.07***
Northeast	0.09	0.31	-0.22***	0.08	0.30	-0.22***
Center-West	0.55	0.44	0.12***	0.60	0.48	0.12***
Southeast	0.28	0.05	0.23***	0.25	0.04	0.21***
South	0.06	0.10	-0.04***	0.06	0.09	-0.04***
Establishment Characteristics						
Firm Size	441.86	698.04	-256.17***	501.76	707.93	-206.17***
Average Log Hourly Wage	2.18	2.09	0.09***	2.27	2.10	0.17***
Sector						
Agriculture	0.02	0.03	-0.01***	0.01	0.02	-0.01***
Extractive industry	0.01	0.01	-0.00***	0.01	0.01	-0.00***
Manufacturing	0.26	0.21	0.05***	0.27	0.22	0.05***
Construction	0.05	0.09	-0.05***	0.05	0.10	-0.05***
Retail	0.26	0.23	0.03***	0.26	0.25	0.01***
Transport & communication	0.13	0.13	-0.00***	0.13	0.13	0.00
Finance & real state	0.13	0.17	-0.04***	0.16	0.18	-0.02***
Education & health	0.09	0.08	0.01***	0.08	0.07	0.01***
Other services	0.05	0.05	0.01***	0.03	0.03	0.00***
N	952,450	558,383		192,996,780	113,674,338	

Notes: The table shows racial gaps (by employer-ascribed race) in mean characteristics of workers who have run for a political office between 2014 and 2024 (our estimation sample, columns (1)-(3)) and all other workers in RAIS (columns (4)-(6)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Mismatch in Ascribed Race, Detailed Racial Categories

Panel A: Mismatch between Self-Declared and Employer-Ascribed Race

	Employer-Ascribed Race					
	White		*Parda*		*Preta*	
Self-Declared Race	%	Skin Tone	%	Skin Tone	%	Skin Tone
White	45.20%	-0.32	9.22%	-0.13	0.49%	0.14
Parda	14.86%	0.07	17.82%	0.21	1.90%	0.55
Preta	2.99%	0.77	4.16%	0.80	3.35%	1.19

Panel B: Mismatch across Subsequent Employers

	Employer-Ascribed Race, Current					
	White		*Parda*		*Preta*	
Employer-Ascribed Race, Previous	%	Skin Tone	%	Skin Tone	%	Skin Tone
White	49.82%	-0.22	10.79%	0.08	1.38%	0.71
Parda	9.57%	0.07	20.71%	0.24	1.73%	0.81
Preta	1.31%	0.72	1.83%	0.75	2.87%	1.11

Notes: The table displays the (mis)match shares for different race measures in our sample, as well as the average skin tone for each cell. This table considered the detailed racial categories available in Brazilian statistics, White, *Parda* (brown), and *Preta* (black), excluding indigenous and Asian individuals. Skin tone is standardized to have a mean of zero and a standard deviation of one across the whole sample (higher values represent darker shades, while lower values represent lighter shades). Panel A presents the distribution of workers by self-declared race (rows) and employer-ascribed race (columns). It considers all worker-year observations ($N = 1,510,833$). Panel B presents the distribution of ascribed race for the same worker across different employers in consecutive years. It considers all workers who changed jobs ($N = 153,366$) and presents the distribution of workers by their ascribed race in the previous employer (rows) and current employer (columns).

Table B.4: Mismatch in Ascribed Race - Robustness

[A] Mismatch between Self-Declared and Employer-Ascribed Race					
Employer-ascribed	White	Non-White	White	Non-White	
Self-declared	White	Non-White	Non-White	White	N
Baseline sample	45.20%	27.25%	17.84%	9.71%	1,510,833
Firms with variation in ascribed race	42.30%	29.24%	18.08%	10.38%	1,328,249
Firms with HR employees	43.48%	28.88%	17.24%	10.41%	910,506
Self-declared race in same period	43.28%	29.23%	16.04%	11.45%	48,316
Councilor candidates	44.97%	27.49%	17.99%	9.55%	1,412,050
Employer-ascribed	Male	Female	Male	Female	
Self-declared	Male	Female	Female	Male	N
Baseline sample	73.75%	24.26%	1.06%	0.93%	1,510,833
[B] Mismatch across Subsequent Employers					
Employer-ascribed, current	White	Non-White	White	Non-White	
Employer-ascribed, previous	White	Non-White	Non-White	White	N
Baseline sample	49.82%	27.12%	10.88%	12.18%	153,366
Firms with variation in ascribed race	45.68%	29.96%	11.51%	12.85%	128,142
Firms with HR employees	47.74%	30.64%	10.22%	11.40%	71,501
All workers in RAIS	50.44%	26.70%	10.78%	12.08%	33,898,971
Employer-ascribed, current	Male	Female	Male	Female	
Employer-ascribed, previous	Male	Female	Female	Male	N
Baseline sample	76.58%	20.76%	1.32%	1.34%	153,366

Notes: The table shows mismatch rates between self-declared and employer-ascribed race (Panel A) and in employer-ascribed race across different employers in subsequent years (Panel B) in alternative subsamples. Specifically, we restrict the sample to firms that do not ascribe the same racial category to all their workers, firms that have at least one employee in an HR occupation, jobs in 2015 and candidates running in 2014 or 2016 (such that self-declared and employer-ascribed race are report at roughly the same time), as well as candidates that run for local councilor positions. Panel B also considers job changes among all workers in RAIS (not limited to the candidate sample). Both panels also report mismatch rates across reported gender instead of race.

Table B.5: Mismatch in Ascribed Race - Aggregating Across Employers

	(1) OLS	(2) OLS	(3) SSIV	(4) OLS	(5) SSIV
Non-White Employer	0.454***				
	(0.002)				
Non-White Employer Mode		0.472***	0.613***		
		(0.002)	(0.003)		
Non-White Employer Mean				0.597***	0.659***
				(0.002)	(0.002)
Constant	0.283***	0.286***	0.240***	0.230***	0.207***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,510,833	1,510,833	1,439,433	1,510,833	1,439,433

Notes: The table reports results from regressions of self-declared race (indicator equal to one if the worker identifies as non-White) on employer-ascribed race (indicator equal to one if the worker is ascribed non-White). While column (1) uses race as ascribed by the current employer, columns (2) to (5) aggregate across employers by considering the mode or mean across all yearly observations of the worker. Columns (3) and (5) implement a split-sample IV (SSIV) approach, in which the mode or mean constructed from one randomly selected half of a worker's yearly observations is instrumented with the mode or mean constructed from the other half of observations. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: Correlates of Employer-Ascribed Race including Employment History Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Skin Tone	0.110***	0.104***	0.082***	0.081***	0.079***	0.068***	0.040***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Non-White Self-Declared							0.244***
							(0.007)
Demographics							
Male		0.037***	0.016***	0.009**	0.001	-0.000	-0.002
		(0.004)	(0.004)	(0.004)	(0.004)	(0.009)	(0.008)
Age		0.007***	0.003*	0.001	-0.000	-0.002	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Age^2		-0.010***	-0.005***	-0.004***	-0.003**	-0.001	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
(Some) High School		-0.001	-0.028***	-0.019***	-0.010**	-0.008	-0.003
		(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)
(Some) University		-0.044***	-0.087***	-0.078***	-0.040***	-0.014	0.000
		(0.006)	(0.005)	(0.006)	(0.006)	(0.017)	(0.015)
Work History							
Tenure		-0.055***	-0.043***	-0.036***	-0.035***	-0.048***	-0.045***
		(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Tenure^2		0.005***	0.005***	0.003**	0.003*	0.002**	0.002*
		(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Final Wage in Previous Job		-0.046***	-0.030***	-0.033***	-0.028***	-0.015***	-0.010**
		(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
Region							
Northeast			-0.057***	-0.060***	-0.062***		
			(0.008)	(0.008)	(0.008)		
Southeast			-0.374***	-0.370***	-0.372***		
			(0.007)	(0.007)	(0.007)		
South			-0.569***	-0.560***	-0.561***		
			(0.007)	(0.007)	(0.007)		
Center-West			-0.191***	-0.188***	-0.188***		
			(0.009)	(0.009)	(0.009)		
Constant	0.370***	0.340***	0.713***	0.710***	0.707***	0.369***	0.258***
	(0.001)	(0.005)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
N	1,510,833	374,036	374,036	374,006	373,905	298,044	298,044
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No	No
Industry × Occupation FE	No	No	No	No	Yes	No	No
Firm × Occupation FE	No	No	No	No	No	Yes	Yes
R-squared	0.057	0.065	0.209	0.213	0.223	0.779	0.793
Adjusted R-squared	0.057	0.065	0.209	0.213	0.222	0.719	0.736

Notes: The table reports regression estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and skin tone, enriching the analysis of Table 2 with work history variables. The sample in columns 2-7 only includes workers whom we observe in multiple establishments. All columns include year fixed effects. The first column displays the unconditional relationship (apart from year fixed effects), while the remaining columns progressively add controls. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7: Correlates of Self-Declared Race

	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone	0.154***	0.149***	0.126***	0.125***	0.122***	0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Demographics						
Male		0.007***	0.002	-0.001	-0.005**	-0.006
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Age		-0.025***	-0.017***	-0.019***	-0.014***	-0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Age^2		-0.011***	-0.008***	-0.007***	-0.007***	-0.006***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Some) High School		-0.033***	-0.054***	-0.050***	-0.031***	-0.025***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
(Some) University		-0.159***	-0.173***	-0.174***	-0.107***	-0.077***
		(0.003)	(0.003)	(0.003)	(0.004)	(0.008)
Region						
Northeast			-0.071***	-0.072***	-0.072***	
			(0.004)	(0.004)	(0.004)	
Southeast			-0.344***	-0.342***	-0.343***	
			(0.004)	(0.004)	(0.004)	
South			-0.535***	-0.529***	-0.531***	
			(0.004)	(0.004)	(0.004)	
Center-West			-0.180***	-0.179***	-0.177***	
			(0.005)	(0.005)	(0.005)	
Constant	0.451***	0.496***	0.815***	0.813***	0.796***	0.483***
	(0.001)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.097	0.110	0.214	0.217	0.226	0.772
Adjusted R-squared	0.097	0.110	0.214	0.217	0.225	0.711

Notes: The table reports regression estimates of the relationship between self-declared race (an indicator equal to one for non-White) and skin tone. All columns include year fixed effects. The first column displays the unconditional relationship (apart from year fixed effects), while the remaining columns progressively add controls. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Correlates of Self-Declared Race including Employment History Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone	0.154***	0.144***	0.125***	0.124***	0.122***	0.115***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Demographics						
Male		0.046***	0.028***	0.024***	0.011**	0.009
		(0.004)	(0.004)	(0.004)	(0.005)	(0.010)
Age		0.002	-0.001	-0.003	-0.004**	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Age^2		-0.013***	-0.009***	-0.008***	-0.007***	-0.006***
		(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
(Some) High School		-0.019***	-0.042***	-0.037***	-0.025***	-0.021**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.009)
(Some) University		-0.104***	-0.140***	-0.136***	-0.090***	-0.058***
		(0.007)	(0.006)	(0.006)	(0.007)	(0.019)
Work History						
Tenure		-0.037***	-0.026***	-0.021***	-0.019***	-0.012***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
$Tenure^2$		0.007***	0.007***	0.006***	0.005***	0.002**
		(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Final Wage in Previous Job		-0.046***	-0.033***	-0.035***	-0.027***	-0.021***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.006)
Region						
Northeast			-0.067***	-0.070***	-0.072***	
			(0.009)	(0.009)	(0.009)	
Southeast			-0.324***	-0.321***	-0.325***	
			(0.008)	(0.008)	(0.008)	
South			-0.515***	-0.507***	-0.510***	
			(0.008)	(0.008)	(0.008)	
North			-0.160***	-0.157***	-0.157***	
			(0.010)	(0.010)	(0.010)	
Constant	0.451***	0.433***	0.763***	0.760***	0.759***	0.454***
	(0.001)	(0.006)	(0.009)	(0.009)	(0.009)	(0.011)
N	1,510,833	374,036	374,036	374,006	373,905	298,044
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.097	0.110	0.214	0.217	0.226	0.772
Adjusted R-squared	0.097	0.110	0.214	0.217	0.225	0.711

Notes: The table reports regression estimates of the relationship between self-declared race (an indicator equal to one for non-White) and skin tone, enriching the analysis of Table 2 with work history variables. The sample in columns 2-7 only includes workers whom we observe in multiple establishments. All columns include year fixed effects. The first column displays the unconditional relationship (apart from year fixed effects), while the remaining columns progressively add controls. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Skin Tone Gaps for Repeated Candidates

	(1) OLS	(2) SSIV
[A] First Stage: Non-White Employer-Ascribed		
Skin Tone	0.152***	0.346***
	(0.001)	(0.005)
[B] Reduced Form: Hourly Wage (log)		
Skin Tone	-0.051***	-0.114***
	(0.002)	(0.007)
N	652,140	652,140
KP F-stat		6,650.87

Notes: The table reports results from regressions of employer-ascribed race (Panel A) and log hourly wages (Panel B) on skin tone. The sample consists of all individuals who ran for political office multiple times and for whom multiple electoral photos are available. Column (1) presents OLS estimates, in which skin tone is measured by averaging across all photos of an individual. Column (2) implements a split-sample IV (SSIV) approach, where the average skin tone from one randomly selected half of an individual's photos is instrumented with the average from the other half of photos. All regressions include year fixed effects. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.10: Wage Disparities, Self-Declared Race

	(1)	(2)	(3)	(4)	(5)	(6)
[A] OLS: Conventional Disparity						
Non-White Self-Declared	-0.163***	-0.107***	-0.070***	-0.077***	-0.043***	-0.019***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Colorism Disparity						
Skin Tone	-0.026***	-0.032***	-0.026***	-0.028***	-0.015***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
[C] First Stage: Race Function						
Skin Tone	0.154***	0.149***	0.126***	0.125***	0.122***	0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
[D] 2SLS: Perception-Normalized Disparity						
Non-White Self-Declared	-0.168***	-0.217***	-0.204***	-0.225***	-0.123***	-0.048***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports different measures of racial wage gaps based on self-declared race. All columns include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. The final column includes all fixed-effects, as well as controls for self-declared race. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced form regressions of log hourly wage on skin tone. Panel C presents the first-stage coefficient from regressing employer-ascribed race on skin tone. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by skin tone. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Details on Data

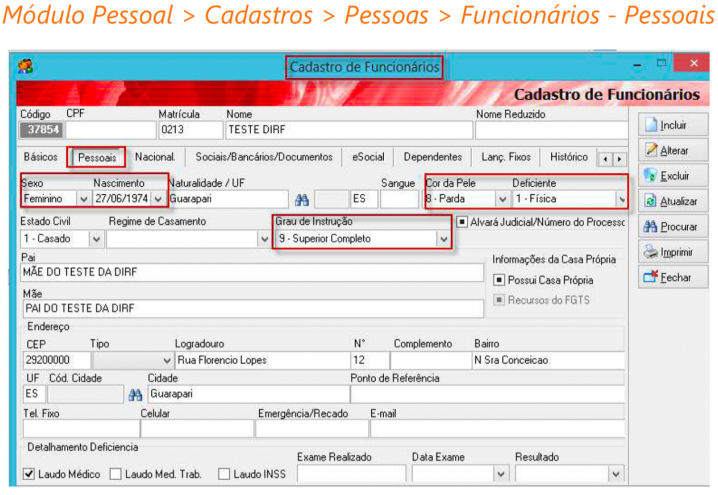
C.1 Employer-Ascribed Race

In RAIS, employers are responsible for reporting information on behalf of their employees. In larger firms, these records are typically submitted by a manager, HR department, or accountant, whereas in smaller firms the owner often completes the reporting. Figure C.1 shows the online interface, maintained by Brazil's Ministry of Labor and Employer, through which the employers enter the information.

Understanding how race is measured in RAIS requires examining how employers obtain information about workers' characteristics. At hiring, workers must present various official documents, including the *Carteira de Trabalho e Previdência Social* (CTPS). The CTPS contains basic identifying information – including name, date of birth, gender, and residence – but does not contain any field of race. Workers must also provide a photograph and proof of required education. Employers then record worker information in the *Livro de Registro dos Empregados* (LRE), an internal employee record maintained at the establishment level. Employers must enter each new worker into the LRE, and the information recorded there is used to comply with a variety of reporting requirements, including RAIS. The required information includes date of birth, hire date, identification number, and various job characteristics. There is also space for a photograph of the worker. Unlike the CTPS, the LRE commonly includes a field for race (cor, literally "color"), although completing this field is not mandatory (Cornwell et al., 2017)).

The official RAIS completion manual offers no specific guidance on how to fill out the race field, and there is no standardized training of those who enter the data. As a result, the race information submitted in RAIS seems to largely reflect employer's perception of the workers' racial identity, based on workplace interaction, photographs, or other worker information (Silveira, 2022; Osório, 2003).

Figure C.1: RAIS Interface



Notes: RAIS interface for reporting employee's information (2017 version).

C.2 Skin Tone

Electoral Photos. We assemble a dataset of nearly three million official photographs of candidates running for political office in Brazil between 2004 and 2024. These photographs are used during elections in Brazil's electronic voting machines, and they typically differ from campaign materials used in advertisements. Electronic voting was introduced by the *Tribunal Superior Eleitoral* (TSE) in 1996 with the aim to reduce the time and costs of vote counting. Voters cast their ballots by typing a candidate's identification number into the machine's keypad. Once the full number is entered, the interface displays the candidate's name, party label, and photograph, after which the voter may either confirm the vote or restart the process. The photograph is thus visible only at the moment of verification, used as a mechanisms to help low-literacy voters cast their ballot and reduce the incidence of misrecorded votes (Fujiwara, 2015).

The submission of a valid photograph is mandatory for a candidacy to be accepted by the TSE. Technical and stylistic guidelines for these photographs are specified in electoral resolutions and have evolved over time. Most recently, Resolution No. 23,609 (2019), Article 27 II, details the following requirements:

- Dimensions: 161 x 225 pixels (W x H); without frame,
- Color depth: 24bpp,
- Background: uniform color,
- Characteristics: frontal (bust) view, with attire suitable for official photography, ensuring the use of ethnic or religious clothing and body paint, as well as accessories necessary for the person with a disability; the use of scenic elements and other adornments is prohibited, especially those that have connotations of electoral propaganda or that induce or hinder the recognition of the candidate by the electorate.

Before that, Resolution No. 23,405 (2014), Article 27 III, imposed similar requirements but specified an 8bpp grayscale format with a uniform, preferably white, background.

Skin Tone Measurement. Next, we provide details on our skin tone measurement algorithm. We begin by detecting the skin part in each photo. For that, we use Google MediaPipe's Image Segmentation tool, which relies on a vision transformer neural network to locate different objects in an image and allows us to segment the skin from non-skin areas of the photo. Specifically, we apply the *multi-class selfie segmentation model*, which classifies pixels as *background*, *hair*, *clothing*, *accessories*, *face-skin*, and *body-skin*.[1] We combine the latter two categories to form the skin mask of each photo. Next, we identify the representative color of the skin area, closely following the algorithm of Adukia et al. (2023). In particular, we use k-means clustering to partition the pixel colors into distinct groups based on their RGB color values. We group the pixels into $k = 5$ clusters but only use the largest three clusters for the skin tone measure, as the smaller clusters often correspond to shadows and non-skin parts such as eyes or lips. For each retained cluster, we compute its centroid,

[1] https://ai.google.dev/edge/mediapipe/solutions/vision/image_segmenter.

which summarizes its characteristic color, and then convert it from the RGB color space to the $L^*a^*b^*$ color space (also referred to as CIELAB).[2] To obtain a single representative skin color for each photo, we average the clusters' L^*, a^*, and b^* values, weighting each cluster by its assigned share of the skin-area pixels. In order to reduce the dimensionality of skin color, we focus on the darkness or lightness of the color – the "perceptual skin tint" (Adukia et al., 2023) – by retrieving only the L^* value from each photo's representative skin color.[3] Finally, we invert the L^* value, such that its scale ranges from 0 (lightest) and 100 (darkest). Figure C.2 shows examples of candidates with measured skin tones at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile of the distribution.

Figure C.2: Skin Tone Examples



Percentile	10	25	50	75	90
Skin tone	23.3	29.9	37.0	46.1	56.7
Race (self)	*Parda* Brown	*Branco* White	*Pardo* Brown	*Preta* Black	*Parda* Brown

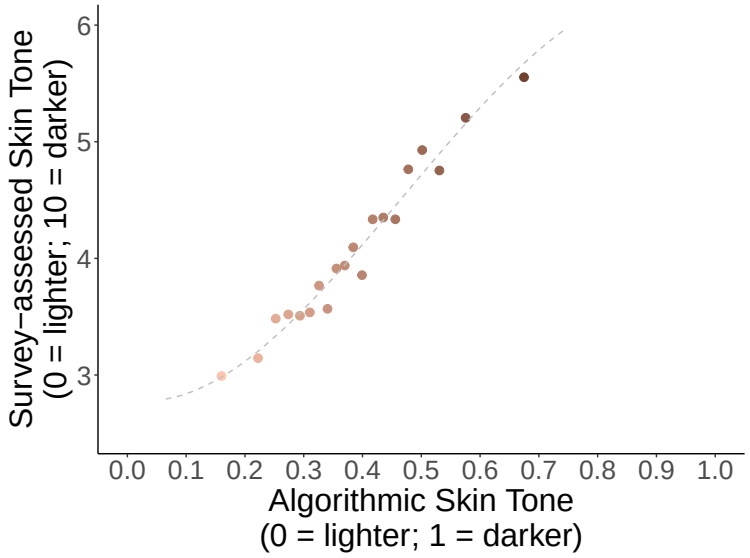
Notes: The figure presents five candidate examples, showing their electoral photographs, the extracted skin tone with its corresponding position in the skin-tone distribution, and the candidates' self-reported race.

Comparison to Human-Assessed Skin Tone. As a validation exercise, we can compare our algorithmic skin tone measure to assessments made by a representative sample of human raters. For that, we draw on data collected by Bueno and Dunning (2017). They conducted an online survey of about 1,100 Brazilian respondents who were shown electoral photos of candidates running in the 2008 and 2010 elections and were asked to rate candidates' skin color on a scale from 1 (lightest) to 10 (darkest). We successfully match 4,826 candidates, each of whom was assessed by an average of three respondents. Appendix Figure C.3 presents a binned scatterplot for the relation between our algorithmic skin tone measure and the mean survey rating for the same photo, showing a strong positive relation across the distribution. The overall correlation between both measures is 0.39. This can be compared to how much human raters diverge in their evaluation of the same photo. The average inter-rater correlation is 0.37. Hence, our algorithm's agreement with the average human perception is on par with the level of agreement observed among humans themselves.

[2]Unlike the RGB color space, the $L^*a^*b^*$ color space is perceptually linear.

[3]In the figures in the paper that depict color differences across the skin tone distribution, we always vary the L^* value for a skin tone with $a^* = 17.2$ and $b^* = 19.1$, which correspond to the sample averages of a^* and b^* across all our colored photos.

Figure C.3: Algorithmic vs. Survey-assessed Skin Tone



Notes: ...

Colored vs. Gray-scale Photos. About half of all candidate photos in our estimation sample (47.1%) are in color, with the other half – especially photos from candidacies in earlier years – being in grayscale.[4] Note that the algorithm is suitable for extracting skin tint from both types of images (following Adukia et al. (2023)). To assess potential differences in measurement quality, in Table C.1 we examine correlations with employer-ascribed race separately for skin tone measured from colored and grayscale photos.[5] In column (1), we replicate our main result from Section 5 when using all available photos of each candidate, while in columns (2) and (3), we only use colored or grayscale photos to measure the skin tone of each candidate (the corresponding binned scatterplots are shown in Figure C.4). For 63.2% (62.2%) of all candidates, we have at least one colored (grayscale) photo (remember that we average the skin tone across photos for individuals who run multiple times), which explains the reduction in sample size in the two columns. We find that the predictive power of skin tone for racial ascription is only modestly higher when using colored photos rather than grayscale photos, which can be explained by larger confounding influence from lighting in grayscale photos. At the same time, comparing columns (1) and (2), we see that the predictive power of colored photos (column (2)) is almost the same as when using all photos. This is because aggregating over more photos to measure the skin tone of each individual improves the measurement, which is highlighted in columns (4) to (6). Here, we focus on candidates who run multiple times and for whom we observe at least one colored and one grayscale photo. As can be seen, averaging across colored and grayscale photos yields a substantially larger correlation with employer-ascribed race compared to using colored

[4]In particular, all photos of candidates until 2016 were in grayscale, in 2018 and 2020 they started to be published in color (depending on location and party), and since then almost all photos are colored.

[5]Note that Bueno and Dunning (2017) have examined photos of candidates running in 2008 and 2010, which are all in grayscale, such that we cannot analyze potential differences in correlations with human-assessed skin tone.

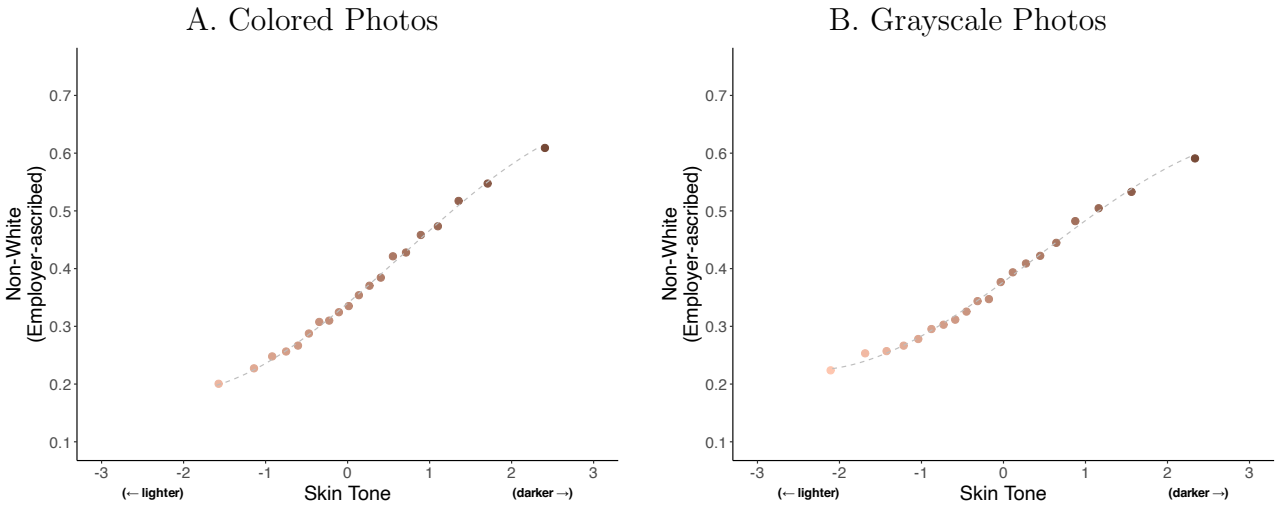
or grayscale photos alone. Overall, Table C.1 shows that incorporating both colored and grayscale photos in our skin tone measurement increases the sample of candidates that we can analyze without compromising measurement quality. Moreover, note that our estimation of perception-normalized disparities (PNDs) in Section 6 precisely accounts for the predictive power of skin tone for racial ascription, adjusting skin tone disparities for any potential noise in how skin tone is measured and how it translates into racial ascriptions.

Table C.1: Relationship between Employer-Ascribed Race and Skin Tone, Colored vs. Grayscale Photos

	All Candidates			Candidates with Colored and Grayscale Photos		
	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone - All Photos	0.107***			0.164***		
	(0.001)			(0.002)		
Skin Tone - Colored Photos		0.111***			0.114***	
		(0.001)			(0.002)	
Skin Tone - Grayscale Photos			0.090***			0.094***
			(0.001)			(0.002)
N	1,480,624	939,770	922,032	381,178	381,178	381,178

Notes: The table reports results from regressions of employer-ascribed race (indicator equal to one if the worker is ascribed non-White) on skin tone measured from all candidate photos vs. only colored photos vs. only grayscale photos. All skin tone measures are standardized using the mean and standard deviation from the distribution of skin tone based on all photos. Columns (1) to (3) use the full sample of candidates, while columns (4) to (6) restrict the sample to candidates who run for office multiple times and for whom we observe at least one colored and at least one grayscale photo. All regressions include year fixed effects. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C.4: Relationship between Employer-Ascribed Race and Skin Tone, Colored and Grayscale Photos



Notes: The figures show binned scatterplots of employer-ascribed race (an indicator equal to one if the worker is ascribed non-White) and skin tone measured from colored photos vs. grayscale photos (corresponding to columns (2) and (3) of Table C.1 but without year fixed effects). All skin tone measures are standardized using the mean and standard deviation from the distribution of skin tone based on all photos. The plots are constructed following the procedure in Cattaneo et al. (2024). For each sample, we divide the skin tone distribution in twenty equal-sized bins and, for each bin, plot the average of employer-ascribed race. We fit third-degree polynomials based on the underlying data of each sample.

D AKM Model

D.1 Empirical Strategy

Estimation Model. This section describes how we leverage workers' moves across establishments to quantify the extent to which ascribed race in RAIS varies depending on the employer, in an exercise in the spirit of Abowd et al. (1999). We estimate the following model using the full RAIS sample:

$$y_{it} = \alpha_i + \psi_{\boldsymbol{j}(i,t)} + X'_{it}\beta + \varepsilon_{it} \tag{1}$$

where y_{it} is an indicator equal to one if worker i at year t is ascribed as non-White (*preto* or *pardo*), and $\boldsymbol{j}(i,t) \in \{1, \ldots, J\}$ is an identity function that returns the firm j at which worker i was employed during year t.

The fixed effects model above divides a worker's employer-ascribed race into four components. First, α_i, the person's race effect, represents the portion of a worker's racial identity that all employers interpret equally. It captures features unambiguously associated with a specific racial group, such as certain skin tones, as well as signals that all employers associate with a particular racial group. Second, the firm's race effect $\psi_{\boldsymbol{j}(i,t)}$ represents the portion of ascribed race that is specific to each employer. This effect may include differences in employers' diligence in race reporting, differences in how employers interpret racial cues, and differences in how workers tend to present to a particular employer. Third, X_{it} is a set of the following time-varying controls: age squared, age cubed, and year, all interacted with workers' education (5 educational levels).[6] Finally, the error term ε_{it} captures all other match-specific factors affecting ascribed race, including idiosyncratic measurement error.

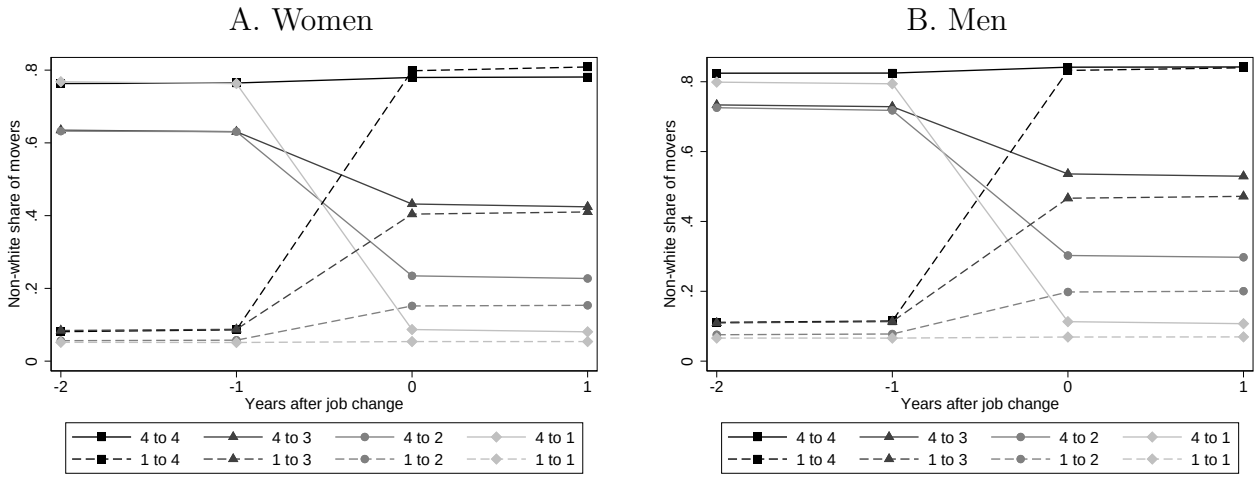
Variance Decomposition. Under the strict exogeneity assumption (discussed below), the person and firm effects in Equation (1) are identified. Estimating the above model allows us to decompose the variance of employer-ascribed race in RAIS into a component that is specific to firms, another that is specific to workers, and their correlation, as proposed by Abowd et al. (1999) in the case of wages:

$$\begin{aligned}\mathbb{V}\left[y_{it}\right] = \mathbb{V}\left[\alpha_i\right] + \mathbb{V}\left[\psi_{\boldsymbol{j}(i,t)}\right] + \mathbb{V}\left[X'_{it}\beta\right] + \mathbb{V}\left[\varepsilon_{it}\right] \\ + 2\mathbb{C}\left[\alpha_i, \psi_{\boldsymbol{j}(i,t)}\right] + 2\mathbb{C}\left[\alpha_i, X'_{it}\beta\right] + 2\mathbb{C}\left[\psi_{\boldsymbol{j}(i,t)}, X'_{it}\beta\right]\end{aligned}$$

where $\mathbb{V}$ and $\mathbb{C}$ are the finite-sample variance and covariance operators, respectively. Following Kline (2024), we present our decomposition in terms of the covariance-adjusted variance in ascribed race, i.e., we scale the variance and covariance of α_i and $\psi_{\boldsymbol{j}(i,t)}$ by $\mathbb{V}\left[y_{it} - X'_{it}\beta\right]$.

[6] To separately identify the person-effects and the time-varying controls, it is necessary to make a normalization assumption. Following Gerard et al. (2021) and Card et al. (2018), we assume that $X'_{it}\beta = 0$ for 40-year-olds. This implies that we measure person effect α_i for 40-year-old workers, which is approximately the age where the average worker achieves the peak of their wage in our sample. Differently from the literature estimating mover's design models outside of the labor context (e.g., Finkelstein et al., 2016; Cantoni and Pons, 2022), we do not control for relative move time because it is too computationally intensive, there is no strong evidence of pre-trends without controlling for them, and this is the usual practice in the labor literature.

Figure D.1: Event Studies around Job Moves



Notes: The figure displays the likelihood of being reported non-White by the employer among workers who moved from establishments in the top and bottom quartiles of establishment race effects to destination establishments in any of the other quartiles. Movers are defined as workers who separated from their origin establishment from 2003-2013, were reemployed in the destination establishment the following year, and were employed in both the origin and destination establishment for at least 2 years.

Identification Assumption and Specification Tests. OLS estimates of Equation (1) will yield unbiased estimates of the person and establishment fixed effects under the assumption of strict exogeneity of this error term, i.e., if $\mathbb{E}\left[\varepsilon_{it}|\boldsymbol{j}(i,s)=j, X_{is}=x\right]=0$ for all workers $i \in \{1,\ldots,N\}$ and all periods $(s,t) \in \{1,\ldots,T\}^2$.

The strict exogeneity identification assumption has two key implications in our context (which mirror the traditional AKM analysis as discussed, for instance, in Kline, 2024). First, worker mobility across firms must be uncorrelated with the time-varying residual component of ascribed race. This assumption is often referred to as *exogenous mobility*. In our setting, it rules out that changes in underlying racial identities are correlated with moves to particular types of firms. For instance, workers who become more likely to be ascribed non-White cannot become more likely to move to *darkening* firms.

As in previous work using similar models to study variation in wages (Card et al., 2013; Gerard et al., 2021), we test this hypothesis by performing an event-study analysis of changes in the ascribed race of workers moving across different types of firms. We split the sample of firms into quartiles of the estimated firm race effects $\hat{\psi}_{\boldsymbol{j}(i,t)}$, and consider the changes in ascribed race for workers moving from firms in the top and bottom quartile to firms in any quartile of the distribution. For that, we follow previous literature and restrict the sample to workers who stayed in the origin job for at least 2 years before the move. Results are displayed in Figure D.1. Reassuringly, we find no evidence of differential trends in ascribed race before the move, supporting the assumption that mobility is exogenous to time-varying components of the residual race change (however, we cannot rule out that sudden changes in racial identity are causing moves.)

The event study plots of Figure D.1 also provide helpful evidence on the second implication of strict exogeneity: that ascribed race is additively separable among firm and worker effects. Additive separability implies that effects of moves from *whitening* to *darkening* firms and vice-versa must be symmetric, and this is exactly what we observe: workers moving from

firms in the top quartile (i.e., firms that tend to *darken* workers) experience a change in the likelihood of being ascribed non-White that is similar in magnitude and opposite in sign to workers moving from firms from the bottom quartile (i.e., firms that tend to *whiten* workers).

D.2 Results

We estimate model (1) separately for female and male workers in the largest connected set of workers of each gender. As in any mover design analysis, worker and firm effects are only identified within the set of establishments linked by worker moves. We focus on the largest connected set of firms in our sample for each worker gender, which comprises 97.6% of firms for the sample of women and 98.8% of firms for the sample of men. Results are shown in Table D.1. The model fits well, with values of the adjusted R^2 between 76.0-78.3% and the root mean squared error (RMSE) ranging from 0.22 to 0.24.

We find that firm-specific reporting behavior accounts for a substantial share of the variance in ascribed race observed in RAIS. The standard deviation of our outcome (an indicator equal to one for workers being ascribed non-White) is 0.471 for women and 0.488 for men. In comparison, the standard deviation of the person effects is 0.338 for women (0.335 for men), while the standard deviation of establishment effects is 0.275 in the female sample (0.270 in the male sample). The implied variance decomposition shows that person effects explain about 46.9-51.5% of the variance in race in RAIS. Variation in establishment-specific race reporting explains 30.3-34.0% of the variance.[7] The remaining variance is largely residual variation, as the covariance between person and firm effects is very low. Figure D.2 displays the distribution of the estimated firm and person race effects.[8]

Limited mobility of workers may lead to biased estimates of the variance and covariance of firm and person effects. Thus, we also report bias-corrected estimates in the leave-one-out-connected set of firms and workers, following the approach of (Kline et al., 2020). As shown in Table D.2, this results in even smaller estimates for the share of variance explained by person effects (32.1-33.3%), while the share explained by firm effects changes only mildly (25.5-27.3%). Moreover, the correlation between person and firm effects gains in importance, now explaining 8.4-9.7% of the variation. Overall, these results suggest an even larger role for establishments in shaping racial ascriptions (when jointly considering firm effects and their covariance with person effects).

These findings are important and novel. First, the worker-specific, portable component of race explains at most half of the variation in ascribed race, indicating that race is not an innate characteristic of workers that is equally perceived across employers. Second, about one-third of the variation in ascribed race is explained by establishment variation, highlighting that firms vary widely in their race perceptions. For comparison, we use the sample sample and estimate an AKM model for log hourly wages, with results shown in Table D.3. Similar to the results in (Gerard et al., 2021), firm effects explain 18.0-18.9% of the variation

[7] Following Kline (2024), we report the decomposition in terms of the covariate-adjusted variance in employer-ascribed race, though results are remarkably similar if we reported unadjusted shares (since the standard deviation of the variation explained by our covariates is low).

[8] Correlating the firm effects estimated in the female and male sample, we find a coefficient of 0.55. The correlation suggests that firms' race reporting applies similarly to different types of workers, providing additional evidence in favor of the additive separability assumption discussed above.

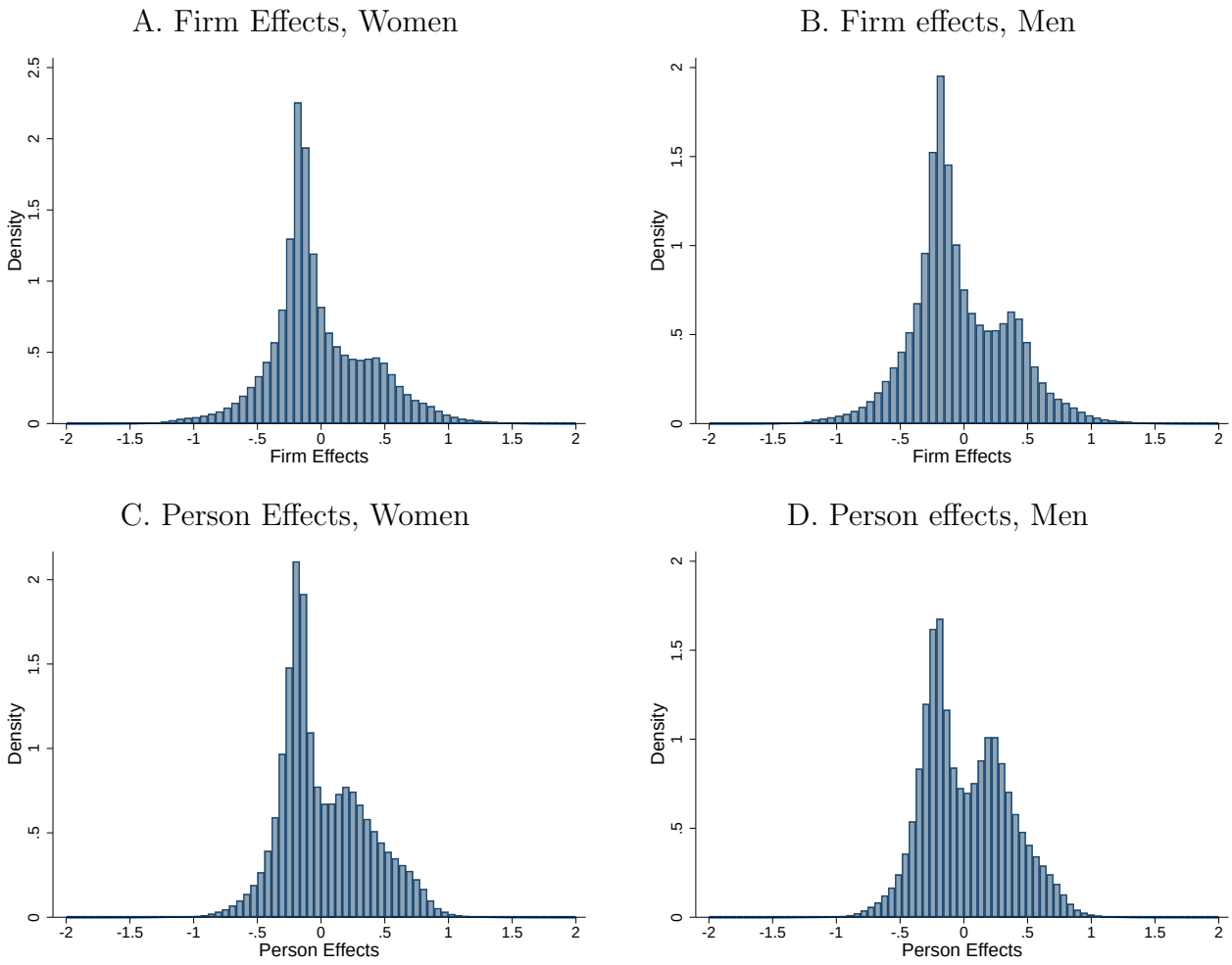
in wages. Thus, firms appear to be even more important in explaining differences in racial ascription than in wage setting.

Table D.1: Summary of Estimated Two-Way Fixed Effects Models

	Women (1)	Men (2)
Largest Connected Set		
Share of sample (across person-year obs.)	0.976	0.988
Share non-White (across person-year obs.)	0.333	0.389
Std. deviation of non-White (across person-year obs.)	0.471	0.488
AKM Decomposition		
SD of person effects (across person-year obs.)	0.338	0.335
SD of establishment effects (across person-year obs.)	0.275	0.270
SD of covariates (across person-year obs.)	0.024	0.029
SD of residuals (across person-year obs.)	0.195	0.216
Correlation of person and establishment effects	-0.031	0.043
Adjusted R^2 of the model	0.783	0.760
Root Mean Squared Error of the model	0.220	0.239
Percentage of covariate-adjusted variance in employer-ascribed race due to:		
Person effect	51.5	46.9
Establishment effect	34.0	30.3
Covariance of person and establishment effects	-2.6	3.3
Number of establishments	3,369,757	3,682,410
Number of workers	19,444,075	29,359,199
Number of person-year observations	108,810,943	183,152,434
Share of workers with any job change	0.639	0.697
Share of workers with any race change	0.259	0.322

Notes: The table summarizes the results from estimating the two-way fixed effect model (1), where the outcome is an indicator equal to one if worker i at year t is ascribed as non-White (*preto* or *pardo*) by their main employer. The models include worker and establishment fixed effects, alongside year dummies interacted with five educational level dummies, and quadratic and cubic age interacted with the educational level dummies. The model is estimated on the largest connected set of firms in our RAIS sample from 2003-2015.

Figure D.2: Distribution of Estimated Person and Establishment Race Effects



Notes: The figures display histograms of the estimated person and establishment effects for the two-way fixed effect models of employer-ascribed race, separately for the samples of female and male workers.

Table D.2: AKM Results, Leave-One-Out Connected Set

	Women (1)	Men (2)
Leave-One-Out Connected Set		
Share of sample (across person-year obs.)	0.894	0.937
Share non-White (across person-year obs.)	0.336	0.392
Std. deviation of non-White (across person-year obs.)	0.472	0.488
AKM Decomposition		
SD of person effects (across person-year obs.)	0.331	0.325
SD of establishment effects (across person-year obs.)	0.262	0.262
Correlation of person and establishment effects	0.013	0.063
Percentage of covariate-adjusted variance in employer-ascribed race due to:		
Person effect	49.0	47.5
Establishment effect	30.8	30.9
Covariance of person and establishment effects	1.0	4.9
KSS (Bias-Corrected) Decomposition		
SD of person effects (across person-year obs.)	0.273	0.268
SD of establishment effects (across person-year obs.)	0.238	0.247
Correlation of person and establishment effects	0.144	0.164
Percentage of covariate-adjusted variance in employer-ascribed race due to:		
Person effect	33.3	32.1
Establishment effect	25.5	27.3
Covariance of person and establishment effects	8.4	9.7
Number of establishments	2,083,141	2,504,285
Number of workers	17,789,210	27,868,425
Number of person-year observations	99,754,855	174,494,003
Share of workers with any job change	0.630	0.694
Share of workers with any race change	0.256	0.321

Notes: The table summarizes the results from estimating the two-way fixed effect model (1), where the outcome is an indicator equal to one if worker i at year t is ascribed as non-White (*preto* or *pardo*) by their main employer for the leave-one-out connected set of moves. The models include worker and establishment fixed effects, alongside year dummies interacted with five educational level dummies, and quadratic and cubic age interacted with the educational level dummies. The table displays both the traditional AKM decomposition and the bias-corrected decomposition from Kline et al. (2020).

Table D.3: Summary of Estimated Two-Way Fixed Effects Models for Log Hourly Wages

	Women (1)	Men (2)
Largest Connected Set		
Share of sample (across person-year obs.)	0.976	0.988
Share non-White (across person-year obs.)	0.333	0.389
Std. deviation of log hourly wage (across person-year obs.)	0.637	0.693
AKM Decomposition		
SD of person effects (across person-year obs.)	0.504	0.523
SD of establishment effects (across person-year obs.)	0.277	0.294
SD of covariates (across person-year obs.)	0.207	0.221
SD of residuals (across person-year obs.)	0.162	0.186
Correlation of person and establishment effects	0.333	0.354
Adjusted R^2 of the model	0.918	0.912
Root Mean Squared Error of the model	0.182	0.206
Percentage of variance in covariate-adjusted log wages due to:		
Person effect	62.6	56.8
Establishment effect	18.9	18.0
Covariance of person and establishment effects	22.8	22.7
Number of establishments	3,369,757	3,682,410
Number of workers	19,444,075	29,359,199
Number of person-year observations	108,810,943	183,152,434
Share of workers with any job change	0.639	0.697
Share of workers with any race change	0.259	0.322

Notes: The table summarizes the results from estimating the two-way fixed effect model (1), where the outcome is the log hourly wages of worker i at year t. The models include worker and establishment fixed effects, alongside year dummies interacted with five educational level dummies, and quadratic and cubic age interacted with the educational level dummies. The model is estimated on the largest connected set of firms in our RAIS sample from 2003-2015.